VYTAUTO DIDŽIOJO UNIVERSITETAS
MATEMATIKOS IR INFORMATIKOS INSTITUTAS

Vaida Bartkutė

POZICINIŲ STATISTIKŲ TAIKYMAS OPTIMALUMO ANALIZEI EKSTREMUMO PAIEŠKOS ALGORITMUOSE

Daktaro disertacija

Fiziniai mokslai (P 000)
Informatika (09 P)
Informatika, sistemų teorija (P 175)

Vilnius, 2007

Disertacija rengta 2003-2007 metais Matematikos ir informatikos institute

Mokslinis vadovas:

Prof. habil. dr. Leonidas Sakalauskas (Matematikos ir informatikos institutas, fiziniai mokslai, informatika – 09 P)

PADĖKA

Disertacijos autorė dėkoja moksliniam vadovui prof. habil. dr. Leonidui Sakalauskui, už per doktorantūros studijas suteiktas nuoširdžias, labai atsakingas bei vertingas mokslines konsultacijas, už patarimus ir pasiūlymus bei nuolatinį skatinimą tobulėti, Matematikos ir informatikos instituto direktoriams prof. habil. dr. Mifodijui Sapagovui bei prof. habil. dr. Gintautui Dzemydai už visapusišką paramą studijuojant doktorantūroje bei suteiktą galimybę dalyvauti tarptautinėse konferencijose, skelbiant tyrimo metu gautus rezultatus, MII duomenų analizės skyriaus operacijų tyrimo sektoriaus darbuotojams bei doktorantams už bendradarbiavimą, pagalbą ir supratimą, disertacijos recenzentams habil. dr. Juliui Žilinskui bei dr. Sauliui Minkevičiui, atidžiai perskaičiusiems disertaciją ir pateikusiems vertingų patarimų bei kritinių pastabų, Utenos kolegijos administracijai padėjusiai atspausdinti šią disertaciją bei visiems draugams ir artimiesiems už jų paramą, kantrybę, supratingumą ir meilę.

Vaida Bartkutė

TURINYS

ŽYMĖJIMAI IR SANTRUMPOS

b.t. (a.s.) – beveik tikrai (*almost surely*);

$P(B)$- įvykio B tikimybė;

M- algoritmo kartojimų skaičius;

N- algoritmo žingsnių skaičius;

α - ekstremaliųjų reikšmių skirstinio formos parametras;

$[\cdot]$ – skaičiaus sveikoji dalis;

K – Lipšico konstanta;

$\mathcal{L}_q$ – funkcijų, tenkinančių Lipšico sąlygą su rodikliu q, aibė;

φ - funkcijos homogeniškumo parametras;

n – funkcijos kintamųjų skaičius;

$\Gamma(x)$ - Gama funkcija;

k, m – pozicinių statistikų skaičius, $\frac{m}{k} \to 0{,}2$;

δ - pasikliovimo lygmuo;

$A = \min_{x \in D} f(x)$- tikslo funkcijos minimali reikšmė;

$A_{N,k}$ - tikslo funkcijos minimalios reikšmės tiesinis įvertis;

$p(\cdot)$ - tankio funkcija;

$I(C)$ - aibės C indikatorius;

V_n - n-mačio rutulio tūris;

$\{\Theta_t\}_{t=0}^{\infty}$ σ-algebrų seka generuota sekos $\{x^t\}_{t=0}^{\infty}$;

$\Re^n$ - n-matė Euklido erdvė;

$\|x\|$ - n-mačio vektoriaus euklidinė norma, $\|x\| = \sqrt{x_1^2 + \ldots + x_n^2}$;

$\|x - y\|$ - atstumas tarp dviejų taškų metrinėje erdvėje;

E_n – n-tos eilės vienetinė matrica;

X~$T(a,b)$- atsitiktinis dydis, tolygiai pasiskirstęs intervale [a, b];

X~ $N(0,1)$- atsitiktinis dydis, pasiskirstęs pagal standartinį normalinį dėsnį;

X~ $N(0, E_n)$- atsitiktinis n-matis vektorius, kurio komponentės yra pasiskirsčiusios pagal standartinį normalinį dėsnį;

X~ $N(\mu, \Sigma)$- atsitiktinis n-matis vektorius, kurio komponentės yra pasiskirsčiusios pagal daugiamatį Gauso dėsnį, μ-vidurkis, Σ - kovariacijų matrica;

X~ $B(x, \iota, \theta)$ - atsitiktinis dydis, pasiskirstęs pagal beta skirstinį su parametrais ι, θ;

X~ $W(x, \alpha, c, A)$ - atsitiktinis dydis, pasiskirstęs pagal ekstremaliųjų reikšmių skirstinį (Veibulo) su parametrais α, c, A.

$Fish_\gamma$, $Fish_{\gamma, n, m}$ – Fišerio skirstinio γ - kvantilis su n ir m laisvės laipsnių skaičiumi.

x^t – vektorius x t-toje iteracijoje;

$x_1, x_2, \ldots, x_n$ – vektoriaus x komponentės;

$\eta_{0,t} \le \eta_{1,t} \le \eta_{2,t} \le \ldots \le \eta_{t,t}$ - optimizavimo metu pasiektų geriausių tikslo funkcijos reikšmių seka (pozicinės statistikos);

ρ - stochastinės aproksimacijos algoritmo žingsnio ilgis arba modeliuojamojo atkaitinimo algoritmo žingsnio aplinkos gylis;

σ - suglodinimo parametras;

T – temperatūra modeliuojamojo atkaitinimo algoritme;

t - žingsnio (iteracijos) numeris;

ν - baigtinių skirtumų parametras;

$g(x, \sigma, \xi)$ - stochastinis gradientas;

$\overline{f}(x, \sigma)$ - suglodinta funkcija;

$\overline{g}(x, \sigma)$ - suglodintos funkcijos gradientas;

$\partial p(y)$ - apibendrintas funkcijos $p(y)$ gradientas;

$h(x)$ - funkcijos su aštriu minimumu apibendrinta Klarko išvestinė pagal kryptį;

$\overline{h}(x)$ - suglodinta $h(x)$ funkcija;

$O(n)$ – viršutinė riba, t.y., sakoma, kad $g(n) = O(f(n))$, jei egzistuoja konstanta $C > 0$, kad $\lim\limits_{n \to 0} \frac{g(n)}{f(n)} < C$;

$o(n)$ – sakoma, kad $g(n) = o(f(n))$, jei $\lim\limits_{n \to 0} \frac{g(n)}{f(n)} = 0$;

Ekstremaliųjų reikšmių skirstinys – I tipo skirstinys (Veibulo);

SPSA - modeliuojamojo pokyčio stochastinė aproksimacija (*Simultaneous Perturbation Stochastic Approximation*);

SPSAL - SPSA su Lipšico suglodinimo operatoriumi (*Simultaneous Perturbation Stochastic Approximation (SPSA) with Lipschitz perturbation operator*);

SPSAU - SPSA su tolygiai pasiskirsčiusiu suglodinimo operatoriumi (*SPSA with Uniform perturbation operator*);

FDSA - baigtinių skirtumų stochastinė aproksimacija (*Finite Difference Stochastic Approximation*);

SA – Modeliuojamojo atkaitinimo algoritmas (*Simulated Annealing*);

MLE – maksimalaus tikėtinumo įvertis (*maximum likelihood estimator*);

AE – analitinis įvertis (*analytical estimator*);

IAE – pagerintas analitinis įvertis (*improved analytical estimator*));

AK – asmeninis kompiuteris.

PAVEIKSLŲ SĄRAŠAS

LENTELIŲ SĄRAŠAS

ALGORITMŲ SĄRAŠAS

1. BENDROJI DARBO CHARAKTERISTIKA

1.1. TYRIMŲ SRITIS

Kuriant skaičiuojamuosius stochastinės ir euristinės paieškos algoritmus tenka spręsti praktines jų stabdymo bei gauto sprendinio optimalumo tyrimo problemas. Sprendžiant šias problemas galima pasinaudoti informacija, gauta iš geriausių pasiektų tikslo funkcijos reikšmių sekos (pozicinių statistikų), sudarytos optimizavimo metu. Pozicinių statistikų taikymas algoritmų optimalumui tirti yra glaudžiai susijęs su algoritmų konvergavimo sąlygomis bei konvergavimo greičio įvertinimu. Tad šio darbo tyrimų sritis yra Markovo tipo optimizavimo algoritmų teorijos bei ekstremaliųjų reikšmių teorijos sankirtoje, apimanti stochastinės bei euristinės paieškos algoritmus, šių algoritmų konvergavimo įvairiose optimizavimo uždavinių klasėse analizę bei optimizavimo metu gautų tikslo funkcijos reikšmių pozicinių statistikų skirstinių statistinės analizės metodus.

1.2. PROBLEMOS AKTUALUMAS

Įvairiems optimizavimo uždaviniams spręsti plačiai taikomi stochastiniai metodai. Šių metodų pritaikomumą lemia jų paprastumas, universalumas bei robastiškumas. Daugelį stochastinių metodų galima aprašyti Markovo tipo algoritmais. Stochastiniai Markovo tipo optimizavimo algoritmai vaizduojami Markovo grandinėmis, kur tikimybė pereiti į naują tašką priklauso tik nuo prieš tai buvusio taško ir funkcijos reikšmės tame taške. Šių algoritmų klasė yra pakankamai plati, apimanti daug žinomų atsitiktinės paieškos algoritmų, tokių kaip stochastinė aproksimacija, modeliuojamojo atkaitinimo algoritmas (*Simulated Annealing*), atsitiktinės paieškos algoritmas (*Rastrigin*) ir kt.

Stochastinės paieškos algoritmų taikymas realaus turinio uždaviniams spręsti reikalauja patikimų bei universalių priemonių, leidžiančių įvertinti gauto sprendinio optimalumą po baigtinio iteracijų skaičiaus, t. y. jo artimumą optimaliam sprendiniui, bei sudaryti taisykles iteraciniam procesui stabdyti, kai gautas sprendinys yra gana arti optimalaus taško arba kai tikslo funkcijos reikšmės pagerėjimas yra mažai tikėtinas. Markovo tipo algoritmų konvergavimas yra gerai ištirtas kai algoritmo iteracijų skaičius be galo didėja. Tačiau rezultatai apie metodo elgesį, kai žingsnių skaičius be galo didėja, ne visada teikia pakankamai informacijos priimti sprendimą po baigtinio optimizavimo algoritmo žingsnių skaičiaus. Priimant sprendimą apie uždavinio sprendinio optimalumą, reikia atsižvelgti į tai, kad informacija, gauta stochastinės paieškos metu, yra atsitiktinė. Todėl toks sprendimas gali būti priimamas statistiniais-tikimybiniais metodais apdorojant informaciją, gautą optimizavimo metu, o tokiu būdu gautos išvados yra patikimos tik su tam tikra tikimybe.

Nagrinėsime optimumo paieškos algoritmus, kai pagrindinė informacija, naudojama algoritmo žingsniams konstruoti, yra funkcijos reikšmės, apskaičiuotos optimizavimo metu tam tikruose taškuose. Plačią tokių algoritmų klasę, kai algoritmo žingsniai yra konstruojami, pasinaudojant tiktai šiomis reikšmėmis, sudaro begradientiniai algoritmai. Nors bandymai pasinaudoti informacija apie funkcijos reikšmes, gautas optimizavimo metu, algoritmo optimalumui įvertinti yra žinomi jau seniai, tačiau taikomųjų stochastinės paieškos algoritmų stabdymo taisyklių kūrimas reikalauja detalesnių tyrimų. Todėl sprendinio optimalumo analizė ekstremumo paieškos algoritmuose, pasinaudojant informacija, gauta optimizavimo metu, yra aktuali praktinė ir teorinė problema. Šios problemos sprendimas yra ypač aktualus kuriant begradientinio optimizavimo metodus.

1.3. TYRIMŲ OBJEKTAS

Disertacijos tyrimų objektas yra Markovo tipo begradientiniai (*derivative-free*) stochastinės paieškos algoritmai diferencijuojamų ir nediferencijuojamų funkcijų klasėse, pozicinių statistikų taikymas sprendinio optimalumo analizei ekstremumo paieškos algoritmuose tirti bei ekstremaliųjų reikšmių skirstinių parametrų vertinimo metodai.

1.4. TYRIMŲ TIKSLAS IR UŽDAVINIAI

Sprendimų apie gauto sprendinio optimalumą analizė yra susijusi su taikomo optimizavimo metodo konvergavimo fakto bei konvergavimo greičio nustatymu bei tyrimu. Daugumos stochastinės paieškos algoritmų konvergavimas yra ištirtas ir nustatytos sąlygos, kai algoritmas konverguoja. Kai kuriems stochastinio optimizavimo algoritmams yra įvertintas konvergavimo greitis bei įrodytas asimptotinis konvergavimas į Gauso dėsnį. Tačiau begradientinės stochastinės paieškos algoritmai dar nėra pakankamai detaliai ištirti, nors taikomiesiems uždaviniams spręsti jie yra taikomi gana dažnai. Bene plačiausiai žinomą ir geriausiai ištirtą stochastinių optimizavimo algoritmų, aprašomų Markovo grandine, klasę sudaro įvairios stochastinės aproksimacijos modifikacijos. Pastaruoju metu literatūroje daug dėmesio yra skiriama modeliuojamojo pokyčio stochastinės aproksimacijos (SPSA) metodams, kadangi šių metodų žingsniams konstruoti pakanka žinoti tik vieną arba kelias tikslo funkcijos reikšmes. Tačiau šio metodo taikymas nediferencijuojamoms funkcijoms optimizuoti taip pat nėra pakankamai gerai ištirtas, šio metodo taikymą riboja tam tikros išankstinės teorinės prielaidos.

Funkcijos reikšmių, gautų optimizavimo metu, taikymas sprendinio optimalumo analizei remiasi geriausių pasiektų tikslo funkcijos reikšmių skirstinio aproksimavimu ekstremaliųjų reikšmių skirstiniu. Funkcijos reikšmių, gautų optimizavimo metu, ekstremaliųjų reikšmių bei pozicinių statistikų savybės yra teoriškai beveik neištirtos. Teoriniai optimalių sprendimų priėmimo modeliai yra labai sudėtingi ir reikalauja kompleksinio algoritmo konvergavimo sąlygų, konvergavimo greičio, konvergavimo į asimptotinį dėsnį, ribinių ekstremaliųjų reikšmių skirstinių ir pan. tyrimo. Kadangi optimalių sprendimų priėmimo algoritmų teorinė analizė yra sudėtinga, kompiuterinis modeliavimas lieka svarbiu tyrimo metodu, leidžiančiu patikrinti ir ištirti hipotezes, keliamas sprendžiant minėtą problemą.

Šio darbo tikslas sudaryti metodus bei algoritmus optimaliems sprendimams priimti po baigtinio stochastinės aproksimacijos iteracijų skaičiaus, taikant pozicines statistikas, gautas optimizavimo metu, bei apibendrinti gautus rezultatus kitiems atsitiktinės paieškos Markovo tipo algoritmams.

Norint pasiekti šį tikslą, reikėjo išspręsti tokius uždavinius:

- Suformuluoti ir ištirti modeliuojamojo pokyčio stochastinės aproksimacijos algoritmus su Lipšico suglodinimo operatoriumi;
- Įrodyti šių algoritmų konvergavimą bei nustatyti konvergavimo sąlygas nediferencijuojamų funkcijų klasėje;
- Nustatyti šių algoritmų konvergavimo greitį nediferencijuojamų funkcijų klasėje;
- Kompiuterinio modeliavimo būdu ištirti sudaryto SPSA algoritmo konvergavimo greitį ir palyginti su teoriniais rezultatais;
- Palyginti sudarytą SPSA algoritmą su kitais stochastinės paieškos algoritmais;
- Pasinaudojant tikslo funkcijos reikšmių, gautų optimizavimo metu, pozicinėmis statistikomis, sudaryti minimalios reikšmės ir jos pasikliautinojo intervalo tiesinius įverčius, priklausančius nuo ekstremaliųjų reikšmių skirstinio formos parametro, nustatomo pagal funkcijos homogeniškumo ekstremumo aplinkoje rodiklį;
- Sudaryti minimalios reikšmės ir jos pasikliautinojo intervalo tiesinius įverčius, priklausančius nuo ekstremaliųjų reikšmių skirstinio formos parametro, įvertinamo maksimalaus tikėtinumo bei analitiniais metodais;
- Kompiuterinio modeliavimo metodu ištirti šių įverčių tinkamumą įvairių stochastinės paieškos algoritmų optimalumo sąlygoms tirti, panaudojant įvairias testines funkcijas;
- Algoritmui stabdyti sudaryti taisyklę – algoritmas stabdomas, kai pasikliautinasis intervalas tampa mažesnis už tam tikrą iš anksto nustatytą reikšmę;
- Pritaikyti gautus rezultatus uždaviniams praktikoje spręsti (opciono kainai modeliuoti ir t.t.).

1.5. MOKSLINIS NAUJUMAS

Darbe gauti šie nauji rezultatai:

- Sudarytas SPSA algoritmas su Lipšico suglodinimo operatoriumi nediferencijuojamo optimizavimo uždaviniams spręsti;

- Nustatytos sudaryto algoritmo konvergavimo sąlygos Lipšico funkcijų klasėje;
- Nustatytas teorinis sudaryto algoritmo konvergavimo greitis Lipšico funkcijų klasėje;
- Kompiuterinio modeliavimo būdu ištirtas įvairių stochastinės aproksimacijos algoritmų konvergavimo greitis;
- Pasiūlyti tikslo funkcijos minimalios reikšmės ir jos pasikliautinojo intervalo tiesiniai įverčiai, pritaikant funkcijos reikšmių, gautų optimizavimo metu, pozicines statistikas, darant prielaidą apie jų pasiskirstymą pagal ekstremaliųjų reikšmių skirstinį $W(x, \alpha, c, A)$;
- Pasiūlyta minimalios reikšmės ir jos pasikliautinojo intervalo tiesinių įverčių koeficientus skaičiuoti, kai ekstremaliųjų reikšmių skirstinio formos parametras α nustatomas pagal funkcijos homogeniškumo rodiklį ekstremumo aplinkoje arba įvertinamas statistiniais metodais;
- Pasiūlytas pagerintas analitinis metodas ekstremaliųjų reikšmių skirstinio formos parametrui įvertinti, kuris kompiuterinio modeliavimo būdu palygintas su kitais įverčiais;
- Atlikti skaičiuojamieji eksperimentai su įvairiomis testinėmis funkcijomis pasiūlytiems įverčiams tirti, kurių rezultatai neprieštarauja hipotezei apie funkcijos reikšmių, gautų optimizavimo metu, pasiskirstymą pagal ekstremaliųjų reikšmių skirstinį;
- Nustatyta, kad tikslo funkcijos minimalią reikšmę bei jos pasikliautinąjį intervalą galima aproksimuoti pasiūlytais įverčiais;
- Kompiuterinio modeliavimo metodu, panaudojant įvairias testines funkcijas, ištirtas šių įverčių tinkamumas kelių Markovo tipo stochastinės paieškos algoritmų optimalumo sąlygoms tirti;
- Skaičiuojamojo eksperimento metu įrodytas pasikliautinojo intervalo mažėjimas didėjant optimizavimo algoritmo iteracijų skaičiui, kuriuo remiantis pasiūlyta algoritmo stabdymo taisyklė – algoritmas stabdomas, kai pasikliautinojo intervalo plotis tampa mažesnis už tam tikrą iš anksto nustatytą reikšmę.

1.6. PRAKTINĖ DARBO REIKŠMĖ

- Sudaryti algoritmai bei programinė įranga nediferencijuojamo optimizavimo uždaviniams spręsti SPSA su Lipšico suglodinimo operatoriumi (SPSAL), SPSA su tolygiai pasiskirsčiusiu suglodinimo operatoriumi (SPSAU), baigtinių skirtumų stochastinės aproksimacijos (FDSA), modeliuojamojo atkaitinimo (SA) metodais;
- Sudaryta testinių uždavinių biblioteka stochastinės paieškos algoritmams testuoti;
- Sudaryti algoritmai bei programinė įranga stochastinių Markovo tipo optimizavimo algoritmų konvergavimui bei optimalių sprendimų priėmimo taisyklėms tirti Monte - Karlo metodu;
- Sudaryti algoritmai bei programinė įranga stochastinių Markovo tipo optimizavimo algoritmų stabdymo taisyklei realizuoti;

- Sudaryta programinė įranga yra realizuota programinių sistemų MathCad, C++, PASCAL, MPICC priemonėmis nuoseklių ir lygiagrečių skaičiavimų kompiuteriams;
- Sudaryta programinė įranga ir algoritmai yra pritaikyti taikomiesiems uždaviniams spręsti:
    - Heston'o stochastinio volatiliškumo modeliui kalibruoti;
    - tankerio pertvarų svoriui optimizuoti;
    - išeminio insulto sričių kompiuterinės tomografijos vaizdų automatinio atpažinimo sistemoms optimizuoti.

1.7. DARBO REZULTATŲ APROBAVIMAS

Tyrimų rezultatai buvo pristatyti ir aptarti šiose nacionalinėse ir tarptautinėse konferencijose:

1. Application of Order Statistics to Stopping of Stochastic Search. *XXI EURO Summer Institute "Stochastic and Heuristic Methods in Optimization"*. Liepos 25 – rugpjūčio 7 d., 2003, Neringa, Lietuva.
2. Vienalaikio trikdžio stochastinės aproksimacijos konvergavimas Lipšico funkcijoms. *Lietuvos matematikų draugijos XLV konferencija*. Birželio 27-28 d., 2004, Kaunas, Lietuva.
3. Application of Order Statistics to Simultaneous Perturbation Stochastic Approximation. *Workshop on Challenges of Continuous Optimization in Theory and Applications*. Liepos 2-4 d., 2004, Rodos, Graikija.
4. Order Statistics for Testing Optimality in Stochastic Optimization. *7th International Conference "Computer data analysis and modeling"*. Rugsėjo 6-10 d., 2004, Minskas, Baltarusija.
5. Modeliuojamojo pokyčio stochastinės aproksimacijos algoritmo konvergavimo greičio tyrimas. *„Informacinės technologijos-2005"* (mokslinė-teminė konferencija). Sausio 26-27 d., 2005, Kaunas, Lietuva.
6. Order Statistics for Optimality Testing in Stochastic Approximation. *55th Session of the International Statistical Institute (ISI)*. Balandžio 5-12 d., 2005, Sidnėjus, Australija.
7. Optimality Testing in Stochastic and Heuristic Algorithms. *Workshop of European Chapter on Metaheuristics "Metaheuristics and Large Scale Optimization"*. Gegužės 19-21 d., 2005, Vilnius, Lietuva.
8. Pozicinių statistikų taikymas modeliuojamojo pokyčio stochastinės aproksimacijos algoritmo optimalumo tyrimui. *„Kompiuterininkų dienos-2005"*. Rugsėjo 15-17 d., 2005, Klaipėda, Lietuva.
9. Application of Stochastic Approximation in Technical Design. *INYS workshop „Optimal Design of Technological Processes"*. Vasario 15-17 d., 2006, Vilnius, Lietuva.

10. Application of Order Statistics to Terminate the Heston Model Calibration Procedure. *The 26th Conference on Applied Statistics in Ireland* (CASI 2006). Gegužės 17–19 d., 2006, Kilarnei, Airija.
11. Three Parameter Estimation of the Weibul Distribution in Termination of Stochastic Algorithms. *9th International Vilnius Conference on Probability Theory and Mathematical Statistics*. Birželio 25–30 d., 2006, Vilnius, Lietuva.
12. Application of Stochastic Approximation to Nondifferentiable Optimization. *Workshop on Advances in Continuous Optimization*. Birželio 30 – liepos 1 d., 2006, Reikjavikas, Islandija.
13. Optimality Testing in Stochastic Algorithms. *21st European Conference on Operational Research* (*EURO XXI*). Liepos 2-5 d., 2006, Reikjavikas, Islandija.
14. Skaičiuojamųjų stochastinių Markovo algoritmų tyrimas, taikant lygiagrečiųjų skaičiavimų metodus. *Lietuvos jaunųjų mokslininkų konferencija "Operacijų tyrimas ir taikymai*". Gegužės 18 d., 2007, Vilnius, Lietuva.
15. Three Parameter Estimation of the Weibul Distribution by Order Statistics of Noncensored Sample. *12th International Conference on Applied Stochastic Models and Data Analysis* (ASMDA 2007). Gegužės 29 – birželio 1 d., 2007, Chania, Kreta, Graikija.
16. Parallel Stochastic Algorithms for Large Scale Derivative-Free Optimization. *2nd Conference on Optimization Methods & Software and 6th EUROPT Workshop on Advances in Continuous Optimization*. Liepos 4-7 d., 2007, Praha, Čekija.
17. Optimal Design by Stochastic Approximation. *22nd European Conference on Operational Research* (EURO XXII). Liepos 8-11 d., 2007, Praha, Čekija.

1.8. DARBO REZULTATŲ PUBLIKAVIMAS

Tyrimo rezultatai publikuoti šiuose moksliniuose leidiniuose:

Užsienio mokslo leidiniuose, įtrauktuose į Mokslinės informacijos instituto pagrindinių žurnalų sąrašą:

1. Bartkute V., Sakalauskas L. Simultaneous Perturbation Stochastic Approximation for Nonsmooth Functions. *European Journal on Operational Research.* Vol.181, No. 3, 2007, p.1174-1188 (ISSN 0377-2217).

Tarptautinių konferencijų darbuose, įtrauktuose į Mokslinės informacijos instituto sąrašą:

1. Bartkutė V., Sakalauskas L. Application of Stochastic Approximation in Technical Design. // *Series on Computers and Operations Research*. Vol. 7. Computer Aided Methods in Optimal Design and Operations, 2006, p. 29-38. (ISBN 981-256-909-X)

Recenzuojamuose Lietuvos ir užsienio leidiniuose:

1. Barkutė V., Sakalauskas L. Vienalaikio trikdžio stochastinės aproksimacijos konvergavimas Lipšico funkcijų klasėje. // *Liet. matemat. rink.* T. 44, spec.nr., 2004, p. 603-608. (ISSN 0132-2818)
2. Bartkutė V., Sakalauskas L. Pozicinių statistikų taikymas optimalumo tyrimui. // *Matematika ir matematinis modeliavimas*. Nr.1, 2005, p. 15–19. (ISSN 1822 – 2757)
3. Bartkutė V., Sakalauskas L. Pozicinių statistikų taikymas tiriant modeliuojamojo pokyčio stochastinės aproksimacijos algoritmo optimalumą. // *Informacijos mokslai*. T. 34, 2005, p. 136-141. (ISSN 1392-0561)
4. Bartkutė V., Sakalauskas L., Felinskas G. Optimality Testing in Stochastic and Heuristic Algorithms. // *Technological and economic development of economy*. Vol. XII , Nr.1, 2006, p. 4-10. (ISSN 1392-8619)
5. Bartkutė V., Sakalauskas L. Pozicinių statistikų taikymas stochastinės aproksimacijos algoritmų optimalumui tirti. // *Jaunųjų mokslininkų darbai.* Nr. 4 (11), 2006, p. 202-210. (ISSN 1648-8776)
6. Bartkutė V., Sakalauskas L. Three parameter estimation of the Weibull distribution by order statistics. // *Recent Advances on Stochastic Modeling and Data Analysis.* World Scientific, 2007, p. 91-101. (ISBN-13 978-981-270-968-4, ISBN-10 981-270-968-1)

Recenzuojamoje tarptautinių konferencijų pranešimų medžiagoje:

1. Bartkutė V., Sakalauskas L. Order Statistics for Testing Optimality in Stochastic Optimization. // *Proceedings of 7th International Conference "Computer data analysis and modeling*". 2004, p. 128-131. (ISBN 985-445-492-4)

Konferencijų pranešimų medžiagoje:

1. Bartkutė V., Sakalauskas L. Modeliuojamojo pokyčio stochastinės aproksimacijos algoritmo konvergavimo greičio tyrimas. // *Konferencijos „Informacinės tecnnologijos 2005" pranešimų medžiaga.* 2005, Kaunas, Lietuva, p. 415-425.
2. Bartkutė V. Application of Order Statistics to Terminate the Heston Model Calibration Procedure. *// Proceedings of the 26th Conference on Applied Statistics in Ireland.* May 17-19 d., 2006, Killarney, Ireland, p. 69-70.

Parengti ir priimti straipsniai:

1. Grigaitis D., Bartkutė V., Sakalauskas L. An Optimization of System for Automatic Recognition of Ischemic Stroke Areas in Computed Tomography Images. // *Informatica,* 2008, No. 1 (*priimtas*).

1.9. DISERTACIJOS STRUKTŪRA

Disertaciją sudaro 7 skyriai, literatūros sąrašas ir priedai.

1-asis skyrius yra įvadinis. Jame pateikiama disertacijos tyrimų sritis, problemos aktualumas, tyrimų objektas, tyrimų tikslas ir uždaviniai, mokslinis naujumas, praktinė darbo reikšmė bei darbo rezultatų aprobavimas ir publikavimas.

2-ajame skyriuje pateikiamas Markovo tipo (markovinių) stochastinių ir euristinių algoritmų analitinis tyrimas bei išsami optimizavimo algoritmų apžvalga. Atsižvelgiant į tikslo funkcijos apibrėžimo srities pobūdį ir į tikslo funkcijos diferencijavimo galimybes, išskiriamos optimizavimo uždavinių klasės. Aptariamas labai svarbus optimizavimo algoritmo kūrimo etapas - stabdymo taisyklės sudarymas, apžvelgiamos įvairios algoritmo stabdymo sąlygos. Toliau šiame skyriuje pateikiama bendra Markovo tipo algoritmų schema, pristatomi gerai žinomi šio tipo stochastinės paieškos algoritmai (stochastinės aproksimacijos bei modeliuojamojo atkaitinimo algoritmai). Pateikiama Markovo tipo algoritmų taikymo nediferencijuojamoms funkcijoms optimizuoti apžvalga bei šių algoritmų stabdymo problema.

3-asis skyrius yra skirtas begradientiniams stochastinio optimizavimo metodams sudaryti bei tirti. Šiame skyriuje yra išnagrinėti SPSA algoritmai Lipšico funkcijų, skaičiuojamų be triukšmo, klasėje kai modeliuojamas pokytis yra absoliučiai tolydus ir aprašomas tankio funkcija, tenkinančia Lipšico sąlygą. Nustatytos ir įrodytos šių algoritmų konvergavimo sąlygos bei nustatytas konvergavimo greitis: $E\left\|x^k - x^*\right\| = O\left(t^{-\beta}\right)$, kai $1 \le \beta < 2$. Kadangi optimizavimo algoritmams Lipšico funkcijų klasėje kurti yra taikomas funkcijos suglodinimas, tai šiame skyriuje yra aprašomas Lipšico funkcijų stochastinis diferencijavimas, t. y. gradientų aproksimavimas jų stochastiniais įverčiais, o tai sudaro sąlygas sukurti metodus, kur nebūtina skaičiuoti gradientų. Atsižvelgiant į stochastinio gradiento įvertinimo būdą, pateikiami ir tarpusavyje palyginami trys skirtingi algoritmai: SPSA algoritmas su Lipšico suglodinimo operatoriumi, SPSA algoritmas su tolygiai pasiskirsčiusiu suglodinimo operatoriumi ir baigtinių skirtumų stochastinės aproksimacijos algoritmas. Skyriaus pabaigoje pateikiami šių algoritmų konvergavimo greičio Monte-Karlo empiriniai mažiausių kvadratų įverčiai, patvirtinantys teorinius rezultatus.

4-ajame skyriuje nagrinėjamas Markovo tipo algoritmų optimizavimo sekų pozicinių statistikų taikymas optimizuojamos funkcijos minimaliai reikšmei bei jos pasikliautinajam intervalui įvertinti, kai iteracijų skaičius yra baigtinis. Pasiūlyti tikslo funkcijos minimalios reikšmės bei šios reikšmės pasikliautinojo intervalo įverčiai yra gana paprasti ir priklauso tik nuo vieno parametro - ekstremaliųjų reikšmių skirstinio formos parametro. Šiame skyriuje nagrinėjamas atvejis, kai ekstremaliųjų reikšmių skirstinio formos parametras yra žinomas ir nustatomas pasinaudojant funkcijos homogeniškumu ekstremumo aplinkoje. Skyriaus pabaigoje pateikiami kompiuterinio skaičiuojamojo eksperimento metu gauti šių įverčių tyrimo Monte-Karlo metodu rezultatai, kai įvairios testinės funkcijos daug kartų minimizuojamos keliais Markovo tipo algoritmais (SPSAL, SPSAU, FDSA). Gauti rezultatai neprieštarauja prielaidai apie funkcijos reikšmių, gautų optimizavimo metu, pasiskirstymą pagal ekstremaliųjų reikšmių skirstinį. Aprašyti rezultatai rodo, kad stochastinės aproksimacijos algoritmams stabdyti galime įvesti taisyklę, t. y. algoritmas yra stabdomas, kai pasikliautinojo intervalo plotis tampa mažesnis už tam tikrą iš anksto užsiduotą reikšmę $\varepsilon > 0$.

5-ajame skyriuje nagrinėjamas atvejis, kai ekstremaliųjų reikšmių skirstinio formos parametras, reikalingas tikslo funkcijos minimalios reikšmės bei jos pasikliautinojo intervalo įverčiams gauti, yra įvertinamas statistiniais metodais. Tuo tikslu išsamiai išnagrinėtas Veibulo skirstinio trijų parametrų vertinimas. Pateikiami šio skirstinio parametrų maksimalaus tikėtinumo bei analitiniai įverčiai. Suformuluotos ir įrodytos maksimalaus tikėtinumo įverčio egzistavimo sąlygos bei nustatytas šio įverčio lokalizavimo intervalas, suformuluotos ir įrodytos analitinio įverčio egzistavimo sąlygos bei įrodytas analitinio įvertinimo algoritmo konvergavimas į Veibulo skirstinio formos parametro įvertį tiesiniu greičiu. Pateikiami kompiuterinio modeliavimo būdu gauti sukurtų Veibulo skirstinio trijų parametrų vertinimo algoritmų efektyvumo tyrimo Monte-Karlo metodu rezultatai. Sudaryti ekstremaliųjų reikšmių skirstinio formos parametro įverčiai pritaikyti tikslo funkcijos minimalios reikšmės įverčiui bei jos pasikliautinajam intervalui gauti.

6-ajame skyriuje aprašoma sudarytų stochastinių nediferencijuojamo optimizavimo algoritmų programinei realizacijai, testavimui bei taikomiesiems uždaviniams spręsti naudojama techninė bei programinė įranga. Šiame skyriuje aprašomas skaičiuojamasis eksperimentas konvergavimo greičiui bei algoritmų stabdymo taisyklei tirti, naudojant lygiagrečiųjų skaičiavimų klasterį, bei pateikiami jo rezultatai. Šio klasterio panaudojimą lėmė nagrinėjamų ir taikomų stochastinės aproksimacijos metodų lėtas konvergavimas, reikalaujantis labai didelio iteracijų skaičiaus, kartu ir ilgo sprendimo laiko, reikalingo didelio matavimo uždaviniams išspręsti. Skyriaus pabaigoje pateikiamas disertacijoje aprašytų algoritmų taikymas praktiniams uždaviniams spręsti (Heston'o stochastinio volatiliškumo modeliui kalibruoti, tankerio pertvarų svoriui optimizuoti bei išeminio insulto sričių kompiuterinės tomografijos vaizdų automatinio atpažinimo sistemoms optimizuoti).

7-ajame skyriuje pateikiami tyrimo „Pozicinių statistikų taikymas optimalumo analizei ekstremumo paieškos algoritmuose“ rezultatai ir išvados.

2. MARKOVO TIPO STOCHASTINIŲ IR EURISTINIŲ ALGORITMŲ ANALITINIS TYRIMAS

2.1. ĮVADAS

Intensyvus kompiuterių tobulinimas ir plitimas sudaro sąlygas taikyti optimizavimo metodus įvairiose praktinės veiklos srityse siekiant maksimalaus (minimalaus) ekonominio ar kitokio efekto. Optimizavimo metodika dažnai taikoma ir aviacijoje, tobulinant raketas ar lėktuvus, ir medicinoje, ieškant geriausių diagnostinių sprendimų, ir transporte, vystant kuo efektyvesnes tvarkaraščių sudarymo strategijas, ir finansuose, priimant kuo pelningesnius investavimo sprendimus, ir daugelyje kitų sričių, kuriose susiduriame su sudėtingomis problemomis. Daugelis gamybos planavimo ir ekonominių uždavinių yra susiję su riboto kiekio išteklių (žaliavų, darbo jėgos, energijos, kuro ir kt.) paskirstymu. Dažnai turimus išteklius galima paskirstyti ir panaudoti ne vienu, o keliais ar daugeliu būdu. Tokiems uždaviniams spręsti taikomi įvairūs optimizavimo metodai.

Sudarant algoritmus, realizuojančius optimizavimo metodus, dažnai daroma prielaida, kad bet kuriame optimizuojamos funkcijos apibrėžimo srities taške galima apskaičiuoti funkcijos reikšmę arba jos gradientą. Šiame darbe apsiribosime dažnai praktikoje pasitaikančiu atveju, kai optimizavimo metu galime apskaičiuoti tiktai funkcijos reikšmes tam tikruose taškuose. Tuomet algoritmo žingsniams konstruoti yra pasinaudojama informacija apie jau apskaičiuotas funkcijos reikšmes. Tokią optimizavimo schemą geriausiai galima aprašyti Markovo tipo (Markoviniais) algoritmais (Жиглявский & Жилинскас (1991)). Markovo tipo atsitiktinės paieškos algoritmai aprašomi Markovo grandine, kur tikimybė pereiti į naują tašką priklauso tik nuo prieš tai buvusio taško ir funkcijos reikšmės tame taške.

Markoviniai algoritmai yra plačiai taikomi nediferencijuojamoms funkcijoms optimizuoti. Šiai algoritmų klasei priklauso labai daug gerai žinomų atsitiktinės paieškos algoritmų, tokių kaip: stochastinės aproksimacijos (Robins & Monro (1951), Kiefer & Wolfowitz (1952), Dvoretzky (1956), Ермольев (1976), Kushner & Yin (2003)), modeliuojamojo atkaitinimo (*Simulated Annealing*) (Metropolis et al. (1953), Yang (2000)), atsitiktinės paieškos (Растригин (1968), Ravindran et al. (2006)) ir kt. algoritmai.

2.2. OPTIMIZAVIMO ALGORITMAI IR JŲ SUDĖTINGUMAS

2.2.1. Optimizavimo uždavinių klasės

Optimizavimo uždavinį nusako skaliarinė funkcija $f : X \to \Re$ ir leistinųjų sprendinių aibė $D \subset X$, čia X yra metrinės erdvės poaibis. Optimizavimo uždavinys gali būti minimizavimo arba maksimizavimo. Sprendžiant minimizavimo (maksimizavimo) uždavinį reikia rasti bent vieną leistinųjų sprendinių aibės elementą $x^* \in D$, tokį, kad $f(x^*) \le f(x)$ ($f(x^*) \ge f(x)$), $\forall x \in D$. Be to maksimizavimo uždaviniai yra gaunami iš minimizavimo uždavinių vietoj funkcijos $f(x)$ imant $-f(x)$, o tai reiškia, kad jei funkcija $f(x)$ taške x^* įgyja mažiausią reikšmę aibėje *D*, tai funkcija $-f(x)$ tame pačiame taške įgyja didžiausią reikšmę toje aibėje. Leistinųjų sprendinių aibė dažnai nusakoma nelygybinių ir lygybinių ribojimų sistema: $\begin{cases} g_i(x) \le 0, 1 \le i \le k \\ g_j(x) = 0, k+1 \le j \le m. \end{cases}$ Funkcija $f(x)$ yra vadinama tikslo funkcija, o $g_i(x)$ - ribojimų funkcijomis.

Bendru atveju optimizavimo (minimizavimo) uždavinį žymėsime taip:

$$f(x) \to \min_{x \in D} \tag{2.1}$$

čia $f: D \to \Re$ yra skaliarinė, aprėžta iš apačios funkcija, $\min_{x \in D} f(x) = f(x^*) = A > -\infty$, $|x^*| < \infty$. Laikysime, kad bet kuriame apibrėžimo srities *D* taške galima apskaičiuoti funkcijos reikšmę (atskiru atveju su atsitiktine arba determinuota paklaida) per baigtinį laiką ir sunaudojant baigtinius resursus. Sprendžiant minimizavimo uždavinį, gautas sprendinys x^* vadinamas lokalaus minimumo tašku, jei egzistuoja tokia taško x^* aplinka $S(x^*)$, kad $f(x^*) \le f(x), x \in S(x^*)$, o jei $f(x^*) \le f(x), x \in \Re^n$, tai gautas sprendinys x^* yra vadinamas globaliuoju sprendiniu. Todėl, ieškant uždavinio globaliojo sprendinio, daug informacijos galima gauti nustatant uždavinio lokaliuosius ekstremumus. O be to globalusis ekstremumas visada yra ir lokalusis.

Disertacijoje daugiausia dėmesio yra skiriama lokalaus optimizavimo uždaviniams, išskyrus kelis atvejus, kai nagrinėjamos daugiaekstremaliosios funkcijos ir tiriami jų globalūs ekstremumai.

Optimizuojamos funkcijos lokalaus sprendinio radimas dažnai yra paprastesnis nei globalaus. Lokaliajam ekstremumui nustatyti paprastai pakanka kompiuterio laiko bei kitokių resursų, kurie kinta polinomiškai nuo uždavinio matavimo skaičiaus. Tačiau daugelyje diskretaus optimizavimo uždavinių globaliajam sprendiniui rasti tenka perrinkti visus galimus variantus, tuo pačiu daugelyje tolydaus optimizavimo uždavinių globaliajam sprendiniui rasti tenka perrinkti visus lokaliuosius ekstremumus, o tai gali pareikalauti didesnio kompiuterio laiko, kintančio eksponentiškai, atsižvelgiant į uždavinio matavimo skaičių *n*.

Pagal tikslo funkcijos apibrėžimo srities pobūdį optimizavimo uždaviniai yra skirstomi į *diskrečiojo* (sveikaskaičio, kombinatorinio) arba *tolydžiojo* optimizavimo uždavinius.

Pagal tikslo funkcijos diferencijuojamumą yra išskiriami *diferencijuojamojo* ir *nediferencijuojamojo* optimizavimo uždaviniai.

Įvairios optimizuojamų funkcijų klasės gali būti įvedamos pasinaudojus Lipšico sąlyga. Yra sakoma, kad funkcija tenkina Lipšico sąlygą su konstanta *K* rodikliu *q*, jei

$$\left|f(x_1)-f(x_2)\right| \le K\|x_1-x_2\|^q . \tag{2.2}$$

$\mathcal{L}_q$ pažymėsime funkcijų, tenkinančių Lipšico sąlygą su rodikliu q, aibę. Jei $q \ge 2$, tai funkcija yra diferencijuojama.

Šiame darbe daugiausia dėmesio skirsime tolydžiojo optimizavimo uždaviniams, kai apibrėžimo sritis yra daugiamatės Euklido erdvės poaibis arba sutampa su visa erdve, o tikslo funkcija priklauso nediferencijuojamųjų funkcijų klasei $\mathcal{L}_1$.

2.2.2 Optimizavimo algoritmai ir metodai

Optimizavimo algoritmu vadinsime algoritmą, skirtą spręsti tam tikros klasės optimizavimo uždavinį. Optimizavimo uždaviniai dažniausiai yra sprendžiami iteraciniu būdu pereinant nuo vieno funkcijos taško prie kito. Tokio algoritmo veikimo rezultatas yra taškų

$$x^1, x^2, \ldots, x^t \tag{2.3}$$

bei funkcijos reikšmių, apskaičiuotų tuose taškuose,

$$f\left(x^1\right), f\left(x^2\right), \ldots, f\left(x^t\right) \tag{2.4}$$

sekos.

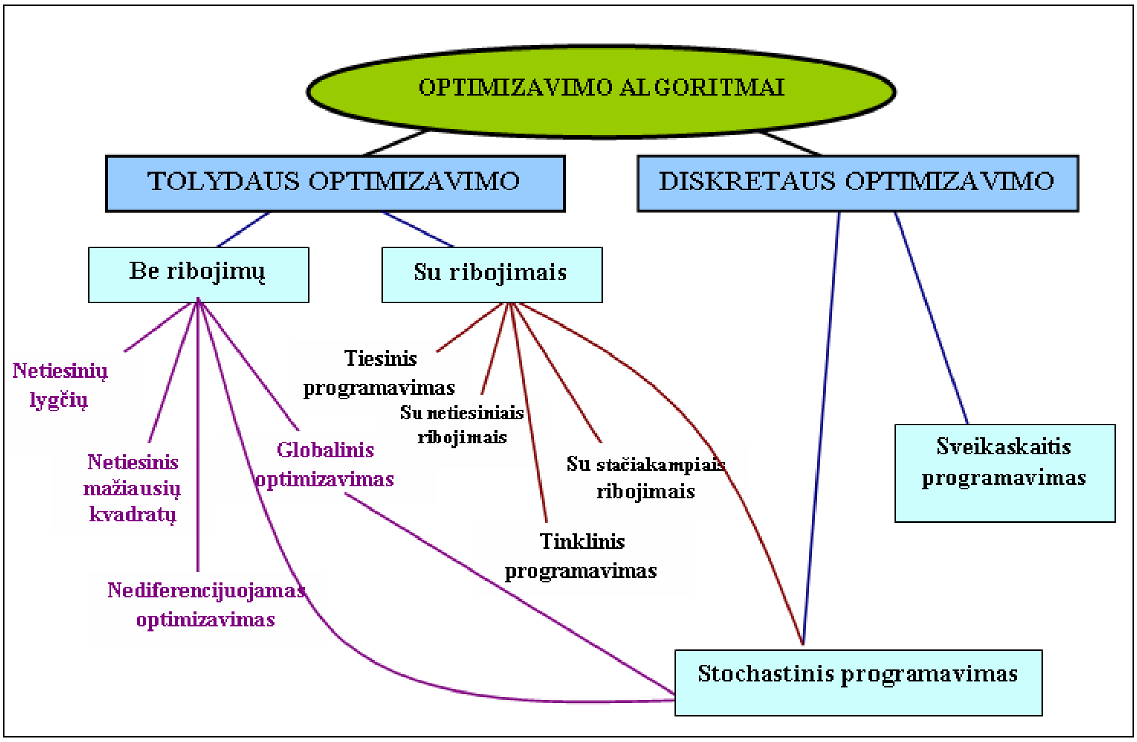
Toks taisyklių arba komandų rinkinys, kuris sprendžia optimizavimo uždavinį per begalinį žingsnių skaičių yra vadinamas *skaičiuojamuoju metodu* (Kleene (1952), Knuth (1973)). Skaičiuojamojo metodo veikimo rezultatas yra begalinės taškų bei funkcijos reikšmių tuose taškuose sekos. Jei skaičiuojamuoju metodu kuriama seka artėja į uždavinio sprendinį $\lim_{t\to\infty}\left\|x^t - x^*\right\| = 0$, tai sakoma, kad metodas yra konverguojantis.

Optimizavimo algoritmai yra kuriami konverguojančių metodų pagrindu. Paverčiant skaičiuojamąjį metodą optimizavimo algoritmu, tenka spręsti dvejopo tipo problemas:

- įvertinti gauto sprendinio atstumą nuo optimalaus, sustabdžius algoritmą bet kuriame žingsnyje;
- parinkti algoritmo stabdymo taisyklę, leidžiančią gauti sprendinį nustatytu tikslumu.

Determinuoto skaičiuojamojo metodo sudėtingumą (Garey & Johnson (1979)) nusako operacijų skaičius arba sprendimo laikas, reikalingas uždaviniui išspręsti *nustatytu tikslumu*. *Stochastinio algoritmo sudėtingumas* yra nusakomas arba vidutiniu žingsnių skaičiumi, reikalingu uždaviniui išspręsti *nustatytu tikslumu*, arba tikimybe, kad uždavinys bus išspręstas *nustatytu tikslumu* po kurio nors žingsnių skaičiaus. Tikslumo kriterijus ir optimizavimo algoritmo stabdymo taisykles aptarsime 2.2.3 skyrelyje.

Atsižvelgiant į optimizavimo uždavinių klasę, algoritmai skirstomi į diskrečiojo optimizavimo (Kouvelis (1997)) arba tolydžiojo optimizavimo su ribojimais arba be jų, tiesinį arba netiesinį programavimą ir pan. Atsižvelgiant į sprendimo schemą, algoritmai būna determinuoti arba stochastiniai, lokaliosios (gradientiniai metodai, stochastinė aproksimacija) arba globaliosios paieškos (modeliuojamojo atkaitinimo, genetiniai algoritmai) ir pan. *Pav. 1* pateikta viešai priimta optimizavimo algoritmų klasifikacija (NEOS Guide (http://www.ece.northwestern.edu/OTC)) pagal sprendžiamų uždavinių klasės pobūdį bei jų sprendimo schemą. Atsižvelgiant į tai, kad kai kuriose taikymo srityse, ypač ekonomikoje, tenka optimizuoti neapibrėžtumo sąlygomis, šiuos neapibrėžtumus įvertinant tikimybių teorijos sąvokomis bei sprendžiamo uždavinio leistinajai sričiai esant nebūtinai baigtinei, ši schema papildoma stochastinio programavimo algoritmais, taikomais tolydaus optimizavimo be ribojimų bei globalaus optimizavimo uždaviniams spręsti.



Pav. 2.1. Optimizavimo algoritmų klasifikacija

2.2.3. Algoritmų stabdymas

Bet kuriam algoritmui turi būti apibrėžta stabdymo taisyklė, kuri po tam tikro žingsnių skaičiaus leidžia sustabdyti skaičiavimus. Stabdymo taisyklės sudarymas yra labai svarbus optimizavimo algoritmo kūrimo etapas. Literatūroje yra skiriamos *stabdymo taisyklės* ir *stabdymo kriterijaus* sąvokos. Stabdymo kriterijus yra apibrėžiamas kaip matematiškai tiksliai suformuluota sąlyga, kurios išpildymas reiškia, kad yra pasiektas optimizavimo tikslas. Stabdymo taisyklė yra apibrėžiama kaip optimizavimo algoritmo dalis, besiremianti vienu ar keliais stabdymo kriterijais, skirta sustabdyti optimizavimo algoritmą (Hutchison & Spall (2005)). Tinkamai sudaryta stabdymo taisyklė turi:

- stabdyti algoritmą, kai paklaida yra visiškai maža;
- stabdyti algoritmą, kai pasiekto sprendinio pagerėjimas yra labai mažas;
- stabdyti algoritmą, pasiekus nustatytą iteracijų skaičių.

Stabdymo taisyklė turi būti siejama su pasiekto sprendinio kokybe. Taikant optimizavimo algoritmus praktiniams uždaviniams spręsti svarbiausios yra pirmos dvi algoritmo stabdymo priežastys. Jei algoritmas yra stabdomas tiktai atlikus nustatytą iteracijų skaičių, tai tokiu atveju yra labai svarbu įvertinti taško, gauto po baigtinio iteracijų skaičiaus, kokybę bei jo skirtumą nuo optimizavimo uždavinio sprendinio.

Be to, algoritmas taip pat gali būti sustabdytas, jei:

- nepakanka kompiuterinių resursų tęsti algoritmo darbą;
- algoritmo darbas yra nutraukiamas paties vykdytojo;
- dėl kompiuterio ir programinės įrangos klaidų.

Tašką x^t, gautą po baigtinio skaičiuojamojo metodo žingsnių skaičiaus t, vadinsime *artiniu*. Žingsnių skaičiui didėjant, skaičiuojamojo metodo artinys, gautas atskiruose metodo žingsniuose, artėja į uždavinio sprendinį. Artinys, gautas sustabdžius skaičiuojamąjį metodą po baigtinio iteracijų skaičiaus, gali būti lygus uždavinio sprendiniui tik apytikriai. Artinio skirtumas nuo uždavinio sprendinio gali būti apibūdinamas arba atstumu $\left\| x^t - x^* \right\|$, arba tikslo funkcijos reikšmių skirtumu $\left| f\left(x^t\right) - f(x^*) \right|$. Diferencijuojamų tikslo funkcijų atveju šis skirtumas gali būti apibūdinamas gradiento norma.

Norint įvertinti gauto artinio kokybę, reikia turėti tam tikrą išankstinę informaciją ir tam tikrus kriterijus, pagal kuriuos galėtume palyginti skirtingus artinius. Skaičiuojamojo metodo konvergavimo faktas rodo, kad atlikus pakankamai didelį žingsnių skaičių, gautas artinys skirsis kiek norima mažai nuo tikslaus uždavinio sprendinio. Tegul $\varepsilon > 0$ yra iš anksto nustatytas skaliarinis dydis, toks, kad jei skirtumas tarp uždavinio sprendinio ir artinio neviršija šio dydžio, tai uždavinio sprendinį galima aproksimuoti gautu artiniu x^t, jei

$$\left\| x^t - x^* \right\| < \varepsilon . \tag{2.5}$$

Optimizavimo uždaviniuose yra siekiama rasti optimalų tašką, todėl jis iš anksto nėra žinomas. Todėl stabdymo taisykles galima sudaryti pasinaudojant optimizavimo metu gautomis taškų arba tikslo funkcijos reikšmių tuose taškuose sekomis.

Sudarysime taisyklę pasinaudodami Koši sąlyga (žr. priedą 1). Iš šios sąlygos seka, kad bet kokiam $\varepsilon > 0$ galima rasti tokį *T*, kad visiems $t \geq T, q \geq T$ yra teisinga $\left\| x^t - x^q \right\| < \varepsilon$. Iš čia gauname labai paprastą algoritmo stabdymo taisyklę, t. y. algoritmas yra stabdomas, kai pasiekus tam tikrą žingsnių skaičių *q* ir atlikus dar papildomai *k* žingsnių, yra išpildoma sąlyga:

$$\left\| x^{q+i} - x^{q+j} \right\| < \varepsilon, \text{ visiems } 0 \leq i \leq k,\ i \leq j \leq k. \tag{2.6}$$

Šiuo atveju išankstinė informacija, reikalinga algoritmo stabdymo taisyklei sukurti, yra nustatytas tikslumas ε bei žingsnių skaičius *k*. Praktiniuose algoritmuose dažnai yra priimama, kad $k = 1$, ir algoritmas yra stabdomas, kai dviejų gretimų artinių skirtumas tampa mažesnis už nustatytą reikšmę. Reikia pabrėžti, kad sąlyga (2.6) seka iš algoritmo konvergavimo fakto, tačiau atvirkščias teiginys, jei nelygybė (2.6) teisinga kokiems nors *q*, *k* ir ε, tai algoritmas konverguoja, ne visada yra teisingas. Todėl norint taikyti taisyklę (2.6) reikia turėti išankstinę informaciją apie algoritmo konvergavimo faktą, kuri dažniausiai yra gaunama teorinių tyrimų metu.

Jei tikslo funkcija yra tolydinė, tai iš taškų sekos konvergavimo gaunamas funkcijos reikšmių, apskaičiuotų tuose taškuose, sekos konvergavimas: $\lim_{t \to \infty} \left| f\left(x^t\right) - f\left(x^*\right) \right| = 0$. Pažymėkime $f^* = f\left(x^*\right)$. Tokiu atveju optimalią funkcijos reikšmę galima aproksimuoti funkcijos reikšme gautame artinyje, jei

$$\left| f\left(x^t\right) - f^* \right| \leq \varepsilon. \tag{2.7}$$

Pasinaudojus Koši sąlyga galima įvesti taisyklę ir algoritmą stabdyti, kai yra išpildoma sąlyga:

$$\left| f\left(x^{q+i}\right) - f\left(x^{q+j}\right) \right| \leq \varepsilon \text{ visiems } 0 \leq i < k,\ i < j \leq k.$$

Jei tikslo funkcija yra diferencijuojama, tai iš Kuhn-Tucker teoremos (Kuhn & Tucker (1951)) seka, kad optimumo taške tikslo funkcijos gradientas yra lygus 0. Taigi tokiu atveju turime, kad gradiento reikšmių seka, apskaičiuota taškuose, gautuose optimizavimo algoritmu, konverguoja į 0. Todėl algoritmas gali būti stabdomas, kai keliuose arba bent viename taške gradientas tampa mažesnis už iš anksto nustatytą reikšmę.

Įvairiems optimizavimo algoritmams stabdyti yra sudaromos kombinuotos stabdymo taisyklės, kuriose panaudojamos aukščiau išvardintos sąlygos. Pvz., algoritmas gali būti stabdomas, jei arba tikslo funkcijos gradientas yra artimas 0, arba funkcijos reikšmės keletą žingsnių iš eilės kito nežymiai (Dennis & Shnabel (1996)).

Algoritmo vykdymo metu daug informacijos apie uždavinio sprendinį teikia geriausių pasiektų funkcijos reikšmių seka. Bendru atveju artinio kokybei įvertinti bei stabdymo taisyklei sudaryti gali būti naudojama tokia informacija:

- geriausios pasiektos funkcijos reikšmės;
- taškai, kuriuose yra pasiektos geriausios funkcijos reikšmės;

- numeriai žingsnių, kuriuose yra pasiekti funkcijos reikšmių pagerėjimai.

Šios informacijos panaudojimas optimizavimo algoritmo stabdymo taisyklei sudaryti yra aktuali ir mažai išnagrinėta praktinė bei teorinė problema. Aptarsime šios problemos kai kuriuos bendrus aspektus.

Tegu optimizavimo metu gautos funkcijos reikšmės yra išrikiuotos didėjimo tvarka. Pažymėkime tokiu būdu gautą seką: $\eta_{0,t} \le \eta_{1,t} \le \eta_{2,t} \le \ldots \le \eta_{t,t}$.

Teiginys 2.1. Jei algoritmu yra gaunama funkcijos reikšmių seka, konverguojanti į optimalią reikšmę, tai bet kuriam baigtiniam $k \ge 0$ seka $\eta_{k,t}$ monotoniškai konverguoja į optimalią reikšmę:

$$\lim_{t\to\infty} \eta_{k,t} = f^* . \tag{2.8}$$

Šis teiginys seka iš konverguojančios sekos bet kurio begalinio posekio konvergavimo į tą pačią ribą.

Sekmuo 2.1. Jei algoritmu yra gaunama konverguojanti funkcijos reikšmių seka, tai

$$\lim_{t\to\infty} \left|\eta_{0,t} - f^*\right| = 0 , \tag{2.9}$$

$$\mathit{lim}_{t\to\infty} \left|\eta_{k,t} - \eta_{0,t}\right| = 0 . \tag{2.10}$$

Iš (2.9) seka tokia taisyklė algoritmui stabdyti: algoritmas stabdomas, jei atlikus nustatytą žingsnių skaičių *k*, pasiekta geriausia funkcijos reikšmė nepasikeitė arba geriausios pasiektos reikšmės pokytis buvo mažesnis už iš anksto nustatytą reikšmę:

$$\left|\eta_{0,q+i} - \eta_{0,q+j}\right| < \varepsilon \text{ visiems } 0 \le i < k ,\ i < j \le k .$$

Iš (2.10) seka tokia stabdymo taisyklė: algoritmas gali būti stabdomas, jei tam tikriems $0 < k \le t$ ir $\varepsilon > 0$ yra teisinga sąlyga:

$$\left|\eta_{k,t} - \eta_{0,t}\right| < \varepsilon . \tag{2.11}$$

Informacija apie geriausias pasiektas tikslo funkcijos reikšmes gali būti panaudota tikslo funkcijos optimaliai reikšmei įvertinti bei nustatyti intervalą, kuriam ši reikšmė priklauso. Pavyzdžiui, iš sekos $\eta_{k,t}$ monotoninio konvergavimo seka toks optimalios reikšmės įvertis:

$$f^* = \eta_{0,t} - \beta_{k,t} \cdot \left(\eta_{k,t} - \eta_{0,t}\right), \tag{2.12}$$

čia $\beta_{k,t}$ tam tikras daugiklis.

Intervalas $\left[f_{apat}; f_{vir\check{s}}\right]$, kuriam priklauso optimali tikslo funkcijos reikšmė, gali būti įvertinamas panašiai:

$$f_{vir\check{s}} = \eta_{0,t} ,\ f_{apat} = \eta_{0,t} - r_{k,t} \cdot \left(\eta_{k,t} - \eta_{0,t}\right). \tag{2.13}$$

Daugiklis $r_{k,t} \ge \beta_{k,t} \ge 0$ yra parenkamas atsižvelgiant į sprendžiamų uždavinių klasę bei pasirinktą algoritmą.

Žinoma, jei algoritmas yra stochastinis, tai įvertį (2.12) bei intervalą (2.13) reikia parinkti naudojant tikimybinius metodus. Geriausių funkcijos reikšmių, pasiektų optimizavimo metu, taikymai algoritmams stabdyti yra detaliau aptariami 2.3.4 skyrelyje.

2.3. MARKOVO TIPO OPTIMIZAVIMO ALGORITMŲ ANALITINIS TYRIMAS

2.3.1 Bendra Markovo tipo optimizavimo algoritmų schema

Markovo tipo (markoviniai) atsitiktinės paieškos algoritmai išsiskiria paprastumu bei interpretacijos vaizdumu. Todėl šie algoritmai yra gerai ištirti teoriškai ir plačiai taikomi optimizavimo uždaviniams spręsti. Markovo tipo algoritmai plačiai taikomi globalaus ekstremumo paieškai, derinant „lokaliuosius“ žingsnius (determinuotus arba atsitiktinius) su „globaliomis“ procedūromis (Жиглявский & Жилинскас (1991), Horst et al. (2000)). Taikomuoju požiūriu svarbu pabrėžti, kad Markovo tipo algoritmai gali būti taikomi ir tuomet, kai tikslo funkcija yra apskaičiuojama su deterministine arba atsitiktine paklaida. Įvairūs markoviniai algoritmai yra nagrinėjami daugelio autorių keletą pastarųjų dešimtmečių. Daugelis deterministinės ir stochastinės paieškos algoritmų gali būti laikomi Markovo tipo algoritmais. Pavyzdžiui, deterministinės gradientinės paieškos (Dennis & Shnabel (1996), Bazaraa et al. (1992)), simplekso metodas (Dantzig (1963), Nelder & Mead (1965),) ir kt. gali būti aprašyti Markovo tipo algoritmais. Stochastinių Markovo tipo optimizavimo algoritmų klasė yra labai plati ir apima stochastinės aproksimacijos (Bhatnagar & Kowshik (2005)), modeliuojamojo atkaitinimo (Yang (2000)), atsitiktinės paieškos „bandymų ir klaidų“ metodus ir kt.

Algoritmo kūrimo pradžioje sudaroma algoritmo schema, kurios atskiri žingsniai gali būti aprašomi Markovo grandinėmis. Ši schema turi tenkinti pagrindines algoritmo diskretumo, žingsnių elementarumo (lokalumo) bei masiškumo savybes (Dičiūnas & Skersys (2003)). Stochastiniuose algoritmuose bazinių operacijų skaičius, reikalingas atskiriems algoritmo žingsniams atlikti, gali būti atsitiktinis, todėl yra reikalaujama, kad tokiuose algoritmuose šis bazinių operacijų skaičius būtų baigtinis su tikimybe 1.

Disertacijoje daugiausia dėmesio skirsime stochastiniams Markovo tipo algoritmams. Bendru atveju Markovo tipo algoritmai gali būti aprašomi Markovo grandine, kur tikimybė pereiti į naują tašką x^{t+1} priklauso tik nuo prieš tai buvusio taško x^t ir funkcijos reikšmės tame taške $\eta^t = f\left(x^t\right)$, t. y.

$$P_{t+1}\left(x^{t+1} \middle| x^1, \eta^1, \ldots, x^t, \eta^t \right) = P_{t+1}\left(x^{t+1} \middle| x^t, \eta^t \right).$$

Aptarsime bendrąją Markovo tipo algoritmų schemą.

Algoritmas 2.1: **Markovo tipo stochastinis algoritmas**

Tikslas: sudaryti tikslo funkcijos optimizavimo seką.

Pradinės sąlygos (*preconditions*)**:** tikslo funkcijos apskaičiavimo algoritmas, atsitiktinio taško, pasiskirsčiusio pagal skirstinį P_t, generavimo algoritmas, atsitiktinių taškų, pasiskirsčiusių pagal skirstinį Q_t, generavimo algoritmas, generuojamų atsitiktinių taškų skaičius $\kappa \geq 1$, stabdymo taisyklės formalizuotas aprašymas.

Galinės sąlygos (*postconditions*)**:** optimizavimo taškų seka, tikslo funkcijos reikšmių tuose taškuose seka.

1. Inicializuoti pradinį tašką x^0, pasiskirsčiusį pagal skirstinį P_0, apskaičiuoti tikslo funkcijos reikšmę η^0 šiame taške, $t = 0$.
2. **While** neišpildyta stabdymo taisyklė **do**
3. Artinio x^t aplinkoje generuoti κ atsitiktinių taškų, pasiskirsčiusių pagal skirstinį $Q_t\left(x^t, \cdot\right)$ bei priklausančių funkcijos apibrėžimo sričiai, kuriuose taip pat yra apskaičiuojamos funkcijos reikšmės z^t {gali būti su paklaida}.
4. Sudaryti naujojo artinio skirstinį $P_{t+1}\left(x^t, \eta^t, \kappa, z^t\right)$.

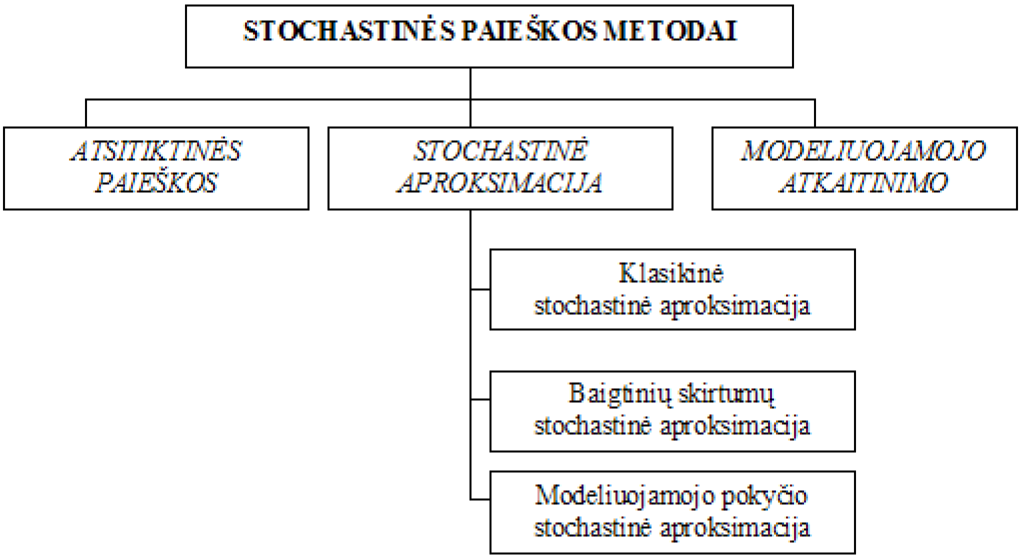
   {*Komentaras*: šis skirstinys priklauso nuo x^t, η^t, sugeneruotų κ atsitiktinių taškų bei funkcijos reikšmių juose z^t }.
5. Generuoti artinį x^{t+1} pagal skirstinį P_{t+1}, apskaičiuoti tikslo funkcijos reikšmę η^{t+1} šiame taške, $t = t + 1$.

   {*Komentaras*: čia $\eta^{t+1} = f\left(x^{t+1}\right) + \xi^{t+1}$, ξ^{t+1} atsitiktinė arba deterministinė paklaida},
6. **done.** ◊

Atsižvelgiant į atsitiktinių taškų generavimo skirstinio Q_t ir naujojo artinio skirstinio gavimo būdą galima gauti skirtingus stochastinės paieškos algoritmus.

2.3.2. Stochastinės paieškos metodai

Šiame darbe apsiribosime Markovo tipo stochastiniais algoritmais, taikomais nediferencijuojamoms, apskaičiuojamoms be paklaidų duotame taške funkcijoms optimizuoti, kai kiekviename žingsnyje tikslo funkcija apskaičiuojama apibrėžtą skaičių kartų. Labiausiai žinomi tokio tipo algoritmai yra atsitiktinės paieškos „bandymų ir klaidų“ metodas, stochastinės aproksimacijos, modeliuojamojo atkaitinimo ir kiti metodai.



Pav. 2.2. Stochastinės paieškos metodai

2.3.2.1. Atsitiktinės paieškos metodai

Atsitiktinės paieškos algoritmai buvo pasiūlyti Andersono (1953) ir vėliau plėtojami Brooks (1958), Karnopp (1963), Растригин (1968) ir kitų autorių. Tokie algoritmai išsiskiria šiomis savybėmis:

- šiuos metodus lengva programuoti ir realizuoti;
- šie metodai gali būti taikomi trūkioms arba nediferencijuojamoms funkcijoms optimizuoti neiškilose srityse;
- šiuos metodus yra lengva modifikuoti ir kombinuoti su kitais ekstremumo paieškos metodais;
- šių metodų konvergavimas optimumo aplinkoje yra lėtas.

Paprasčiausia atsitiktinė paieška yra susijusi su tiesioginės atrankos procedūromis, kurių metu yra generuojamos atsitiktinių taškų imtys (Reklaitis et al. (1986)). Tokia imtis yra gaunama sugeneravus N atsitiktinių taškų

$$x_i^j = a_i + \xi_i^j \left(b_i - a_i \right), \quad i = \overline{1,n}, \quad j = \overline{1,N},$$

čia $a = \left(a_1, a_2, ..., a_n \right)$, $b = \left(b_1, b_2, ..., b_n \right)$ yra optimizuojamų parametrų apatinių ir viršutinių rėžių vektoriai, ξ_i - atsitiktiniai dydžiai, tolygiai pasiskirstę intervale [0, 1]. Toliau yra apskaičiuojamos tikslo funkcijos reikšmės šiuose taškuose ir pasirenkamas taškas, kuriame tikslo funkcijos reikšmė yra mažiausia. Šis taškas gali būti priimtas kaip optimizavimo uždavinio sprendinys arba gali būti generuojamos kitos atsitiktinės imtys, reguliuojant apatinius ir viršutinius kintamųjų rėžius bei imties tūrį. Kadangi šis metodas išsiskiria lėtu konvergavimu ir norint gauti optimizavimo uždavinio sprendinio artinį priimtinu tikslumu, reikia generuoti labai dideles atsitiktinių taškų imtis.

Šio metodo efektyvumui padidinti galima taikyti reikšmingų imčių arba išsluoksniuotosios imties metodus (Korn & Korn (2003)). Kituose algoritmuose siūloma geriausio žinomo taško aplinkoje generuoti atsitiktinių taškų, pasiskirsčiusių pagal Gauso dėsnį su tam tikra dispersija, imtis. Dispersija yra laikoma metodo parametru, kuris yra mažinamas arba didinamas atsižvelgiant į tai, ar pavyko pagerinti tikslo funkcijos reikšmę (Heuckroth et al. (1976)). Kai kuriuose atsitiktinės paieškos algoritmuose geriausio žinomo taško aplinkoje yra generuojami atsitiktiniai taškai, tolygiai pasiskirstę tam tikro spindulio sferos paviršiuje, opotimizavimo metu reguliuojant sferos spindulio ilgį ir pereinant į geresnį pasiektą tašką (Растригин (1968), Schumer & Steiglitz (1972)). Atsitiktinės paieškos algoritmai dėl jų paprastumo yra dažnai taikomi realaus turinio optimizavimo uždaviniams spręsti (Reklaitis et al. (1986)), taip pat jų pagrindu yra kuriami euristinės paieškos metodai (Rardin (1998)).

2.3.2.2. Stochastinės aproksimacijos metodai

Stochastinė aproksimacija yra vienas iš geriausiai žinomų ir ištirtų stochastinės optimizacijos metodų. Pirmaisiais šios klasės metodais yra laikomi Robbins-Monro (Robbins & Monro (1951)) ir Kiefer-Wolfowitz algoritmai (Kiefer & Wolfowitz (1952)). Robbins-Monro įvedė stochastinę aproksimaciją kaip bendrą metodą regresijos lygčiai spręsti, kai funkcijos reikšmės yra apskaičiuojamos su atsitiktine paklaida. Kiefer-Wolfowitz sudarė šio metodo modifikaciją tikslo funkcijos minimumui rasti aproksimuojant gradientą baigtinių skirtumų metodu. 1992 m. J. Spall (Spall (1992)) įvedė modeliuojamojo pokyčio stochastinės aproksimacijos (*Simulated Perturbation Stochastic Approximation* (SPSA)) metodą, kuris nuo Kiefer-Wolfowitz metodo skyrėsi tuo, kad gradientui įvertinti naudojamos tiktai dvi tikslo funkcijos reikšmės.

Stochastinės aproksimacijos metodų tyrimo rezultatų apžvalginius aprašymus galima rasti šiuose darbuose: Dvoretzky (1956), Wasan (1969), Poliak (1987), Kushner & Yin (2003) ir kituose. Įvairių stochastinės paieškos metodų efektyvumo palyginimą galima rasti Chin (1997).

Stochastinės aproksimacijos metodai yra grindžiami tikslo funkcijos suglodinimo ir stochastinio gradiento idėjomis. Bendru atveju ši idėja atrodo taip: yra įvedama funkcijų seka $f(x,\sigma)$, konverguojanti į $f(x)$, kai $\sigma \to 0$. Tada optimizavimo uždavinys (2.1) sprendžiamas minimizuojant šios sekos funkcijas ir atitinkamu būdu keičiant suglodinimo parametrą σ. Stochastinis gradientas yra atsitiktinis vektorius $g(x,\sigma,\xi)$, kurio vidurkis sutampa su suglodintos funkcijos gradientu $\overline{g}(x,\sigma) = E\left(g(x,\sigma,\xi)\right)$, čia $E(\cdot)$-atsitiktinio dydžio vidurkis.

Bendru atveju stochastinės aproksimacijos metodų schema yra tokia:

$$x^{t+1} = x^t - \rho_t \cdot g^t, \quad t = 1, 2, ..., \tag{2.14}$$

čia $g^t = g\left(x^t, \sigma_t, \xi^t\right)$ stochastinis gradientas, ξ^t atsitiktinis vektorius, $\sigma_t > 0$ suglodinimo parametras, $\rho_t > 0$ algoritmo žingsnio ilgis *t*-oje iteracijoje. Ši schema yra tokia pat skirtingiems stochastinės aproksimacijos metodams. Skiriasi tik stochastinio gradiento įverčiai. Stochastinio gradiento įvertinimo atskirus atvejus aptarsime skyrelyje 3.2.2.

Algoritmu 2.2 aprašytas optimizavimo taškų sekos sudarymas stochastinės aproksimacijos metodu.

Algoritmas 2.2: **Stochastinės aproksimacijos algoritmas**

Tikslas: sudaryti tikslo funkcijos stochastinės aproksimacijos seką.

Pradinės sąlygos (*preconditions*)**:** tikslo funkcijos apskaičiavimo algoritmas, stochastinio gradiento įvertinimo algoritmas, formalizuotas stabdymo taisyklės aprašymas, optimizavimo žingsnių skaičius *N*, pozicinių statistikų skaičius *k*, formalizuotas optimizavimo žingsnio ilgio ρ bei suglodinimo parametro σ reguliavimo taisyklės aprašymas, pradinis taškas x^0.

Galinės sąlygos (*postconditions*)**:** optimizavimo taškų seka x^t, tikslo funkcijos reikšmių tuose taškuose seka $f\left(x^t\right)$, geriausių pasiektų tikslo funkcijos reikšmių masyvas η^t.

1. Inicializuoti pradinį tašką x_0, optimizavimo žingsnio ilgį ρ_0, suglodinimo parametrą σ_0, užpildyti masyvą η^t pradinėmis reikšmėmis, t = 0.
2. **While** neišpildyta stabdymo taisyklė **do**
3. Pagal pasirinktą optimizavimo žingsnio ilgio bei suglodinimo parametro reguliavimo taisyklę perskaičiuojame parametrus ρ_t ir σ_t.
4. Apskaičiuoti tikslo funkcijos reikšmę $\eta_t = f\left(x^t\right)$.
5. Perskaičiuoti *k* geriausių pasiektų tikslo funkcijos reikšmių masyvą
$$\eta^t = \left(\eta_{(0)}, \dots, \eta_{(k)}\right).$$
6. Apskaičiuoti tikslo funkcijos stochastinį gradientą g^t taške x^t
{čia $g^t = g\left(x^t, \sigma_t, \xi^t\right)$, ξ^t atsitiktinis vektorius (algoritmai 3.1, 3.2, 3.3)}.
7. Apskaičiuoti artinį $x^{t+1} = x^t - \rho_t \cdot g^t$, $t = t + 1$.
8. **done**. ◊

Nors stochastinės aproksimacijos metodams, kaip ir kitiems atsitiktinės paieškos metodams, būdingas paprastumas, tačiau jų teorinis tyrimas yra gana sudėtingas uždavinys. Tiriant stochastinės aproksimacijos metodus, pirmiausia yra nustatomos šių metodų konvergavimo sąlygos, po to tiriamas ir nustatomas šių metodų konvergavimo greitis ir galiausiai tiriamas artinio, teikiamo šiais algoritmais, konvergavimas į normalinį dėsnį.

Stochastinės aproksimacijos algoritmų konvergavimui įrodyti bei konvergavimo sąlygoms Lipšico funkcijų klasėje nustatyti yra skirta labai daug darbų (Wasan (1969), Poliak (1987), Dupac (1988), Kushner & Yin (2003)). Paprastai yra siekiama įrodyti, kad algoritmo žingsniuose konstruojamas sprendinys konverguoja beveik tikrai (b. t.) (*almost surely*) į optimizavimo uždavinio lokalųjį sprendinį, kai žingsnių skaičius be galo didėja:

$$P\left(\lim_{t\to\infty}\left\|x^t - x^*\right\| = 0\right) = 1.$$

Jei metodo konvergavimas yra įrodytas, toliau nustatomas metodo konvergavimo greitis. Reguliuojant tam tikru būdu stochastinės aproksimacijos algoritmo parametrus σ_t, ρ_t (žr. 2.14), jei tikslo funkcija skaičiuojama be paklaidų ar triukšmo, galima pasiekti konvergavimo greitį $E\left\|x^t - x^*\right\|^2 = O\left(\frac{1}{t}\right)$ (Wasan (1969), Poliak (1987), Dupac (1988)). Jei tikslo funkcija apskaičiuojama su atsitiktine paklaida, tai konvergavimo greitis sumažėja ir yra lygus $O\left(\frac{1}{t^{\beta}}\right), 0 < \beta < 1$ (Граничин & Поляк (2003)). Nors daugelio stochastinės aproksimacijos algoritmų konvergavimas Lipšico klasėje yra įrodytas, tačiau nediferencijuojamo optimizavimo algoritmų konvergavimo greitis dar nėra pakankamai gerai ištirtas.

Daugeliui stochasinės aproksimacijos algoritmų difrenencijuojamų funkcijų klasėje yra įrodyta, kad sprendinys x^t, konstruojamas algoritmo žingsnių metu, yra asimptotiškai pasiskirstęs pagal daugiamatį Gauso dėsnį $N\left(\mu^t, \Sigma^t\right)$, čia $\mu^t \to x^*, t \to \infty$ (Wasan (1969), Kushner & Yin (2003)). Nediferencijuojamo optimizavimo atveju konvergavimas į asimptotinį dėsnį dar nėra pakankamai gerai ištirtas.

Reikia pabrėžti, kad algoritmų, paremtų stochastinės aproksimacijos metodais, stabdymo taisyklės sudarymas priklauso nuo metodo konvergavimo greičio įvertinimo ir konvergavimo į ribinį dėsnį sąlygų nustatymo.

2.3.2.3. Modeliuojamojo atkaitinimo metodas

Modeliuojamasis atkaitinimas (*Simulated Annealing* (SA)) - optimizavimo metodas, kurio matematinis modelis atspindi dalelių aibės fizikinėje sistemoje elgseną mažinant temperatūrą (Metropolis et al. (1953)). Šis metodas remiasi fizikinio proceso – atkaitinimo (grūdinimo) idėja, kuri yra susijusi su atkaitinimo procesu metalurgijoje, kurio metu tam tikra medžiaga yra kaitinama, o po to vėsinama siekiant padidinti medžiagos kristalų dydį ir pašalinti jų defektus. Kaitinimas priverčia daleles pereiti atsitiktinai iš pradinės būsenos (vidinės energijos lokalaus minimumo) į aukštesnės energijos būsenas; lėtas aušinimas suteikia daugiau galimybių rasti mažesnės nei pradinė vidinės energijos formą. Tikimybė pereiti iš vienos būsenos E_1 į kitą E_2 yra lygi :

$$p = \begin{cases} 1, & E_1 - E_2 < 0 \\ e^{\frac{-(E_1 - E_2)}{k \cdot T}}, & E_1 - E_2 \geq 0 \end{cases}$$

čia k – Bolcmano konstanta, T – sistemos temperatūra. Remiantis analogija šiam fizikiniam procesui kiekviename SA algoritmo žingsnyje yra pereinama į kaimyninį atsitiktinį tašką su tikimybe, priklausančia nuo atitinkamų funkcijos reikšmių skirtumo ir algoritmo globalaus parametro T (vadinamo temperatūra), kuris yra tolygiai mažinamas viso proceso metu. Ši priklausomybė yra tokia, kad kai T yra aukšta, tai pereinama į naują tašką atsitiktinai, o kai T artėja į 0, tikimybė pereiti į „blogesnį“ tašką mažėja. Todėl šiam parametrui reguliuoti įvedama speciali temperatūros atnaujinimo (aušinimo) funkcija. Taip pat įvedamas specialus nuo temperatūros priklausantis tikimybinis patvirtinimo kriterijus, patvirtinantis kiekviename šio metodo žingsnyje surastus geresnius nei esamas sprendinius bei su nedidele tikimybe patvirtinantis ir blogesnius sprendinius. O tai suteikia galimybę ištrūkti iš lokaliųjų optimumo taškų ieškant globalaus optimumo.

Šio algoritmo tikimybę rasti globalų sprendinį galima padidinti, įvedus leistiną minimalų atstumą ρ nuo einamojo taško, kuriuo yra generuojami nauji sprendiniai (Yang (2000)). Pirmuosiuose algoritmo žingsniuose šis atstumas gali būti pakankamai didelis, o po to nuosekliai vis mažinamas. Modeliuojamojo atkaitinimo algoritmo taikymo diskretaus ir tolydaus optimizavimo uždaviniams spręsti rezultatus galima rasti daugelyje darbų (Dorea (1997), Locatelly (2000a), Locatelly (2000b)). Kirkpatrick et al. (1983) atspausdino šio algoritmo taikymo kombinatorinio optimizavimo uždaviniams spręsti rezultatus. Collins et al. (1988) pateikė apžvalgą apie šio metodo taikymą įvairiems optimizavimo uždaviniams spręsti.

Modeliuojant atkaitinimą svarbu parinkti tinkamą temperatūros mažinimo formulę. Įvairios temperatūros atnaujinimo funkcijos apžvelgiamos Kirkpatrick et al. (1983), Locatelly (2000b), Yang (2000).

Atkaitinimo procesas teoriškai turėtų būti tęsiamas tol, kol galutinė temperatūra tampa lygi 0. Tačiau atkaitinimą galima baigti ir anksčiau, kai tikslo funkcijos pagerėjimas tampa mažai tikėtinas. Pvz., kai tikslo funkcijos reikšmė nemažėja gana ilgą laiko tarpą arba kai, atlikus gana daug bandymų, priimtų teigiamų sprendimų (t. y. perėjimų iš vieno sprendinio į kitą) skaičius tampa mažesnis už tam tikrą slenkstį. Dažnai algoritmo stabdymo sąlyga yra laikoma fiksuotas atliekamų bandymų skaičius, t. y. atkaitinimo schemos „ilgis“.

Modeliuojamojo atkaitinimo algoritmui realizuoti yra taikomos įvairios schemos. Disertacijoje pateiksime šio algoritmo schemą, pasiūlytą (Yang (2000)). Taikant taškų generavimo mechanizmą bei Metropolio patvirtinimo kriterijų SA algoritmu sudaromos dvi atsitiktinių taškų sekos. Taškų seka $\{z^t, t \geq 0\}$ generuojama pagal dėsnį su tankio funkcija:

$$p(x, T_t) = \prod_{j=1}^{n} \frac{T_t^{1/m}}{2m \cdot \left[\left|x_j\right| + T_t\right]^{\frac{m+1}{m}}}, \quad x \in \Re^n.$$

Iš tankio funkcijos išraiškos gauname, kad vektoriaus z^t j-toji komponentė yra pasiskirsčiusi pagal dėsnį:

$$P(x, T_t) = \begin{cases} \dfrac{1}{2 \cdot \left(1 - \dfrac{x}{T_t}\right)^{\frac{1}{m}}}, \; x < 0 \\ 1 - \dfrac{1}{2 \cdot \left(\dfrac{x}{T_t} + 1\right)^{\frac{1}{m}}}, \; x > 0 \end{cases}.$$

Spręsdami lygtį $P(x, T_t) = \varsigma$, gauname atsitiktinio vektoriaus z^t j-tąją komponentę:

$$z_j^t = \begin{cases} T_t \cdot \left(1 - \left(\dfrac{1}{2\varsigma}\right)^m\right), \; 0 < \varsigma < \dfrac{1}{2}, \\ T_t \cdot \left(\left(\dfrac{1}{2(1-\varsigma)}\right)^m - 1\right), \; \dfrac{1}{2} \leq \varsigma < 1 \end{cases},$$

čia $\varsigma \sim T[0,1]$. (2.15)

Seka $\{x^t, t \geq 0\}$ sudaroma taikant Metropolio patvirtinimo kriterijų (Metropolis et al. (1953)). Šios abi sekos priklauso nuo temperatūros $\{T_t, t \geq 0\}$ bei aplinkos gylio $\{\rho^t, t \geq 0\}$, kurie perskaičiuojami tokiu būdu:

$$\rho^t = \frac{\rho_0}{t^{\frac{\gamma \cdot m}{\lambda \cdot (m+1) \cdot n}}},$$

$$T_t = \frac{T_0}{t^{\frac{m}{n}}},$$

čia $m \geq 1$, $\lambda > 1$, $0 < \gamma < \min\left\{\lambda, \dfrac{m}{n}\right\}$, $t \geq 1$.

Disertacijoje nagrinėjama modeliuojamojo atkaitinimo metodo modifikacija aprašyta algoritmu 2.3:

Algoritmas 2.3: **Modeliuojamojo atkaitinimo (*Simulated Annealing*) algoritmas**

Tikslas: sudaryti tikslo funkcijos globalaus optimizavimo seką.

Pradinės sąlygos (*preconditions*)**:** tikslo funkcijos apskaičiavimo algoritmas, pradinis taškas, atkaitinimo proceso temperatūros apskaičiavimo funkcija, aplinkos gylio ρ_t apskaičiavimo funkcija, atsitiktinių vektorių z^i generavimo algoritmas, stabdymo taisyklės formalizuotas aprašymas.

Galinės sąlygos (*postconditions*)**:** optimizavimo taškų seka x^t, tikslo funkcijos reikšmių tuose taškuose seka $f\left(x^t\right)$, geriausių pasiektų tikslo funkcijos reikšmių masyvas η^t.

Pradinių duomenų inicializavimas

1. Inicializuoti pradinį tašką $y^0 = x^0 \in D \subset \Re^n$, pradinę temperatūrą $T_0 > 0$, pradinį aplinkos gylį $\rho_0 > 0$.

2. Apskaičiuoti pradinę tikslo funkcijos reikšmę $\eta_0 = f(x^0)$, $t = 0$.

3. **While** neišpildyta stabdymo taisyklė **do**

4. Pagal (2.15) generuoti atsitiktinį vektorių z^t.

5. **While** $\left(y^t \notin D\right) \vee \left(\left|z_j^t\right| < \rho_t,\ 1 \le j \le n\right)$ **do**

{*Komentaras*: z_j^t yra *j*-ta vektoriaus z^t komponentė}.

6. generuoti naują bandomąjį tašką $y^t = x^t + z^t$,

7. **done**,

8. apskaičiuoti tikslo funkcijos reikšmę naujame taške $\eta_{t+1} = f(y^t)$.

Taikyti Metropolio kriterijų taškui x^{t+1} patvirtinti:

9. generuoti atsitiktinį skaičių κ, tolygiai pasiskirsčiusį intervale [0,1], ir skaičiuoti taško y^t patvirtinimo tikimybę

$$P(y^t, x^t, T_t) = \min\left\{1,\quad e^{\frac{\eta_t - \eta_{t+1}}{T_t}}\right\}.$$

10. **If** $\kappa \le P\left(y^t, x^t, T_t\right)$ **then** $x^{t+1} = y^t$

else liekame einamajame taške $x^{t+1} = x^t$ ir $\eta_{t+1} = \eta_t$.

11. Perskaičiuoti k geriausių pasiektų tikslo funkcijos reikšmių masyvą $\eta^t = \left(\eta_{(0)}, \dots, \eta_{(k)}\right)$,

12. $t = t + 1$, atnaujinti temperatūrą T_{t+1} ir aplinkos gylį ρ_{t+1}.

13. **done.** ◊

Yra žinoma nemažai darbų SA algoritmo konvergavimui tirti. Yra įrodyta, kad atitinkamai reguliuojant temperatūros T_t ir aplinkos gylio ρ_t parametrus, galima užtikrinti SA konvergavimą į globalųjį minimumą (Gidas (1985), Granville et al. (1994), Yang (2000)). Tačiau konvergavimo greitis bei ribinis sprendinio skirstinys dar nėra ištirti.

2.3.3. Markovo tipo algoritmų taikymas nediferencijuojamoms funkcijoms optimizuoti

Optimizavimo uždavinių, kai tikslo funkcija yra nediferencijuojama, sprendimas yra aktuali teorinė ir praktinė problema. Inžinerinio projektavimo, vadybos uždaviniuose bei statistikoje, vertinant parametrus mažiausios absoliutinės paklaidos metodu, ir pan., dažnai sutinkamas nediferencijuojamų funkcijų optimizavimas. Todėl tolydžioms tikslo funkcijoms, kurios nėra diferencijuojamos arba kurių išvestines yra sudėtinga apskaičiuoti arba aproksimuoti, optimizuoti yra taikomi begradientiniai optimizavimo metodai (*derivative-free methods*) (Bagirov & Ugon (2006)).

Dažnai taikomų nediferencijuojamo optimizavimo algoritmų klasę sudaro funkcijų, tenkinančių Lipšico savybę, stochastiniai Markovo tipo optimizavimo algoritmai. Lipšico funkcijų klasė yra gana plati, apimanti daugelį tolydžiųjų funkcijų, nagrinėjamų matematiniame programavime. Be to, Lipšico funkcijoms būdingos svarbios savybės, kurios leidžia įvesti stochastinį gradientą ir sukurti konverguojančias skaičiuojamąsias procedūras.

Stochastinės aproksimacijos algoritmai buvo vieni pirmųjų pasiūlyti optimizavimo uždaviniams spręsti, kai tikslo funkcija yra nediferencijuojama (Kiefer & Wolfowitz (1952), Ермольев (1976), Михалевич et al. (1987) ir t.t.). Šių algoritmų efektyvumas žymia dalimi priklauso nuo naudojamo stochastinio gradiento aproksimavimo būdo ir remiasi stochastinio gradiento arba kvazigradiento idėja bei tam tikromis optimizavimo žingsnio reguliavimo taisyklėmis, užtikrinančiomis konvergavimą. Efektyvius skaitmeninius nediferencijuojamo optimizavimo algoritmus leidžia sukurti modeliuojamojo pokyčio stochastinės aproksimacijos (*Simultaneous Perturbation Stochastic Approximation* (SPSA)) metodas, kadangi pagal šį metodą stochastiniam gradientui įvertinti kiekvienoje iteracijoje pakanka žinoti tik vieną arba dvi tikslo funkcijos reikšmes.

Žinomi SPSA algoritmai remiasi tikslo funkcijos suglodinimu naudojant Bernulio modelį (Spall (1992)) arba suglodinimo operatorių, tolygiai pasiskirsčiusį daugiamačiame hiperkube (Михалевич et al. (1987)). Disertacijoje taip pat įvedamas SPSA algoritmas, kuriame suglodinimo operatorius yra Lipšico, bei įrodomas šio algoritmo konvergavimas Lipšico funkcijų klasėje (Bartkutė &

Sakalauskas (2004b)). Šis modelis leidžia išplėsti suglodinimo operatorių klasę, įjungiant į ją praktiniu požiūriu svarbių atvejų, pvz., gabalais tiesinius suglodinimo tankius, duomenų poligonus ir t. t. Be to, ši prielaida leidžia aproksimuoti tikslo funkciją dukart glodžiai diferencijuojamomis funkcijomis bei pritaikyti reikšmingų imčių metodą.

2.3.4. Markovo tipo algoritmų stabdymas

Dažniausiai taikoma Markovo tipo algoritmų stabdymo taisyklė: algoritmas yra vykdomas iš anksto nustatytą iteracijų skaičių. Tačiau tokia stabdymo taisyklė nėra efektyvi, nes iš anksto yra sunku parinkti iteracijų skaičių, kuris garantuotų gauto rezultato paklaidą duotu tikslumu. Atlikus skaičiavimus gali pasirodyti, kad užsiduoto iteracijų skaičiaus nepakako pakankamai tiksliai įvertinti optimizavimo uždavinio sprendinį arba, atvirkščiai, iteracijų buvo atlikta per daug.

Pasinaudojus informacija, gauta optimizavimo metu ir pritaikius teorinio konvergavimo tyrimo išvadas tos informacijos analizei, galima sudaryti kitas optimizavimo algoritmo, aprašomo Markovo grandine, stabdymo taisykles. Konvergavimo tyrimo metu, kaip jau buvo minėta, yra sprendžiami tokie uždaviniai:

- sudaryto metodo konvergavimo įrodymas;
- konvergavimo greičio nustatymas;
- asimptotinio konvergavimo į ribinį dėsnį tyrimas.

Pritaikius šių uždavinių rezultatus stabdymo taisyklei sudaryti, galima pasinaudoti sprendinio asimptotiniu konvergavimu į normalinį dėsnį. Pavyzdžiui, yra žinoma, kad jei stochastinės aprokasimacijos algoritmo žingsnio ilgis yra reguliuojams tokiu būdu: $\rho_t = \frac{a}{t^\beta}$, tai vektorius $t^{\frac{\beta}{2}} \cdot \left(x^t - x^*\right), 0 < \beta \le 1$ yra asimptotiškai pasiskirstęs pagal daugiamatį Gauso dėsnį su nuliniu vidurkio vektoriumi ir tam tikra kovariacijų matrica Σ^* (Kushner & Yin (2003)). Kadangi optimizavimo uždavinio sprendinys x^* bei matrica Σ^* iš anksto nėra žinomi, tai autoriai, taikantys asimptotinį sprendinio normališkumą stabdymo taisyklei sudaryti, dažniausiai bando šiuos parametrus aproksimuoti pasinaudodami optimizavimo metu gautais rezultatais. Pavyzdžiui, ieškomas sprendinys yra aproksimuojamas optimizavimo metu gautų sprendinių vidurkiu: $x^{t,*} = \frac{\sum_{i=j_t}^{t} x^t}{t - j_t + 1}$, o matrica Σ^* imties kovariacijų matrica: $\Sigma^{t,*} = \frac{1}{t - j_t - n + 1} \cdot \sum_{i=j_t}^{t} \left(x^i - x^{t,*}\right)^T \cdot \left(x^i - x^{t,*}\right)$, čia j_t parenkamas taip, kad sprendiniai, naudojami imties vidurkiui ir kovariacijų matricai įvertinti, būtų pakankamai arti optimalaus sprendinio, o šių sprendinių būtų pakankamai daug, kad būtų galima pasinaudoti asimptotiškumu. Toliau įvedama pasikliovimo sritis $D_\gamma = \left\{ x : \frac{t^{1-\beta} \cdot \left(x - x^{t,*}\right) \cdot \left(\Sigma^{t,*}\right)^{-1} \cdot \left(x - x^{t,*}\right)^T}{n} \leq Fish_\gamma \right\}$. Algoritmas yra stabdomas, jei iteracijos taškas x^t priklauso šiai sričiai. Šis metodas buvo taip pat pritaikytas modeliuojamojo pokyčio stochastinės aproksimacijos algoritmui stabdyti (Hutchison & Spall (2005)). Nesunku pastebėti, kad $x^{t,*} \underset{t \to \infty}{\to} x^*$ ir $\Sigma^{t,*} \underset{t \to \infty}{\to} \Sigma^*$, tačiau konvergavimo greitis yra labai lėtas, todėl tenka atlikti gana didelį iteracijų skaičių. Kita vertus, nėra aišku, kaip parinkti indeksą j_t. Todėl šis algoritmo stabdymo metodas yra labiau įdomus teoriniu požiūriu, kuriant baigtinius optimizavimo algoritmus.

Kuriant algoritmo stabdymo taisykles galima pasinaudoti informacija, gaunama iš geriausių optimizavimo metu pasiektų tikslo funkcijos reikšmių. Bendri šios informacijos taikymo algoritmui stabdyti aspektai yra panagrinėti skyrelyje 2.2.3.

Tegul Markovo tipo optimizavimo algoritmu yra sprendžiamas uždavinys (2.1) Lipšico funkcijų klasėje (2.2). Nemažinant bendrumo galima laikyti, kad optimizavimo metu yra konstruojamos taškų bei funkcijos reikšmių sekos: $x_0, x_1, ..., x_t, ...$ ir $\mathrm{H} = \{\eta_0, \dots, \eta_N\}$, čia $\eta_t = f(x_t), t = 0, 1, \dots$. Daug svarbios informacijos apie tikslo funkcijos minimalią reikšmę ir esamo sprendinio atstumą iki optimumo teikia $k+1$ geriausių funkcijos reikšmių

$$\eta_{(0)}, \dots, \eta_{(k)} \tag{2.16}$$

paimtų iš išrikiuotos sekos $\eta_{(0)} \leq \eta_{(1)} \leq \dots \leq \eta_{(N)}$ (Жиглявский & Жилинскас (1991), David & Nagaraja (2003)). Seka H yra atsitiktinė, todėl šios reikšmės vadinamos *pozicinėmis statistikomis*. Žinomi kelių autorių darbai, kuriuose analizuojamas pozicinių statistikų taikymas funkcijos minimalios (maksimalios) reikšmės arba funkcijos ekstremumo padėčiai įvertinti. Šios problemos nagrinėjimo pradininku galima laikyti J. Mockų (1967), kuris pritaikė statistinių išvadų teoriją optimizavimo algoritmų optimalumui tirti, vėliau teoriniai nagrinėjimai, paremti ekstremaliųjų reikšmių teorija, buvo išplėtoti A. Žilinsko ir A. Zhigljavsky (1991). H. Chen (1996) geriausias funkcijos reikšmes pritaikė maksimumo taškui nustatyti. Aiex et al. (2002) kompiuterinio modeliavimo būdu tyrė kombinatorinio optimizavimo algoritmuose žingsnių skaičių, reikalingą nustatytai funkcijos reikšmei pasiekti, ir parodė, kad šis žingsnių skaičius gali būti modeliuojamas eksponentiniu skirstiniu.

Toliau nagrinėsime pozicinių statistikų tinkamumą įvairiems determinuotų Lipšico funkcijų Markovo tipo stochastinio optimizavimo algoritmams stabdyti. Kadangi stochastinė aproksimacija yra vienas iš geriausiai žinomų ir ištirtų stochastinės optimizacijos metodų, disertacijoje pagrindinis dėmesys yra skiriamas šiems metodams sudaryti, konvergavimo analizei bei stabdymo taisyklėms sukurti. Gauti rezultatai gali būti nesunkiai apibendrinami modeliuojamojo atkaitinimo bei kitiems Markovo tipo algoritmams. Disertacijoje pristatysime metodą minimaliai tikslo funkcijos reikšmei bei jos pasikliautinajam intervalui įvertinti, naudojant ekstremaliųjų reikšmių teoriją nepriklausomiems vienodai pasiskirsčiusiusiems atsitiktiniams dydžiams. Sudaryto metodo tinkamumą tirsime kompiuterinio modeliavimo būdu, kurio rezultatai yra pateikiami 4 skyriuje.

2.4. IŠVADOS

1. Markovo tipo stochastiniai algoritmai yra gerai žinomi ir plačiai taikomi determinuotojo ir stochastinio programavimo uždaviniams spręsti. Tačiau aktuali teoriniu ir praktiniu požiūriu šių algoritmų stabdymo problema yra beveik neištirta.
2. Dažniausiai yra taikoma Markovo tipo algoritmų stabdymo taisyklė: algoritmas yra vykdomas iš anksto nustatytą iteracijų skaičių. Tačiau tokia stabdymo taisyklė nėra efektyvi, nes iš anksto sunku parinkti iteracijų skaičių, kuris garantuotų gauto rezultato paklaidą duotu tikslumu.
3. Markovo tipo algoritmų stabdymo taisyklės sudaromos konverguojančioms sekoms, t.y. jų sudarymas bei optimalumo tyrimas remiasi algoritmo konvergavimo tyrimo rezultatais. Šiuo metu yra nustatytos daugelio stochastinės aproksimacijos metodų konvergavimo sąlygos bei asimptotinio konvergavimo į normalinį dėsnį sąlygos. Taip pat yra nustatytos kai kurių modeliuojamojo atkaitinimo algoritmo modifikacijų konvergavimo sąlygos.
4. Geriausių optimizavimo metu pasiektų tikslo funkcijos reikšmių teikiama informacija gali būti pritaikyta minimaliai tikslo funkcijos reikšmei bei jos pasikliautinajam intervalui įvertinti. Šių reikšmių skirstinys, reikalingas šiems įverčiams gauti, teoriškai dar nėra ištirtas.

3. NEDIFERENCIJUOJAMO OPTIMIZAVIMO STOCHASTINIŲ METODŲ KONVERGAVIMO TYRIMAS

3.1. ĮVADAS

Stochastinių Markovo tipo optimizavimo algoritmų konvergavimo tyrimas teikia daug informacijos apie patį algoritmą ir yra svarbus etapas sudarant algoritmo stabdymo taisykles. Kadangi stochastinės aproksimacijos metodai apima plačią Markovo tipo algoritmų klasę detaliau panagrinėsime šios klasės algoritmų konvergavimo aspektus. Literatūroje pastaruoju metu yra skiriama nemažai dėmesio modeliuojamojo pokyčio stochastinės aproksimacijos metodams (SPSA) (http://www.jhuapl.edu/SPSA). Kiekviename šių metodų žingsnyje (2.14) stochastiniam gradientui įvertinti pakanka apskaičiuoti tik vieną arba dvi tikslo funkcijos reikšmes. Šie metodai yra ištirti ir pritaikyti diferencijuojamoms funkcijoms, apskaičiuojamoms su triukšmu arba be jo, optimizuoti. Tačiau šių metodų taikymas nediferencijuojamoms funkcijoms optimizuoti dar nėra pakankamai gerai išnagrinėtas. Be to, pradiniame šių metodų variante (Spall (1992)) yra įvesti ribojimai, kurie stipriai sumažina šių metodų modifikacijų galimybes. Pavyzdžiui, šis variantas nėra taikytinas nediferencijuojamoms funkcijoms optimizuoti. Disertacijoje įvesta ir nagrinėjama SPSA modifikacija su Lipšico suglodinimo operatoriumi, kuri leidžia išvengti šių ribojimų ir sukurti begradientinius optimizavimo metodus Lipšico funkcijoms optimizuoti.

SPSA algoritmai dažniausiai nagrinėjami diferencijuojamų funkcijų klasėje naudojant įvairius suglodinimo operatorius. J. Spall (1992) pasiūlė SPSA metodą su Bernulio suglodinimo modeliu diferencijuojamoms funkcijoms su triukšmu arba be jo optimizuoti ir parodė šio metodo efektyvumą lyginant su standartine baigtinio skirtumo aproksimacija. Михалевич et al. (1987) pasiūlė metodą n-mačiame hiperkube tolygiai suglodintoms funkcijoms optimizuoti, kurį galima priskirti SPSA metodams su tolygiai aprėžta suglodinimo operatoriaus tankio funkcija.

SPSA metodų konvergavimo sąlygos, kai funkcija skaičiuojama su triukšmu, yra tiriamos daugelio autorių. Kai kurių autorių darbuose įvedamos tris kartus diferencijuojamos tikslo funkcijos (Spall (1992), Gerencsér (1999)), silpnesnės konvergavimo sąlygos yra nustatytos Sadegh (1997), Fu & Hill (1997), Граничин & Поляк (2003). Chen et al. (1999) tyrė SPSA metodo konvergavimą ir savo darbe įvedė diferencijuojamas tikslo funkcijas, kurių gradientas tenkina Lipšico sąlygą. J. Spall (1992) įrodė SPSA asimptotinį konvergavimą į normalinį dėsnį diferencijuojamoms funkcijoms, skaičiuojamoms su triukšmu. Visais išvardintais atvejais yra įrodytas konvergavimas diferencijuojamoms funkcijoms bei iškiloms funkcijoms, skaičiuojamoms su triukšmu. Šioms funkcijoms yra gautas SPSA konvergavimo greičio įvertis $E\left\|x^k - x^*\right\|^2 = O\left(\frac{1}{t^\gamma}\right)$, čia rodiklis γ, gautas įvairių autorių, yra skirtingas ir didžiausia jo pasiekta reikšmė yra $\frac{2}{3}$.

Kaip jau buvo minėta SPSA algoritmai su Lipšico suglodinimo operatoriumi literatūroje nėra nagrinėti. Įvairių suglodinimo operatorių palyginimo arba geriausio suglodinimo operatoriaus parinkimo uždaviniai taip pat dar nėra ištirti.

Nediferencijuojamų funkcijų atveju Kushner & Yin (2003), Михалевич et al. (1987) suformulavo teoremą apie Kiefer–Wolfowitz (Kiefer & Wolfowitz (1952)) tipo stochastinės aproksimacijos algoritmų konvergavimą, kai tikslo funkcija skaičiuojama be triukšmo. Nediferencijuojamų funkcijų, apskaičiuojamų be triukšmo, stochastinės aproksimacijos konvergavimo greitis išvis nėra nustatytas.

Žinomi stochastinės aproksimacijos algoritmų tyrimų rezultatai nėra pakankamai bendri, kad galiotų visiems SPSA algoritmams. Disertacijoje suformuluota ir įrodyta bendresnė teorema apie SPSA metodo konvergavimą nediferencijuojamoms funkcijoms, skaičiuojamoms be triukšmo.

Šiame skyriuje yra išnagrinėti SPSA algoritmai Lipšico funkcijų, skaičiuojamų be triukšmo, klasėje, kai modeliuojamas pokytis yra absoliučiai tolydus ir aprašomas tankio funkcija, tenkinančia Lipšico sąlygą. Gradiento įverčiui gauti yra pasiūlyta naudoti tikėtinumo santykio (*Likelihood Ratio* (LR)) metodą (Rubinstein & Shapiro (1993), Ermoliev et al. (1995), Sakalauskas (2002)). Naudojant šį metodą, tikslo funkcija yra suglodinama su Lipšico suglodinimo operatoriumi, po to suglodintos funkcijos gradientas yra išreiškiamas vidurkiu, kuriam įvertinti pakanka apskaičiuoti vieną arba dvi tikslo funkcijos reikšmes. Tokių algoritmų nagrinėjimas leido aproksimuoti minimizuojamas funkcijas du kartus diferencijuojamomis funkcijomis, nustatyti optimizavimo algoritmo konvergavimo sąlygas ir įrodyti konvergavimo greitį $E\left\|x^k - x^*\right\|^2 = O\left(\frac{1}{t^\gamma}\right)$, kai $1 \le \gamma < 2$. Nustatytos sudaryto SPSA metodo konvergavimo sąlygos yra gana paprastos ir padeda parinkti algoritmo žingsnio parametrus, leidžiančius gauti konverguojančias sekas, reikalingas stabdymo taisyklei (2.11) ir minimalios funkcijos reikšmės (2.12) bei pasikliautinojo intervalo (2.13) įverčiams sukurti.

Visi šiame skyriuje pateikiami rezultatai buvo pristatyti konferencijose: *Lietuvos matematikų draugijos XLV konferencija* (2004, LŽŪU, Kaunas), tarptautiniame seminare (*Workshop on Challenges of Continuous Optimization in Theory and Applications*, 2004, Rodos, Graikija) ir „*Informacinės technologijos-2005*“ (2005, KTU, Kaunas) bei publikuoti šiuose moksliniuose leidiniuose: Bartkutė & Sakalauskas (2004b), Bartkutė & Sakalauskas (2005c) bei Bartkutė & Sakalauskas (2007a).

3.2. LIPŠICO FUNKCIJŲ STOCHASTINIS DIFERENCIJAVIMAS

3.2.1. Lipšico funkcijų suglodinimas

Kai tikslo funkcijos yra neiškilos neglodžios funkcijos, optimizavimo uždaviniai yra labai sudėtingi. Šiems uždaviniams spręsti yra įvedamos apibendrintai diferencijuojamos funkcijos ir Lipšico funkcijos (žr. priedą 1).

Optimizavimo algoritmams Lipšico funkcijų klasėje kurti yra taikomas funkcijos suglodinimas. Šiuo būdu galima aproksimuoti suglodintų funkcijų gradientus jų stochastiniais įverčiais ir sukurti metodus, kur nebūtina skaičiuoti gradientų (Wazan (1969), Ермольев (1976), Михалевич et al. (1987), Spall (2003), Kushner & Yin (2003)). Bendru atveju ši idėja atrodo taip: yra įvedama funkcijų seka

$$\overline{f}\left(x,\sigma\right)=Ef\left(x+\sigma\xi\right), \tag{3.1}$$

konverguojanti į $f(x)$, kai $\sigma\to 0$, čia ξ yra atsitiktinis vektorius, $\sigma\geq 0$ suglodinimo parametras. Tada (2.1) sprendžiamas minimizuojant šios sekos funkcijas, atitinkamu būdu keičiant suglodinimo parametrą σ.

Suglodintą funkciją galima užrašyti daugialypiu integralu (Rubinstein & Shapiro (1993), Ermoliev et al (1995), Sakalauskas (2002)):

$$\overline{f}(x,\sigma)=\int\limits_{\Re^n} f(x+\sigma y)\cdot p(y)dy=\frac{1}{\sigma^n}\int\limits_{\Re^n} f(y)\cdot p\left(\frac{y-x}{\sigma}\right)dy. \tag{3.2}$$

Priklausomai nuo suglodinimo operatoriaus galima gauti skirtingas suglodintas funkcijas. Panagrinėsime kelis funkcijos suglodinimo pavyzdžius.

Pavyzdys 3.1. Turime funkciją $f(x)=|x|$. Šios funkcijos apibendrintas gradientas yra (žr. priedą 1):

$$\overline{g}(x)=\begin{cases}1, & x>0,\\ [-1;1], & x=0,\\ -1, & x<0.\end{cases}$$

a) jei suglodinimo tankis $p(\xi)=\begin{cases}\frac{1}{2}, & |\xi|\leq 1,\\ 0, & \text{kitais atvejais}\end{cases}$, tai pagal (3.2) nesunku apskaičiuoti, kad suglodinta funkcija yra

$$\overline{f}(x,\sigma)=\begin{cases}\frac{1}{2}\left(\frac{x^2}{\sigma}+\sigma\right), & |x|\leq\sigma \text{ ir } \sigma\neq 0,\\ |x|, & \text{kitu atveju}\end{cases},$$

o suglodintos funkcijos gradientas:

$$\overline{g}(x,\sigma)=\begin{cases}\dfrac{x}{\sigma}, & |x|\le\sigma \text{ ir } \sigma\ne 0,\\ sign(x), & \text{kitu atveju.}\end{cases}$$

b) jei suglodinimo tankis yra Lipšico, pvz.:

$$p(y)=\begin{cases}1-|y|, & if\ |y|\le 1,\\ 0, & if\ |y|>1,\end{cases},$$

tai suglodinta funkcija yra

$$\overline{f}(x,\sigma)=\begin{cases}\dfrac{\sigma}{3}+\dfrac{x^2\cdot(x+3\sigma)}{3\sigma^2}, & x<0 \text{ ir } x>\text{-}\sigma,\\ -x, & x<\text{-}\sigma,\\ x, & x>\sigma,\\ \dfrac{\sigma}{3}-\dfrac{x^2\cdot(x-3\sigma)}{3\sigma^2}, & x>0 \text{ ir } x<\sigma,\end{cases}$$

o jos gradientas:

$$\overline{g}(x,\sigma)=\begin{cases}\dfrac{2x}{\sigma}+\dfrac{x^2}{\sigma^2}, & x<0 \text{ ir } x>\text{-}\sigma,\\ -1, & x<-\sigma,\\ 1, & x>\sigma,\\ \dfrac{2x}{\sigma}-\dfrac{x^2}{\sigma^2}, & x>0 \text{ ir } x<\sigma.\end{cases}$$

Ši suglodinta funkcija yra diferencijuojama du kartus. Jos antros eilės išvestinė yra:

$$\frac{d^2\overline{f}(x,\sigma)}{dx^2}=\begin{cases}\dfrac{2}{\sigma}+\dfrac{2x}{\sigma^2}, & x<0 \text{ ir } x>\text{-}\sigma,\\ 0, & x<-\sigma,\\ 0, & x>\sigma,\\ \dfrac{2}{\sigma}-\dfrac{2x}{\sigma^2}, & x>0 \text{ ir } x<\sigma.\end{cases}$$

c) jei suglodinimo tankis yra Gauso

$$p(y)=\left(\frac{1}{2\pi}\right)^{\frac{1}{2}}\exp\left(-\frac{|y|^2}{2}\right),$$

tai suglodinta funkcija yra

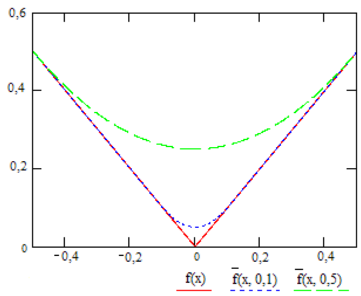
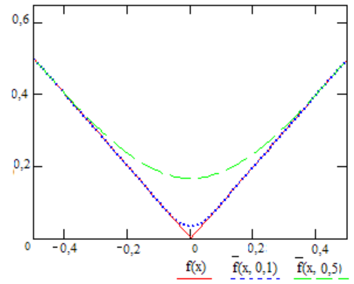
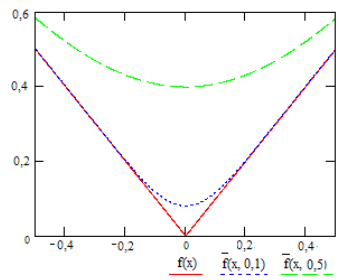
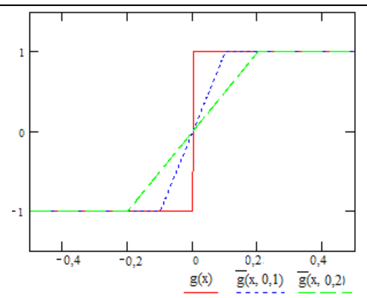
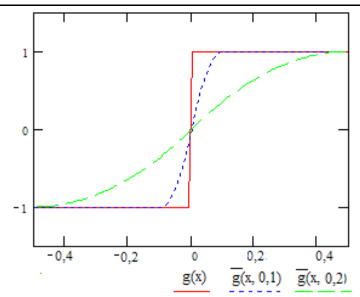
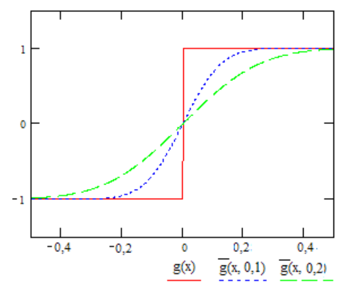
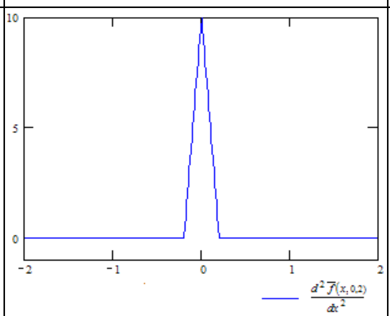
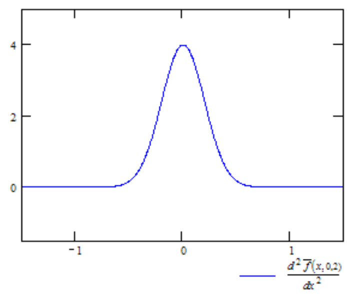
$$\overline{f}(x,\sigma)=\begin{cases}x\cdot\Phi\left(\dfrac{x}{\sigma}\right)+\sigma\cdot\sqrt{\dfrac{2}{\pi}}\cdot e^{-\frac{x^2}{2\sigma^2}}, & \sigma\ne 0,\\ |x|, & \textit{priešingu atveju},\end{cases}$$

ir jos gradientas yra $\overline{g}(x,\sigma)=\Phi\left(\frac{x}{\sigma}\right)$, kai $\Phi(y)=\frac{1}{\sqrt{2\pi}}\cdot\int\limits_{-y}^{y} e^{-\frac{t^2}{2}}\,dt$. Tokiu būdu suglodinta funkcija yra diferencijuojama daug kartų. Pavyzdžiui, jos antros eilės išvestinė yra

$$\frac{d^2\overline{f}(x,\sigma)}{dx^2}=\frac{1}{\sqrt{\pi}}\cdot\frac{\sqrt{2}\cdot e^{-\frac{x^2}{2\sigma^2}}}{\sigma}$$

Lentelėje 3.1 yra pateikti funkcijos, nagrinėjamos pavyzdyje 3.1, jos apibendrinto gradiento ir suglodintos funkcijos pirmos ir antros eilės išvestinių grafikai. Suglodintos funkcijos antros eilės išvestinės grafikai iliustruoja antros eilės išvestinės egzistavimą kai suglodinimo tankis yra Lipšico arba Gauso.

Lentelė 3.1. Funkcija $f(x)=|x|$ ir jos apibendrintas gradientas, suglodinta funkcija bei jos pirmos ir antros eilės išvestinės priklausomai nuo pasirinkto suglodinimo tankio.

	Suglodinimo tankis $p(\xi)=\begin{cases}\frac{1}{2}, & \lvert\xi\rvert\le 1,\\ 0, & \text{kitais atvejais.}\end{cases}$	Suglodinimo tankis Lipšico	Suglodinimo tankis Gauso
Pradinė ir suglodinta funkcija			
Pradinės ir suglodintos funkcijų apibendrinti gradientai			
Suglodintos funkcijos antros eilės išvestinė	neegzistuoja		

3.2.2. Suglodintų funkcijų diferencijavimas

Tegul tikslo funkcija ir suglodinimo tankis tenkina Lipšico sąlygą:

$$f(y) \in \mathcal{L}_1,\ p(y) \in \mathcal{L}_1.$$

Nesunku pastebėti, kad $\overline{f}(x,\sigma)$ tenkina Lipšico sąlygą pagal σ:

$$\left|\overline{f}(x,\sigma_1) - \overline{f}(x,\sigma_2)\right| \le E\left|f(x+\sigma_1\xi) - f(x+\sigma_2\xi)\right| \le C \cdot K \cdot \left|\sigma_1 - \sigma_2\right|, \tag{3.3}$$

čia $C = E\|\xi\|$.

Tegul $\partial p(y)$ yra apibendrintas p gradientas. Nemažindami bendrumo galime laikyti, kad $\|\partial f(y)\| \le K$ ir $\|\partial p(y)\| \le K$, $y \in \Re^n$.

Apibrėžimas 3.1. (Ермольев (1976)). Tegul $\partial f(x)$ yra lokaliai Lipšico funkcijos apibendrintas gradientas (žr. priedą 1). Atsitiktinis vektorius $g(x,\xi) \in \Re^n$ yra vadinamas stochastiniu gradientu, jei $Eg(x,\xi) = \partial f(x)$, $E\|g(x,\xi)\|^2 < \infty$.

Panagrinėsime tikėtinumo santykio metodą ir mato keitimo procedūros taikymą suglodintai funkcijai diferencijuoti. Naudojant šį metodą yra įvedama tankio funkcija $\psi(y)$, kuri bendru atveju gali sutapti su $p(y)$. Suformuluosime lemą, kuri leidžia įvertinti suglodintos funkcijos gradientą bei jos stochastinį gradientą.

Lema 3.1. (Bartkutė & Sakalauskas (2007a)). *Tegul $f(x)$ yra Lipšico funkcija su konstanta K. Tarkime, kad Ψ yra absoliučiai tolydus matas su aprėžta tankio funkcija $\psi : \Omega \to \Re_+$, kurio nenulinių reikšmių sritis sutampa su P: $dom(\Psi) = \{y \mid \psi(y) > 0\} \equiv P$.*

Tegul $\int_P \|y\|^2 \cdot \frac{\|\partial p(y)\|^2}{\psi(y)} dy = A < \infty$, $\int_P \|\partial p(y)\| \cdot dy = L < \infty$. Tada suglodintos funkcijos (3.2) gradientas gali būti išreiškiamas tokiu būdu:

$$\overline{g}(x,\sigma) = \frac{\partial \overline{f}(x,\sigma)}{\partial x} = E(g(x,\sigma,\xi)), \tag{3.4}$$

čia $g(x,\sigma,\xi)$ yra stochastinio gradiento įvertis:

$$g(x,\sigma,\xi) = \frac{(f(x+\sigma\xi) - f(x))}{\sigma} \cdot \frac{\partial p(\xi)}{\psi(\xi)}, \tag{3.5}$$

čia

$$E(\|g(x,\sigma,\xi)\|)^2 \le K^2 \cdot A. \tag{3.6}$$

Suglodintos funkcijos (3.2) hesianas:

$$V(x,\sigma)=\frac{\partial^2 \overline{f}(x,\sigma)}{\partial x^2}=\frac{1}{\sigma}\cdot E\left(\frac{\partial f(x+\sigma\xi)\cdot(\partial p(\xi))^T}{\psi(\xi)}\right), \tag{3.7}$$

čia $\|V(x,\sigma)\|\le\frac{K\cdot L}{\sigma}$.

Įrodymas.

▷ Iš Rademacherio teoremos (žr. priedą 1) seka, kad tankio funkcijos apibendrinto gradiento daugiareikšmiškumo taškų aibės Borelio matas lygus nuliui. Kadangi apibendrintas gradientas yra aprėžtas, tai išdiferencijavus antrąjį formulės (3.2) integralą ir pasinaudojus teorema 22.1 (teorema 22.1, Михалевич et al. (1987)) suglodintos funkcijos gradientą galime išreikšti vidurkio pavidalu:

$$\frac{\partial \overline{f}(x,\sigma)}{\partial x}=\frac{1}{\sigma^{n+1}}\cdot\int_{\Re^n} f(y)\cdot\partial p\left(\frac{y-x}{\sigma}\right)dy=\frac{1}{\sigma}\cdot\int_{\Re^n} f(x+\sigma y)\cdot\partial p(y)dy=$$

$$=\frac{1}{\sigma}\cdot\int_P f(x+\sigma y)\cdot\frac{\partial p(y)}{\psi(y)}\cdot\psi(y)dy=\frac{1}{\sigma}\cdot E\left(f(x+\sigma\xi)\cdot\frac{\partial p(\xi)}{\psi(\xi)}\right), \tag{3.8}$$

Toliau, diferencijuodami tapatybę x atžvilgiu $\int_{\Re^n} p(y)dy\equiv\int_{\Re^n} p(y+x)dy=1$, gausime tapatybę

$$\int_{\Re^n}\partial p(y+x)dy\equiv\int_{\Re^n}\partial p(y)dy=0,$$

iš kurios išplaukia, kad

$$\int_{\Re^n}\partial p(y)dy=\int_P\frac{\partial p(y)}{\psi(y)}\cdot\psi(y)dy=E\left(\frac{\partial p(\xi)}{\psi(\xi)}\right)=0. \tag{3.9}$$

Todėl (3.4) seka iš (3.8) ir (3.9).

Pastebėsime, kad suglodintos funkcijos gradientas yra aprėžtas $\|\overline{g}(x,\sigma)\|\le K$. Toliau,

$$E(g(x,\sigma,\xi))^2=E\left(\frac{(f(x+\sigma\xi)-f(x))}{\sigma}\cdot\frac{\partial p(\xi)}{\psi(\xi)}\right)^2=\int_{dom(\Psi)}\frac{(f(x+\sigma y)-f(x))^2}{\sigma^2}\cdot\frac{\|\partial p(y)\|^2}{\psi(y)}dy\le$$

$$\le K^2\cdot\int_{dom(\Psi)}\|y\|^2\cdot\frac{\|\partial p(y)\|^2}{\psi(y)}dy\le K^2\cdot A.$$

Diferencijuodami (3.8) galutinę išraišką x atžvilgiu, gausime suglodintos funkcijos hesiano įvertį:

$$\|V(x,\sigma)\|=\frac{1}{\sigma}\cdot\left\|E\left(\frac{\partial f(x+\sigma\xi)\cdot(\partial p(\xi))^T}{\psi(\xi)}\right)\right\|\le\frac{1}{\sigma}\cdot E\frac{\left\|\partial f(x+\sigma\xi)\cdot(\partial p(\xi))^T\right\|}{\psi(\xi)}\le$$

$$\le\frac{1}{\sigma}\cdot E\frac{\|\partial f(x+\sigma\xi)\|\cdot\|\partial(p(\xi))\|}{\psi(\xi)}\le\frac{K\cdot L}{\sigma}.$$

◁

Suformuluosime lemą, leidžiančią funkciją $f(x)$ aproksimuoti suglodinta funkcija $\overline{f}(x,\sigma)$, kai $\sigma \to 0$.

Lema 3.2. (Bartkutė & Sakalauskas (2007a)). *Tegul* $f(x)$ *yra Lipšico funkcija su konstanta K,* $x \in \Re^n$*, be to kiekvienam* $\varepsilon > 0$ *galime rasti tokį* $\delta > 0$ *kad su visais y,* $\|y - x\| \le 2\delta$*, kur* $f(x)$ *yra diferencijuojama įprastine prasme, nelygybė* $\min_{z \in \partial f(x)} \left\| \frac{\partial f(y)}{\partial y} - z \right\| \le \frac{\varepsilon}{2}$ *galioja tolygiai x atžvilgiu. Tegul dar* $\sigma_i \to 0$ *ir* $E\|\xi\|^2 < \infty$*. Tuomet bet kuriam* $\varepsilon > 0$ *galima rasti tokį k, kad visiems* $i \ge k$ *ir y,* $\|y - x\| \le \delta$*, galioja nelygybė* $\min_{z \in \partial f(x)} \left\| \overline{g}(y, \sigma_i) - z \right\| \le \varepsilon$ *tolygiai x atžvilgiu.*

Įrodymas.

$\triangleright$ Žymėsime $I(A)$ aibės A indikatorių. Tegul $z_1 \in \partial f(x)$ toks, kad

$$\left\| E\left(\left(\partial f(y + \sigma_i \xi) - z_1\right) \cdot I\left(\|\xi\| \le \frac{\delta}{\sigma_i} \right) \right) \right\| = \min_{z \in \partial f(x)} \left\| E\left(\left(\partial f(y + \sigma_i \xi) - z\right) \cdot I\left(\|\xi\| \le \frac{\delta}{\sigma_i} \right) \right) \right\|,$$

o k yra toks, kad $\sigma_i \le \frac{\delta}{2} \cdot \sqrt{\frac{\varepsilon}{K \cdot E\|\xi\|^2}}$ visiems $i \ge k$. Tuomet turime, kad

$$E\left(\left\|\partial f(y + \sigma_i \xi) - z_1\right\| \cdot I\left(\|\xi\| > \frac{\delta}{\sigma_i} \right) \right) \le 2K \cdot P\left(\|\xi\| > \frac{\delta}{\sigma_i} \right) \le \frac{2 \cdot K \cdot \sigma_i^2 \cdot E\|\xi\|^2}{\delta^2} \le \frac{\varepsilon}{2}. \quad (3.10)$$

Tegul $H : \Re^n \to \partial f(x)$ yra selektorius, tai yra tokia mati vienareikšmė funkcija, kad $\left\| \frac{\partial f(y)}{\partial y} - H(y) \right\| = \min_{z \in \partial f(x)} \left\| \frac{\partial f(y)}{\partial y} - z \right\|$ (žr. teoremą apie selektorius Kuratowski (1968)). Tuomet, kai $\|y - x\| \le \delta$, seka, kad

$$\min_{z \in \partial f(x)} \left\| \overline{g}(y, \sigma_i) - z \right\| \le \left\| E\left(\partial f(y + \sigma_i \xi) - z_1\right) \right\| =$$

$$= \left\| E\left(\left(\partial f(y + \sigma_i \xi) - z_1\right) \cdot I\left(\|\xi\| \le \frac{\delta}{\sigma_i} \right) \right) + E\left(\left(\partial f(y + \sigma_i \xi) - z_1\right) \cdot I\left(\|\xi\| > \frac{\delta}{\sigma_i} \right) \right) \right\| \le$$

$$\le \min_{z \in \partial f(x)} \left\| E\left(\left(\partial f(y + \sigma_i \xi) - z\right) \cdot I\left(\|\xi\| \le \frac{\delta}{\sigma_i} \right) \right) \right\| + \left\| E\left(\left(\partial f(y + \sigma_i \xi) - z_1\right) \cdot I\left(\|\xi\| > \frac{\delta}{\sigma_i} \right) \right) \right\| \le \varepsilon \text{ pagal (3.10) ir}$$

$$\frac{\varepsilon}{2} \ge E\left(\left(\min_{z \in \partial f(x)} \left\| \left(\partial f(y + \sigma_i \xi) - z\right) \right\| \right) \cdot I\left(\|\xi\| \le \frac{\delta}{\sigma_i} \right) \right) = E\left\| \left(\partial f(y + \sigma_i \xi) - H(y + \sigma_i \xi)\right) \cdot I\left(\|\xi\| \le \frac{\delta}{\sigma_i} \right) \right\| \ge$$

$$\ge \left\| E\left(\left(\partial f(y + \sigma_i \xi) - E\left(H(y + \sigma_i \xi) \middle| \|\xi\| \le \frac{\delta}{\sigma_i} \right) \right) \cdot I\left(\|\xi\| \le \frac{\delta}{\sigma_i} \right) \right) \right\| \ge$$

$$\geq \min_{z\in\partial f(x)}\left(\left\| E\left(\left(\partial f\left(y+\sigma_i\xi\right)-z\ \right)\cdot I\left(\|\xi\|\leq\frac{\delta}{\sigma_i}\right)\right)\right\|\right)$$ kadangi $E\left(H\left(y+\sigma_i\xi\right)\middle|\|\xi\|\leq\frac{\delta}{\sigma_i}\right)\in\partial f(x)$ dėka apibendrinto gradiento atvaizdžio iškilumo. ◁

Sekmuo 3.1. Jei $\left\{x^t\right\}\to x$, $\overline{g}\left(x^t,\sigma_t\right)\to g$, $\sigma_t\to 0$, $t\to\infty$, tai $g\in\partial f(x)$.

Šios dvi lemos leidžia suglodintos funkcijos gradientus išreikšti tūriniais integralais, kurie gali būti skaičiuojami Monte-Karlo metodu, ir jiems skaičiuoti nereikia žinoti tikslo funkcijos apibendrinto gradiento $\partial f(x)$. Pastebėsime, kad formulė (3.5) gradiento įverčiui skaičiuoti reikalauja tik dviejų funkcijos reikšmių. Šie rezultatai praplečia žinomų SPSA metodų klasę.

Panagrinėsime kai kurias stochastinio diferencijavimo formules, išplaukiančias iš aukščiau įrodytų lemų. Jeigu $\psi(y)=p(y)$, tai stochastinis gradientas gali būti įvertinamas tokiu būdu:

$$g(x,\sigma,\xi)=\frac{\left(f(x+\sigma\xi)-f(x)\right)\cdot\partial\ln\left(p(\xi)\right)}{\sigma}, \tag{3.11}$$

čia tankio funkcijos logaritmo apibendrintas gradientas apibrėžiamas taip:

$$\partial\ln\left(p(y)\right)=\begin{cases}\dfrac{\partial p(y)}{p(y)}, & if\ \ y\in P\\ 0, & if\ \ y\notin P.\end{cases}$$

Taip pat suglodintos funkcijos gradientą galima išreikšti, diferencijuojant formulės (3.2) pirmąjį integralą, kurioje reikia žinoti tikslo funkcijos apibendrintą gradientą :

$$\overline{g}(x,\sigma)=\frac{\partial\,\overline{f}(x,\sigma)}{\partial x}=E\partial f\left(x+\sigma\xi\right). \tag{3.12}$$

Įvertindami pastarojoje formulėje tikslo funkcijos gradientą skirtuminiu būdu, gausime stochastinio integralo įverčius, kuriems apskaičiuoti tereikia žinoti tikslo funkcijos reikšmes kai kuriuose taškuose (žr. 3.17).

Įvesdami įvairius suglodinimo operatorius galime gauti skirtingus suglodintos funkcijos gradiento įvertinimo būdus.

Pavyzdys 3.2. Tegul suglodinimo tankis yra pasiskirstęs vienetiniame rutulyje su tankio funkcija

$$p(y)=\begin{cases}W_n\cdot\left(1-\|y\|\right), & if\ \ \|y\|\leq 1,\\ 0, & if\ \ \|y\|>1.\end{cases}, \tag{3.13}$$

čia $W_n=\dfrac{1}{\int\limits_{\|y\|\leq 1}\left(1-\|y\|\right)dy}=\dfrac{n\cdot(n+1)\cdot\Gamma\left(\frac{n}{2}\right)}{(2\pi)^{\frac{n}{2}}}$.

▷

$$\int_{\|y\|\le 1}(1-\|y\|)dy=\int_{\|y\|\le 1}dy-\int_{\|y\|\le 1}\|y\|dy=\begin{bmatrix}\|y\|=r\\ dy=r^{n-1}dr\end{bmatrix}=\frac{(2\pi)^{\frac{n}{2}}}{\Gamma\left(\frac{n}{2}\right)}\left(\int_0^1 r^{n-1}dr-\int_0^1 r^n dr\right)=\frac{(2\pi)^{\frac{n}{2}}}{\Gamma\left(\frac{n}{2}\right)}\cdot\left(\frac{1}{n}-\frac{1}{n+1}\right)$$

$\triangleleft$

Tada suglodintos funkcijos gradientas gali būti išreiškiamas tokiu būdu

$$\begin{aligned}\overline{g}(x,\sigma)&=\frac{1}{\sigma}\int_{\Re^n}f(x+\sigma\cdot y)\partial p(y)dy=\frac{W_n}{\sigma}\int_{\|y\|\le 1}f(x+\sigma\cdot y)\cdot\frac{y}{\|y\|}dy=\\&=\frac{W_n}{\sigma}\int_{\|y\|\le 1}\left(f(x+\sigma\cdot y)-f(x)\right)\cdot\frac{y}{\|y\|}dy=E\left(\frac{\left(f(x+\sigma\xi)-f(x)\right)\cdot\xi}{\sigma\cdot\|\xi\|}\right)\end{aligned}\tag{3.14}$$

kai atsitiktinis vektorius ξ yra tolygiai pasiskirstęs vienetiniame rutulyje $\psi(y)=\begin{cases}\frac{1}{V_n}, & if\ \|y\|\le 1,\\ 0, & if\ \|y\|>1.\end{cases}$,

čia V_n yra n-mačio rutulio tūris.

Kadangi tankio funkcija (3.13) yra Lipšico, suglodinta funkcija yra diferencijuojama du kartus (*Lema 3.1*). Kai n=1 šios funkcijos ir jos išvestinių grafikai yra pateikiami *lentelėje 3.1*.

Pavyzdys 3.3. Jei suglodinimo tankis yra Gauso

$$p(y)=\left(\frac{1}{2\pi}\right)^{\frac{n}{2}}\exp\left(-\frac{\|y\|^2}{2}\right),$$

tada pagal (3.4) suglodintos funkcijos gradientas gali būti išreiškiamas tokiu būdu

$$\overline{g}(x,\sigma)=E\left(\frac{\left(f(x+\sigma\xi)-f(x)\right)\cdot\xi}{\sigma}\right),$$

čia ξ yra pasiskirstęs pagal daugiamatį standartinį Gauso dėsnį.

Tokiu būdu suglodinta funkcija yra diferencijuojama daug kartų. Kai n=1 šios funkcijos ir jos išvestinių grafikai taip pat yra pateikiami *lentelėje 3.1*.

Taikydami įvairius būdus suglodintos funkcijos gradientui įvertinti, gausime skirtingus stochastinio gradiento įverčius, kuriuos galima taikyti sudarant SPSA algoritmus (Bartkutė & Sakalauskas (2006a), Bartkutė & Sakalauskas (2006b)):

1) <u>*SPSAL*</u> -SPSA algoritmas su Lipšico suglodinimo operatoriumi (*SPSA with Lipschitz perturbation operator*), čia stochastinis gradientas įvertinamas tokiu būdu:

$$g(x,\sigma,\xi)=\frac{\left(f(x+\sigma\xi)-f(x)\right)\cdot\xi}{\sigma\cdot\|\xi\|},\tag{3.15}$$

ξ -atsitiktinis pokytis, tolygiai pasiskirstęs vienetiniame rutulyje.

2) *SPSAU* – SPSA algoritmas su tolygiai pasiskirsčiusiu suglodinimo operatoriumi (*SPSA with Uniform perturbation operator*), čia stochastinis gradientas įvertinamas tokiu būdu (Михалевич et al. (1987)):

$$g(x,\sigma,\xi)=\frac{(f(x+\sigma\cdot\xi)-f(x-\sigma\cdot\xi))\cdot\xi}{2\sigma}, \qquad (3.16)$$

ξ -atsitiktinis pokytis, tolygiai pasiskirstęs daugiamačiame kube $[-1;1]^n$

3) *FDSA* -baigtinių skirtumų stochastinės aproksimacijos algoritmas (*Finite Difference Stochastic Approximation algorithm*), įvertinant stochastinį gradientą skirtuminiu būdu (Михалевич et al. (1987)):

$$g_i(x,\sigma,\xi,\upsilon)=\frac{f(x+\sigma\cdot\xi+\upsilon\cdot\varepsilon_i)-f(x+\sigma\cdot\xi-\upsilon\cdot\varepsilon_i)}{2\upsilon}, \qquad (3.17)$$

čia ξ yra tas pats kaip ir (3.15), $\varepsilon_i=(0,0,0,....,1,.....,0)$ yra vektorius, kurio visos komponentės lygios 0, išskyrus *i*-tąją, kuri lygi 1, o υ ir σ yra baigtinių skirtumų bei suglodinimo parametrai, $i=1,2,...,n$.

Algoritmais 3.1, 3.2 ir 3.3 aprašomas stochastinio gradiento įvertinimas aukščiau aprašytais metodais.

***Algoritmas 3.1*: Stochastinio gradiento įvertinimas (SPSA su Lipšico suglodinimo operatoriumi)**

Tikslas: įvertinti stochastinį gradientą.

Pradinės sąlygos (*preconditions*)**:** tikslo funkcijos apskaičiavimo algoritmas, kintamųjų skaičius *n*, iteracijos numeris *t*, suglodinimo parametras σ_t, vektorius x^t .

Galinės sąlygos (*postconditions*)**:** stochastinio gradiento vektorius g^t.

1. Generuoti *n*-matį vektorių $\nu^t \sim N(0, E_n)$ ir apskaičiuoti šio vektoriaus normą $\left\|\nu^t\right\|$;
2. Apskaičiuoti *n*-matį vektorių, tolygiai pasiskirsčiusį *n*-mačio vienetinio rutulio paviršiuje: $\xi^t=\frac{\nu^t}{\left\|\nu^t\right\|}$;
3. Apskaičiuoti vektoriaus, tolygiai pasiskirsčiusio *n*-mačiame vienetiniame rutulyje, normą: $r^t=s^{\frac{1}{n}},\ s\sim T(0,1)$;
4. Apskaičiuoti tikslo funkcijos reikšmę $f(x^t)$;
5. Apskaičiuoti tikslo funkcijos reikšmę $f^t=f(x^t+\sigma_t\cdot r^t\cdot\xi^t)$;

6. Apskaičiuoti stochastinio gradiento įvertį $g^t = g^t\left(x^t, \sigma_t, \xi^t\right) = \frac{\left(f^t - f\left(x^t\right)\right) \cdot \xi^t}{\sigma_t}$.◊

Algoritmas 3.2: **Stochastinio gradiento įvertinimas (SPSA su tolygiai pasiskirsčiusiu suglodinimo operatoriumi)**

Tikslas: įvertinti stochastinį gradientą.

Pradinės sąlygos (*preconditions*)**:** tikslo funkcijos apskaičiavimo algoritmas, kintamųjų skaičius *n*, iteracijos numeris *t*, suglodinimo parametras σ_t, vektorius x^t .

Galinės sąlygos (*postconditions*)**:** stochastinio gradiento vektorius g^t.

1. Generuoti *n*-matį vektorių ξ^t, kurio komponentės yra pasiskirčiusios tolygiai T(-1, 1);
2. Apskaičiuoti tikslo funkcijos reikšmes $f\left(x^t + \sigma_t \cdot \xi^t\right)$ ir $f\left(x^t - \sigma_t \cdot \xi^t\right)$.
3. Apskaičiuoti stochastinio gradiento įvertį

$$g^t = g^t\left(x^t, \sigma_t, \xi^t\right) = \frac{\left(f\left(x^t + \sigma_t \cdot \xi^t\right) - f\left(x^t - \sigma_t \cdot \xi^t\right)\right) \cdot \xi^t}{2\sigma_t}.◊$$

Algoritmas 3.3: **Stochastinio gradiento įvertinimas (baigtinių skirtumų stochastinė aproksimacija)**

Tikslas: įvertinti stochastinį gradientą.

Pradinės sąlygos (*preconditions*)**:** tikslo funkcijos apskaičiavimo algoritmas, kintamųjų skaičius *n*, iteracijos numeris *t*, suglodinimo parametras σ_t, vektorius x^t, baigtinių skirtumų parametras υ_t .

Galinės sąlygos (*postconditions*)**:** stochastinio gradiento vektorius g^t.

1. Generuoti *n*-matį vektorių $\nu^t \sim N\left(0, E_n\right)$ ir apskaičiuoti šio vektoriaus normą $\left\|\nu^t\right\|$;
2. Apskaičiuoti *n*-matį vektorių, tolygiai pasiskirsčiusį *n*-mačio vienetinio rutulio paviršiuje: $\xi^t = \frac{\nu^t}{\left\|\nu^t\right\|}$;
3. Apskaičiuoti vektoriaus, tolygiai pasiskirsčiusio *n*-mačiame vienetiniame rutulyje, normą: $r^t = s^{\frac{1}{n}}$, $s \sim T\left(0, 1\right)$;
4. **for** *i*=1 to *n* **do**
5. Apskaičiuoti tikslo funkcijos reikšmę taškuose $x^t + \sigma_t \cdot \xi^t + \upsilon_t \cdot \varepsilon_i$ ir $x^t + \sigma_t \cdot \xi^t - \upsilon_t \cdot \varepsilon_i$

{*Komentaras*: $\varepsilon_i = (0,0,0,....,1,.....,0)$ yra vektorius, kurio visos komponentės lygios 0, išskyrus *i*-tąją, kuri lygi 1};

6. Apskaičiuoti stochastinio gradiento įvertį

$$g_i^t = g_i^t\left(x^t, \sigma_t, \xi^t, \upsilon_t\right) = \frac{f\left(x^t + \sigma_t \cdot \xi^t + \upsilon_t \cdot \varepsilon_i\right) - f\left(x^t + \sigma_t \cdot \xi^t - \upsilon_t \cdot \varepsilon_i\right)}{2\upsilon_t}$$

{*Komentaras*: g_i^t yra stochastinio gradiento g^t vektoriaus, i-toji komponentė };

7. **done.** ◊

3.3. SPSA METODO KONVERGAVIMAS

Panagrinėkime seką (2.14), kai $g^t = g\left(x^t, \sigma_t, \xi^t\right)$ stochastinio gradiento, apskaičiuoto pagal (3.5), reikšmė taške x^t, $\xi_1, \xi_2, ...$ yra nepriklausomos ξ kopijos, ρ_t - skaliarinis daugiklis, o σ_t suglodinimo parametro reikšmė iteracijoje t, x^0 - pradinis taškas. Ši seka konverguoja į (2.1) sprendinį beveik tikrai (b. t.) (t.y. bet kuriam $\varepsilon > 0$, $\lim_{t\to\infty} P\left(\left\|x^t - x^*\right\| \le \varepsilon\right) = 1$), kai iteracijų skaičius neaprėžtai didėja, jei yra išpildytos tam tikros sąlygos.

Pažymėkime X^* $f(x)$ stacionarių taškų aibę:

$$X^* = \{x | 0 \in \partial f(x)\},$$

o F^* jos stacionarių reikšmių aibę:

$$F^* = \{z | z = f(x), x \in X^*\}.$$

Teorema 3.1. (Bartkutė & Sakalauskas (2007a)). *Jei funkcija* $f : \Re^n \to \Re$ *:*

A) tenkina Lipšico sąlygą su konstanta K,

B) yra aprėžta iš apačios,

C) $\underline{\lim}_{x\to\infty} \inf_{g \in \partial f(x)} \|g\| > 0$ (*apibendrintas gradientas negali būt lygus nuliui begalybėje*),

D) bet kuriam $\varepsilon > 0$ *galima rasti tokį* $\delta > 0$*, kad visiems y,* $\|y - x\| \le 2\delta$*, kur funkcija* $f(x)$ *diferencijuojama įprastine prasme, tolygiai x atžvilgiu galioja* $\min_{z \in \partial f(x)} \left\| \frac{\partial f(y)}{\partial y} - z \right\| \le \frac{\varepsilon}{2}$,

E) funkcijos stacionarių reikšmių aibė F^* *neturi vidinių taškų,*

ir

$$\int_{dom(\Psi)} \|y\|^4 \cdot \frac{\|\partial p(y)\|^4}{(\psi(y))^3} dy < \infty,$$

$$\int_P \|\partial p(y)\| \cdot dy < \infty, \; E\|\xi\|^2 < \infty,$$

$\{\rho_t\}$, $\{\sigma_t\}$ *yra neneigiamų skaičių sekos, tokios, kad*

$$\sum_{t=1}^{\infty} \rho_t = \infty, \; \sum_{t=1}^{\infty} \rho_t^2 < \infty, \; \sigma_t \to 0, \; \frac{|\sigma_t - \sigma_{t-1}|}{\rho_t} \to 0, \; \frac{\rho_t}{\sigma_t} \to 0,$$

tada $\lim_{t \to \infty} x^t \in X^*$ *b. t.*

Įrodymas.

$\triangleright$ Teoremos įrodymas yra pagrįstas žinomais stochastinės aproksimacijos konvergavimo faktais (Wasan (1969), Ермольев (1976), Gupal & Norkin (1977), Nurminski (1979), Михалевич et al. (1987), Dupac (1988), Kushner and Yin (2003) ir pan.). Dėl atskirų detalių, susijusių su *lemų 3.1* ir *3.2* taikymu, pateikiame išsamų algoritmo konvergavimo įrodymą.

Iš (3.3) turime

$$\bar{f}\left(x^{t+1}, \sigma_{t+1}\right) \le \bar{f}\left(x^{t+1}, \sigma_t\right) + C \cdot K \cdot |\sigma_{t+1} - \sigma_t|. \qquad (3.18)$$

Iš Lagrandžo formulės (Dieudonné (1960)), (3.3), (3.5), (3.6) ir *Lemos 3.1* seka:

$$\bar{f}\left(x^{t+1}, \sigma_t\right) = \bar{f}\left(x^t, \sigma_t\right) + \left(\bar{g}\left(x^t + \tau\left(x^{t+1} - x^t\right), \sigma_t\right) - \bar{g}\left(x^t, \sigma_t\right)\right)^T \cdot \left(x^{t+1} - x^t\right) + \bar{g}\left(x^t, \sigma_t\right)^T \cdot \left(x^{t+1} - x^t\right) \le$$

$$\le \bar{f}\left(x^t, \sigma_t\right) + \bar{g}\left(x^t, \sigma_t\right)^T \cdot \left(x^{t+1} - x^t\right) + \frac{K \cdot L}{\sigma_t} \cdot \left\|x^{t+1} - x^t\right\|^2 = \bar{f}\left(x^t, \sigma_t\right) - \rho_t g\left(x^t, \sigma_t, \xi_{t+1}\right)^T \bar{g}\left(x^t, \sigma_t\right) +$$

$$+ \frac{K \cdot L \cdot \rho_t^2}{\sigma_t} \left\|g\left(x^t, \sigma_t, \xi^t\right)\right\|^2 = \bar{f}\left(x^t, \sigma_t\right) + \frac{K \cdot L}{\sigma_t} \rho_t^2 \left\|g^t\right\|^2 - \rho_t \left\|\bar{g}^t\right\|^2 + \rho_t \left(\bar{g}^t\right)^T \left(\bar{g}^t - g^t\right) \le$$

$$\le \bar{f}\left(x^t, \sigma_t\right) + \frac{K^3 \cdot L \cdot A}{\sigma_t} \rho_t^2 - \rho_t \left\|\bar{g}^t\right\|^2 + \frac{K \cdot L}{\sigma_t} \rho_t^2 \cdot \left(\left\|g^t\right\|^2 - E\left\|g^t\right\|^2\right) + \rho_t \left(\bar{g}^t\right)^T \left(\bar{g}^t - g^t\right), \qquad (3.19)$$

čia $0 \le \tau \le 1$, ir trumpumo dėlei pažymėta, kad $\bar{g}^t = \bar{g}\left(x^t, \sigma_t\right)$, $g^t = g\left(x^t, \sigma_t, \xi_{t+1}\right)$.

Pažymėsime $\{\Theta_t\}_{t=0}^{\infty}$ σ-algebrų seką generuotą sekos $\left\{x^t\right\}_{t=0}^{\infty}$.

Pirmiausia parodysime, kad egzistuoja aprėžtas baigtinis posekis, konverguojantis į X^*. Tarkime priešingai, t.y. galima rasti tokius skaičius $\bar{s}$ ir δ, kad aibių $\left\{x \,\middle|\, \left\|x^s - x\right\| \le 2\delta\right\}$ seka neturi sankirtos su X^*, kai $s \ge \bar{s}$. Iš *Lemos 3.2* seka, kad galima rasti tokį $\varepsilon > 0$, kad $\left\|\bar{g}^s\right\| \ge \varepsilon > 0$ tolygiai x atžvilgiu, jei $s \ge \bar{s}$, ir $\bar{s}$ yra pakankamai didelis. Tegul $\bar{s}$ yra toks, kad yra teisinga nelygybė, kai $i \ge \bar{s}$: $\frac{C \cdot K \cdot |\sigma_{i+1} - \sigma_i|}{\rho_i} + \frac{\rho_i \cdot K^3 \cdot L \cdot A}{\sigma_i} \le \frac{\varepsilon^2}{2}$. Tada iš (3.18) ir (3.19) gauname:

$$\overline{f}\left(x^{t+1},\sigma_{t+1}\right)\le \bar{f}\left(x^{s},\sigma_{s}\right)-\frac{\varepsilon^{2}}{2}\sum_{i=s}^{t}\rho_{i}-\sum_{i=s}^{t}\rho_{i}\left(\frac{\varepsilon^{2}}{2}-\frac{C\cdot K\cdot\left|\sigma_{i+1}-\sigma_{i}\right|}{\rho_{i}}-\frac{\rho_{i}\cdot K^{3}\cdot L\cdot A}{\sigma_{i}}\right)+$$

$$+\sum_{i=s}^{t}\rho_{i}^{2}\cdot\frac{K\cdot L}{\sigma_{i}}\cdot\left(\left\|g^{i}\right\|^{2}-E\left(\left\|g^{i}\right\|^{2}|\Theta_{i}\right)\right)+\sum_{i=s}^{t}\rho_{i}\left(\overline{g}^{i}\right)^{T}\left(\overline{g}^{i}-g^{i}\right)\le$$

$$\le \bar{f}\left(x^{s},\sigma_{s}\right)-\frac{\varepsilon^{2}}{2}\sum_{i=s}^{t}\rho_{i}+\sum_{i=s}^{t}\rho_{i}\left(\overline{g}^{i}\right)^{T}\left(\overline{g}^{i}-g^{i}\right)+\sum_{i=s}^{t}\rho_{i}^{2}\cdot\frac{K\cdot L}{\sigma_{i}}\cdot\left(\left\|g^{i}\right\|^{2}-E\left(\left\|g^{i}\right\|^{2}|\Theta_{i}\right)\right). \qquad (3.20)$$

Kadangi $\sum_{i=0}^{t}\rho_{i}^{2}\left(E\left(\left(g^{i}\right)^{T}|\Theta_{i}\right)\cdot\overline{g}^{i}\right)^{2}\le\sum_{i=0}^{t}\rho_{i}^{2}\left(E\left(\left\|g^{i}\right\|^{2}|\Theta_{i}\right)\right)^{2}\le K^{4}\cdot A^{2}\cdot\sum_{i=0}^{t}\rho_{i}^{2}$ ir $\sum_{k=1}^{\infty}\rho_{k}^{2}<\infty$, iš P4. Lemos (žr. priedą 1) išplaukia, kad $\sum_{i=0}^{\infty}\rho_{i}\left(\overline{g}^{i}\right)^{T}\cdot\left(\overline{g}^{i}-g^{i}\right)$ b.t. konverguoja. Panašiai įrodome, kad $\sum_{i=0}^{\infty}\rho_{i}^{2}\cdot\frac{K\cdot L}{\sigma_{i}}\cdot\left(\left\|g^{i}\right\|^{2}-E\left(\left\|g^{i}\right\|^{2}|\Theta_{i}\right)\right)$ b.t. konverguoja, nes:

$$E\left\|g^{t}\right\|^{4}=E\left(\left\|\frac{f\left(x^{t}+\sigma_{k}\xi\right)-f\left(x^{t}\right)}{\sigma_{t}}\cdot\frac{\partial\,ln\,p(\xi)}{\psi(\xi)}\right\|^{4}\right)\le K^{4}\cdot\int_{dom(P)}\left\|y\right\|^{4}\cdot\frac{\left\|\partial\,ln\,p(y)\right\|^{4}}{\left(\psi(y)\right)^{3}}dy<\infty .$$

Kadangi du paskutiniai dėmenys formulėje (3.20) yra aprėžti beveik tikrai, gauname prieštaravimą, nes $\sum_{i=s}^{\infty}\rho_{i}=\infty$ ir $\overline{f}(x,\sigma)>-\infty$.

Taigi, turi egzistuoti begalinis posekis, konverguojantis į X^{*}.

Dabar tarkime, kad sekoje (2.14) egzistuoja, arba aprėžtas posekis, konverguojantis į $x'\notin X^{*}$, arba posekis, konverguojantis į begalybę.

Pirma įrodysime, kad, jei sekoje $\left\{x^{t}\right\}_{t=0}^{\infty}$ yra aprėžtas posekis, konverguojantis į tašką x', čia $\inf_{g\in\partial f(x')}\left\|g\right\|>0$, tai egzistuoja toks δ_{0}, kad esant bet kokiems $\delta\in\left(0,\ \delta_{0}\right]$, galima rasti tokias indeksų sekas $\left\{l_{s}\right\}_{s=0}^{\infty}$, $\left\{k_{s}\right\}_{s=0}^{\infty}$, $l_{s}<k_{s}$, kad $\left\|x^{i}-x'\right\|\le 2\delta$ dėl visų $i\in\left[l_{s},k_{s}-1\right]$, ir:

$$\overline{\lim_{s\to\infty}}\,f\left(x^{k_{s}}\right)<\lim_{s\to\infty}f\left(x^{l_{s}}\right). \qquad (3.21)$$

Iš tikrųjų, iš prielaidų apie posekių, konverguojančių į $x'\notin X^{*}$ ir į $x\in X^{*}$, egzistavimą, išplaukia, kad galime rasti tokias indeksų sekas l_{s}, k_{s}, $l_{s}<k_{s}$, kad pakankamai mažam δ turime, kad $\left\|x^{l_{s}}-x'\right\|<\delta$, $\left\|x^{k_{s}}-x'\right\|>2\delta$ ir $\left\|x^{i}-x'\right\|\le 2\delta$, kai $i\in\left[l_{s},k_{s}-1\right]$. Dabar įrodysime, kad iš tokių indeksų sekų egzistavimo išplaukia (3.21). Turime, kad $x^{k_{s}}=x^{l_{s}}-\sum_{i=l_{s}}^{k_{s}-1}\rho_{i}g^{i}=x^{l_{s}}-\sum_{i=l_{s}}^{k_{s}-1}\rho_{i}\overline{g}^{i}-\sum_{i=l_{s}}^{k_{s}-1}\rho_{i}\left(g^{i}-\overline{g}^{i}\right)$. Iš *P4. Lemos* (žr. priedą 1) išplaukia, kad

$\sum_{i=0}^{\infty} \rho_i \cdot \left(g^i - \overline{g}^i\right)$ b. t. konverguoja. Reiškia $\left\| \sum_{i=l_s}^{k_s-1} \rho_i \left(g^i - \overline{g}^i\right) \right\| < \frac{\delta}{2}$, kai $s \geq \overline{s}$, jei $\overline{s}$ yra pakankamai didelis. Todėl $\delta \leq \left\| x^{k_s} - x^{l_s} \right\| \leq \left\| \sum_{i=l_s}^{k_s-1} \rho_i \overline{g}^i \right\| + \left\| \sum_{i=l_s}^{k_s-1} \rho_i \left(g^i - \overline{g}^i\right) \right\| \leq \left\| \sum_{i=l_s}^{k_s-1} \rho_i \overline{g}^i \right\| + \frac{\delta}{2}$.

Tad $\left\| \sum_{i=l_s}^{k_s-1} \rho_i \overline{g}^i \right\| \geq \frac{\delta}{2}$ ir $\sum_{i=s}^{k_s-1} \rho_i \geq \frac{\delta}{2 \cdot \sqrt{K}}$, nes $\left\| \overline{g}^i \right\|^2 < K$. Tuomet iš (3.20) gauname įvertį:

$$\overline{f}\left(x^{k_s}, \sigma_{k_s}\right) \leq \overline{f}\left(x^{l_s}, \sigma_{l_s}\right) - \frac{\varepsilon^2 \cdot \delta}{8 \cdot \sqrt{K}}.$$

Iš *Lemos 3.2* seka tolygus konvergavimas x atžvilgiu: $\overline{f}(x, \sigma) \to f(x)$, kai $\sigma \to 0$, iš kurio išplaukia (3.21). Pastebėsime, kad bet kuriems skaičiams f' ir f'', tokiems, kad

$$\overline{\lim_{s\to\infty}} f\left(x^{k_s}\right) < f' < f'' < \lim_{s\to\infty} f\left(x^{l_s}\right),$$

seka $f\left(x^t\right)$ kerta intervalą (f', f'') be galo daug kartų. Tuomet galima rasti tokius du posekius $\left\{x^{r_s}\right\}$, ir $\left\{x^{p_s}\right\}$, kuriems

$$f\left(x^{r_i}\right) \leq f', \; f\left(x^{r_i+1}\right) > f', \tag{3.22}$$

$$f\left(x^{p_i}\right) > f'', \; f\left(x^t\right) > f', \; r_s < t < p_s. \tag{3.23}$$

Nemažinant bendrumo seką $\left\{x^{r_s}\right\}$ galima laikyti konverguojančia, priešingu atveju vietoj šios sekos galima paimti jos konverguojantį posekį. Iš sekos $\left\{x^{r_s}\right\}$ konvergavimo, funkcijos *f* tolydumo, (3.22) ir $\rho_i \to 0$ turime, kad $\lim_{s\to\infty} f\left(x^{r_s}\right) = f'$. Tuomet iš (3.22), (3.23) išplaukia, kad

$$\overline{\lim_{s\to\infty}} f\left(x^{p_s}\right) \geq \lim_{s\to\infty} f\left(x^{r_s}\right). \tag{3.24}$$

Kadangi F^* neturi vidinių taškų, tai f' galime parinkti tokį, kad $f' \notin F^*$. Dabar galime išvesti (3.21), kai seka $\{r_s\}$ atitinka $\{l_s\}$. Tačiau pastaruoju atveju (3.24) prieštarauja (3.21).

Jei sekoje $\left\{x^t\right\}_{t=0}^{\infty}$ yra posekis, konverguojantis į begalybę, panašiai įrodome, kad egzistuoja toks δ_0, kad, kai $\delta \in (0, \delta_0]$, galima rasti tokias indeksų sekas $\{l_s\}_{s=0}^{\infty}$, $\{k_s\}_{s=0}^{\infty}$, kad $\inf_{x\in X^*} \left\| x^{l_s} - x \right\| > 2 \cdot \delta$, $\inf_{x\in X^*} \left\| x^{k_s} - x \right\| < \delta$, ir $\inf_{x\in X^*} \left\| x^i - x \right\| \geq \delta$ dėl visų $i \in [l_s, k_s - 1]$, iš kurių egzistavimo seka $\overline{\lim_{s\to\infty}} f\left(x^{k_s}\right) < \lim_{s\to\infty} f\left(x^{l_s}\right)$, toliau analogiškai aukščiau išnagrinėtam atvejui nustatome atitinkamų indeksų r_s, p_s egzistavimą, vedantį į prieštaravimą. ◃

Analogiškos konvergavimo sąlygos nustatytos SPSAU ir FDSA metodams (Михалевич et al. (1987)).

3.4. SPSA METODO KONVERGAVIMO GREITIS

Stochastinės aproksimacijos konvergavimo greičio tyrimai, kai tikslo funkcija yra diferencijuojama, pateikiami daugelio autorių darbuose (Wasan (1969), Nurminski (1979), Михалевич et al (1987), Kushner & Yin (2003) ir t.t.). Kai tikslo funkcija yra du kartus diferencijuojama ir skaičiuojama be triukšmo nustatytas konvergavimo greitis $O\left(\frac{1}{t}\right)$ (Dupac (1988)). Tuo atveju kai tikslo funkcija yra diferencijuojama ir skaičiuojama su atsitiktine paklaida konvergavimo greitis yra didesnis: $O\left(\frac{1}{t^{\gamma}}\right), 0<\gamma<\frac{2}{3}$ (Poliak (1987), Граничин & Поляк (2003), Kushner & Yin (2003)). Yra žinoma, kad tam tikrais atvejais algoritmo konvergavimo greitis, kai tikslo funkcija turi aštrų minimumą, gali būti didesnis nei glodžių funkcijų (Poliak (1987), Nesterov (2004)).

Disertacijoje įrodysime, kad SPSA metodas gali pasiekti konvergavimo greitį $O\left(\frac{1}{t^{\gamma}}\right)$, kai $1 \le \gamma < 2$. Šiame darbe konvergavimo greičio tyrimas susideda iš dviejų dalių: optimizavimo sekos (2.14) konvergavimo į suglodintos funkcijos (3.1) minimumo tašką ir suglodintos funkcijos minimumo konvergavimo į tikslo funkcijos minimumo tašką, kai $\sigma_t \to 0$.

Panagrinėkime Lipšico funkciją, turinčią aštrų minimumą taške x^* su konstanta $\mu > 0$ (žr. Poliak (1987)):

$$f(x)-f(x^*) \ge \mu\|x-x^*\|, \text{ kiekvienam } x \in \Re^n. \quad (3.25)$$

Pažymėkime $h:\Re^n \to \Re$ Klarko apibendrintą išvestinę pagal kryptį taške $x^* \in \Re^n$, t.y. tokį atvaizdavimą, kad (Clarke (1975), Rockafellar (1979), žr. priedą 1):

$$h(y) = \max_{z\in\partial f(x^*)}(z \cdot y), \; y \in \Re^n. \quad (3.26)$$

Įveskime funkciją, aprašančią tikslo funkcijos savybes aštraus minimumo aplinkoje:

$$\bar{h}(x) = E(h(x+\xi)). \quad (3.27)$$

Lema 3.3. Tegul $f(x)$ yra Lipšico funkcija. Tada funkcija $h:\Re^n \to \Re$, apibrėžta (3.26), yra iškila kūgio pavidalo Lipšico funkcija. Jei funkcija $f(x)$ turi aštrų minimumą su konstanta μ, tada

$$h(y) \ge \mu \cdot \|y\|, \; y \in \Re^n. \quad (3.28)$$

Įrodymas.

$\triangleright$ Apibendrinos išvestinės pagal kryptį atvaizdavimo h iškilumo ir kūgiškumo savybės yra įrodytos Clarke (1983). Iš apibrėžimo (3.26) turime:

$$h(y_1)-h(y_2) = \max_{z\in\partial f(x^*)}(z\cdot y_1) - \max_{z\in\partial f(x^*)}(z\cdot y_2) = (z_1\cdot y_1)-(z_2\cdot y_2) \le (z_1\cdot(y_1-y_2)) \le K\cdot\|y_2-y_1\|,$$

čia $(z_1 \cdot y_1) = \max\limits_{z \in \partial f(x^*)} (z \cdot y_1)$ ir $(z_2 \cdot y_2) = \max\limits_{z \in \partial f(x^*)} (z \cdot y_2)$.

Panašiai įrodome, kad:

$$h(y_2) - h(y_1) \le K \cdot \|y_2 - y_1\|.$$

Iš čia seka Lipšico sąlyga:

$$|h(y_2) - h(y_1)| \le K \cdot \|y_2 - y_1\|.$$

(3.28) įrodymas lengvai seka iš apibendrintos išvestinės pagal kryptį apibrėžimo (Clarke (1983), Михалевич et al. (1987)) ir aštraus minimumo savybės (3.25):

$$h(y) = \lim_{\delta \to 0_+} \sup_{\substack{\|v\| \le \delta \\ 0 < \sigma \le \delta}} \frac{f(x^* + v + \sigma \cdot y) - f(x^* + v)}{\sigma} \ge \lim_{\sigma \to 0_+} \frac{f(x^* + \sigma \cdot y) - f(x^*)}{\sigma} \ge \mu \cdot \|y\|. \ \triangleleft$$

Kadangi $h(\cdot)$ yra iškila Lipšico funkcija, tai jos vidurkis $\overline{h}(x)$ yra du kartus diferencijuojama funkcija ir jos gradientas gali būti įvertintas pagal *lemą 3.1.*

Pažymėkime funkcijas $h(x,\sigma) = \dfrac{f(x^* + \sigma \cdot (x - x^*)) - f(x^*)}{\sigma}$ ir $\overline{h}(x,\sigma) = E(h(x + \xi, \sigma))$.

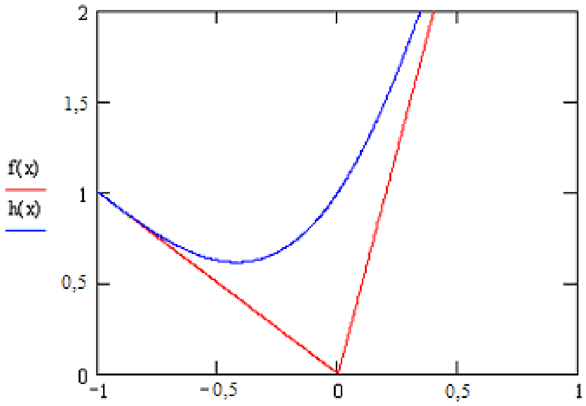
Bendru atveju tikslo funkcijos minimumas x^* skiriasi nuo funkcijos $\overline{h}(x)$ minimumo taško y^*. Tokiu būdu galime įvesti sąryšį tarp suglodintos funkcijos $\overline{f}(x,\sigma)$ minimumo taško x_σ^* ir tikslo funkcijos minimumo taško x^*:

$$x_\sigma^* = x^* + \sigma(y^* - x^*) + o(\sigma).$$

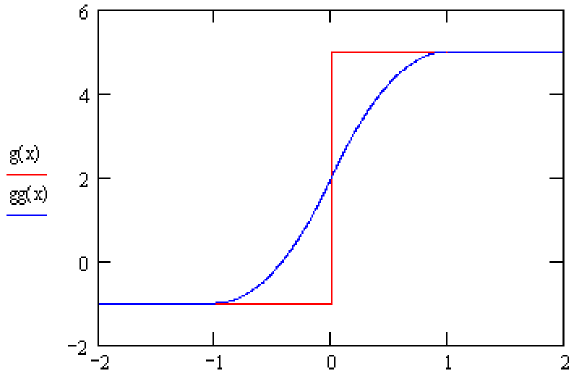
Pavyzdys 3.4. Tegul turime funkciją $f(x) = \max(-m \cdot x, K \cdot x)$, $K \ge m > 0$, o suglodinimo tankis, apibrėžtas kaip pavyzdyje 3.1 b). Nesunku įsitikinti, kad $\overline{h}(x)$ yra:

$$h(x,\sigma) = \begin{cases} -m \cdot x, & x \le -1, \\ -m \cdot \left(-\dfrac{1}{6} + \dfrac{x}{2} - \dfrac{x^2 \cdot (x+3)}{6} \right) + K \cdot \left(\dfrac{1}{6} + \dfrac{x}{2} + \dfrac{x^2 \cdot (x+3)}{6} \right), & x > -1 \text{ ir } x \le 0, \\ K \cdot x, & x > 1, \\ m \cdot \left(\dfrac{1}{6} - \dfrac{x}{2} - \dfrac{x^2 \cdot (x-3)}{6} \right) - K \cdot \left(\dfrac{x^2 \cdot (x+3)}{6} - \dfrac{1}{6} - \dfrac{x}{2} \right), & x < 1 \text{ ir } x \ge 0. \end{cases}$$

o jos gradientas
$$\frac{dh(x,\sigma)}{dx} = \begin{cases} -m, & x \le -1, \\ -m \cdot \left(\dfrac{1}{2} - \dfrac{x^2}{2} - x \right) + K \cdot \left(\dfrac{1}{2} + \dfrac{x^2}{2} + x \right), & x > -1 \text{ ir } x \le 0, \\ K, & x > 1, \\ m \cdot \left(x - \dfrac{x^2}{2} - \dfrac{1}{2} \right) - K \cdot \left(\dfrac{x^2}{2} - \dfrac{1}{2} - x \right), & x < 1 \text{ ir } x \ge 0. \end{cases}$$



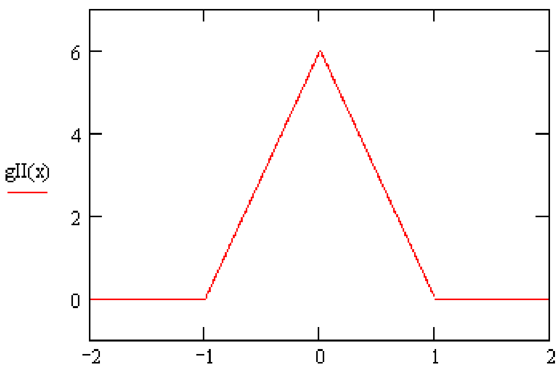
Pav. 3.1. Funkcija $f(x) = \max(-m \cdot x, K \cdot x)$ ir Lipšico tankiu suglodinta funkcija



Pav. 3.2. Funkcijos $f(x) = \max(-m \cdot x, K \cdot x)$ ir suglodintos funkcijos apibendrinti gradientai

Ši funkcija yra du kartus diferencijuojama. Jos antros eilės išvestinė yra:

$$\frac{d^2 h(x,\sigma)}{dx^2} = \begin{cases} 0, & x \le -1, \\ (K+m)\cdot(1+x), & x > -1 \text{ ir } x \le 0, \\ 0, & x > 1, \\ (K+m)\cdot(1-x), & x < 1 \text{ ir } x \ge 0. \end{cases}$$



Pav. 3.3. Suglodintos funkcijos $f(x) = \max(-m \cdot x, K \cdot x)$ antros eilės išvestinė

Nesunku įsitikinti, kad funkcijos $\overline{h}(x)$ minimumo taškas yra lygus $y^* = \sqrt{\frac{2m}{K+m}} - 1$, o šios funkcijos hesianas minimumo taške yra lygus $H = \left. \frac{d^2 \overline{h}(x)}{dx^2} \right|_{x=y^*} = \sqrt{2m \cdot (K+m)}$.

Suformuluosime ir įrodysime lemą apie funkcijos $\overline{h}(x,\sigma)$ konvergavimą į $\overline{h}(x)$, kai $\sigma \to 0$.

Lema 3.4. Tegul *f(x)* yra pusiauglodi funkcija (žr. priedą 1). Tada $\lim_{\sigma \to 0} \overline{h}(x,\sigma) = \overline{h}(x)$.

Įrodymas.

$\triangleright$ Lemos įrodymas seka iš glodžios funkcijos išvestinės pagal kryptį bei funkcijos $\overline{h}(x,\sigma)$ apibrėžimų ir Lebego teoremos:

$$\lim_{\sigma \to 0} \overline{h}(x,\sigma) = \lim_{\sigma \to 0} E\left(\frac{f\left(x^* + \sigma \cdot \left(x - x^*\right) + \sigma \cdot \xi\right) - f\left(x^*\right)}{\sigma} \right) =$$

$$= E\left(\lim_{\sigma \to 0} \left(\frac{f\left(x^* + \sigma \cdot \left(x - x^*\right) + \sigma \cdot \xi\right) - f\left(x^*\right)}{\sigma} \right) \right) = E\left(h(x+\xi)\right) = \overline{h}(x). \ \triangleleft$$

Suformuluosime ir įrodysime lemą apie tikslo funkcijos ir funkcijos $\overline{h}(x)$ minimumo taškų sąryšį.

Lema 3.5. Tegul x_σ^* yra suglodintos funkcijos (3.1) minimumo taškas, čia $f(x)$ yra pusiauglodi funkcija. Tuomet

$$\lim_{\sigma \to 0} \frac{x_\sigma^* - x^*}{\sigma} = y^* - x^*,$$

čia y^* yra funkcijos $\overline{h}(x)$ minimumo taškas.

Įrodymas.

$\triangleright$ Pastebėsime, kad funkcija $\overline{h}(x,\sigma)$ yra tolydinė, aprėžta iš apačios ir neaprėžtai didėjanti, kai $x \to \infty$. Tai reiškia, kad funkcija turi minimumą baigtiniame taške y_σ^*. Nesunku pastebėti, kad funkcijų $\overline{h}(x,\sigma)$ ir $\overline{f}(x,\sigma)$ minimumo taškai y_σ^* ir x_σ^* yra susieti sąryšiu: $x_\sigma^* = x^* + \sigma\left(y_\sigma^* - x^*\right)$.

Lemą galima laikyti įrodyta, kadangi iš *lemos 3.4* išplaukia, kad $y^* = \lim_{\sigma \to 0} y_\sigma^*$. $\triangleleft$

Suformuluosime ir įrodysime pagalbinę lemą. Lemos įrodymas yra panašus į Chung lemą (Chung (1954),Wasan (1969)).

Lema 3.6. (Bartkutė & Sakalauskas (2007a)). Tegul $\{u_t\}_1^\infty$ yra neneigiamų skaičių seka, tokia, kad tam tikriems $t \geq t_0$ galioja:

$$u_{t+1}^2 \leq \left(u_t \cdot \left(1 - \frac{c_t}{t^p}\right) + \frac{l_t}{t^s} \right)^2 + \frac{d}{t^r},$$

čia $c_t > 0$, $l_t > 0$, $c_t \to c$, $l_t \to l$, $d, l > 0, c > \frac{r-p}{2} > 0, s > \frac{r+p}{2}$. Tuomet

$$u_t^2 \le \frac{d}{2c} \cdot \frac{1}{t^{r-p}} + O\left(\frac{1}{t^{2c}}\right).$$

Įrodymas.

$\rhd$ Tarkime, kad $c_t \ge c' = c - \frac{\varepsilon}{2}$, $l_t \le l' = l + \frac{\varepsilon}{2}$ esant tam tikriems $t \ge t_0$ ir tam tikram mažam $\varepsilon > 0$. Iš Chung lemos (Wasan (1969), priedas 3, lema 6) gauname, kad

$$\frac{d}{t^r} \le \frac{d}{2\cdot(c-\varepsilon)}\left(\frac{1}{(t+1)^{r-p}} - \left(1 - \frac{2\cdot(c-\varepsilon)}{t^p}\right)\frac{1}{t^{r-p}}\right). \tag{3.29}$$

Tegul t_0 yra toks, kad $u_t^2 \le \frac{d}{2\cdot(c-\varepsilon)} \cdot \frac{1}{t^{r-p}}$ visiems $t \ge t_0$.

Tada iš (3.29), lemos prielaidos ir pastarosios prielaidos seka:

$$u_{t+1}^2 - \frac{d}{2(c-\varepsilon)} \cdot \frac{1}{(t+1)^{r-p}} \le$$

$$\le u_t^2\left(1 - \frac{c - \frac{\varepsilon}{2}}{t^p}\right)^2 + 2\cdot u_t\left(1 - \frac{c-\frac{\varepsilon}{2}}{t^p}\right)\cdot\frac{l+\frac{\varepsilon}{2}}{t^s} + \frac{\left(l+\frac{\varepsilon}{2}\right)^2}{t^{2s}} - \frac{d}{2(c-\varepsilon)}\cdot\left(1 - \frac{2(c-\varepsilon)}{t^p}\right)\frac{1}{t^{r-p}} \le$$

$$\le \left(1 - \frac{c-\frac{\varepsilon}{2}}{t^p}\right)^2\left(u_t^2 - \frac{d}{2(c-\varepsilon)}\cdot\frac{1}{t^{r-p}}\right) - \frac{d}{2\cdot(c-\varepsilon)}\cdot\frac{\varepsilon}{t^r} + o\left(\frac{1}{t^r}\right) \le 0.$$

Vadinasi, $u_t^2 \le \frac{d}{2(c-\varepsilon)} \cdot \frac{1}{t^{r-p}}$ yra teisinga visiems $t \ge t_0$, jei tai yra teisinga tam tikram dideliam t_0. Toliau, tarkime priešingai, visiems pakankamai dideliems *t*:

$$u_t^2 \ge \frac{d}{2(c-\varepsilon)} \cdot \frac{1}{t^{r-p}} + O\left(\frac{1}{t^{2(c-\varepsilon)}}\right).$$

Tuomet, taikydami (3.29) po nesudėtingų pertvarkymų, kai *t* yra didelis, gausime:

$$u_{t+1}^2-\frac{d}{2(c-\varepsilon)}\cdot\frac{1}{(t+1)^{r-p}}\le$$

$$\le\left(1-\frac{c-\varepsilon}{t^p}\right)^2\left(u_t^2-\frac{d}{2(c-\varepsilon)}\cdot\frac{1}{t^{r-p}}\right)-\frac{u_t}{t^p}\left(u_t\cdot\varepsilon\cdot\left(1-\frac{c-\frac{3\varepsilon}{4}}{t^p}\right)-2\left(1-\frac{c}{t^p}\right)\cdot\frac{l+\frac{\varepsilon}{2}}{t^{s-p}}\right)+$$

$$+\frac{\left(l+\frac{\varepsilon}{2}\right)^2}{t^{2s}}+\frac{d(c-\varepsilon)}{2\cdot t^{r+p}}\le\left(1-\frac{c-\varepsilon}{t^p}\right)^2\left(u_t^2-\frac{d}{2(c-\varepsilon)}\cdot\frac{1}{t^{r-p}}\right)-\varepsilon\cdot\frac{d}{2(c-\varepsilon)}\cdot\frac{1}{t^r}+o\left(\frac{1}{t^r}\right)\le$$

$$\le\left(1-\frac{c-\varepsilon}{t^p}\right)^2\left(u_t^2-\frac{d}{2(c-\varepsilon)}\cdot\frac{1}{t^{r-p}}\right).$$

Pažymėsime u'_{t+1} pastarosios nelygybės kairiąją pusę, o dešiniąją pusę $u'_t\left(1-\frac{c-\varepsilon}{t^p}\right)^2$. Tada $0\le u'_{t+1}\le u'_t\left(1-\frac{c-\varepsilon}{t^p}\right)^2$. Ši seka yra aprėžta iš apačios, monotoninė bei konverguojanti. Tada

$$u'_{t+1}\le u'_m\cdot\prod_{i=m}^{t}\left(1-\frac{c-\varepsilon}{i}\right)^2\le u'_m\cdot e^{-2(c-\varepsilon)\cdot\sum_{i=m}^{t}\frac{1}{i}}=O\left(\frac{1}{t^{2(c-\varepsilon)}}\right),\ m>c.$$

Todėl abiem atvejais: $u_t^2\le\inf_{\varepsilon>0}\left(\frac{d}{2(c-\varepsilon)}\cdot\frac{1}{t^{r-p}}+O\left(\frac{1}{t^{2(c-\varepsilon)}}\right)\right)\le\frac{d}{2c}\cdot\frac{1}{t^{r-p}}+O\left(\frac{1}{t^{2c}}\right)$. $\triangleleft$

Suformuluosime ir įrodysime teoremą, SPSA metodo konvergavimo greičiui įvertinti, kai tikslo funkcijos turi aštrų minimumą (3.25).

Teorema 3.2. (Bartkutė & Sakalauskas (2007a)). *Tegul yra išpildytos teoremos 3.1 sąlygos, be to funkcija $f(x)$ yra pusiauglodi ir taške x^* turi vienintelį minimumą, kuris yra aštrus su konstanta $\mu>0$. Tarkime, kad sąlyga $\lim_{\sigma\to 0}\frac{\partial^2\overline{h}(y,\sigma)}{\partial y^2}=\frac{\partial^2\overline{h}(y)}{\partial y^2}>0$ yra išpildyta visiems y iš tam tikros mažos taško y^* aplinkos.*

Jei

$$\rho_t=\frac{a}{t},\ a>0,\ \sigma_t=\frac{b}{t^\beta},\ b>0,\ 0<\beta<1,\ \frac{a}{b}>\frac{1+\beta}{2\cdot H},$$

tada

$$E\left(\left\|x^{t+1}-x^{*}_{\sigma t+1}\right\|^{2}\right)\leq\frac{A\cdot K^{2}\cdot a\cdot b}{H}\cdot\frac{1}{t^{1+\beta}}+o\left(\frac{1}{t^{\frac{2aH}{b}}}\right), \tag{3.30}$$

kai $t\to\infty$*, čia* x^{t+1} *yra apibrėžiamas pagal (2.14),* y^{*} *funkcijos* $\overline{h}(y)$ *minimumo taškas,*

$$H=\left\|\left(\left.\frac{\partial^{2}\overline{h}(y)}{\partial y^{2}}\right|_{y=y^{*}}\right)^{-1}\right\|^{-1},\qquad A=\int_{dom(P)}\|y\|^{2}\cdot\frac{\|\partial p(y)\|^{2}}{\psi(y)}\,dy.$$

Įrodymas.

⊳ Kadangi tarėme, kad tikslo funkcija turi tiktai aštrų minimumą, seka (2.14) konverguoja b.t. į šį minimumo tašką. Iš (2.14) ir optimalumo sąlygos $\overline{g}\left(x^{*}_{\sigma t},\sigma_{t}\right)=0$ turime, kad:

$$\left\|x^{t+1}-x^{*}_{\sigma t+1}\right\|^{2}=\left\|x^{t}-x^{*}_{\sigma t+1}-\rho_{t}\cdot\left(\overline{g}\left(x^{t},\sigma_{t}\right)-\overline{g}\left(x^{*}_{\sigma t},\sigma_{t}\right)\right)-\rho_{t}\cdot\left(g\left(x^{t},\sigma_{t},\xi_{t+1}\right)-\overline{g}\left(x^{t},\sigma_{t}\right)\right)\right\|^{2}=$$

$$=\left\|x^{t}-x^{*}_{\sigma t+1}-\rho_{t}\cdot\left(\overline{g}\left(x^{t},\sigma_{t}\right)-\overline{g}\left(x^{*}_{\sigma t},\sigma_{t}\right)\right)\right\|^{2}-$$
$$-2\rho_{t}\cdot\left(g\left(x^{t},\sigma_{t},\xi_{t+1}\right)-\overline{g}\left(x^{t},\sigma_{t}\right)\right)^{T}\cdot\left(x^{t}-x^{*}_{\sigma t+1}-\rho_{t}\cdot\left(\overline{g}\left(x^{t},\sigma_{t}\right)-\overline{g}\left(x^{*}_{\sigma t},\sigma_{t}\right)\right)\right)+$$
$$+\rho_{t}^{2}\cdot\left\|g\left(x^{t},\sigma_{t},\xi_{t+1}\right)-\overline{g}\left(x^{t},\sigma_{t}\right)\right\|^{2}.$$

Suvidurkinę abi šios lygybės puses, dėka (3.6) gausime:

$$E\left(\left\|x^{t+1}-x^{*}_{\sigma t+1}\right\|^{2}|\Theta_{t}\right)=$$
$$=\left\|x^{t}-x^{*}_{\sigma t+1}-\rho_{t}\left(\overline{g}\left(x^{t},\sigma_{t}\right)-\overline{g}\left(x^{*}_{\sigma t},\sigma_{t}\right)\right)\right\|^{2}+\rho_{t}^{2}E\left(\left\|g\left(x^{t},\sigma_{t},\xi_{t+1}\right)-\overline{g}\left(x^{t},\sigma_{t}\right)\right\|^{2}|\Theta_{t}\right)\leq$$
$$\leq\left\|x^{t}-x^{*}_{\sigma t+1}-\rho_{t}\left(\overline{g}\left(x^{t},\sigma_{t}\right)-\overline{g}\left(x^{*}_{\sigma t},\sigma_{t}\right)\right)\right\|^{2}+\rho_{t}^{2}\cdot K^{2}\cdot A.$$

Pastebėsime, kad $\dfrac{\partial^{2}\overline{f}(x,\sigma)}{\partial x^{2}}=\dfrac{1}{\sigma}\cdot\dfrac{\partial^{2}\overline{h}\left(\frac{(x-x^{*})}{\sigma}+x^{*},\sigma\right)}{\partial x^{2}}$.

Kadangi suglodinta funkcija yra du kartus diferencijuojama, tai iš Lagrandžo formulės ir pastarosios lygybės turime įvertį:

$$\left\|\overline{g}\left(x^{t},\sigma_{t}\right)-\overline{g}\left(x^{*}_{\sigma t},\sigma_{t}\right)\right\|^{2}\geq\frac{H_{t}^{2}}{\sigma_{t}^{2}}\cdot\left\|x^{t}-x^{*}_{\sigma t}\right\|^{2},$$

čia $H_{t}=\min\limits_{0\leq\tau\leq1}\left\|\left(\left.\dfrac{\partial^{2}\overline{h}(y,\sigma)}{\partial y^{2}}\right|_{y=\frac{x^{t}-x^{*}+\tau\cdot\left(x^{t}-x^{*}_{\sigma t}\right)}{\sigma_{t}}+x^{*}}\right)^{-1}\right\|^{-1}$.

Iš pastarojo įverčio ir *lemos 3.5* turime:

$$
\begin{aligned}
&E\left(\left\|x^{t+1}-x^*_{\sigma_{t+1}}\right\|^2|\Theta_t\right)\le \\
&\le\left(\left\|x^t-x^*_{\sigma_t}\right\|\cdot\left(1-\frac{\rho_t}{\sigma_t}\cdot H_t\right)+(\sigma_t-\sigma_{t+1})\cdot\left(\left\|y^*-x^*\right\|+o(\sigma_t)\right)\right)^2+\rho_t^2\cdot K^2\cdot A,
\end{aligned}
\tag{3.31}
$$

nes pagal teoremos sąlygą $H_t \to H = \left\|\left(\left.\frac{\partial^2 \bar{h}(y)}{\partial y^2}\right|_{y=y^*}\right)^{-1}\right\|^{-1}$, $t\to\infty$.

Po elementarių pertvarkymų, apskaičiavę abiejų (3.31) pusių vidurkius, gausime tokią nelygybę:

$$
E\left(\left\|x^{t+1}-x^*_{\sigma_{t+1}}\right\|^2\right)\le\left(\sqrt{E\left\|x^t-x^*_{\sigma_t}\right\|^2}\cdot\left(1-\frac{\rho_t}{\sigma_t}\cdot H_t\right)+(\sigma_t-\sigma_{t+1})\cdot\left(\left\|y^*-x^*\right\|+o(\sigma_t)\right)\right)^2+\rho_t^2\cdot K^2\cdot A\,.
$$

Galiausiai, dėka *lemos 3.6* ir pasirinkto teoremos prielaidose σ_t ir ρ_t kitimo dėsnio, gausime įvertį:

$$
E\left(\left\|x^{t+1}-x^*_{\sigma_{t+1}}\right\|^2\right)\le\frac{A\cdot K^2\cdot a\cdot b}{H}\cdot\frac{1}{t^{1+\beta}}+o\left(\frac{1}{t^{\frac{2aH}{b}}}\right). \qquad \triangleleft
$$

Pažymėsime, kad konvergavimo greitis (3.30) nustatytas, kai stochastinis gradientas įvertinamas pagal (3.5) ir tikslo funkcija skaičiuojama be triukšmo. Kai tikslo funkcija yra skaičiuojama su atsitiktine paklaida, konvergavimo greitis iš esmės mažėja (Poliak (1987), Граничин & Поляк (2003) ir kt.).

Gauta konvergavimo greičio įverčio išraiška (3.30) yra konstruktyvi, kadangi konstantos joje yra išreiškiamos per tikslo funkcijos charakteristikas bei metodo parametrus a ir b, kuriuos atitinkamai derinant galima užtikrinti metodo konvergavimą. Kaip matome iš pateiktos konvergavimo greičio formulės, konvergavimo greičio eilė nepriklauso nuo uždavinio dimensijos. Metodo parametrus galima parinkti taip pat nepriklausančius nuo uždavinio dimensijos.

Pateiksime keletą pavyzdžių, kuriuose apskaičiuojamos konvergavimo greičio įverčio konstantos skirtingiems suglodinimo operatoriams.

Pavyzdys 3.5. Tarkime $f(x)=\mu\cdot\|x\|$, $x\in\Re^n$. Tai konvergavimo greičio įverčio konstantos yra lygios:

a) jei suglodinimo tankis yra Gauso:

$$\frac{A\cdot K^2}{H}=\frac{K^2}{\mu^2}\cdot\frac{\int\limits_{\Re^n}\|y\|^2\cdot\|\partial\, ln(p(y))\|^2\cdot p(y)dy}{\int\limits_{\Re^n}\|y\|\cdot\left(\frac{\partial^2 p(y)}{\partial y^2}-\left(\frac{\partial p(y)}{\partial y}\right)\cdot\left(\frac{\partial p(y)}{\partial y}\right)^T\right)_{1,1}\cdot p(y)dy}=\frac{K^2\cdot E\|\xi\|^4}{\mu^2\cdot E\left(\|\xi\|\cdot\left(\xi_1^2-1\right)\right)}=$$

$$=\frac{K^2\cdot\Gamma\left(\frac{n}{2}\right)}{\mu^2\cdot\Gamma\left(\frac{n+1}{2}\right)}\cdot\frac{n^2\cdot(n+2)}{\sqrt{2}}=\frac{K^2\cdot n^{\frac{5}{2}}}{\mu^2}+O(1). \tag{3.32}$$

b) jei suglodinimo tankis yra toks kaip *pavyzdyje 3.1* b):

$$\frac{A\cdot K^2}{H}=\frac{K^2}{\mu^2}\cdot\frac{\int\limits_{\|y\|\le 1}\|y\|^2 dy}{\int\limits_{\|y\|\le 1}\frac{y_1^2}{\|y\|^2}dy}=\frac{K^2\cdot n^2}{\mu^2\cdot(n+2)}=\frac{K^2\cdot n}{\mu^2}+O(1). \tag{3.33}$$

Kaip jau buvo pažymėta, SPSA metodo konvergavimo greitis priklauso nuo suglodinimo operatoriaus. Tačiau klausimas apie optimalaus Lipšico suglodinimo operatoriaus savybes dar nėra ištirtas ir gali būti tolimesnių tyrimų objektas.

3.5. STOCHASTINĖS APROKSIMACIJOS KONVERGAVIMO GREIČIO TYRIMAS KOMPIUTERINIO MODELIAVIMO BŪDU

Optimizavimo metodo konvergavimo greičio tyrimas yra sudėtingas uždavinys, reikalaujantis detalių teorinių tyrimų bei intensyvaus kompiuterinio modeliavimo. Literatūroje galima rasti daug rezultatų skirtų teoriniam stochastinės aproksimacijos konvergavimui tirti, tačiau konvergavimo greičio kompiuterinio modeliavimo studijų yra pateikiama labai nedaug. Šiame skyrelyje pateikiami sukurto SPSA metodo konvergavimo greičio tyrimo Monte-Karlo metodu rezultatai, palyginant juos su kitais stochastinės aproksimacijos metodais. Kompiuterinio skaičiuojamojo eksperimento metu aukščiau išvardintais algoritmais daug kartų minimizuojamos kelios testinės funkcijos. Konvergavimo greičiui testuoti pasirinkta testinių funkcijų su aštriu minimumu klasė $f(x)=\sum_{i=1}^{n}b_i|x_i|+A$, čia b_i atsitiktiniai tolygiai pasiskirstę intervale $[\mu, K]$, $K>\mu>0$, skaičiai. Tokių testinių funkcijų pasirinkimą lėmė tai, kad nediferencijuojamos funkcijos minimumo aplinkoje gali būti aproksimuojamos funkcijomis su aštriu minimumu.

Konvergavimo greičiui palyginti testinės funkcijos su aštriu minimumu buvo minimizuojamos SPSAL, SPSAU ir FDSA metodais *M*=100 kartų, vieno nusileidimo metu atliekant *N*=10000 minimizavimo žingsnių, kai $n=2, 4$ ir $\mu=2$, *K*=5, sekos (2.14) koeficientus nustatant tokiu būdu:

$\rho_t = min\left(c, \frac{a}{t}\right)$, $\sigma_t = min\left(d, \frac{b}{t^{\beta}}\right)$, čia $\beta = 0{,}5,\ 0{,}75,\ 0{,}9$, o a, b, c, d empiriškai parenkami koeficientai skirtingi įvairiems stochastinės aproksimacijos algoritmams. Tiriant šių metodų konvergavimo greitį didelio matavimo uždaviniuose, buvo atliktas skaičiuojamasis eksperimentas kai $n = 100$. Kadangi stochastiniai metodai pasižymi lėtu konvergavimu, reikalaujančiu didelio optimizavimo žingsnių skaičiaus ir tuo pačiu ilgo kompiuterinio laiko tokiems uždaviniams spręsti, pastarasis eksperimentas buvo atliktas, taikant lygiagrečiųjų skaičiavimų klasterį. Šis eksperimentas aprašytas 6 skyriuje, o konvergavimo greičio eilės empiriniai įverčiai pateikti *lentelėje 3.3*.

Kaip matome iš (3.15), (3.16) SPSA metodai stochastiniam įverčiui įvertinti reikalauja skaičiuoti tik dvi tikslo funkcijos reikšmes, o FDSA metodas (3.17) reikalauja tikslo funkciją skaičiuoti $2n$ kartų (Михалевич et al. (1987), Граничин & Поляк (2003) ir t.t.). Tikslo funkcijos skaičiavimų skaičius yra dažnai naudojamas optimizavimo algoritmo sudėtingumo matas. Skaičiavimų sudėtingumą nusakant šiuo matu, palygindami optimizavimo metodų efektyvumą, turime atsižvelgti į tai, kad skirtinguose metoduose per tą patį iteracijų skaičių tikslo funkcija gali būti apskaičiuota skirtingą skaičių kartų. Tokiu būdu, atlikus N žingsnių SPSA metodu tikslo funkciją tenka skaičiuoti $2N$ kartų, o skaičiuojant FDSA metodu toks tikslo funkcijos reikšmių skaičius bus apskaičiuotas per $\frac{N}{n}$ iteracijų.

Žemiau pateiktu algoritmu 3.4 sudaroma lygtis konvergavimo greičio įverčiams gauti mažiausių kvadratų metodu, kuri vėliau sprendžiama dalijimo pusiau metodu (algoritmas 3.5).

***Algoritmas 3.4*: Funkcijos konvergavimo greičio įverčiams gauti mažiausių kvadratų metodu sudarymas**

Tikslas: sudaryti funkciją konvergavimo greičio įverčiams gauti mažiausių kvadratų metodu.
Pradinės sąlygos (*preconditions*)**:** optimizavimo žingsnių skaičius N, einamojo taško atstumo nuo optimalaus taško masyvas $m4$ (žr. algoritmą 4.3, 4 skyriuje), optimizavimo žingsnių kartotinumas d, optimizavimo sekos konverguojančio posekio pradžia d‘.
Galinės sąlygos (*postconditions*)**:** funkcija $F(\gamma)$.

1. sum1 =0, sum2 =0, sum3 =0, sum4 =0;
   Apskaičiuojame sumas
2. **for** i =trunc(d‘/d) to trunc(N/d) **do**
3. sum1 =sum1+ $m4[i] \cdot \frac{\ln(i)}{i^{\gamma}}$; sum2 =sum2+ $\frac{m4[i]}{i^{\gamma}}$; sum3 =sum3+ $\frac{m4[i]}{i^{2\gamma}}$; sum4 =sum4+ $\frac{1}{i^{2\gamma}}$;
4. **done**;
5. Apskaičiuojame funkciją: $F(\gamma) = \frac{sum1 \cdot sum4}{sum2 \cdot sum3} - 1$. ◊

Algoritmas 3.5. **Stochastinės paieškos metodų konvergavimo greičio tyrimas**

Tikslas: siekiant ištirti stochastinės paieškos metodų konvergavimo greitį, dalijimo pusiau metodu sprendžiama lygtis $F(\gamma) = 0$.

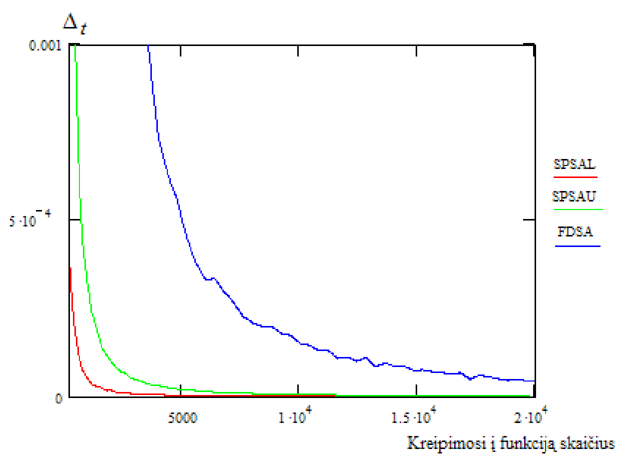
Pradinės sąlygos (*preconditions*)**:** visas optimizavimo žingsnių skaičius *N*, optimizavimo sekos konverguojančio posekio pradžia *d*‘, dalijamo pusiau intervalo pradžia *a* ir pabaiga *b*.

Galinės sąlygos (*postconditions*)**:** empirinė konvergavimo greičio eilė γ.

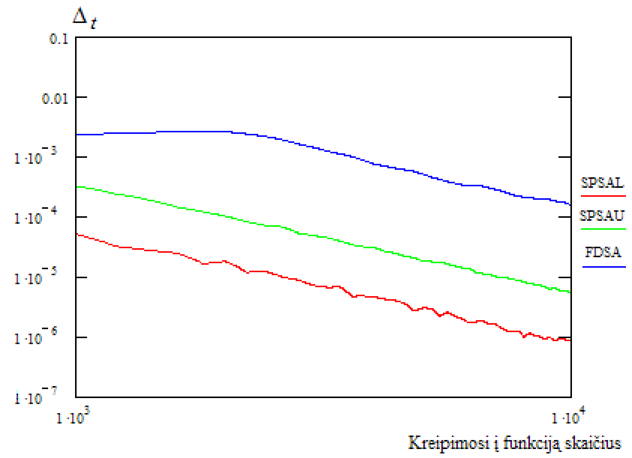
```
a =1;  b =2;  d‘=100;
    while d’ ≤ N  do
        γ = (a + b) / 2
        {Komentaras:  čia a ir b dalijamo pusiau intervalo pradžia ir pabaiga}
        Apskaičiuojame funkcijos reikšmę F(γ) (algoritmas 3.4).
        if b − a > 0.00001  then do
            if F(γ) < 0  then  a = γ
                else  b = γ ;
                    done;
        d‘=d‘+1;
    done. ◊
```

Monte-Karlo įverčio $\Delta_t = E\left\|x^t - x^*\right\|^2$ priklausomybė nuo tikslo funkcijos skaičiavimo skaičiaus, kai $n = 2$, pateikta *pav. 3.4*, ši priklausomybė logaritminėje skalėje pavaizduota *pav. 3.5*. *Lentelėse 3.2* ir *3.3* pateikti konvergavimo greičio eilės teoriniai ir Monte-Karlo empiriniai mažiausių kvadratų įverčiai.

Iš žemiau pateiktų rezultatų matome, kad sukurtas SPSA algoritmas su Lipšico suglodinimo operatoriumi pasirodė esantis ne blogesnis, o kartais net efektyvesnis nei gerai literatūroje žinomas baigtinių skirtumų stochastinės aproksimacijos algoritmas. Taip pat šis metodas nagrinėjamu atveju pasirodė efektyvesnis nei SPSA su tolygiai pasiskirsčiusiu suglodinimo operatoriumi. Efektyvumas ypač pastebimas, esant nedideliam kintamųjų skaičiui ($n = 2, 4$). Nors tiriamų algoritmų konvergavimo greičio eilė teoriškai yra vienoda, tačiau iš pateiktų priklausomybių matome, kad SPSAL metodas konverguoja sparčiau nei SPSAU ar FDSA metodai. Iš *pav. 3.5* pateiktų SPSAL, SPSAU ir FDSA metodų konvergavimo greičio eilės priklausomybių matome, kad gautos kreivės yra artimos tiesinėms funkcijoms, o tai patvirtina gautus teorinius rezultatus, t.y. jei išlogaritmuosime gautą teorinį konvergavimo greičio įvertį (3.30), pastebėsime, kad konvergavimo greičio teorinis įvertis kinta kaip tiesinė funkcija. SPSAL metodo konvergavimo spartumą parodo tai, kad SPSAL metodo konvergavimo greičio eilės įverčio kreivė yra žemiausiai nei kitų palyginimui paimtų metodų. Šių kreivių padėtį įtakoja tam tikros konstantos, išreiškiamos per tikslo funkcijos charakteristikas bei metodo parametrus (*teorema 3.2*). Iš *lentelėse 3.2* ir *3.3* pateiktų stochastinės paieškos metodų konvergavimo greičio eilės įverčių, gautų mažiausių kvadratų metodu matome, kad kompiuterinio modeliavimo metu gauta konvergavimo greičio eilė yra artima teorinei konvergavimo greičio eilei $\gamma = 1 + \beta$, kai sprendžiamo uždavinio dimensija bei β yra keičiami. Tačiau SPSA algoritmų efektyvumo priklausomybė nuo kintamųjų skaičiaus dar reikalauja papildomų tyrimų



Pav. 3.4. SPSAL, SPSAU ir FDSA metodų konvergavimo greičio eilė, n =2, β = 0,9



Pav. 3.5. SPSAL, SPSAU ir FDSA metodų konvergavimo greičio eilė(logaritminėje skalėje), n =2, β = 0,9

Lentelė 3.2. Stochastinės paieškos metodų konvergavimo greičio eilės įverčiai, kai n=2, 4, β = 0,5; 0,75; 0,9

	γ		
	β = 0,5	β = 0,75	β = 0,9
Teorinis: $\gamma = 1 + \beta$	1,5	1,75	1,9
Empirinis			
SPSAL (SPSA algoritmas su Lipšico suglodinimo operatoriumi)			
$n = 2$	1,5139	1,7478	1,9094
$n = 4$	1,4180	1,7443	1,9590
SPSAU (SPSA algoritmas su tolygiai pasiskirsčiusiu suglodinimo operatoriumi)			
$n = 2$	1,5073	1,7348	1,8885
$n = 4$	1,5515	1,7845	1,9981
FDSA (Baigtinių skirtumų stochastinės aproksimacijos algoritmas)			

$n=2$	1,5141	1,7685	1,9172
$n=4$	1,5024	1,7506	1,9062

Lentelė 3.3. Stochastinės paieškos metodų konvergavimo greičio eilės įverčiai, kai n=100, $\beta = 0{,}75$

Teorinis greitis: $\gamma = 1+\beta$	**SPSAL**	**SPSU**	**FDSA**
n=100	1,7055	1,7215	1,7303

3.6. REZULTATAI IR IŠVADOS

1. Lipšico funkcijų klasėje įvestas suglodinimo operatorius su Lipšico tankio funkcija leidžia gauti du kartus diferencijuojamas suglodintas funkcijas.
2. Sukurtas SPSA algoritmas funkcijoms, suglodintoms su Lipšico suglodinimo operatoriumi, optimizuoti, leidžia praplėsti žinomų SPSA metodų klasę.
3. Įrodytas sudaryto algoritmo konvergavimas ir nustatytos algoritmo konvergavimo sąlygos Lipšico funkcijų klasėje, keičiant žingsnio ilgį bei suglodinimo parametrą (*teorema 3.1*).
4. Sudaryto SPSAL algoritmo konvergavimo greitis yra $O\left(\frac{1}{k^{\gamma}}\right)$, $1<\gamma<2$, pusiauglodžių funkcijų klasėje. Konvergavimo greičio eilė nepriklauso nuo uždavinio dimensijos (*teorema 3.2*).
5. Nustatytos konstruktyvios konvergavimo greičio konstantos, išreiškiamos per tikslo funkcijos skaitmenines charakteristikas bei metodo parametrus. Atitinkamai derinant šiuos parametrus, galima gauti konverguojančias sekas.
6. Monte-Karlo metodu įvertinta konvergavimo greičio eilė atitinka teorines išvadas.

4. POZICINIŲ STATISTIKŲ TAIKYMAS TIKSLO FUNKCIJOS MINIMALIAI REIKŠMEI ĮVERTINTI

4.1. ĮVADAS

Optimizavimo uždaviniams spręsti taikomi įvairūs algoritmai, konverguojantys iteracijų skaičiui augant į begalybę. Tačiau praktiniu požiūriu yra svarbu priimti sprendimus apie gauto sprendinio optimalumą bei naudojamo algoritmo stabdymą po baigtinio iteracijų skaičiaus.

Šiame skyriuje nagrinėjamas Markovo tipo algoritmų optimizavimo sekų pozicinių statistikų taikymas optimizuojamos funkcijos minimaliai reikšmei bei jos pasikliautinajam intervalui įvertinti. Šie įverčiai teikia svarbią informaciją optimizavimo algoritmo realizavimo metu, nustatant tolesnio optimizavimo arba stabdymo perspektyvumą.

Žinomi kelių autorių darbai, kuriuose analizuojamas pozicinių statistikų taikymas funkcijos minimaliai (maksimaliai) reikšmei arba funkcijos ekstremumo padėčiai įvertinti. Šios problemos nagrinėjimo vienas iš pradininkų yra J. Mockus (Моцкус (1976)), pritaikęs statistinių išvadų teoriją optimizavimo algoritmų optimalumui tirti. Vėliau teoriniai nagrinėjimai, paremti ekstremaliųjų reikšmių teorija, buvo išplėtoti A. Žilinsko ir A. Zhigljavsky (Жиглявский & Жилинскас (1991)). H. Chen (1996) geriausias funkcijos reikšmes pritaikė maksimumo taškui nustatyti.

Ekstremaliųjų reikšmių teorijos taikymas optimaliems sprendimams priimti remiasi prielaida, kad tikslo funkcijos reikšmių, gautų optimizavimo metu, pozicinės statistikos yra pasiskirsčiusios pagal ekstremaliųjų reikšmių skirstinį. Reikia pastebėti, kad nors šis skirstinys yra labiau išnagrinėtas nepriklausomų vienodai pasiskirsčiusių dydžių atveju, juo paremti modeliai gali būti taikomi ir tais atvejais, kai nebūtinai yra išpildytos nepriklausomumo bei homogeniškumo prielaidos (Галамбош (1984)). Pavyzdžiui, ekstremaliųjų reikšmių ribinis skirstinys gali būti taikomas nepriklausomų atsitiktinių dydžių su skirtingomis dispersijomis ekstremaliųjų reikšmių skirstiniams modeliuoti (David & Nagaraja (2003)).

Kita vertus, yra įrodyta, kad funkcijos reikšmių seka, gauta stochastinės aproksimacijos būdu, asimptotiškai konverguoja į nepriklausomų Gauso dydžių su skirtingomis dipersijomis seką.

Teorema (Wasan (1969), Dupac (1988)): tegul seka $\{x^t\}$ yra konstruojama pagal (2.14). Jei g^t yra stochastinis gradientas, suglodinta funkcija yra glodžiai diferencijuojama ir $\rho_t = \frac{a}{t}$, $0 < a_1 \le a \le a_2$, a_1, a_2 tam tikros konstantos, tai $\sqrt{t} \cdot \left(x^t - x^*\right)$ yra vektorius asimptotiškai pasiskirstęs pagal normalųjį dėsnį su nuliniu vidurkių vektoriumi bei tam tikra kovariacine matrica.

Be to, SPSA metodo su Bernulio suglodinimo operatoriumi asimptotinį konvergavimą į normalųjį dėsnį yra įrodęs J. Spall (2003).

Taigi, iš gerai žinomo paminėto fakto, kad atsitiktinių dydžių su skirtingomis dispersijomis ekstremaliųjų reikšmių skirstinys yra Veibulo (David & Nagaraja (2003)) ir pateiktos teoremos seka, kad teorinių išvadų, prieštaraujančių ekstremaliųjų reikšmių skirstinio taikymui optimaliems sprendimams priimti, nėra žinoma. Be to, pozicinių statistikų, gautų optimizavimo metu, skirstinys nėra teoriškai išnagrinėtas. Todėl kompiuterinio modeliavimo taikymas yra vienas iš būdų patikrinti ekstremaliųjų reikšmių tinkamumą optimizavimo metodų optimalumui tirti.

Norint taikyti ekstremaliųjų reikšmių skirstinius optimizavimo uždavinių sprendinio optimalumui tirti, reikia žinoti šių skirstinių parametrus, kurių svarbiausias yra skirstinio formos parametras α. Daugelyje optimizavimo uždavinių šį parametrą galima parinkti remiantis funkcijos homogeniškumu optimalaus taško aplinkoje (Haan (1981), Жиглявский (1985)). Kita vertus, šį parametrą taip pat galima įvertinti nagrinėjant Veibulo skirstinio parametrų statistinio vertinimo uždavinį (žr. 5 skyrių).

Toliau nagrinėsime pozicinių statistikų taikymą tikslo funkcijos minimalios reikšmės įverčiams gauti, kai ekstremaliųjų reikšmių skirstinio formos parametras yra žinomas ir nustatomas pasinaudojant funkcijos homogeniškumu ekstremumo aplinkoje. Pažymėsime, kad tikslo funkcijos pozicinių statistikų, gautų optimizavimo metu, skirstinių analitinis tyrimas yra sudėtingas, dar neišspręstas uždavinys. Todėl sukurtų metodų tinkamumą tikslo funkcijos minimaliai reikšmei įvertinti Markovo tipo optimizavimo algoritmuose ištirsime kompiuterinio modeliavimo būdu. Gauti rezultatai buvo pristatyti tarptautinėse konferencijose: *VII tarptautinėje konferencijoje "Computer data analysis and modeling"* (2004, Minskas, Baltarusija), *tarptautinio statistikos instituto* (*ISI*) *55-oje sesijoje* (2005, Sidnėjus, Australija), *seminare Workshop of European Chapter on Metaheuristics "Metaheuristics and Large Scale Optimization"* (2005, Vilnius), *"Kompiuterininkų dienos-2005"* (KU, Klaipėda), *Operacijų tyrimo 21-oje konferencijoje* (*EURO XXI*) (2006, Reikjavike, Islandijoje) bei publikuoti moksliniuose leidiniuose: Bartkutė & Sakalauskas (2004a), Bartkutė & Sakalauskas (2005a), Bartkutė & Sakalauskas (2005b), Bartkutė et al. (2006), Bartkutė & Sakalauskas (2006b).

4.2. OPTIMIZAVIMO SEKOS OPTIMALUMO TYRIMO METODAS

Stochastiniuose ir euristiniuose optimizavimo algoritmuose dažna yra stabdymo problema. Pasinaudodami pozicinėmis statistikomis įvesime taisyklę Markovo tipo algoritmams stabdyti. Trumpai apžvelgsime metodą, leidžiantį įvertinti minimalios reikšmės pasikliautinąjį intervalą naudojant pozicines statistikas, kai iteracijų skaičius N yra baigtinis.

Tarkime, turime seką $H = \{\eta_1, \dots, \eta_N\}$, kurios elementai yra funkcijos reikšmės $\eta_k = f(x_k)$, gautos optimizavimo metu. Norint įvertinti parametro $\min_{x \in \Re^n} f(x) = f(x^*) = A$ pasikliautinąjį intervalą, iš sekos H pakanka išrinkti tiktai k+1 pozicinių statistikų: $\eta_{(0)}, \dots, \eta_{(k)}$, čia $k = k(N)$, $\frac{k^2}{N} \to 0$, $N \to +\infty$. Toliau panagrinėsime pozicinių statistikų taikymą minimalios reikšmės tiesiniam įverčiui bei tikslo funkcijos minimalios reikšmės pasikliautinajam intervalui gauti. Šie įverčiai yra gana paprasti ir priklauso tik nuo vieno parametro: ekstremaliųjų reikšmių skirstinio formos parametro α. Šis parametras gali būti nustatomas dviem būdais:

- pasinaudojant funkcijos homogeniškumu minimumo aplinkoje;
- įvertinant statistiniais metodais.

Šio parametro įvertinimą pasinaudojant funkcijos homogeniškumu išnagrinėsime skyrelyje 4.2.3, o jo statistinius įverčius pateiksime 5 skyriuje.

4.2.1. Tikslo funkcijos minimalios reikšmės tiesinis įvertis

Tikslo funkcijos minimaliai reikšmei įvertinti galima pasinaudoti skirstinio ribinės reikšmės įvertinimo metodu (*end-point estimation* (Hall (1982))). Tada tiesinis parametro A įvertis yra (Hall (1982), Bartkutė & Sakalauskas (2004a)):

$$A_{N,k} = \sum_{i=0}^{k} a_i \eta_{(i)}, \tag{4.1}$$

čia k daug mažesnis nei N, $a_0, \dots, a_k$ koeficientai, tenkinantys lygybę $\sum_{i=0}^{k} a_i = 1$.

Tiesiniam parametro A įverčiui skaičiuoti koeficientų a_i seką galime apibrėžti trimis skirtingais būdais (Bartkutė & Sakalauskas (2004a)):

I būdas $\quad a = (1 + c_k,\ 0,\ \dots,\ 0\ ,\ -c_k)$,

čia $c_k = \dfrac{\Gamma\left(1 + \frac{1}{\alpha}\right)}{\dfrac{\Gamma\left(1 + k + \frac{1}{\alpha}\right)}{k!} - \Gamma\left(1 + \frac{1}{\alpha}\right)}$, α ekstremaliųjų reikšmių skirstinio formos parametras.

Po algebrinių pertvarkymų galime gauti patogesnę skaičiavimams koeficiento c_k išraišką:

$$c_k = \frac{1}{\prod_{i=0}^{k}\left(1 + \frac{1}{i \cdot \alpha}\right) - 1}. \tag{4.2}$$

Paprasta šių koeficientų išraiška, nagrinėjama (Hall (1982)), aproksimuoja (4.2), kai $\alpha \approx 1$:

$$c_k^* = \frac{2\alpha - 1}{k}. \tag{4.3}$$

II būdas $\quad a = \left(\lambda^T \Lambda^{-1} \lambda\right)^{-1} \Lambda^{-1} \lambda$, čia $\lambda = (1, 1, \ldots, 1)^T$,

čia Λ - simetrinė matrica, kurios elementai apskaičiuojami taip $\lambda_{ij} = \dfrac{\Gamma\left(\frac{2}{\alpha} + i + 1\right)\Gamma\left(\frac{1}{\alpha} + j + 1\right)}{\Gamma\left(\frac{1}{\alpha} + i + 1\right)\Gamma(j+1)}$, $(j \geq i)$, α ekstremaliųjų reikšmių skirstinio formos parametras.

III būdas $\quad a = \Lambda^{-1}\left(\dfrac{\left(b^T \Lambda^{-1} b\right)\lambda - \left(\lambda^T \Lambda^{-1} b\right)b}{\left(b^T \Lambda^{-1} b\right)\left(\lambda^T \Lambda^{-1} \lambda\right) - \left(\lambda^T \Lambda^{-1} b\right)^2}\right)$,

čia $b = (b_0,\ b_1,\ \ldots,\ b_k)$, $b_i = \dfrac{\Gamma\left(\frac{1}{\alpha} + i + 1\right)}{\Gamma(i+1)}$, α ekstremaliųjų reikšmių skirstinio formos parametras.

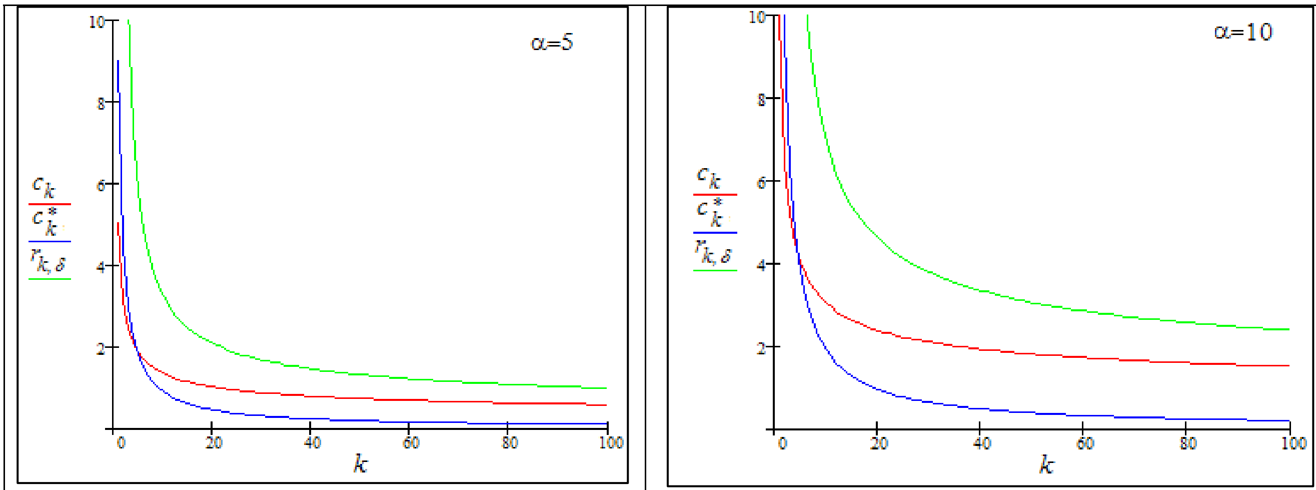
4.2.2. Tikslo funkcijos minimalios reikšmės pasikliautinasis intervalas

Dažnai optimizavimo uždaviniuose ekstremalioms reikšmėms vertinti naudojami vienpusiai pasikliautinieji intervalai (Жиглявский (1985), Bartkutė & Sakalauskas (2004a)). Tada tikslo funkcijos minimalios reikšmės vienpusis pasikliautinasis intervalas gali būti įvertinamas tokiu būdu:

$$[\,\eta_{(0)} - r_{k,\delta} \cdot \left(\eta_{(k)} - \eta_{(0)}\right), \eta_{(0)}\,], \tag{4.4}$$

čia $r_{k,\delta} = \dfrac{\left(1 - \delta^{\frac{1}{k}}\right)^{\frac{1}{\alpha}}}{1 - \left(1 - \delta^{\frac{1}{k}}\right)^{\frac{1}{\alpha}}}$, δ pasikliovimo lygmuo.

Pav. 4.1 pateiktos koeficientų c_k, c_k^* ir $r_{k,\delta}$ priklausomybės, priklausomai nuo pozicinių statistikų skaičiaus *k*, kai $\alpha = 5; 10,\ \delta = 0{,}95$. Iš šių priklausomybių matome, kad kai pozicinių statistikų skaičius yra didelis, tikslo funkcijos minimalios reikšmės įvertis $A_{N,k} = \eta_{(0)} - c_k\left(\eta_{(k)} - \eta_{(0)}\right)$, gaunamas skaičiuojant koeficientus c_k pagal (4.3) formulę, yra arti geriausios pasiektos tikslo funkcijos reikšmės $\eta_{(0)}$, o tuo pačiu ir pasikliautinojo intervalo apatiniojo rėžio. Didinant parametrą α, ši tendencija dar labiau ryškėja. Todėl tikslo funkcijos minimalią reikšmę galima įvertinti tiksliau, taikant įvertį, gautą skaičiuojant koeficientus c_k pagal (4.2) formulę.



Pav. 4.1. Koeficientų c_k, c_k^* ir $r_{k,\delta}$ priklausomybės, didėjant pozicinių statistikų skaičiui, kai

$\alpha = 5; 10,\ \ \delta = 0{,}95$

Tikslo funkcijos minimalios reikšmės dvipusis pasikliautinasis intervalas (Жиглявский (1985)):

$$[A_{N,k} - \varepsilon, A_{N,k} + \varepsilon], \tag{4.5}$$

čia $\varepsilon = k^{\frac{1}{\alpha}} \cdot \left(\eta_{(k)} - \eta_{(0)}\right) \cdot \sqrt{\dfrac{a^T \cdot \Lambda \cdot a}{\delta}}$, Λ ir a apskaičiuojami pagal 76 psl. pateiktas formules.

4.2.3. Homogeniškų funkcijų ekstremaliųjų reikšmių skirstinio formos parametras

Norint pasinaudoti skyrelių 4.2.1 ir 4.2.2 formulėmis tiesiniam įverčiui ir pasikliautiniesiems intervalams įvertinti, reikia žinoti ekstremaliųjų reikšmių skirstinio formos parametrą α. Tegul tikslo funkcija yra homogeniška minimumo taško aplinkoje su parametru φ:

$$\left| f(x) - f\left(x^*\right) \right| = O\left(\left\| x - x^* \right\|^{\varphi} \right). \tag{4.6}$$

Tuomet ekstremaliųjų reikšmių skirstinio parametrą galime įvertinti pasinaudodami funkcijos homogeniškumu:

$$\lim_{N \to \infty} P\left(a_N \cdot \min_i \left(\eta_{i,N} - A \right) \le y \right) = 1 - e^{-y^{\alpha}} \tag{4.7}$$

čia a_N normavimo konstanta.

Tada

$$\alpha = \frac{n}{\varphi}, \tag{4.8}$$

φ funkcijos homogeniškumo parametras, taškai (2.3), kuriuose apskaičiuotos funkcijos reikšmės, yra nepriklausomi vienodai pasiskirstę bei yra išpildytos dar tam tikros papildomos sąlygos (Haan (1981), Жиглявский (1985)).

Iš (4.6) gauname, kad jei funkcija turi aštrų minimumą (3.25), tai $\varphi = 1$, $\alpha = n$.

Funkciją minimumo aplinkoje išskleisime Teiloro eilute:

$f(x) = A + \frac{(x - x^*) \cdot f''(x^*) \cdot (x - x^*)}{2} + o\left(\|x - x^*\|^2\right)$, $f''(x^*)$ antros eilės išvestinių matrica ir $f'(x^*) = 0$. Jei minimumo taškas yra vienintelis, tai matrica $f''(x^*)$ yra teigiamai apibrėžta. Pastaruoju atveju galima parinkti kintamųjų mastelius, kad tikslo funkcija būtų homogeniška minimumo aplinkoje su homogeniškumo parametru $\varphi = 2$. Tokiu atveju $\alpha = \frac{n}{2}$.

4.3. KOMPIUTERINIO MODELIAVIMO REZULTATAI

Pozicinių statistikų (2.16) skirstinio teorinis tyrimas, reikalingas sukurtam metodui pagrįsti, yra sudėtingas uždavinys. Todėl funkcijos minimalios reikšmės ir jos pasikliautinojo intervalo įverčių, t. y. (4.1), (4.4) ir (4.5), tinkamumas algoritmo stabdymo taisyklei sudaryti buvo tiriamas Monte-Karlo metodu.

4.3.1. Testinių funkcijų sąrašas

Kompiuterinio skaičiuojamojo eksperimento metu įvairios testinės funkcijos buvo daug kartų minimizuojamos keliais Markovo tipo algoritmais (SPSAL, SPSAU, FDSA bei modeliuojamojo atkaitinimo). Vieną ekstremumą turinčios tikslo funkcijos (funkcijų su aštriu minimumu klasė, CB3, Rozen Suzuki) buvo minimizuojamos stochastinės aproksimacijos metodais, o daug ekstremumų turinčios funkcijos (Branino, Beale, Rastrigino) buvo minimizuojamos modeliuojamojo atkaitinimo metodu.

Testinė funkcija 1. (Funkcijų su aštriu minimumu klasė) (Bartkutė & Sakalauskas (2006b))

$f(x) = \sum_{k=1}^{n} b_k |x_k| + A$, n = 2, 4, 100, minimali reikšmė 0, pradinis taškas $\overline{x_i} = 1, i = 1,...,n$, koeficientai b_k yra tolygiai pasiskirstę intervale [μ, K].

Testinė funkcija 2. (CB3) (Luksan & Vlcek (1999))

$f(x) = max\left(x_1^4 + x_2^2, (2 - x_1)^2 + (2 - x_2)^2, 2e^{-x_1 + x_2}\right)$, n = 2, minimali reikšmė 2, pradinis taškas $\overline{x_1} = 2, \overline{x_2} = 2$.

Testinė funkcija 3. (Rosen Suzuki) (Luksan & Vlcek (1999))

$$f(x) = \max\{f_1(x), f_1(x)+10f_2(x), f_1(x)+10f_3(x), f_1(x)+10f_4(x)\},$$

$$f_1(x) = x_1^2 + x_2^2 + 2x_3^2 + x_4^2 - 5x_1 - 5x_2 - 21x_3 + 7x_4;$$

$$f_2(x) = x_1^2 + x_2^2 + x_3^2 + x_4^2 + x_1 - x_2 + x_3 - x_4 - 8;$$

$$f_3(x) = x_1^2 + 2x_2^2 + x_3^2 + 2x_4^2 - x_1 - x_4 - 10;$$

$f_4(x) = x_1^2 + x_2^2 + x_3^2 + 2x_1 - x_2 - x_4 - 5$, $n = 4$, minimali reikšmė -44, pradinis taškas $\overline{x_i} = 0,\ i = 1,...,n$, minimumo taškas $x^* = (0, 1, 2, -1)$.

Testinė funkcija 4. (Branino)

$f(x) = \left(x_2 - \frac{5}{4\pi^2}x_1^2 + \frac{5}{\pi}x_1 - 6\right)^2 + 10\left(1 - \frac{1}{8\pi}\right)\cos(x_1) + 10$, n=2, leistinoji sritis $-5 \le x_1 \le 10,\quad 0 \le x_2 \le 15$ ir globalus minimumas 0,397887.

Testinė funkcija 5. (Beale)

$$f(x_1, x_2) = (1{,}5 - x_1 + x_1 \cdot x_2)^2 + \left(2{,}25 - x_1 + x_1 \cdot x_2^2\right)^2 + \left(2{,}625 - x_1 + x_1 \cdot x_2^3\right)^2,$$

leistinoji sritis $-4{,}5 \le x_1, x_2 \le 4{,}5$, pradinis taškas $\overline{x_i} = 1,\ i = 1,...,n$, globalus minimumas 0.

Testinė funkcija 6. (Rastrigino)

$f(x) = 10n + \sum_{i=1}^{n}\left(x_i^2 - 10\cos(2\pi \cdot x_i)\right)$, leistinoji sritis $-5{,}12 \le x_i \le 5{,}12,\ n = 2$, globalus minimumas 0.

4.3.2. Hipotezės apie Pareto skirstinio tinkamumą tikrinimas

Panagrinėsime statistikos

$$D_{N,k} = \frac{\eta_{0,N} - A}{\eta_{k,N} - \eta_{0,N}} \tag{4.9}$$

skirstinį. Pasinaudoję Renji sąryšiu (Галамбош (1984)), gauname, kad atsitiktinis dydis $X = \left(\eta_{0,N} - A\right)^{\frac{1}{\alpha}}$ yra asimptotiškai pasiskirstęs pagal eksponentinį dėsnį, o atsitiktinis dydis $Y = \left(\eta_{k,N}\right)^{\frac{1}{\alpha}}$ pasiskirstęs pagal Erlango *k+1* stadijų skirstinį (Johnson & Kotz (1970)), čia *X* ir *Y* yra nepriklausomi. Įsitikinsime, kad statistika $D_{N,k}$ yra asimptotiškai pasiskirsčiusi pagal Pareto skirstinį. Pažymėję $w = \left(1 + \frac{1}{u}\right)^{\alpha} - 1,\ w > 0$, turime:

$$F_k(u) = P(D_{N,k} \le u) = P\left(\frac{X}{Y} \le u\right) = P\left(\frac{Y}{X} > w-1\right) = \int_0^{\infty}\int_{w\cdot x}^{\infty} e^{-x} \cdot \frac{y^{k-1}}{(k-1)!} \cdot e^{-y} dxdy = \begin{bmatrix} y = xt \\ dy = xdt \end{bmatrix} =$$

$$= \int_0^{\infty}\int_{w}^{\infty} e^{-x} \cdot \frac{(xt)^{k-1}}{(k-1)!} \cdot e^{-xt} x \cdot dxdt = \int_w^{\infty} \frac{t^{k-1}}{(k-1)!} \int_0^{\infty} e^{-x(1+t)} \cdot x^k dxdt = \begin{bmatrix} x(1+t) = v \\ dx = \dfrac{dv}{1+t} \end{bmatrix} =$$

$$= \frac{1}{(k-1)!} \int_w^{\infty} t^{k-1} \cdot \left(\frac{1}{1+t}\right)^{k+1} \int_0^{\infty} e^{-v} \cdot v^k dvdt = k \cdot \int_w^{\infty} t^{k-1} \cdot \left(\frac{1}{1+t}\right)^{k+1} dt = 1 - \left(\frac{w}{w+1}\right)^k .$$

Iš čia gauname, kad

$$F_k(u) = 1 - \left(1 - \left(\frac{u}{1+u}\right)^{\alpha}\right)^k, u \ge 0, \tag{4.10}$$

čia $w = \left(1 + \frac{1}{u}\right)^{\frac{1}{\alpha}}$.

Norėdami patikrinti ar statistika $D_{N,k}$, kai pozicinės statistikos yra gautos stochastinės aproksimacijos metodais, yra pasiskirsčiusi pagal skirstinį (4.10), sekantį iš ekstremaliųjų reikšmių teorijos nepriklausomiems a. d., tikrinsime statistinę hipotezę

$$\mathrm{H_0}: \ P\left(\frac{\eta_{0,N} - A}{\eta_{k,N} - \eta_{0,N}} \le u\right) = F_k(u).$$

Šią hipotezę tikrinsime pasinaudodami ω^2 kriterijumi. Skaičiuojamojo eksperimento metu kiekviena testinė funkcija buvo optimizuojama keliais disertacijoje nagrinėjamais optimizavimo metodais M=500 kartų po N=10000 žingsnių. Kiekvienu atveju statistinės išvados buvo daromos naudojant statistinę imtį iš 500 statistikų $D_{N,k}$, gautų kiekvieno nusileidimo pabaigoje. *Lentelėje 4.1.* pateikiami empirinio kriterijaus ω^2 tyrimo rezultatai kelių stochastinės paieškos metodų atveju, kai $n \le 100$, o kritinė hipotezės atmetimo reikšmė lygi 0,46 (reikšmingumo lygmuo 0,05).

Šio uždavinio sprendimo etapai aprašyti algoritme 4.1.

Algoritmas 4.1: **Hipotezės apie Pareto skirstinio tinkamumą tikrinimas**

Tikslas: patikrinti hipotezę apie statistikos $D_{N,k}$ pasiskirstymą pagal Pareto dėsnį.

Pradinės sąlygos (*preconditions*)**:** optimizavimo kartojimų (Monte-Karlo imčių) skaičius M, pozicinių statistikų skaičius k, ekstremaliųjų reikšmių skirstinio parametras α.

Galinės sąlygos (*postconditions*)**:** empirinio $\boldsymbol{\omega^2}$ kriterijaus reikšmė.

1. Sudarome statistinę imtį.
2. **for** i =1 iki M **do**

3. pagal algoritmus 2.2 arba 2.3 sudarome geriausių pasiektų tikslo funkcijos reikšmių masyvą η^t.

4. pagal (4.9) apskaičiuojame statistiką $D_{N,k}^i$.

5. **done.**

6. Išrikiuojame gautą statistikų seką;

{*Komentaras*: Suformuluojame statistinę hipotezę: H$_0$: $D_{N,k} \sim F_k(u)$,

H$_a$: $D_{N,k} \nsim F_k(u)$.}

7. Apskaičiuojame kriterijų:

$$\omega_{empyr}^2 = \frac{1}{12 \cdot M} + \sum_{i=1}^{M} \left(1 - \left(1 - \left(\frac{D_{N,k}^i}{1 + D_{N,k}^i} \right)^{\alpha} \right)^k - \frac{i - 0.5}{M} \right)^2 .$$

{*Komentaras*: priimame sprendimą - hipotezė H$_0$: atmetama, jeigu $\omega_{empyr}^2 > \omega_{krit}^2 = 0{,}46$, ir hipotezė H$_0$: neatmetama, jeigu $\omega_{empyr}^2 \le \omega_{krit}^2 = 0{,}46$, kai reikšmingumo lygmuo δ=0,05}◊

Lentelė 4.1. Hipotezės apie Pareto skirstinio tinkamumą tikrinimo rezultatai (N=10000, M=500)

Testinės funkcijos	***Algoritmas***	***ω^2 kriterijus (kritinė reikšmė 0,46, δ=0,05)***		
Funkcijų su aštriu minimumu klasė μ=2, K=5		***n*=2**	***n*=10**	***n*=100**
	SPSAL	0,4149	0,3137	0,2706
	SPSAU	0,0714	0,1658	0,1842
	FDSA	0,1234	0,3818	0,2611
CB3	SPSAL	0,2792		
	SPSAU	0,0861		
	FDSA	0,2828		
Rosen Suzuki	SPSAL	0,3843		
	SPSAU	0,3209		
	FDSA	0,4245		
Branino	SA	0,14972		
Beale		0,11049		
Rastrigino		0,22196		

Taigi, *lentelėje 4.1* pateiktos kompiuterinio modeliavimo būdu gautos empirinio kriterijaus ω^2 reikšmės neprieštarauja hipotezei, apie statistikos $D_{N,k}$ pasiskirstymą pagal Pareto dėsnį.

4.3.3. Tikslo funkcijos minimalios reikšmės ir jos pasikliautinojo intervalo įverčių tyrimas Monte-Karlo metodu

Kompiuterinio skaičiuojamojo eksperimento metu kiekviena 4.3.1 skyrelyje pateikta testinė funkcija buvo minimizuojama M=500 kartų, vieno nusileidimo metu buvo atliekama N=10000 minimizavimo žingsnių, optimizavimo sekos (2.14) koeficientai buvo parenkami tokiu būdu: $\rho_t = n \cdot min\left(a; \frac{b}{t}\right)$, $\sigma_t = \sqrt{\frac{(n+2)(n+3)}{n \cdot (n+1)}} \cdot min\left(c; \frac{d}{t^{\beta}}\right)$, čia $\beta = 0{,}75$, koeficientai a, b, c, d parenkami skirtingai įvairiems stochastinės aproksimacijos algoritmams.

Tikslo funkcijos minimalios reikšmės bei jos pasikliautinojo intervalo vertinimas aprašytas algoritme 4.2.

Algoritmas 4.2: **Tikslo funkcijos minimalios reikšmės ir jos pasikliautinojo intervalo įvertinimo algoritmas**

Tikslas: įvertinti tikslo funkcijos minimalią reikšmę bei jos pasikliautinąjį intervalą.

Pradinės sąlygos (*preconditions*)**:** iteracijos numeris t, geriausių optimizavimo metu iki iteracijos t pasiektų tikslo funkcijos reikšmių masyvas η^t, šių reikšmių skaičius k.

Galinės sąlygos (*postconditions*)**:** minimalios tikslo funkcijos reikšmės ir jos pasikliautinojo intervalo įverčiai $A_{N,k}$, $A_{apat.}$, $A_{virš}$.

1. pagal formules (4.1) ir (4.2) apskaičiuojame tikslo funkcijos minimalios reikšmės tiesinį įvertį $A_{N,k}$;
2. pagal formulę (4.4) apskaičiuojame minimalios reikšmės pasikliautinojo intervalo viršutinįjį ir apatinįjį rėžius [$A_{apat.}$; $A_{virš}$].

   {*Komentaras*: stabdymo taisyklė – algoritmas yra stabdomas, jei $A_{virš.} - A_{apat.} \le \varepsilon$ ir priimamas sprendimas, kad minimalios reikšmės pasikliautinasis intervalas yra [$A_{apat.}$; $A_{virš}$], o minimalios reikšmės įvertis $A_{N,k}$.} ◊

Tikslo funkcijos minimalios reikšmės ir jos pasikliautinojo intervalo įverčių tyrimas Monte-Karlo metodu pateiktas algoritme 4.3. Tyrimo metu buvo tiriamos tikslo funkcijos minimalios reikšmės, jos pasikliautinojo intervalo įverčių bei minimalios reikšmės patekimo į pasikliautinąjį intervalą tikimybės ir jos pasikliautinojo intervalo priklausomybės nuo iteracijų skaičiaus. Tikimybės pasikliautinojo intervalo viršutinis ir apatinis rėžiai gaunami sprendžiant lygtis (Большев & Смирнов (1983)):

$$B\left(A_{apat.}, l, M - l + 1\right) = 1 - \delta, \tag{4.11}$$

$$B\left(A_{virs.}, l+1, M-l\right)=\delta, \tag{4.12}$$

čia δ yra pasikliovimo lygmuo, l tikslo funkcijos minimalios reikšmės patekimo į pasikliautinąjį intervalą skaičius, $B(x, \iota, \theta)$ beta skirstinys su parametrais ι, θ.

Tikslo funkcijos minimalios reikšmės ir jos pasikliautinojo intervalo įverčių tyrimas Monte-Karlo metodu aprašytas algoritme 4.3.

Algoritmas 4.3: **Tikslo funkcijos minimalios reikšmės ir jos pasikliautinojo intervalo įverčių tyrimo Monte-Karlo metodu algoritmas**

Tikslas: sudaryti tikslo funkcijos minimalios reikšmės, jos pasikliautinojo intervalo įverčių bei minimalios reikšmės patekimo į pasikliautinąjį intervalą tikimybės ir jos pasikliautinojo intervalo priklausomybes nuo iteracijų skaičiaus.

Pradinės sąlygos (*preconditions*)**:** tikslo funkcijos optimizavimo algoritmas, tikslo funkcijos minimalios reikšmės ir jos pasikliautinojo intervalo įvertinimo algoritmas 4.2, iteracijos numeris *t*, geriausių optimizavimo metu pasiektų tikslo funkcijos reikšmių masyvas η^t, šių reikšmių skaičius *k*, tikslo funkcijos minimali reikšmė *A* bei minimumo taškas x^*.

{*Komentaras*: tikslo funkcijai optimizuoti gali būti taikomas algoritmas 2.2 arba 2.3}.

Galinės sąlygos (*postconditions*)**:** tikslo funkcijos minimalios reikšmės vidurkio masyvas $m1$, minimalios reikšmės pasikliautinojo intervalo įverčių vidurkių masyvai $m2$ ir $m3$ bei einamojo taško atstumo iki optimalaus taško kvadrato vidurkio masyvas *m4*, o taip pat minimalios reikšmės patekimo į pasikliautinąjį intervalą tikimybės masyvas *m5* bei šios tikimybės pasikliautinojo intervalo apatis ir viršutinis rėžiai $m6[t]$ ir $m7[t]$.

for *t*=1 iki *N* **do**
    $m1[t]$= 0, $m2[t]$= 0, $m3[t]$= 0, $m4[t]$= 0, $m5[t]$= 0;
  done
for *i*=1 to *M* **do**
  for *t*=1 iki *N* **do**
    sudaryti geriausių optimizavimo metu pasiektų iki iteracijos *t* tikslo funkcijos reikšmių masyvą η^t;
    pagal (4.1) ir (4.4) apskaičiuoti tiesinį minimalios reikšmės įvertį $A_{N,k}$ bei pasikliautinojo intervalo viršutinįjį ir apatinįjį rėžius;
    $m1[t]$= $m1[t]$+ $A_{N,k}$; $m2[t]$= $m2[t]$+$A_{\text{apat.}}$; $m3[t]$= $m3[t]$+$A_{\text{virš.}}$;

$m4[t] = m4[t] + \left\| x^t - x^* \right\|^2$;

if $A \le \eta_{(0)} - r_{k,\delta} \cdot \left(\eta_{(k)} - \eta_{(0)} \right)$ **then** $m5[t] = m5[t] + 1$;

done

done

for t=1 iki N **do**

pagal formules (4.11) ir (4.12), kai $l = m5[t]$, apskaičiuojame minimalios reikšmės patekimo į jos pasikliautinąjį intervalą tikimybės apatinįjį ir viršutinįjį rėžius $m6[t]$, $m7[t]$;

$m1[t] = \dfrac{m1[t]}{M}$, $m2[t] = \dfrac{m2[t]}{M}$, $m3[t] = \dfrac{m3[t]}{M}$, $m4[t] = \dfrac{m4[t]}{M}$; $m5[t] = \dfrac{m5[t]}{M}$;

done. ◊

Lentelėje 4.2 pateikti minimalios reikšmės A tiesinių įverčių $A_{N,k}$ (4.1) vidurkiai, dvipusio pasikliautinojo intervalo (4.5) Monte-Karlo įverčių vidurkiai, tikimybės patekti tikslo funkcijos minimaliai reikšmei į pasikliautinąjį intervalą (4.4) Monte-Karlo įverčiai p bei jos pasikliautinojo intervalo rėžiai, kai tiesiniam parametro A įverčiui skaičiuoti koeficientų a_i seka yra skaičiuojama pagal aukščiau pateiktus 3 būdus (75-76 psl.).

Lentelė 4.2. Funkcijos su aštriu minimumu (n=2) minimalios reikšmės tiesinių įverčių bei dvipusio pasikliautinojo intervalo Monte-Karlo įverčių modeliavimo rezultatai (δ = 0,95)

Koeficientų a_i skaičiavimo būdai:	***I būdas***	***II būdas***	***III būdas***
$A_{N,k}$	0,000015	0,0000034	0,0000011
Dvipusis pasikliautinasis intervalas:			
apatinis rėžis	-0,000071	-0,000069	-0,000072
viršutinis rėžis	0,000101	0,000076	0,000077
p	0,974	0,974	0,976
Patekimo tikimybės p pasikliautinasis intervalas:			
apatinis rėžis	0,97397	0,97381	0,9759
viršutinis rėžis	0,9926	0,9927	0,9939

Lentelėse 4.3 - 4.6 pateikti minimalios reikšmės A tiesinių įverčių $A_{N,k}$ (4.1) vidurkiai, vienpusio pasikliautinojo intervalo (4.4) Monte-Karlo įverčių vidurkiai, tikimybės patekti tikslo funkcijos minimaliai reikšmei į pasikliautinąjį intervalą (4.4) Monte-Karlo įverčiai p bei jos pasikliautinasis intervalas. *Pav. 4.2* pateikti funkcijos su aštriu minimumu (n=10) minimalios reikšmės įverčio pasikliautinojo intervalo viršutinis ir apatinis rėžiai, įvertinti Monte-Karlo metodu, o *pav. 4.3* yra pateikti funkcijos minimaliai reikšmei patekti į pasikliautinąjį intervalą tikimybės p

viršutinė ir apatinė ribos, taip pat įvertintos Monte-Karlo metodu. Kitoms testinėms funkcijoms priklausomybės yra analogiškos.

Lentelė 4.3. Funkcijos su aštriu minimumu minimalios reikšmės ir vienpusio pasikliautinojo intervalo įverčių modeliavimo rezultatai

A=0,μ=2, K=5, $\delta = 0{,}95$		$A_{N,k}$	***Vienpusis pasikliautinasis intervalas***		***p***	***Patekimo tikimybės p pasikliautinasis intervalas***	
	Algoritmas		***Apatinis rėžis***	***Viršutinis rėžis***		***Apatinis rėžis***	***Viršutinis rėžis***
n=2	SPSAL	0,0000013	-0,0001	0,00007	0,95	0,8976	0,9801
	FDSA	0,0000004	-0,0005	0,0004	0,94	0,8849	0,9736
	SPSAU	0,000008	-0,0003	0,00022	0,958	0,9401	0,9717
n=10	SPSAL	0,000073	-0,0177	0,01902	0,94	0,9167	0,9543
	FDSA	-0,000005	-0,00068	0,00072	0,954	0,9354	0,9684
	SPSAU	-0,00058	-0,01992	0,02009	0,954	0,9354	0,9684
n=100	SPSAL	0,000672	0,00082	0,16357	0,944	0,9197	0,9548
	FDSA	-0,003333	-0,14148	0,16215	0,946	0,9316	0,9564

Lentelė 4.4. Funkcijos su aštriu minimumu, kai *n*=10, minimalios reikšmės ir vienpusio pasikliautinojo intervalo įverčių modeliavimo rezultatai

A=0,μ=2, K=5		$A_{N,k}$	***Vienpusis pasikliautinasis intervalas***		***p***	***Patekimo tikimybės p pasikliautinasis intervalas***	
Pasikliovimo lygmuo	***Algorit-mas***		***Apatinis rėžis***	***Viršutinis rėžis***		***Apatinis rėžis***	***Viršutinis rėžis***
$\delta = 0{,}9$	SPSAL	0,000073	-0,01318	0,01902	0,89	0,8676	0,9057
	FDSA	-0,000005	-0,00051	0,00072	0,892	0,8719	0,9094
	SPSAU	-0,000578	-0,01503	0,0200863	0,892	0,8719	0,9094
$\delta = 0{,}975$	SPSAL	0,000073	-0,022005	0,01902	0,974	0,9559	0,9861
	FDSA	-0,000005	-0,00085	0,00072	0,968	0,9485	0,9816
	SPSAU	-0,00058	-0,02465	0,02009	0,978	0,9610	0,9890
$\delta = 0{,}99$	SPSAL	0,000073	-0,02766	0,01902	0,99	0,9740	0,9974
	FDSA	-0,000005	-0,00107	0,00072	0,984	0,9655	0,9941
	SPSAU	-0,00058	-0,03083	0,02009	0,994	0,9801	0,9991

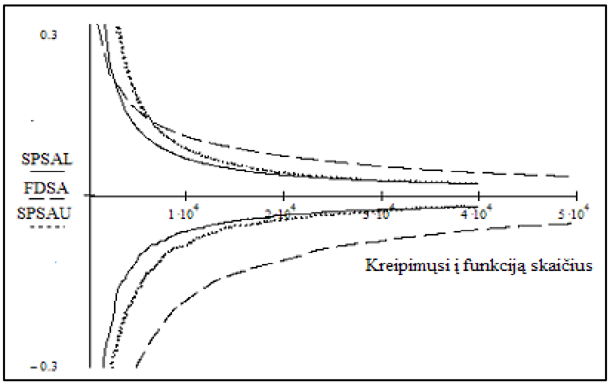
Lentelė 4.5. CB3 funkcijos minimalios reikšmės ir vienpusio pasikliautinojo intervalo įverčių modeliavimo rezultatai

A=2, n=2		$A_{N,k}$	***Vienpusis pasikliautinasis intervalas***		***p***	***Patekimo tikimybės p pasikliautinasis intervalas***	
Pasikliovimo lygmuo	***Algorit-mas***		***Apatinis rėžis***	***Viršutinis rėžis***		***Apatinis rėžis***	***Viršutinis rėžis***
$\delta = 0{,}9$	SPSAL	2,0000008	1,99998	2,00002	0,892	0,8719	0,9094
	SPSAU	2,000006	1,9994	2,00059	0,906	0,8869	0,9223
	FDSA	2,0000006	1,99999	2,00001	0,896	0,8762	0,9131
$\delta = 0{,}95$	SPSAL	2,0000008	1,99997	2,00002	0,95	0,9304	0,9650
	SPSAU	2,000006	1,9992	2,00059	0,948	0,9281	0,9628
	FDSA	2,0000006	1,99998	2,00001	0,948	0,92807	0,9628
$\delta = 0{,}975$	SPSAL	2,0000008	1,99996	2,00002	0,968	0,9485	0,9816
	SPSAU	2,000006	1,9989	2,0006	0,976	0,9584	0,9875
	FDSA	2,0000006	1,99998	2,00001	0,976	0,9584	0,9875
$\delta = 0{,}99$	SPSAL	2,0000008	1,9999	2,00002	0,986	0,9683	0,9953
	SPSAU	2,000006	1,9986	2,00059	0,986	0,9683	0,9953
	FDSA	2,0000006	1,99997	2,00001	0,992	0,97696	0,9984

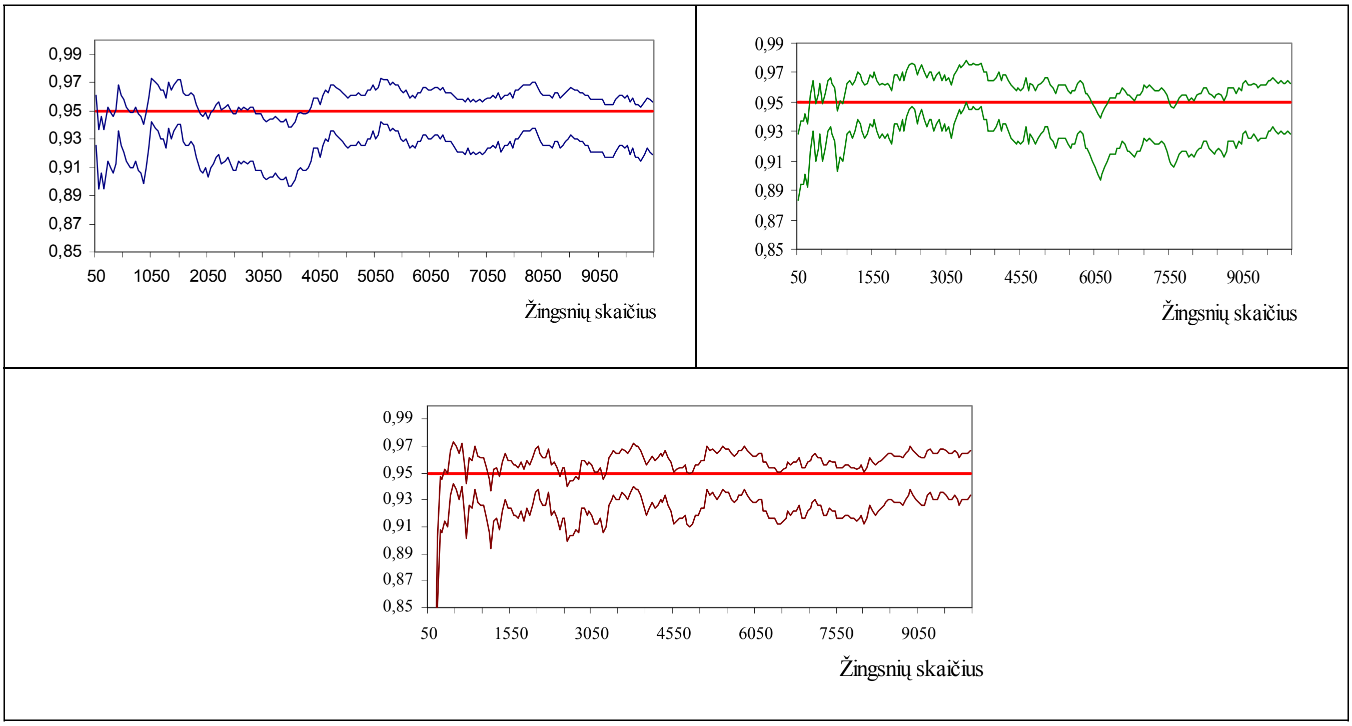
Lentelė 4.6. Rosen Suzuki funkcijos minimalios reikšmės ir vienpusio pasikliautinojo intervalo įverčių modeliavimo rezultatai

A= -44, n=4		$A_{N,k}$	***Vienpusis pasikliautinasis intervalas***		***p***	***Patekimo tikimybės p pasikliautinasis intervalas***	
Pasikliovimo lygmuo	***Algorit-mas***		***Apatinis rėžis***	***Viršutinis rėžis***		***Apatinis rėžis***	***Viršutinis rėžis***
$\delta = 0{,}9$	SPSAL	-44,00001	-44,0121	-43,0905	0,89	0,8801	0,9051
	SPSAU	-43,9991	-44,0180	-43,9761	0,878	0,8570	0,9064
	FDSA	-44,0002	-44,0085	-43,99	0,9	0,8805	0,9168
$\delta = 0{,}95$	SPSAL	-44,00001	-44,01512	-43,0905	0,95	0,9302	0,958
	SPSAU	-43,9991	-44,0249	-43,9761	0,946	0,9123	0,9508
	FDSA	-44,0002	-44,0116	-43,99	0,952	0,9331	0,9667
$\delta = 0{,}975$	SPSAL	-44,00001	-44,0243	-43,0905	0,973	0,939	0,9792
	SPSAU	-43,9991	-44,0315	-43,9761	0,96	0,9389	0,9754

	FDSA	-44,0002	-44,0145	-43,99	0,972	0,9535	0,9846
δ = 0,99	SPSAL	-44,00001	-44,036	-43,0905	0,986	0,9634	0,9963
	SPSAU	-43,9991	-44,0402	-43,9761	0,984	0,9655	0,9942
	FDSA	-44,00012	-44,0183	-43,99	0,988	0,9711	0,9964



Pav. 4.2. Tikslo funkcijos minimalios reikšmės pasikliautinieji rėžiai (funkcija su aštriu minimumu, n=10, A=0, δ=0,95)



Pav. 4.3. Minimalios reikšmės patekimo į pasikliautinąjį intervalą tikimybės viršutinis ir apatinis rėžiai (funkcija su aštriu minimumu, n=10, A=0, δ=0,95, FDSA, SPSAU, SPSAL)

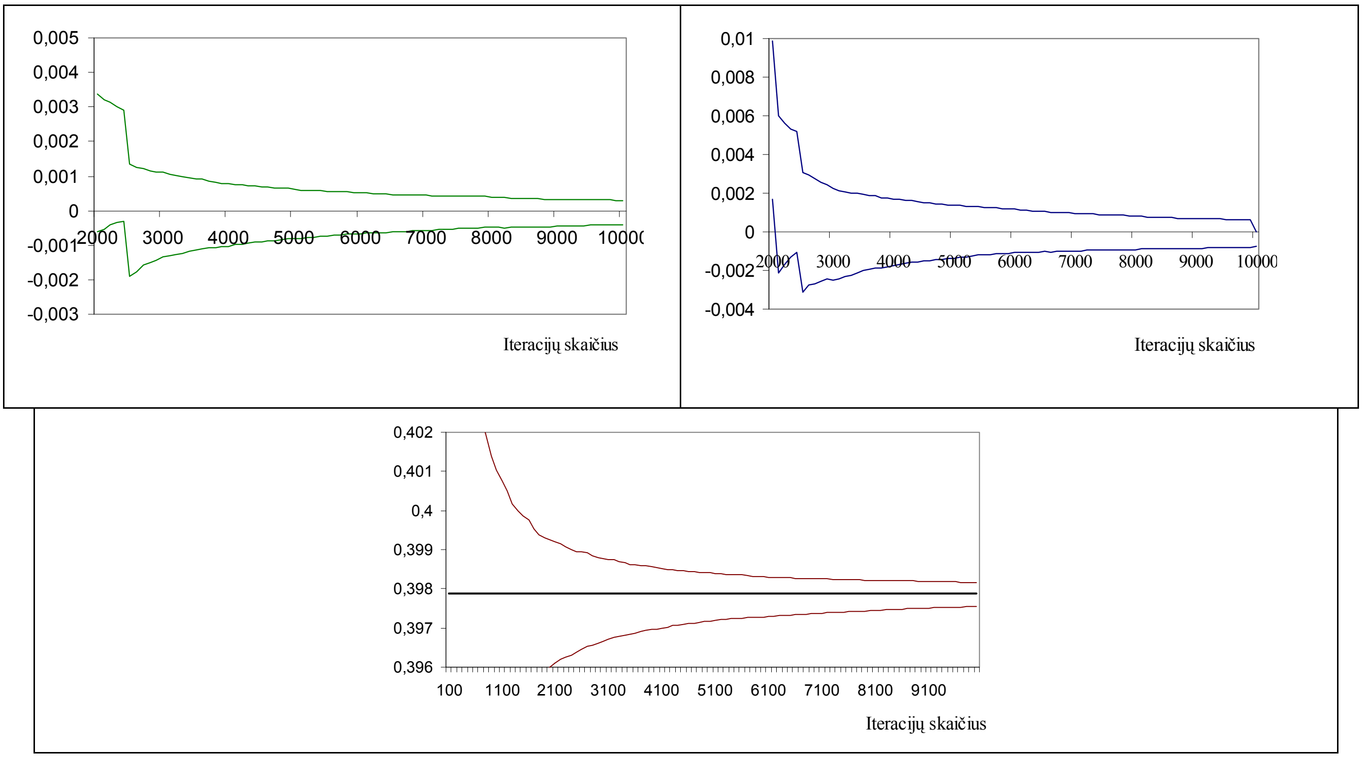
Iš pateiktų priklausomybių (*pav. 4.2* ir *4.3*) bei lentelėse (*lentelės 4.2 - 4.6*) esamų rezultatų matome, kad formulės (4.1), (4.4) bei (4.5) gana gerai aproksimuoja tikslo funkcijos minimalią reikšmę bei jos pasikliautinąjį intervalą. *Pav. 4.2* pateiktos priklausomybės iliustruoja šio intervalo mažėjimą didėjant iteracijų skaičiui, o *pav. 4.3* ir *lentelėse 4.2 - 4.6* pateikti tikimybės minimaliai tikslo funkcijos reikšmei patekti į pasikliautinąjį intervalą Monte – Karlo įverčiai bei jos pasikliautinieji rėžiai iliustruoja tai, kad didėjant iteracijų skaičiui tikimybė minimaliai tikslo funkcijos reikšmei patekti į pasikliautinąjį intervalą artėja prie teorinės tikimybės (δ=0,9; δ=0,95; δ=0,975; δ=0,99), o tai sąlygoja sudarytų įverčių tinkamumą sprendinio, gauto stochastinės aproksimacijos algoritmais, optimalumo analizei. Šie rezultatai leidžia sudaryti taisyklę stochastinės aproksimacijos algoritmams stabdyti, t.y. algoritmas yra stabdomas, kai pasikliautinojo intervalo plotis tampa mažesnis už tam tikrą iš anksto užsiduotą reikšmę $\varepsilon > 0$.

Analogiški skaičiavimai buvo atlikti bei gautos analogiškos išvados modeliuojamojo atkaitinimo metodu minimizuojant daug ekstremumų turinčias testines funkcijas, t.y. sprendžiant globalaus optimizavimo uždavinius.

Lentelė 4.7. Daugiaekstremaliųjų testinių funkcijų minimalios reikšmės ir vienpusio pasikliautinojo intervalo įverčių modeliavimo rezultatai

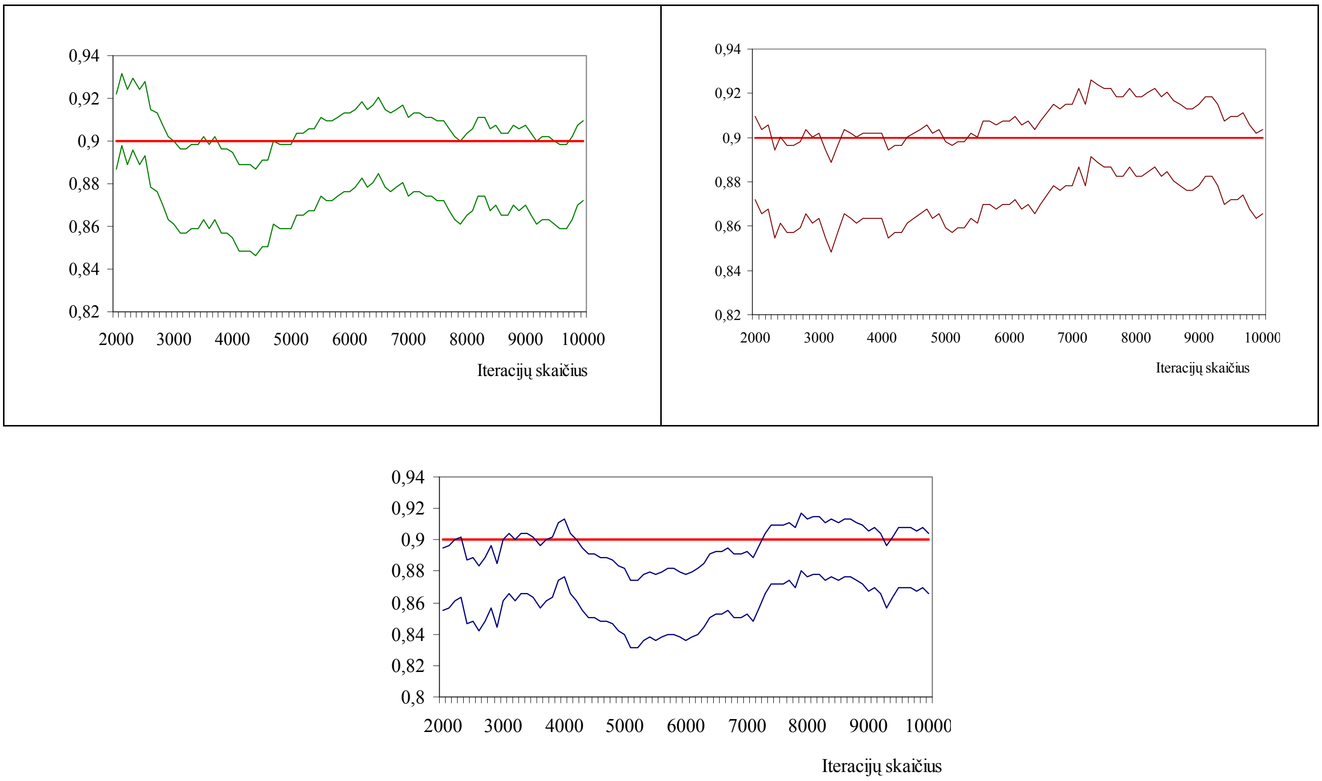
		$A_{N,k}$	***Pasikliautinasis intervalas***		***p***	***Patekimo tikimybės p pasikliautinasis intervalas***	
Testinė funkcija	***Pasikliovimo tikimybė***		***Apatinis rėžis***	***Viršutinis rėžis***		***Apatinis rėžis***	***Viršutinis rėžis***
Branino funkcija *A*=0,397887, *n*=2	δ = 0,9	0,39791	0,3975	0,39816	0,886	0,86549	0,90385
	δ = 0,95	0,39791	0,3973	0,39816	0,95	0,93035	0,965
	δ = 0,975	0,39791	0,3971	0,39816	0,98	0,96352	0,9904
	δ = 0,99	0,39791	0,3969	0,39816	0,992	0,97696	0,9984
Beale funkcija *A*=0, *n*=2	δ = 0,9	0,000032	-0,00039	0,00032	0,882	0,86123	0,90015
	δ = 0,95	0,000032	-0,00064	0,00032	0,948	0,92807	0,96284
	δ = 0,975	0,000032	-0,00086	0,00032	0,966	0,94611	0,98007
	δ = 0,99	0,000032	-0,00118	0,00032	0,98	0,96011	0,9917
Rastrigino funkcija *A*=0, *n*=2	δ = 0,9	0,0000478	-0,00078	0,00062	0,886	0,90385	0,86549
	δ = 0,95	0,0000478	-0,00123	0,00062	0,948	0,92807	0,96284
	δ = 0,975	0,0000478	-0,0017	0,00062	0,97	0,95099	0,98312
	δ = 0,99	0,0000478	-0,00234	0,00062	0,984	0,96552	0,99416

Kaip ir stochastinės aproksimacijos atveju *lentelėje 4.7* pateikiami minimalios reikšmės *A* tiesinių įverčių $A_{N,k}$ (4.1) vidurkiai, šios reikšmės pasikliautinojo intervalo (4.4) Monte-Karlo įverčių vidurkiai, tikimybės patekti tikslo funkcijos minimaliai reikšmei į pasikliautinąjį intervalą (4.4) Monte-Karlo įvertis *p* bei jo pasikliautinasis intervalas aukščiau išvardintas daug ekstremumų turinčias tikslo funkcijas (Branino, Beale, Rastrigino) minimizuojant modeliuojamojo atkaitinimo metodu.



Pav. 4.4. Tikslo funkcijos globalaus minimumo pasikliautinieji rėžiai

(—— Beale funkcija, —— Rastrigino funkcija, —— Branino funkcija)

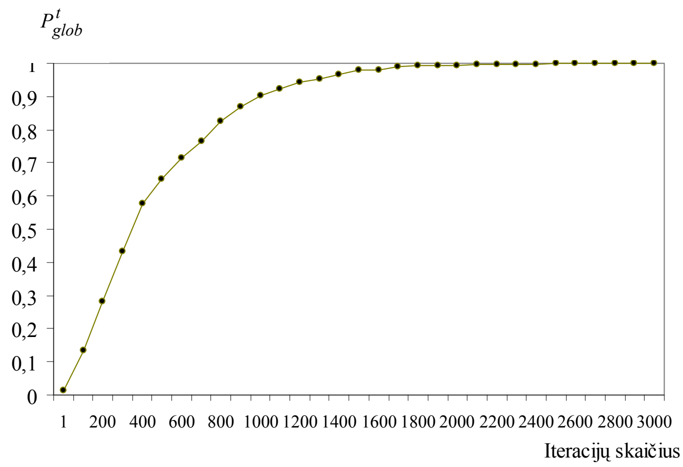


Pav. 4.5. Minimalios reikšmės patekimo į pasikliautinąjį intervalą tikimybės viršutinis ir apatinis rėžiai (—— Beale funkcija, —— Branino funkcija, —— Rastrigino funkcija)

Pav. 4.4 pateikti tikslo funkcijos minimalios reikšmės pasikliautinojo intervalo viršutinis ir apatinis rėžiai, įvertinti Monte-Karlo metodu. *Pav. 4.5* yra pateikti funkcijos minimalios reikšmės patekimo į pasikliautinąjį intervalą tikimybės p viršutinė ir apatinė ribos, taip pat įvertintos Monte-

Karlo metodu. Paveiksluose ir lentelėse pateikti rezultatai rodo, kad sudaryti metodai gali būti taikomi sprendimams apie minimalią tikslo funkcijos reikšmę bei jos pasikliautinąjį intervalą priimti su iš anksto užsiduota pasikliovimo tikimybe ne tik lokalaus bet ir globalaus optimizavimo uždaviniuose.

Pav. 4.6 pateikiama tikimybės patekti į globalaus minimumo traukos zoną $P_{glob}^t = P\left(x^t \in A_{glob}\right)$ įverčio priklausomybė nuo iteracijų skaičiaus Rastrigino funkcijai. Nagrinėjama testinė funkcija turi 36 lokaliuosius ekstremumus leistinoje srityje. Einamojo optimizavimo taško traukos zona buvo nustatyta kaip artimiausias lokalinio minimumo taškas Euklido metrikoje. Tegul funkcija turi keletą lokaliųjų ekstremumo taškų $x_1^*, x_2^*, \ldots, x_q^*$. Nemažinant bendrumo galima laikyti, kad globalusis minimumas sutampa su x_1^*. Tuomet globalinio optimumo traukos zona nustatoma tokiu būdu: $A_{glob} = \left\{x \middle| \left\| x - x_1^* \right\| \le \left\| x - x_i^* \right\|, i = 2, \ldots, q \right\}$. Gauta priklausomybė iliustruoja tiriamos modeliuojamojo atkaitinimo algoritmo modifikacijos konvergavimo globališkumą. Kitaip tariant šiuo algoritmu gaunamos sekos yra konverguojančios. Panašios priklausomybės gautos ir kitoms testinėms funkcijoms.



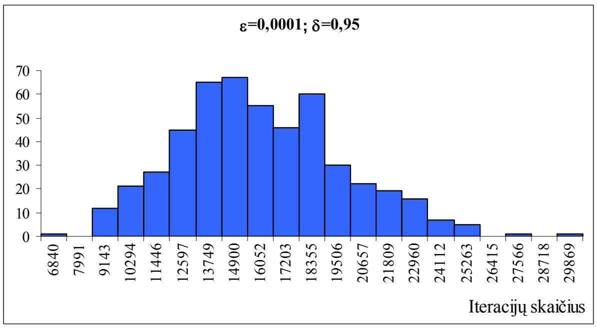
Pav. 4.6. Patekimo į globalaus minimumo traukos zoną tikimybė

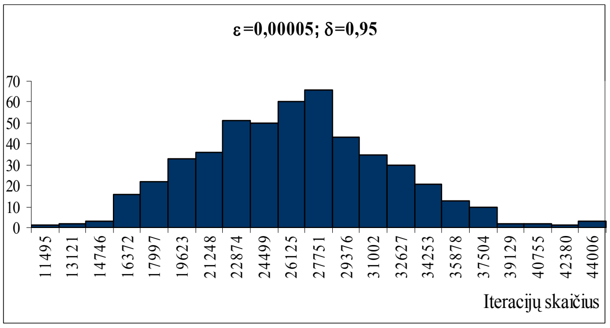
4.3.4. Stabdymo taisyklės kompiuterinis tyrimas

Kaip jau buvo minėta 2 skyriuje, stabdymo problema yra viena iš pagrindinių sudarant stochastinius bei euristinius optimizavimo algoritmus. Šiems algoritmams realizuoti reikia sukurti stabdymo taisykles, leidžiančias stabdyti skaičiavimus po baigtinio iteracijų skaičiaus, užtikrinant gauto sprendinio kokybę. Stochastinės aproksimacijos algoritmų stabdymo problemą pirmieji iškėlė Kiefer & Wolfowitz (1952).

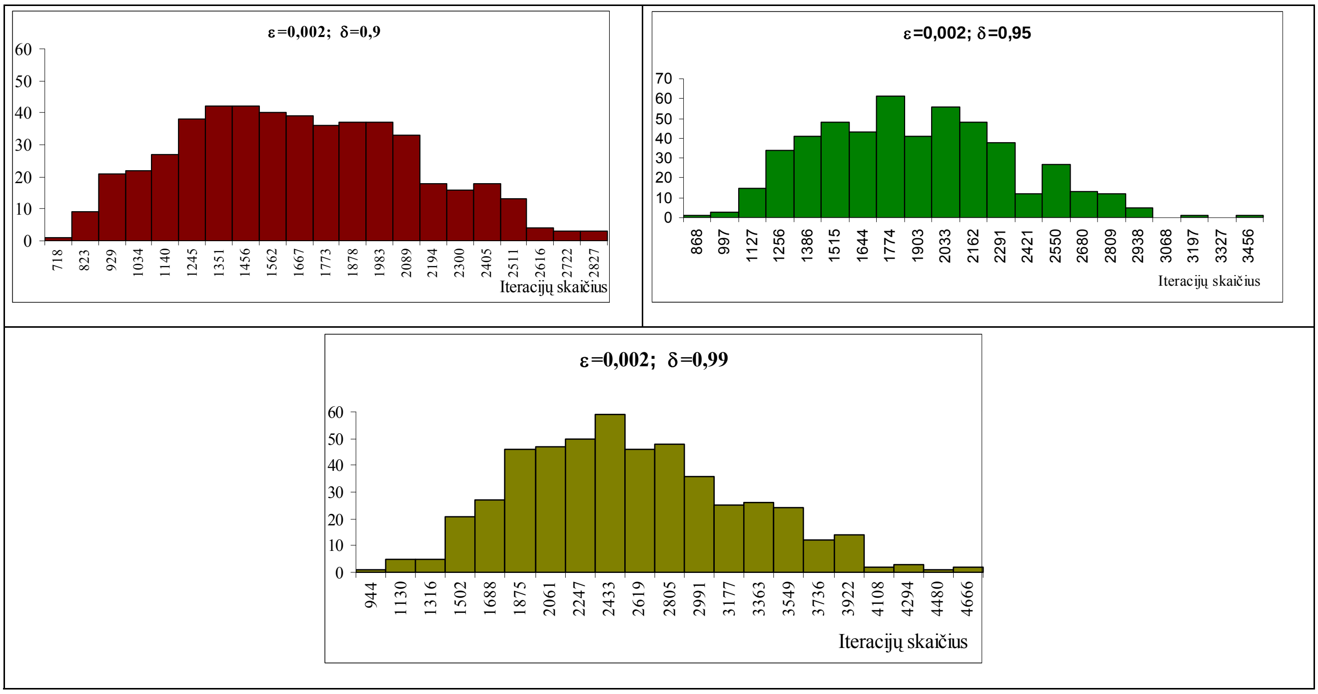
Įvairūs algoritmo stabdymo taisyklės sudarymo būdai yra aprašyti 2 skyriuje. Šiame skyrelyje aptarsime stabdymo taisyklę, kuriai sudaryti pasinaudosime informacija, gaunama iš geriausių optimizavimo metu pasiektų tikslo funkcijos reikšmių. Remiantis prieš tai skyrelyje aprašytais metodais, stochastinės aproksimacijos algoritmas yra stabdomas, kai pasikliautinojo intervalo (4.4 ir 4.5) plotis tampa mažesnis už iš anksto užsiduotą reikšmę $\varepsilon > 0$. Optimizavimo uždavinio sprendiniu tokiu atveju galima laikyti tašką, kuriame optimizavimo metu buvo pasiekta geriausia tikslo funkcijos reikšmė.

Pav. 4.7 ir *4.8* pateiktose histogramose matome algoritmo žingsnių skaičiaus, reikalingo algoritmui sustoti, sklaidą priklausomai nuo iš anksto užsiduoto vienpusio pasikliautinojo intervalo pločio ir pasirinkto pasikliovimo lygmens, SPSAL metodu minimizuojant M=500 kartų funkciją su aštriu minimumu (n=2). Iš pateiktų rezultatų galime apskaičiuoti, kad kai pasikliautinojo intervalo plotis $\varepsilon = 0{,}002$ ir pasikliovimo lygmuo yra $\delta = 0{,}9$, tai vidutinis iteracijų skaičius, reikalingas algoritmui sustoti, lygus 1609,086 iteracijų, kai pasikliovimo lygmuo pasirenkamas $\delta = 0{,}95$, optimizavimo algoritmas stabdomas vidutiniškai po 1821,4 žingsnių, o kai pasikliovimo lygmuo parenkamas $\delta = 0{,}99$ algoritmas sustoja po vidutiniškai 2466,626 iteracijų. Kai pasikliautinojo intervalo plotis yra mažinamas, t.y. $\varepsilon = 0{,}0001$ ir $\varepsilon = 0{,}00005$, algoritmas atitinkamai stabdomas vidutiniškai po 15525,49 ir 25537,23 žingsnių. Analogiški rezultatai gaunami ir kitoms testinėms funkcijoms bei kitiems optimizavimo algoritmams.

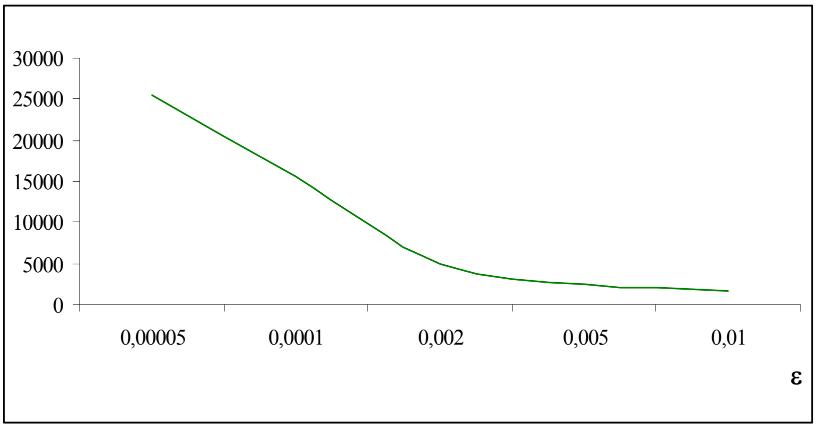




Pav. 4.7. Iteracijų skaičius reikalingas algoritmui stabdyti, priklausomai nuo iš anksto užsiduoto pasikliautinojo intervalo pločio



Pav. 4.8. Iteracijų skaičius reikalingas algoritmui stabdyti, priklausomai nuo pasikliovimo lygmens



Pav. 4.9. Vidutinis iteracijų skaičius reikalingas algoritmui stabdyti, priklausomai nuo iš anksto užsiduoto pasikliautinojo intervalo pločio, $\delta = 0{,}95$

Pav. 4.9 iliustruoja vidutinio iteracijų skaičiaus, reikalingo algoritmui sustoti, didėjimą priklausomai nuo pasikliautinojo intervalo pločio.

4.4. REZULTATAI IR IŠVADOS

1. Sukurtas metodas ir sudarytas algoritmas tikslo funkcijos minimaliai reikšmei bei jos pasikliautinajam intervalui įvertinti, naudojant optimizavimo sekų, gautų Markovo tipo stochastiniais algoritmais, pozicines statistikas.
2. Sudarytas algoritmas pritaikytas stochastinės aproksimacijos bei modeliuojamojo atkaitinimo optimizavimo metodams.
3. Sukurtas metodas tikslo funkcijos minimaliai reikšmei bei jos pasikliautinajam intervalui įvertinti, naudojant optimizavimo sekų, gautų Markovo tipo stochastiniais algoritmais,

pozicines statistikas, gali būti taikomas sprendimams apie minimalią tikslo funkcijos reikšmę bei jos pasikliautinąjį intervalą priimti su iš anksto užsiduota pasikliovimo tikimybe.

4. Minimalios reikšmės ir jos pasikliautinojo intervalo tiesinių įverčių koeficientai gali būti skaičiuojami, pritaikius ekstremaliųjų reikšmių teorijos rezultatus nepriklausomiems vienodai pasiskirsčiusiusiems atsitiktiniams dydžiams, kai ekstremaliųjų reikšmių skirstinio formos parametras nustatomas pagal tikslo funkcijos homogeniškumo rodiklį minimumo aplinkoje.
5. Gauti rezultatai leidžia pasiūlyti optimizavimo algoritmo stabdymo taisyklę, naudojant optimizavimo metu pasiektas geriausias tikslo funkcijos reikšmes, ir stabdyti algoritmą kai pasikliautinojo intervalo plotis tampa mažesnis už tam tikrą iš anksto užsiduotą reikšmę.

5. EKSTREMALIŲJŲ REIKŠMIŲ SKIRSTINIO PARAMETRŲ NUSTATYMAS IR TAIKYMAS STOCHASTINIŲ ALGORITMŲ OPTIMALUMUI TIRTI

5.1. ĮVADAS

Veibulo skirstinys (Weibull (1951)) yra vienas iš ekstremaliųjų reikšmių skirstinių, naudojamų Markovo tipo algoritmų optimalumui tirti (Haan (1981), Zhigljavsky (1990), Bartkutė & Sakalauskas (2004a)). Ekstremaliųjų reikšmių skirstinius taikyti optimizuojamos funkcijos maksimaliai (minimaliai) reikšmei įvertinti pasiūlė J. Mockus (Моцкус (1976)), teorinius optimizavimo algoritmų nagrinėjimus, paremtus ekstremaliųjų reikšmių teorija, plėtojo A. Žilinskas ir A. Zhigljavsky (Жиглявский & Жилинскас (1991)), o pozicines statistikas taikyti regresijos funkcijos maksimumo taškui nustatyti pasiūlė H. Chen (1996).

Tačiau, norint taikyti ekstremaliųjų reikšmių skirstinius optimizavimo uždavinių sprendinio optimalumui tirti, reikia žinoti šių skirstinių parametrus, kurių svarbiausias skirstinio formos parametras α. Metodas parametrui α parinkti, priklausomai nuo tikslo funkcijos homogeniškumo ekstremumo aplinkoje, buvo išnagrinėtas 4 skyriuje. Šį parametrą taip pat galima įvertinti statistiniais metodais, sprendžiant trijų parametrų Veibulo skirstinio įvertinimo uždavinį. Pastarasis uždavinys buvo nagrinėjamas kelių autorių, tačiau aiškus šios problemos sprendinys nėra gautas. Yra žinomi autorių darbai (Rockette et al. (1974), Lemon (1975), Hirose (1991), ir t.t.) susiję su šio skirstinio formos, padėties ir skalės parametrų įvertinimu. Dažniausiai trijų parametrų Veibulo skirstinio parametrams įvertinti taikomi iteraciniai skaičiuojamieji metodai (Hirose (1991), Bartolucci et al. (1999)).

Disertacijoje palyginsime maksimalaus tikėtinumo ir analitinius metodus Veibulo skirstinio parametrams įvertinti. Gauti rezultatai buvo pristatyti tarptautinėse konferencijose: *9-toje tarptautinėje tikimybių teorijos ir matematinės statistikos konferencijoje Vilniuje*, *12-toje tarptautinėje taikomųjų stochastinių modelių ir duomenų analizės konferencijoje* (ASMDA 2007), (Chania, Kreta, Graikija) bei publikuoti moksliniame leidinyje Bartkutė & Sakalauskas (2007b).

5.2. VEIBULO SKIRSTINIO TRIJŲ PARAMETRŲ VERTINIMAS MAKSIMALAUS TIKĖTINUMO METODU

Taikant įverčius (4.1), (4.4) ir (4.5), reikia nustatyti ekstremaliųjų reikšmių skirstinio formos parametrą α. Šio parametro vertinimas statistiniais metodais yra suvedamas į trijų parametrų Veibulo skirstinio vertinimo uždavinį. Panagrinėsime šio uždavinio sprendimą maksimalaus tikėtinumo metodu.

Trijų parametrų Veibulo skirstinio pasiskirstymo funkciją galima užrašyti tokiu būdu:

$$F\left(x,\alpha,c,A\right)=1-e^{-c\cdot\left(x-A\right)^{\alpha}}=c\left(x-A\right)^{\alpha}+o\left(\left(x-A\right)^{\alpha}\right),\quad \alpha>0,\ x\geq A,\ c>0, \tag{5.1}$$

čia c, A ir α atitinkamai skirstinio skalės, padėties ir formos parametrai. Nagrinėsime maksimalaus tikėtinumo metodo taikymą šio skirstinio parametrams įvertinti. Taikyti standartinį maksimalaus tikėtinumo metodą trijų parametrų Veibulo skirstinio vertinimui yra problematiška, kadangi maksimalaus tikėtinumo įvertis ne visada egzistuoja, o taip pat jo skaitmeninis realizavimas yra sudėtingas (Murthy et al. (2004), Dress et al. (2004), Blischke (1974), Zanakis & Kyparisis (1986)). Siekiant įveikti minėtus sunkumus taikysime maksimalaus tikėtinumo metodo modifikaciją (Hall (1982), Smith (1987), Жиглявский & Жилинскас (1991)).

Tarkime m pozicinių statistikų yra gautos iš N atsitiktinių dydžių populiacijos, kurioje atsitiktiniai dydžiai pasiskirstę vienodai su skirstiniu F(·). Pozicinių statistikų tikėtinumo funkcija užrašoma tokiu būdu:

$$F_{\eta(0),\eta(2),\ldots,\eta(k)}\left(x_1,x_2,\ldots,x_k\right)=\frac{N!}{\left(N-k\right)!}\cdot\left(1-F\left(x_k\right)\right)^{N-k}\cdot\prod_{l=1}^{k}p\left(x_l\right), \tag{5.2}$$

čia $p(\cdot)$ –skirstinio tankio funkcija. Iš tikėtinumo funkcijos išvestinių, prilyginus jas nuliui, gauname lygtis, kurias turi tenkinti vertinamų parametrų įverčiai. Tačiau šios lygtys neturi išreikštinių sprendinių, todėl jas tenka spręsti sudėtingais skaitmeniniais metodais.

Galima pasiūlyti paprastesnį būdą. Kadangi pozicinių statistikų elgesį labiausiai apibūdina skirstinio pavidalas minimalios reikšmės aplinkoje, pakanka nagrinėti skirstinį, kuriame įvertintas tiktai skirstinio (5.1) skleidinio Teiloro eilute pirmasis narys (Hall (1982)):

$$F\left(x,\alpha,c_0,A\right)=1-c_0\left(x-A\right)^{\alpha}. \tag{5.3}$$

Užrašysime tikėtinumo funkciją

$$L(\eta_{(0)},\ldots,\eta_{(k)};A,c_o;\alpha)=\frac{N!}{\left(N-k\right)!}\cdot\left(c_0\alpha\right)^k\left(1-c_0\left(\eta_{(k)}-A\right)^{\alpha}\right)^{N-k}\prod_{l=0}^{k}\left(\eta_{(l)}-A\right)^{\alpha-1}.$$

Diferencijuodami šios funkcijos logaritmą

$$\ln\left(L(\eta_{(0)},\ldots,\eta_{(k)};A,c_0;\alpha)\right)=$$

$$=\ln\left(\frac{N!}{\left(N-k\right)!}\right)+k\cdot\ln\left(c_0\alpha\right)+\left(N-k\right)\cdot\ln\left(1-c_0\left(\eta_{(k)}-A\right)^{\alpha}\right)+\left(\alpha-1\right)\cdot\sum_{l=0}^{k}\ln\left(\eta_{(l)}-A\right),$$

gausime išvestines:

$$\frac{\partial\ln\left(L(\eta_{(0)},\ldots,\eta_{(k)};A,c_o;\alpha)\right)}{\partial\alpha}=\frac{k}{\alpha}+\left(N-k\right)\cdot\frac{c_0\left(\eta_{(k)}-A\right)^{\alpha}\cdot\ln\left(\eta_{(k)}-A\right)}{1-c_0\left(\eta_{(k)}-A\right)^{\alpha}}+\sum_{l=0}^{k}\ln\left(\eta_{(l)}-A\right);$$

$$\frac{\partial\ln\left(L(\eta_{(0)},\ldots,\eta_{(k)};A,c_o;\alpha)\right)}{\partial c_0}=\frac{k}{c_0}-\left(N-k\right)\cdot\frac{\left(\eta_{(k)}-A\right)^{\alpha}}{1-c_0\left(\eta_{(k)}-A\right)^{\alpha}};$$

$$\frac{\partial \ln\left(L(\eta_{(0)}, ..., \eta_{(k)}; A, c_o; \alpha)\right)}{\partial A} = (N-k)\cdot\frac{c_0\alpha\left(\eta_{(k)}-A\right)^{\alpha-1}}{1-c_0\left(\eta_{(k)}-A\right)^{\alpha}} - (\alpha-1)\cdot\sum_{l=0}^{k}\frac{1}{\eta_{(l)}-A}.$$

Ir prilyginę jas nuliui gauname parametrų $\hat{\alpha}$, c_0 įverčius

$$\hat{\alpha} = \frac{k}{\sum_{l=0}^{k}\ln\left(1+\beta_l\left(\hat{A}\right)\right)}, \tag{5.4}$$

$$c_0 = \frac{k}{N\left(\eta_{(k)}-\hat{A}\right)^{\alpha}}, \tag{5.5}$$

čia $\beta_l\left(\hat{A}\right) = \dfrac{\eta_{(k)}-\eta_{(l)}}{\eta_{(l)}-\hat{A}}$, o $\hat{A}$ yra minimalus lygties

$$\frac{1}{\sum_{l=0}^{k}\ln\left(1+\beta_l\left(\hat{A}\right)\right)} - \frac{1}{\sum_{l=0}^{k}\beta_l\left(\hat{A}\right)} = \frac{1}{k} \tag{5.6}$$

sprendinys, kai $\hat{A} < \eta_{(1)}$.

Pažymėkime

$$y = \frac{\eta_{(k)}-\eta_{(0)}}{\eta_{(0)}-\hat{A}} \tag{5.7}$$

ir įveskime funkciją:

$$F(y) = \frac{1}{\sum_{l=0}^{k}\ln\left(1+\dfrac{\eta_{(k)}-\eta_{(l)}}{\frac{1}{y}\cdot\left(\eta_{(k)}-\eta_{(0)}\right)+\eta_{(l)}-\eta_{(0)}}\right)} - \frac{1}{\sum_{l=0}^{k}\dfrac{\eta_{(k)}-\eta_{(l)}}{\frac{1}{y}\cdot\left(\eta_{(k)}-\eta_{(0)}\right)+\eta_{(l)}-\eta_{(0)}}} - \frac{1}{k}. \tag{5.8}$$

Pažymėsime, kad skirstinys (5.1) yra absoliučiai tolydusis, todėl sąlyga, kad $\eta_{(i)} \neq \eta_{(j)}, i \neq j,\ i, j = 1,...,m$ yra teisinga su tikimybe 1.

Į tikėtinumo funkciją galime įstatyti $\hat{\alpha}$ ir c_0 reikšmes, apskaičiuotas pagal (5.4) ir (5.5) formules bei $\hat{A}$ išsireiškę iš (5.8) ir šią funkciją nagrinėti priklausomai nuo y. Tuomet nesunku įsitikinti, kad šios funkcijos išvestinė ir $F(y)$ yra priešingo ženklo, bei lygi nuliui tuose pačiuose taškuose kaip ir $F(y)$. Lygtis $F(y) = 0$ gali turėti arba neturėti baigtinio sprendinio y^*. Jei baigtinis sprendinys egzistuoja, tai maksimalaus tikėtinumo įvertis atitinka tokį y^*, kuriame funkcija $F(y)$ keičia ženklą iš „-" į „+".

Norėdami ištirti sprendinio egzistavimą, turime ištirti funkcijos $F(y)$ ir jos pirmos eilės išvestinės elgesį nulio aplinkoje ir begalybėje. Diferencijuodami (5.8) gauname:

$F'(y) =$

$$= \frac{1+\left(\eta_{(k)}-\eta_{(0)}\right)\cdot\sum_{j=1}^{k}\frac{\eta_{(k)}-\eta_{(j)}}{\left(\eta_{(k)}-\eta_{(0)}+y\cdot\left(\eta_{(j)}-\eta_{(0)}\right)\right)^2}}{y^2\cdot\left(1+\sum_{j=1}^{k}\frac{\eta_{(k)}-\eta_{(j)}}{\eta_{(k)}-\eta_{(0)}+y\cdot\left(\eta_{(j)}-\eta_{(0)}\right)}\right)^2} - \frac{\frac{1}{1+y}\cdot\left(1+\sum_{j=1}^{k}\frac{\eta_{(k)}-\eta_{(j)}}{\eta_{(k)}-\eta_{(0)}+y\cdot\left(\eta_{(j)}-\eta_{(0)}\right)}\right)}{\left(\ln(1+y)+\sum_{j=1}^{k}\ln\left(1+\frac{y\cdot\left(\eta_{(k)}-\eta_{(j)}\right)}{\eta_{(k)}-\eta_{(0)}+y\cdot\left(\eta_{(j)}-\eta_{(0)}\right)}\right)\right)^2} \tag{5.9}$$

Nesunku įsitikinti, kad

$$\lim_{y\to\infty} F(y) = -\frac{1}{k} \tag{5.10}$$

ir

$$\mathit{lim}_{y\to\infty} F'(y) = -0\,. \tag{5.11}$$

Apskaičiavę atitinkamas ribas, gauname, kad

$$\lim_{y\to 0} F(y) = \frac{1}{2}\cdot\frac{\sum_{j=0}^{k}\left(\eta_{(k)}-\eta_{(j)}\right)^2}{\left(\sum_{j=0}^{k}\left(\eta_{(k)}-\eta_{(j)}\right)\right)^2} - \frac{1}{k} \tag{5.12}$$

ir

$$\lim_{y\to 0} F'(y) = \frac{\frac{2}{3}\cdot\sum_{j=0}^{k}\left(\eta_{(k)}-\eta_{(j)}\right)\cdot\sum_{j=0}^{k}\left(\eta_{(k)}-\eta_{(j)}\right)^3-\frac{3}{4}\cdot\left(\sum_{j=0}^{k}\left(\eta_{(k)}-\eta_{(j)}\right)^2\right)^2}{\left(\sum_{j=0}^{k}\left(\eta_{(k)}-\eta_{(j)}\right)\right)^3\cdot\left(\eta_{(k)}-\eta_{(0)}\right)}. \tag{5.13}$$

Pažymėsime $y_{\max} = \arg\max_{0\le y<\infty} F(y)$.

Sąlyga 5.1*. Tegul* $\eta_{(0)} < \eta_{(2)} < \ldots < \eta_{(k)}$. *Jei* $F(y_{\max}) > 0$, $0 < y_{\max} < \infty$ *ir*

$$\frac{1}{2}\cdot\frac{\sum_{j=0}^{k}\left(\eta_{(k)}-\eta_{(j)}\right)^2}{\left(\sum_{j=0}^{k}\left(\eta_{(k)}-\eta_{(j)}\right)\right)^2} \le \frac{1}{k}, \tag{5.14}$$

tada lygtis $F(y)=0$ *turi sprendinį,* $0 < y^* < \infty$, *kuriame funkcija* $F(y)$ *keičia ženklą iš "-" į "+".*

Sąlyga 5.2*. Tegul* $\eta_{(0)} < \eta_{(2)} < \ldots < \eta_{(k)}$. *Jei* $F(y_{\max}) > 0$, $0 < y_{\max} < \infty$ *ir*

$$\frac{8}{9}\cdot\sum_{j=0}^{k}\left(\eta_{(k)}-\eta_{(j)}\right)\cdot\sum_{j=0}^{k}\left(\eta_{(k)}-\eta_{(j)}\right)^3-\left(\sum_{j=0}^{k}\left(\eta_{(k)}-\eta_{(j)}\right)^2\right)^2 \le 0, \tag{5.15}$$

tada funkcija F(y) turi minimumą taške $0 < y < \infty$.

Šios sąlygos seka iš formulių (5.10) - (5.13).

Kompiuterinio modeliavimo būdu įsitikinome, kad funkcija $F(y)$ turi tiktai vieną minimumo tašką $0 < y_{max} < \infty$.

Galima išskirti 3 funkcijos $F(y)$ kitimo atvejus. *Lentelėje 5.1* pateikti pozicinių statistikų rinkiniai, su kuriais lygtis $F(y) = 0$ gali turėti arba neturėti sprendinio. Lentelėje pateiktas I-asis pozicinių statistikų rinkinys, su kuriuo funkcija $F(y)$ yra gaunama tokia, kad lygtis $F(y) = 0$ turi sprendinį y =3,303455, su II –uoju pozicinių statistikų rinkiniu gauta lygtis sprendinio neturi, todėl $y = \infty$, su III –uoju pozicinių statistikų rinkiniu gauta lygtis turi sprendinį tik begalybėje, todėl y = 0. Žemiau pateikti *pav. 5.1 – 5.3* iliustruoja minėtus funkcijos $F(y)$ kitimo atvejus. Šie rezultatai ir sąlygos 5.1 ir 5.2 leidžia lokalizuoti lygties $F(y) = 0$ sprendinio egzistavimo intervalą. Funkcijos $F(y)$ maksimumo arba minimumo taškus, reikalingus sprendiniui lokalizuoti, o taip pat ir pačiam lygties (5.6) sprendiniui gauti galima taikyti Niutono arba dalijimo pusiau metodus. Jei lygtį (5.6) spręstume Niutono metodu pradėdami nuo taško $y_0 = 0$, tai, panaudoję išraiškas (5.8), (5.9), (5.12), (5.13), galime apskaičiuoti kitos iteracijos tašką:

$$y_1 = \frac{\frac{4}{3}\left(\eta_{(k)} - \eta_0\right) \cdot \sum_{l=0}^{k}\left(\eta_{(k)} - \eta_{(l)}\right)^3 \cdot \left(\frac{\sum_{l=0}^{k}\left(\eta_{(k)} - \eta_{(l)}\right)^2}{2 \cdot \left(\sum_{l=0}^{k}\left(\eta_{(k)} - \eta_{(l)}\right)\right)^2} - \frac{1}{k} \right)}{\left(\sum_{l=0}^{k}\left(\eta_{(k)} - \eta_{(l)}\right)^2\right)^2 - \frac{8}{9} \cdot \sum_{l=0}^{k}\left(\eta_{(k)} - \eta_{(l)}\right) \cdot \sum_{l=0}^{k}\left(\eta_{(k)} - \eta_{(l)}\right)^3} \tag{5.16}$$

Šį tašką yra patogu naudoti iteraciniuose algoritmuose kaip pradinį tašką.

Įstatę gautą sprendinį y^* į formulę (5.7), gauname parametrų α, c bei A įverčius:

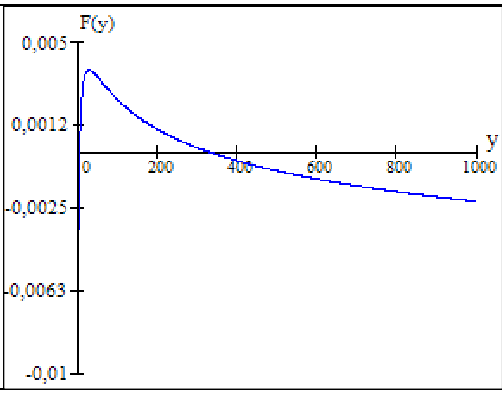
$$\hat{A} = \eta_0 - \frac{\eta_{(k)} - \eta_{(0)}}{y^*}, \tag{5.17}$$

$$\hat{\alpha} = \frac{k}{\sum_{j=0}^{k} \ln\left(1 + \frac{\eta_{(k)} - \eta_{(j)}}{\frac{1}{y^*} \cdot \left(\eta_{(k)} - \eta_{(0)}\right) + \eta_{(j)} - \eta_{(0)}} \right)}, \tag{5.18}$$

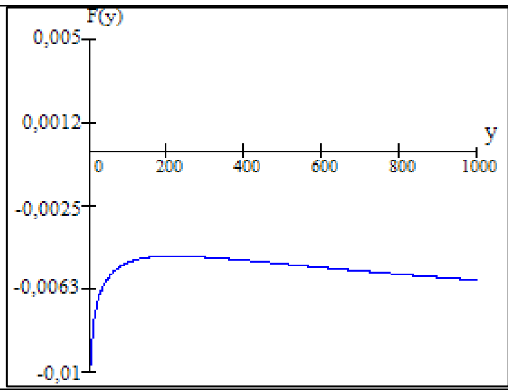
$$\hat{c} = \frac{k}{N \cdot \left(\eta_{(k)} - \eta_{(0)}\right)^{\hat{\alpha}} \cdot \left(1 + \frac{1}{y^*}\right)^{\hat{\alpha}}}. \tag{5.19}$$

Lentelė 5.1. Pozicinių statistikų rinkiniai

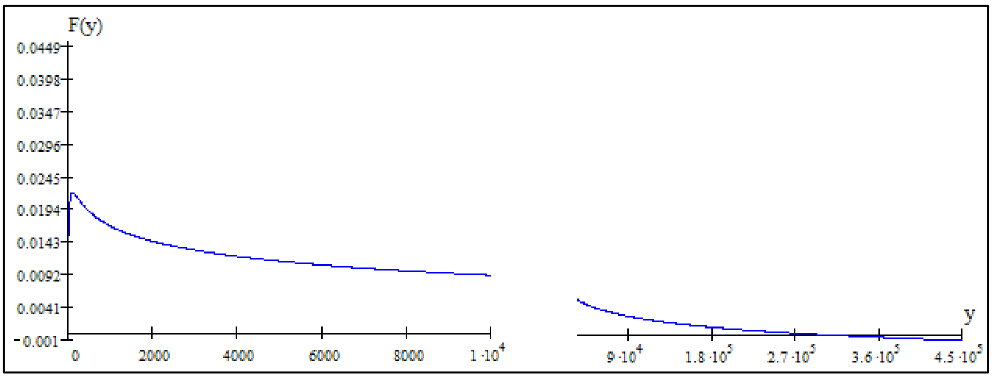
I	II	III
y=3,303455; α=2,382 F(0)= -0,00698; F'(0)= 0,0048245	y=∞; α=0 F(0)= -0,010431; F'(0)= 0,0009832	y=0; α=∞ F(0)= 0,0024354; F'(0)= 0,0078306
0,03557803	0,057444	0,024413834
0,03666162	0,062526	0,03888799
0,04933906	0,066576	0,069626927
0,07390800	0,091872	0,083658429
0,09326787	0,094020	0,111542977
0,14029756	0,096033	0,121783627
0,14767105	0,099611	0,130026012
0,15360406	0,120455	0,137666128
0,15703936	0,131640	0,148265311
0,15737907	0,150618	0,150428607
0,16227232	0,156487	0,151967254
0,16902877	0,178250	0,154760327
0,17252765	0,183191	0,155451889
0,18011925	0,183622	0,161724615
0,18494679	0,184540	0,162388212
0,18656994	0,186545	0,162827479
0,19601424	0,189034	0,164857189
0,19837062	0,193821	0,165373759
0,20460386	0,208171	0,168139369
0,21022217	0,208247	0,172904561



Pav. 5.1. Funkcija $F(y)$, atitinkanti I –ąjį pozicinių statistikų rinkinį



Pav. 5.2. Funkcija $F(y)$, atitinkanti II –ąjį pozicinių statistikų rinkinį



Pav. 5.3. Funkcija $F(y)$, atitinkanti III –ąjį pozicinių statistikų rinkinį

Jei funkcijos reikšmė maksimumo taške yra neigiama, tai lygtis (5.6) sprendinio neturi (II atvejis). Tokiu atveju siūloma priimti, kad $\hat{A} = \eta_{(0)}$, $y^* = \infty$ ir $\alpha = 0$. Jei sąlyga (5.14) neteisinga (III atvejis), tai lygtis (5.6) turi sprendinį, kurį atitinka tik tikėtinumo funkcijos minimumas. Pastaruoju atveju tikėtinumo funkcija neturi maksimumo taško intervale $0 < y < \infty$. Tokiu atveju taip pat galima priimti, kad $\hat{A} = \infty$, $y^* = 0$ ir $\alpha = \infty$.

Algoritmu 5.1 aprašytas trijų parametrų Veibulo skirstinio vertinimas maksimalaus tikėtinumo metodu. Maksimalaus tikėtinumo įverčio kompiuterinio modeliavimo rezultatai pateikti skyrelyje 5.4.

***Algoritmas 5.1*: Trijų parametrų Veibulo skirstinio vertinimas maksimalaus tikėtinumo metodu**

Tikslas: įvertinti Veibulo skirstinio parametrus maksimalaus tikėtinumo metodu.

Pradinės sąlygos (*preconditions*)**:** išrikiuota seka $\eta = \left(\eta_{(0)}, \eta_{(1)}, ..., \eta_{(k)}\right)$, pozicinių statistikų skaičius *k*.

Galinės sąlygos (*postconditions*)**:** Veibulo skirstinio trijų parametrų (formos, mastelio, postūmio) maksimalaus tikėtinumo įverčiai.

1. Inicializuoti pradinį tašką y_0 (5.16).
2. **If** neišpildyta sąlyga (5.14) **then** lygtis (5.8) sprendinio neturi ir $\hat{A} = \infty$, $y^* = 0$, $\alpha = \infty$
   else
   if neišpildyta sąlyga (5.15) **then** dalijimo pusiau būdu randame lygties (5.8) sprendinį *y*.
   else dalijimo pusiau metodu randame lygties (5.8) sprendinį *y*.
3. Taške *y* apskaičiuojame funkcijos reikšmę *f*.
4. **If** *f* >0 **then** dalijimo pusiau metodu randame lygties (5.8) sprendinį y.
   else $\hat{A} = \eta_{(0)}$, $y^* = \infty$ ir $\alpha = 0$.
5. Sprendimo priėmimas $y^* = y$. Apskaičiuojame Veibulo skirstinio parametrų įverčius pagal (5.17) – (5.19).

5.3. VEIBULO SKIRSTINIO TRIJŲ PARAMETRŲ VERTINIMAS ANALITINIAIS METODAIS

5.3.1. Analitinis įvertis

Weiss (1971), Haan (1981) pasiūlė paprastą analitinį ekstremaliųjų reikšmių skirstinio formos parametro įvertį

$$\hat{\alpha} = \frac{\ln\left(\frac{k}{m}\right)}{\ln\left(\frac{\eta(k) - \eta(0)}{\eta(m) - \eta(0)}\right)}, \tag{5.20}$$

čia $k \to \infty, \frac{m}{k} \to 0{,}2$. Empiriniai tyrimai rodo, kad šis įvertis yra paslinktas (žr. skyrelį 5.5). Geresnėmis savybėmis pasižymi tolimesniame skyrelyje nagrinėjamas įvertis.

5.3.2. Pagerintas analitinis įvertis

Kadangi įvertis (5.20) yra paslinktas ir tas poslinkis didėja, kai α yra didinamas, geresni rezultatai gali būti gaunami, kai vietoj $\eta_{(0)}$ yra naudojamas minimalios reikšmės tiesinis įvertis (4.1):

$$A_{N,k} = \eta_{(0)} - c_k\left(\eta_{(k)} - \eta_{(0)}\right) \tag{5.21}$$

Dar kartą pažymėsime, kad jei skirstinys (5.3) yra absoliučiai tolydusis, tai sąlyga $\eta_{(i)} \neq \eta_{(j)}, i \neq j,\ i, j = 1,...,m$ yra teisinga su tikimybe 1. Tokiu būdu tikslesnis įvertis yra gaunamas sprendžiant lygtį:

$$\alpha^* = \frac{\ln\left(\frac{k}{m}\right)}{\frac{\ln\left(\frac{k}{m}\right)}{\hat{\alpha}} + \ln\left(\frac{1+c_k}{1+c_m}\right)} = \frac{\ln\left(\frac{k}{m}\right)}{\ln\left(\frac{\eta(k) - \eta(0)}{\eta(m) - \eta(0)}\right) + \ln\left(\frac{1+c_k}{1+c_m}\right)},$$

čia $\hat{\alpha}$ apskaičiuojamas pagal (5.20).

Siekdami ištirti šios lygties sprendinio egzistavimą, įvesime pažymėjimą:

$$f(z, \alpha) = \frac{\ln\left(\frac{k}{m}\right)}{\ln(z) + \omega\left(\frac{1}{\alpha}\right)}, \tag{5.22}$$

čia $z = \frac{\eta(k) - \eta(0)}{\eta(m) - \eta(0)}$, $\omega(x) = \ln\left(\frac{1 - \frac{1}{\prod_{j=1}^{m}\left(1 + \frac{x}{j}\right)}}{1 - \frac{1}{\prod_{j=1}^{k}\left(1 + \frac{x}{j}\right)}}\right)$, $x = \frac{1}{\alpha}$.

Todėl pagerintas analitinis įvertis turi tenkinti lygtį:

$$\alpha^* = f\left(z, \alpha^*\right) \tag{5.23}$$

Šiai lygčiai spręsti yra patogu taikyti iteracinį metodą:

$$\alpha_{t+1} = f(z, \alpha_t) \tag{5.24}$$

čia α_0 -pradinis taškas.

Suformuluosime ir įrodysime lemą lygties (5.23) sprendinio egzistavimo sąlygoms nustatyti.

Lema 5.1. *Lygtis* (5.23) *turi sprendinį, jei* $z > \dfrac{\sum_{j=1}^{k}\frac{1}{j}}{\sum_{j=1}^{m}\frac{1}{j}}$.

Įrodymas.

▹ Funkcijos $\omega(x)$ išvestinė yra

$$\omega'(x) = \frac{\sum_{j=1}^{m}\left(\frac{1}{x+j}\right)}{\prod_{j=1}^{m}\left(1+\frac{x}{j}\right)-1} - \frac{\sum_{j=1}^{k}\left(\frac{1}{x+j}\right)}{\prod_{j=1}^{k}\left(1+\frac{x}{j}\right)-1}. \tag{5.25}$$

Įrodysime, kad santykio $\dfrac{\sum_{j=1}^{m}\left(\frac{1}{x+j}\right)}{\prod_{j=1}^{m}\left(1+\frac{x}{j}\right)-1}$ reikšmė didėja, kai yra didinamas *m*.

Išdiferencijavę turėsime:

$$S_m(x) = \left(\prod_{j=1}^{m}\left(1+\frac{x}{j}\right)\right)' = \sum_{j=1}^{m}\left(\frac{1}{x+j}\right)\cdot\prod_{j=1}^{m}\left(1+\frac{x}{j}\right) \ge 0,$$

$$S_m'(x) = \prod_{j=1}^{m}\left(1+\frac{x}{j}\right)\left(\left(\sum_{j=1}^{m}\frac{1}{x+j}\right)^2 - \sum_{j=1}^{m}\left(\frac{1}{x+j}\right)^2\right) \ge 0. \tag{5.26}$$

Toliau, iš Lagranžo formulės ir

$\left(\prod_{j=1}^{m}\left(1+\frac{x}{j}\right)-1\right) = x\cdot S_m(\xi) \le x\cdot S_m(x) = x\cdot\sum_{j=1}^{m}\left(\frac{1}{x+j}\right)\cdot\prod_{j=1}^{m}\left(1+\frac{x}{j}\right)$, $0 \le \xi \le x$ seka, kad

$$\frac{\sum_{j=1}^{m+1}\left(\frac{1}{x+j}\right)}{\prod_{j=1}^{m+1}\left(1+\frac{x}{j}\right)-1} = \frac{\left(\prod_{j=1}^{m}\left(1+\frac{x}{j}\right)-1\right)\cdot\frac{1}{x+m+1} - \frac{x}{m+1}\cdot\sum_{j=1}^{m}\left(\frac{1}{x+j}\right)\cdot\prod_{j=1}^{m}\left(1+\frac{x}{j}\right)}{\left(\prod_{j=1}^{m}\left(1+\frac{x}{j}\right)\cdot\left(1+\frac{x}{m+1}\right)-1\right)\left(\prod_{j=1}^{m}\left(1+\frac{x}{j}\right)-1\right)} + \frac{\sum_{j=1}^{m}\left(\frac{1}{x+j}\right)}{\prod_{j=1}^{m}\left(1+\frac{x}{j}\right)-1} \le$$

$$\le \frac{\sum_{j=1}^{m}\left(\frac{1}{x+j}\right)}{\prod_{j=1}^{m}\left(1+\frac{x}{j}\right)-1}.$$

Taigi $\omega'(x) \ge 0$, nes $m < k$. Pritaikę Lopitalio taisyklę, gauname, kad:

$$\lim_{x\to 0} \omega(x) = \ln\left(\frac{\sum_{j=1}^{m}\frac{1}{j}}{\sum_{j=1}^{k}\frac{1}{j}}\right) < 0. \tag{5.27}$$

Nesunku įsitikinti, kad $\lim_{x\to\infty} \omega(x) = 0$. Taigi, matome, kad funkcija $\omega(x)$ yra monotoniškai didėjanti. Iš žemiau pateiktų rezultatų:

$$\lim_{\alpha\to 0} f(z,\alpha) = \frac{\ln\left(\frac{k}{m}\right)}{\ln z}, \quad \lim_{\alpha\to\infty} f(z,\alpha) = \frac{\ln\left(\frac{k}{m}\right)}{\ln z + \ln\left(\frac{\sum_{j=1}^{m}\frac{1}{j}}{\sum_{j=1}^{k}\frac{1}{j}}\right)}$$

seka, kad, kai α didėja, funkcija $f(z, \alpha)$ yra monotoniškai didėjanti.

Be abejo, jei $z > \frac{\sum_{j=1}^{k}\frac{1}{j}}{\sum_{j=1}^{m}\frac{1}{j}}$ funkcijos $f(z,\alpha)$ grafikas turi susikirtimo tašką su kreive, žyminčia α kitimą. ◁

Lema 5.2. *Funkcijos* (5.22) *išvestinė yra aprėžta lygties* (5.23) *sprendinio taške:*

$$f'\left(z,\alpha^*\right) \le \frac{1}{2}\cdot\frac{\sum_{j=m+1}^{k}\frac{1}{j} - \frac{\sum_{j=1}^{k}\frac{1}{j^2}}{\sum_{j=1}^{k}\frac{1}{j}} + \frac{\sum_{j=1}^{m}\frac{1}{j^2}}{\sum_{j=1}^{m}\frac{1}{j}}}{\ln\left(\frac{k}{m}\right)} \le \frac{1}{2}. \tag{5.28}$$

Įrodymas.

▷ Turime, kad

$$f'(z,\alpha)=\frac{\ln\left(\frac{k}{m}\right)\cdot\frac{1}{\alpha^2}}{\left(\ln(z)+\omega\left(\frac{1}{\alpha}\right)\right)^2}\cdot\omega'\left(\frac{1}{\alpha}\right).$$

Įsitikinsime, kad

$$\lim_{x\to 0}\omega'(x)=\lim_{x\to 0}\left(\frac{\sum_{j=1}^{m}\left(\frac{1}{x+j}\right)}{\prod_{j=1}^{m}\left(1+\frac{x}{j}\right)-1}-\frac{\sum_{j=1}^{k}\left(\frac{1}{x+j}\right)}{\prod_{j=1}^{k}\left(1+\frac{x}{j}\right)-1}\right)=$$

$$=\lim_{x\to 0}\left[\frac{\sum_{j=1}^{m}\frac{1}{j}-x\sum_{j=1}^{m}\frac{1}{j^2}+o\left(x^2\right)}{x\cdot\sum_{j=1}^{m}\frac{1}{j}+\frac{x^2}{2}\cdot\left(\left(\sum_{j=1}^{m}\frac{1}{j}\right)^2-\sum_{j=1}^{m}\frac{1}{j^2}\right)+x\cdot o\left(x^2\right)}-\frac{\sum_{j=1}^{k}\frac{1}{j}-x\sum_{j=1}^{k}\frac{1}{j^2}+o\left(x^2\right)}{x\cdot\sum_{j=1}^{k}\frac{1}{j}+\frac{x^2}{2}\cdot\left(\left(\sum_{j=1}^{k}\frac{1}{j}\right)^2-\sum_{j=1}^{k}\frac{1}{j^2}\right)+x\cdot o\left(x^2\right)}\right]=$$

$$=\left[\frac{\sum_{j=1}^{k}\frac{1}{j}}{2}+\frac{\sum_{j=1}^{m}\frac{1}{j^2}}{2\sum_{j=1}^{m}\frac{1}{j}}-\frac{\left(\sum_{j=1}^{m}\frac{1}{j}\right)}{2}-\frac{\sum_{j=1}^{k}\frac{1}{j^2}}{2\left(\sum_{j=1}^{k}\frac{1}{j}\right)}\right]=\frac{\sum_{j=1+m}^{k}\frac{1}{j}}{2}+\frac{\sum_{j=1}^{m}\frac{1}{j^2}}{2\sum_{j=1}^{m}\frac{1}{j}}-\frac{\sum_{j=1}^{k}\frac{1}{j^2}}{2\sum_{j=1}^{k}\frac{1}{j}}.$$

Dabar įrodysime kad $\omega'(x)$ yra monotoniškai mažėjanti funkcija. Tai sektų iš nelygybės:

$$\omega''(x)=W(k,x)-W(m,x)\le 0,$$

čia $W(i,x)=\dfrac{\sum_{j=1}^{i}\frac{1}{(x+j)^2}\cdot\left(\prod_{j=1}^{i}\left(1+\frac{x}{j}\right)-1\right)+\prod_{j=1}^{i}\left(1+\frac{x}{j}\right)\cdot\left(\sum_{j=1}^{i}\frac{1}{x+j}\right)^2}{\left(\prod_{j=1}^{i}\left(1+\frac{x}{j}\right)-1\right)^2},\ i=1,2,3,...,\infty.$

Tačiau pakanka įrodyti, kad $W(i,x)$ yra mažėjanti priklausomai nuo i funkcija, būtent $W(i+1,x)<W(i,x)$. O tai seka iš nelygybės:

$$\sum_{j=1}^{i}\frac{1}{x+j}\le x\cdot\left(\frac{\left(\sum_{j=1}^{i}\frac{1}{x+j}\right)^2}{\prod_{j=1}^{i}\left(1+\frac{x}{j}\right)-1}+\frac{1}{2}\cdot\left(\sum_{j=1}^{i}\frac{1}{(x+j)^2}+\left(\sum_{j=1}^{i}\frac{1}{x+j}\right)^2\right)\right).$$

Iš to, kad $\omega'(x)$ monotoniškai didėja, t.y.

$$\lim_{x\to\infty} \omega'(x) = 0\,,\ \lim_{x\to 0} \omega'(x) = \frac{\sum_{j=m+1}^{k} \frac{1}{j}}{2} + \frac{\sum_{j=1}^{k} \frac{1}{j^2}}{2\cdot \sum_{j=1}^{k} \frac{1}{j}} - \frac{\sum_{j=1}^{m} \frac{1}{j^2}}{2\cdot \sum_{j=1}^{m} \frac{1}{j}} \le \frac{1}{2}\cdot \ln\left(\frac{k}{m}\right),$$

seka:

$$f'(z,\alpha) \le \frac{\frac{1}{2}\cdot \ln^2\left(\frac{k}{m}\right)\cdot \frac{1}{\alpha^2}}{\left(\ln(z) + \omega\left(\frac{1}{\alpha}\right)\right)^2}\,. \tag{5.29}$$

Ši nelygybė ir lygtis (5.23) ir įrodo šią lemą. ◁

Teorema 5.1. *Tarkime, kad* $z > \frac{\sum_{j=1}^{k} \frac{1}{j}}{\sum_{j=1}^{m} \frac{1}{j}}$. *Tada bet kokiam* $\frac{1}{2} \le \nu < 1$ *egzistuoja toks* $\varepsilon > 0$, *kad seka* (5.24) *konverguoja į* (5.23) *sprendinį tiesiniu greičiu*: $\left|\alpha_t - \alpha^*\right| = O\left(\nu^t\right)$, jei $\left|\alpha_0 - \alpha^*\right| \le \varepsilon$.

Įrodymas.

▷ Iš teoremos sąlygos ir lemos 5.1 seka, kad egzistuoja lygties (5.23) sprendinys α^*. Iš lemos 5.2 turime, kad bet kokiam $\frac{1}{2} < \nu < 1$ egzistuoja toks $\varepsilon > 0$, kad $\left|f'(\alpha)\right| \le \nu$, jei $\left|\alpha - \alpha^*\right| \le \varepsilon$.

Taigi, iš Lagrandžo teoremos ir (5.23) bei (5.25) turime:

$$\left|\alpha_t - \alpha^*\right| \le \left|f'\left(z, \alpha^* + \tau\left(\alpha_{t-1} - \alpha^*\right)\right)\right| \cdot \left|\alpha_{t-1} - \alpha^*\right| \le \nu \cdot \left|\alpha_{t-1} - \alpha^*\right| \le \nu^t \cdot \left|\alpha_0 - \alpha^*\right|,$$

kadangi $\left|f'\left(z, \alpha^* + \tau\left(\alpha_{t-1} - \alpha^*\right)\right)\right| \le \nu$, $t = 1,2,...$. ◁

Sekos (5.24) pradine reikšme rekomenduojama parinkti: $\alpha_0 = \dfrac{\ln\left(\frac{k}{m}\right)}{\ln(z) + \ln\left(\dfrac{\sum_{j=1}^{m} \frac{1}{j}}{\sum_{j=1}^{k} \frac{1}{j}}\right)}$.

Jei $z < \dfrac{\sum_{j=1}^{k} \frac{1}{j}}{\sum_{j=1}^{m} \frac{1}{j}}$, tai lygtis (5.23) neturi sprendinio. Tokiu atveju $\alpha^* = \hat{\alpha}$.

5.4. KOMPIUTERINIO MODELIAVIMO REZULTATAI

5.4.1. Veibulo skirstinio parametrų įverčių kompiuterinio modeliavimo rezultatai

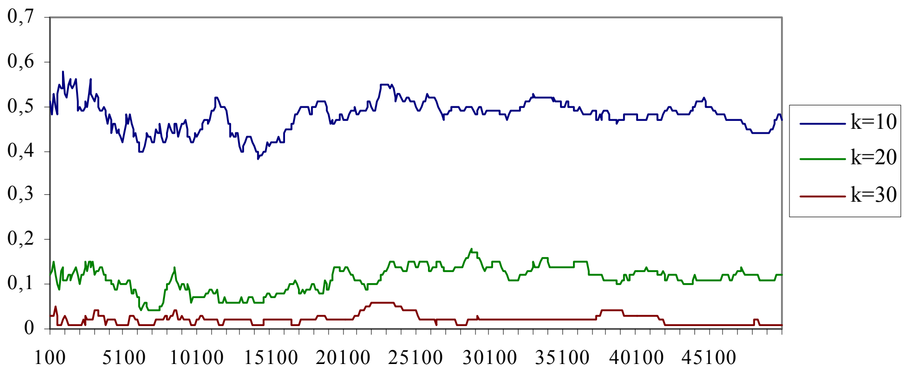
Šiame skyrelyje pateiksime Veibulo skirstinio parametrų įverčių kompiuterinio modeliavimo rezultatus. Skirstinio parametrų įverčiams skaičiuoti buvo generuojama M=100 atsitiktinių sekų, pasiskirsčiusių pagal Veibulo skirstinį (c=1, A=0) ir kiekvienai sugeneruotai sekai skaičiuojami parametrų įverčiai. *Lentelėje 5.2* pateikti Veibulo skirstinio parametrų α, A ir c įverčių vidurkiai, priklausomai nuo imties tūrio bei pozicinių statistikų skaičiaus k, gauti Monte-Karlo metodu.

Lentelė 5.2. Veibulo skirstinio parametrų Monte-Karlo įverčių vidurkiai (MLE – maksimalaus tikėtinumo įverčiai (*maximum likelihood estimators*), AE – analitiniai įverčiai (*analytical estimators*), IAE – pagerinti analitiniai įverčiai (*improved analytical estimators*))

***k*=100, *m*=20, α=2,5, c=1, A=0**				
N=1000	α	**dispersija**	**A**	**c**
MLE	2,4254	0,2776	0,0144	1,032
AE	2,0302	0,0904	0,0195	1,2072
IAE	2,6107	0,4560	-0,0090	1,0346
N=10000				
MLE	2,4342	0,3747	0,0038	1,1917
AE	2,0462	0,1507	0,0063	1,6540
IAE	2,7188	1,1151	-0,0077	1,1673
N=20000				
MLE	2,3774	0,4143	0,0046	1,4267
AE	1,9687	0,1063	0,0072	1,9706
IAE	2,4981	0,5099	-0,0007	1,3352
N=50000				
MLE	2,3979	0,5743	0,0030	1,3964
AE	1,9538	0,0961	0,0052	2,2493
IAE	2,4581	0,4129	0,0001	1,4110
***k*=500, *m*=100, α=5,0, c=1, A=0**				
N=1000	α	**dispersija**	**A**	**c**
MLE	4,9313	0,7662	0,0129	0,9881
AE	3,4307	0,1428	0,1157	0,9241
IAE	5,1715	1,7762	-0,0272	1,0286
N=10000				

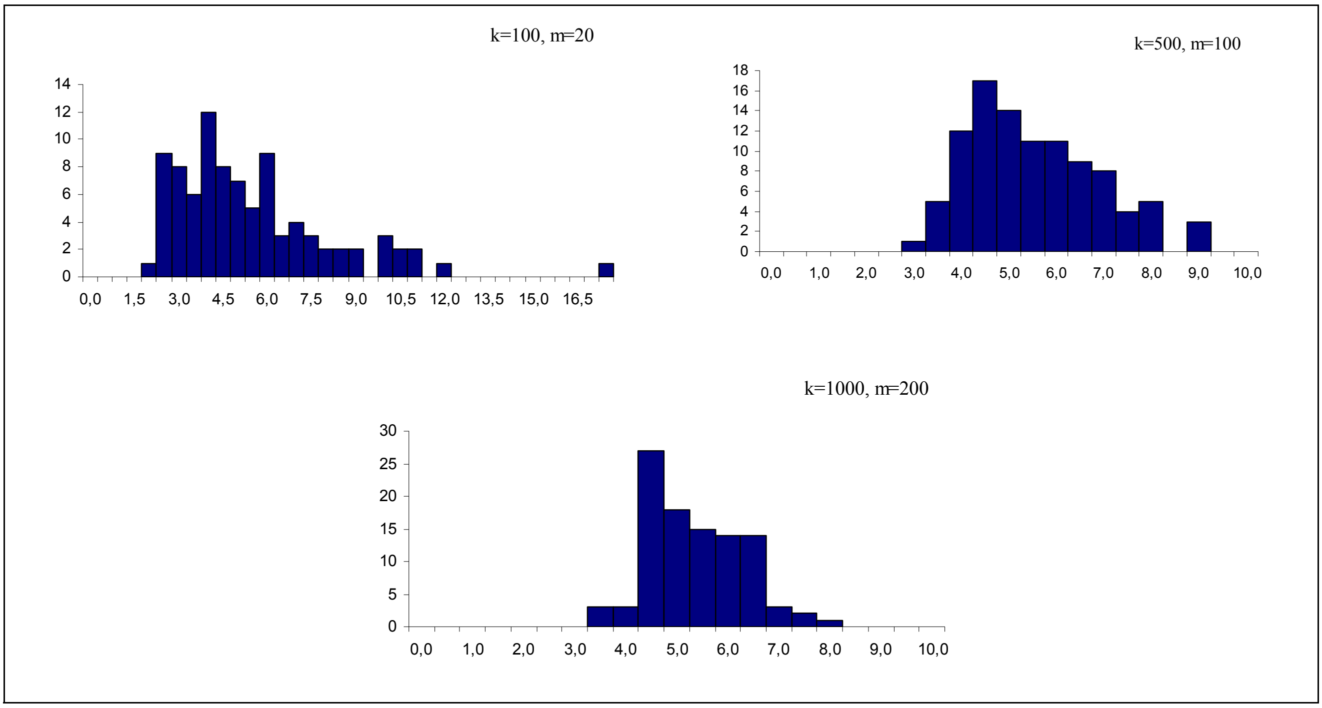
MLE	5,0345	0,9592	-0,0035	1,0141
AE	3,4632	0,1715	0,0676	1,1513
IAE	5,2958	1,8477	-0,0284	1,0373
N=20000				
MLE	5,0011	1,2829	0,00006	1,0190
AE	3,4105	0,1493	0,0667	1,2233
IAE	5,1052	1,5999	-0,0089	1,0278
N=50000				
MLE	5,0931	1,0688	-0,0036	1,0156
AE	3,4121	0,1515	0,0564	1,3350
IAE	5,1278	1,9152	-0,0074	1,0335

Pateikti rezultatai rodo, kad maksimalaus tikėtinumo bei pagerintas analitinis įverčiai gana gerai įvertina Veibulo skirstinio parametrus, kai sekos tūris ir pozicinių statistikų skaičius yra didinami. Norint gauti patikimus Veibulo skirstinio parametrų įverčius, kai formos parametras α didėja, reikia imti daugiau pozicinių statistikų ir generuoti didesnes atsitiktinių dydžių sekas, t.y. reikia didinti M ir N. Kai pozicinių statistikų skaičius k yra mažas, maksimalaus tikėtinumo įvertis gali neegzistuoti. Šio įverčio egzistavimo sąlygos yra aptartos 5.2 skyrelyje. *Paveiksle 5.4* pateikta maksimalaus tikėtinumo įverčio neegzistavimo tikimybė priklausomai nuo iteracijų skaičiaus kai pozicinių statistikų skačius yra nedidelis (k =10, 20, 30). Iš šio paveikslo matome, kad maksimalaus tikėtinumo įverčio neegzistavimo tikimybė mažėja k yra didinami. Didinant pozicinių statistikų skaičių šios tikimybės įvertis virsta 0.

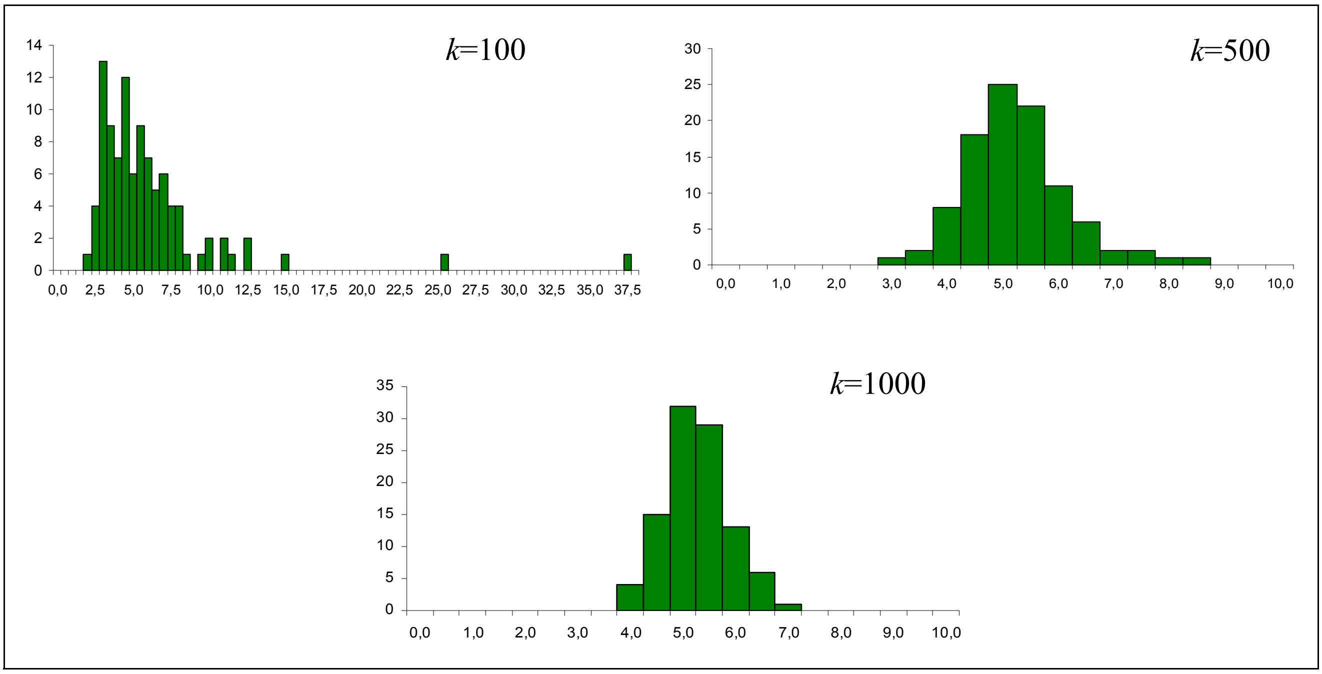


Pav. 5.4. Maksimalaus tikėtinumo įverčio neegzistavimo tikimybė priklausomai nuo iteracijų ir pozicinių statistikų skaičiaus, α=5,0

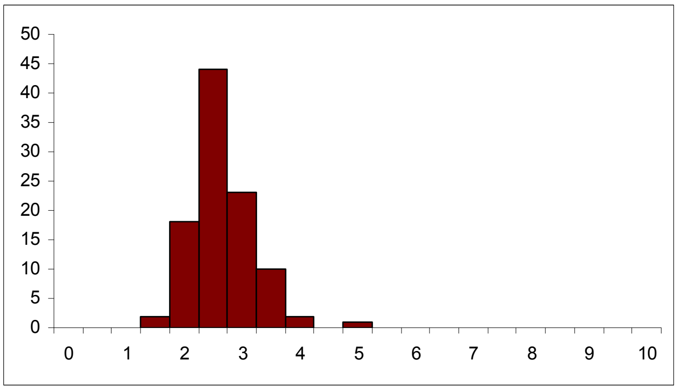
Paveiksluose 5.5 – 5.7 pateiktos įverčių, gautų pagerintu analitiniu arba maksimalaus tikėtinumo metodais, sklaidos histogramos. Iš šių paveikslų matome, kad didinant pozicinių statistikų skaičių šių įverčių išsibarstymas mažėja. Kaip ir reikėjo tikėtis maksimalaus tikėtinumo įverčio išsibarstymas yra mažesnis nei analitiniu metodu gautų įverčių ir didėjant pozicinių statistikų skaičiui gali būti aprašomas normaliniu dėsniu.



Pav. 5.5. Parametro α pagerinto analitinio įverčio histogramos, priklausomai nuo pozicinių statistikų skaičiaus (α=5, 100 sugeneruotų sekų)



Pav. 5.6. Parametro α maksimalaus tikėtinumo įverčio histogramos, priklausomai nuo pozicinių statistikų skaičiaus (α=5, 100 sugeneruotų sekų)



Pav. 5.7. Parametro α maksimalaus tikėtinumo įverčio histograma (α=2,5, k=100, 100 sugeneruotų sekų)

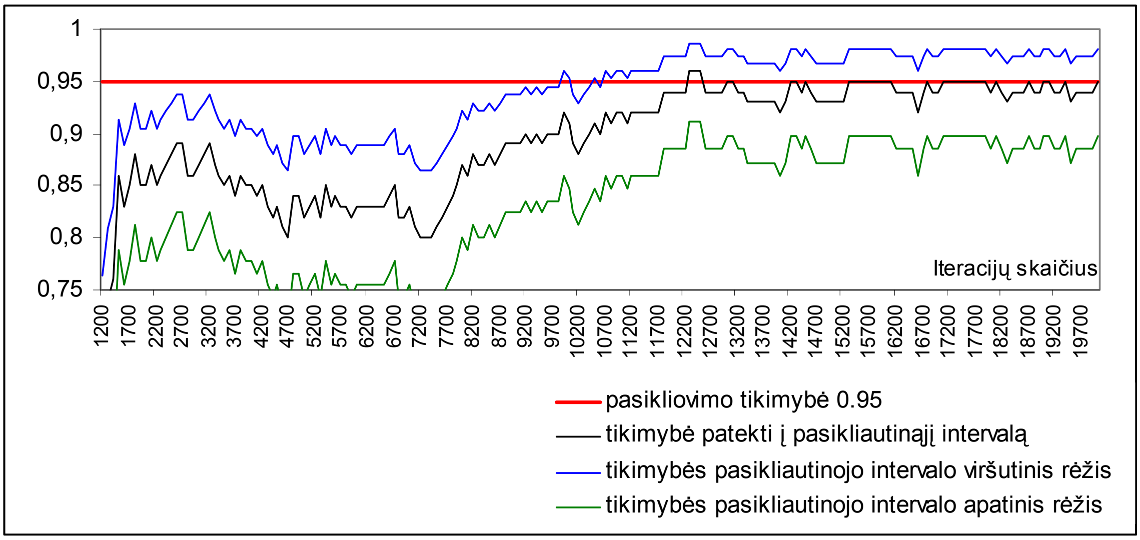
5.4.2. Optimizavimo sekos optimalumo tyrimo metodas, kai ekstremaliųjų reikšmių skirstinio parametras yra nežinomas

Šiame skyrelyje pateikiami ekstremaliųjų reikšmių skirstinio formos parametro įverčių tinkamumo stochastinės paieškos sprendinio optimalumui tirti skaičiuojamojo eksperimento rezultatai.

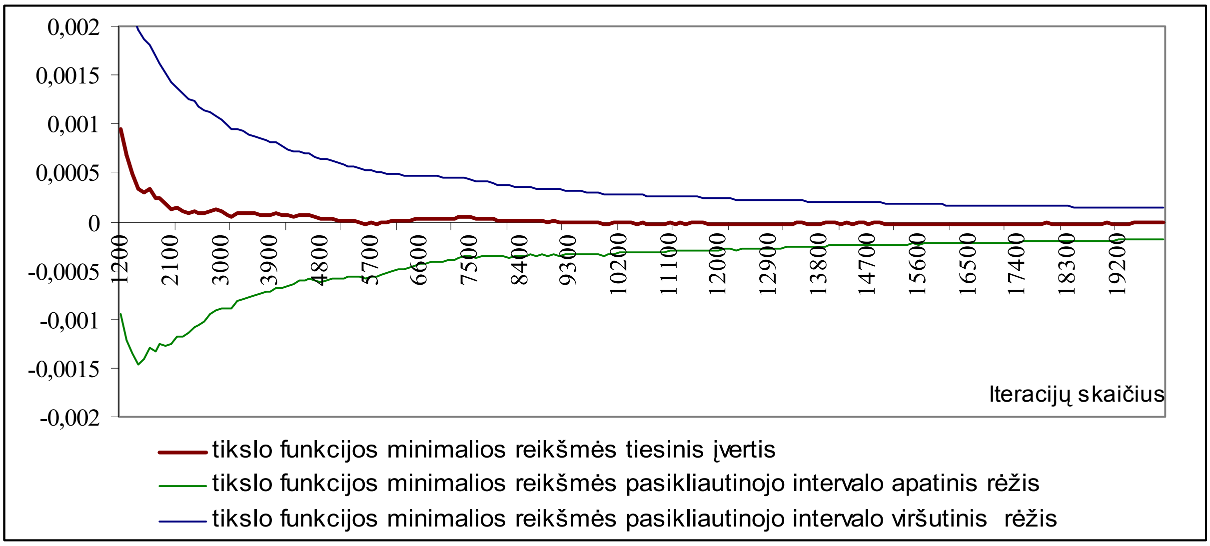
Lentelė 5.3. Tikslo funkcijos minimalios reikšmės, jos vienpusio pasikliautinojo intervalo, tikimybės minimaliai reikšmei patekti į pasiklautinąjį intervalą Monte-Karlo įverčių vidurkiai bei tikimybės pasikliautinasis intervalas, kai ekstremaliųjų reikšmių skirstinio formos parametras α įvertinamas statistiniais metodais, pasinaudojant geriausiomis pasiektomis optimizavimo metu tikslo funkcijos reikšmėmis (funkcija su aštriu minimumu, n=2, k=500, m=100, N=20000, M=100)

A=0,μ=2, K=5, $\delta = 0.95$	α ***įverčiai***	$A_{N,k}$	***Vienpusis pasikliautinasis intervalas***		***p***	***Patekimo tikimybės p pasikliautinasis intervalas***	
			Apatinis rėžis	***Viršutinis rėžis***		***Apatinis rėžis***	***Viršutinis rėžis***
MLE	1,9843	-0,000018	-0,00018	0,000139	0,95	0,89763	0,9801
AE	1,8919	0,0000048	-0,000138	0,000139	0,87	0,8007	0,9215
IAE	2,0220	-0,000029	-0,000197	0,000139	0,95	0,8976	0,9801

Lentelėje 5.3 pateikiami ekstremaliųjų reikšmių skirstinio formos parametro maksimalaus tikėtinumo ir analitiniai įverčiai, gauti pasinaudojus geriausiomis optimizavimo metu pasiektomis tikslo funkcijos reikšmėmis, lyginant šiuos įverčius su ekstremaliųjų reikšmių skirstinio formos parametru, įvertinamu pasinaudojant funkcijos homogeniškumu minimumo aplinkoje ($\alpha = \frac{n}{\beta} = 2$), o taip pat tikslo funkcijos minimalios reikšmės bei jos pasikliautinojo intervalo Monte-Karlo įverčių modeliavimo rezultatai.



Pav. 5.8. Tikslo funkcijos minimalios reikšmės patekimo į pasikliautinąjį intervalą tikimybė ir jos pasikliautinasis intervalas (funkcija su aštriu minimumu, n=2, A=0, δ=0.95, SPSAL, MLE)



Pav. 5.9. Tikslo funkcijos minimalios reikšmės tiesinis įvertis ir jos pasikliautinieji rėžiai (funkcija su aštriu minimumu, n=2, SPSAL, MLE)

Iš *lentelėje 5.3* bei *pav. 5.8* ir *5.9* pateiktų rezultatų matome, kad kai pozicinių statistikų skaičius bei iteracijų skaičius yra dideli, ekstremaliųjų reikšmių skirstinio formos parametro maksimalaus tikėtinumo ir analitiniai įverčiai, taikomi tikslo funkcijos minimaliai reikšmei įvertinti, artėja į įvertį (4.8), gaunamą pasinaudojant funkcijos homogeniškumu minimumo aplinkoje. Analogiški rezultatai gaunami, kai tikslo funkcija minimizuojama kitais disertacijoje išnagrinėtais optimizavimo metodais (SPSAU, FDSA) bei kai ekstremaliųjų reikšmių parametras vertinamas pagerintu analitiniu metodu. Apibendrinant skaičiuojamojo eksperimento metu gautus rezultatus, galima teigti, kad tikslo funkcijos minimalią reikšmę galima įvertinti pasinaudojus skyrelyje 4.2 įvestais įverčiais, kai pozicinių statistikų skaičius ir iteracijų skaičius didėja. Taigi, jei nėra kitų būdų, nežinomas ekstremaliųjų reikšmių skirstinio formos parametras gali būti įvertinamas statistiniais metodais.

5.5. REZULTATAI IR IŠVADOS

1. Sudarytas algoritmas Veibulo skirstinio parametrams (formos, padėties ir skalės) įvertinti maksimalaus tikėtinumo metodu, naudojant necenzūruotos imties pozicines statistikas;
2. Suformuluotos ir įrodytos maksimalaus tikėtinumo įverčio egzistavimo sąlygos bei nustatytas šio įverčio lokalizavimo intervalas;
3. Sukurtas analitinis metodas, apibrėžtas bet kokiam pozicinių statistikų rinkiniui, ir sudarytas jį realizuojantis algoritmas Veibulo skirstinio parametrams (formos, padėties ir skalės) įvertinti, taikant tikslo funkcijos minimalios reikšmės tiesinį įvertį ir naudojant necenzūruotos imties pozicines statistikas;
4. Suformuluotos ir įrodytos pagerinto analitinio įverčio egzistavimo sąlygos;
5. Įrodytas analitinio algoritmo konvergavimas į Veibulo skirstinio formos parametro įvertį tiesiniu greičiu;
6. Sukurtas analitinis metodas kompiuterinio modeliavimo metu pademonstravo pranašumą lyginant jį su Haan (1981) pasiūlytu analitiniu Veibulo skirstinio formos parametro įverčiu;
7. Maksimalaus tikėtinumo ir sukurtas analitinis įverčiai, gauti kompiuterinio modeliavimo būdu, yra pagrįsti bei tinkami Veibulo skirstinio parametrams vertinti;
8. Tikslo funkcijos minimalią reikšmę galima įvertinti pasinaudojus tiesiniais įverčiais, kai ekstremaliųjų reikšmių parametras vertinamas pagerintu analitiniu arba maksimalaus tikėtinumo metodais, o pozicinių statistikų skaičius ir iteracijų skaičius yra pakankamai dideli.

6. STOCHASTINIŲ NEDIFERENCIJUOJAMO OPTIMIZAVIMO ALGORITMŲ PROGRAMINĖ REALIZACIJA BEI TAIKOMŲJŲ UŽDAVINIŲ SPRENDIMAS

6.1. STOCHASTINIŲ NEDIFERENCIJUOJAMO OPTIMIZAVIMO ALGORITMŲ PROGRAMINĖ REALIZACIJA

6.1.1. Techninė ir programinė įranga

Sudarytų stochastinių nediferencijuojamo optimizavimo algoritmų programinei realizacijai, testavimui bei taikomiesiems uždaviniams spręsti buvo naudojami MII lokalinio tinklo AK, Tarptautinio teorinės fizikos centro ICTP (Triestas, Italija) tinklo AK, superkompiuterių centro CINECA (Bolonija, Italija) lygiagrečių skaičiavimų klasteris.

Programinei įrangai realizuoti bei testuoti daugiausia buvo naudojamas MII lokalinio tinklo AK su AMD Athlon 64 1.8 GHz procesoriumi, RAM 512 MB, 120 GB.

Sukurtų algoritmų praktinio taikymo galimybės finansinio opciono kainai skaičiuoti buvo tiriamos ICTP lokalinio tinklo PK: Pentium 4 2.0/2.4 Ghz, CPU 256/512 MB RAM, 40 GB kietasis diskas ir Pentium 4 2.6 Ghz CPU 512 MB RAM 80 GB kietasis diskas (http://scs.ictp.it/facilities/hardware).

Kadangi nagrinėjamų ir taikomų stochastinės aproksimacijos metodų lėtas konvergavimas reikalauja labai didelio iteracijų skaičiaus, o tuo pačiu ir didelio sprendimo laiko, reikalingo uždaviniui išspręsti, todėl konvergavimo greičiui bei algoritmų stabdymo taisyklei tirti didelio matavimo uždaviniuose buvo pasinaudota superkompiuterių centro CINECA lygiagrečių skaičiavimų klasteriu CLX (klasteris – lygiagrečioji arba paskirstytoji sistema, kurią sudaro keli tarpusavyje susieti kompiuteriai, naudojama kaip vieningas unifikuotas kompiuterinis resursas). Tokio klasterio panaudojimas leido ne tik išspręsti optimizavimo problemas reikiamu tikslumu, bet ir padidinti skaičiavimų spartą. CLX tai IBM Linux Cluster 1350, sudarytas iš 512 dviprocesorinių IBM X335 mazgų (nodes), kiekvienas skaičiavimų mazgas sudarytas iš 2 Xeon Pentium IV procesorių, visi skaičiavimų mazgai yra 2GB atminties (1GB vienam procesoriui), prisijungimo ir aptarnaujantis skaičiavimų mazgas sudarytas iš 2.8GHz procesorių. CLX gali būti pasiekiamas 6.1 TFlops skaičiavimų greitis (https://hpc.cineca.it/docs/user-guide-zwiki/CLXUserGuide).

Šiame skyriuje pateiksime trumpą analitinę apžvalgą tokių programinių priemonių, kurios tiesiogiai buvo pritaikytos kompiuteriniams modeliams kurti. Taikytas programines priemones galima suskirstyti į dvi grupes: matematines sistemas ir universalias programavimo kalbas.

Matematinės sistemos: skaičiavimams atlikti dažnai buvo naudojamos matematinės sistemos Mathematica ir MathCad. MathCad 2001 matematinė sistema pasirinkta, todėl, kad ši sistema išsiskiria nesudėtinga vartotojo sąsaja, standartinių funkcijų ir simbolinio skaičiavimo bei programavimo priemonėmis, galimybe naudotis *Windows*, puikia grafika ir labai didele pavyzdžių biblioteka. Mathematicos pasirinkimą lėmė tai, kad tai galinga kompiuterinė matematinė sistema su plačiomis skaitinių ir simbolinių skaičiavimų galimybėmis, turinti galingas grafikos ir animacijos galimybes, o be to komandų sintaksė panaši į C kalbos sintaksę.

Visos aukščiau aptartos priemonės neleidžia kurti paleidžiamųjų programų (*.exe failų), o tai reiškia, kad norint kurti konkretų kompiuterinį modelį turime pasinaudoti kokiu nors programiniu paketu. Programiniams paketams kurti buvo pasirinktos universaliosios programavimo kalbos tokios kaip MS Visual C++, Free Pascal.

Programavimo sistemos:

C++ programavimo kalba leidžia geriausiai išnaudoti kompiuterinius resursus. Šia kalba parašytos programos yra kompaktiškos ir greitai vykdomos, o be to jos gali būti lengvai, su nedideliais pataisymais arba visai be jų, perkeliamos į kitas skaičiavimo sistemas, pvz.: iš Windows operacinės sistemos perkelti programos veikimą į UNIX operacinę sistemą.

FreePascal (versija 1.0.6 1.9) nemokama programavimo sistema, naudojanti 32 bitų technologiją ir kalbos atžvilgiu suderinama su Turbo Pascal 7.0 ir Delphi 5 dialektais. Ši programavimo sistema yra atviro kodo, turinti patogią grafinę sąsają, be to ją lengva įdiegti. Įdiegus išsyk galima pradėti dirbti. Nereikia nieko papildomai nustatyti ar įdiegti papildomų priemonių. Turi visas svarbiausias funkcines galimybes, tokias kaip: pažingsninio derinimo, reiškinių reikšmių stebėjimo ir jų reikšmių pasikeitimo, priskyrimo, skaitymo stebėjimo galimybes.

Dažniausiai sukurtos programos buvo testuojamos operacinėje sistemoje Windows 2000, o lygiagrečiųjų skaičiavimų testai atliekami LINUX operacinėje sistemoje.

6.1.2. Skaičiuojamųjų stochastinių Markovo tipo algoritmų tyrimas, taikant lygiagrečiųjų skaičiavimų metodus

Lygiagretūs algoritmai – tai algoritmai kuriuose yra išskirtos veiksmų grupės, kurias galima vykdyti tuo pačiu metu. Lygiagretusis kompiuteris tokį algoritmą gali vykdyti sparčiau, negu įprastą nuoseklų algoritmą, kadangi nepriklausomas lygiagrečias algoritmo šakas gali paskirstyti tarp atskirų jo turimų procesorių (Čiegis (2001)). Todėl superkompiuterių bei lygiagrečiųjų skaičiavimų panaudojimas leidžia ne tik išspręsti optimizavimo problemas reikiamu tikslumu, bet ir padidinti skaičiavimų spartą, įgauti didesnius kompiuterinius pajėgumus ir išspręsti didesnius uždavinius.

6.1.2.1. Skaičiuojamojo eksperimento Monte – Karlo metodu išlygiagretinimas

Lygiagretiems algoritmams programuoti buvo pasirinktas MPI (*Message-Passing Interface*) standartas. Programos buvo rašomos standartine C++ kalba ir tik jas kompiliuojant prijungiama MPI biblioteka. Naudodami MPI biblioteką, lygiagretųjį algoritmą skaidome į procesus, kurie gali būti vykdomi lygiagrečiai. Žemiau pateikti rezultatai buvo gauti pasinaudojus FP6 projekto HPC-EUROPA (RII3-CT-2003-506079) lėšomis superkompiuterių centre CINECA (Bolonija, Italija) ir pristatyti tarptautinėje konferencijoje *2nd Conference on Optimization Methods & Software and 6th EUROPT Workshop on Advances in Continuous Optimization* (Praha, Čekija, 2007) ir *Lietuvos jaunųjų mokslininkų konferencijoje* "Operacijų tyrimas ir taikymai" (Vilnius, 2007).

Išlygiagretinant Monte-Karlo metodą laikysime, kad visos užduotys yra nepriklausomos ir visų užduočių sudėtingumas yra vienodas. Todėl kiekvienam procesoriui paskirta vykdyti $\left[\frac{M}{p}\right]$ algoritmo realizacijų, čia $[\cdot]$ skaičiaus sveikoji dalis. Jei algoritmo realizavimų skaičius M be liekanos nesidalija iš procesorių skaičiaus p, tai liekaną išdalijame pirmiesiems procesoriams, t.y. pirmieji $M - \left[\frac{M}{p}\right] \cdot p$ valdomi procesoriai (darbininkai) atlieka dar po vieną papildomą algoritmo realizaciją. Valdantysis procesorius taip pat apskaičiuoja statistinius algoritmo charakteristikų įverčius.

Lentelė 6.1. Laikas (*sekundėmis*), per kurį išsprendžia uždavinį SPSAL lygiagretusis algoritmas naudojant p procesorių

n \ p	1	10	20	100
2	7,4718	0,7589	0,2438	0,1269
10	17,988	1,8185	1,5210	0,2527
50	69,877	7,0307	4,0176	0,7485
100	134,77	13,562	6,8319	1,4086

Skaičiuojamojo eksperimento metu buvo tiriamas pagrindinių MPI konstrukcijų naudojimo efektyvumas, šiuo būdu sukurtų lygiagrečiųjų programų efektyvumas ir spartinimo koeficientas. Spartinimo koeficientas (lygiagrečiojo algoritmo pagreitėjimas) sprendžiant uždavinį su p procesorių (Čiegis (2001)):

$$S_p = \frac{T_1}{T_p},$$

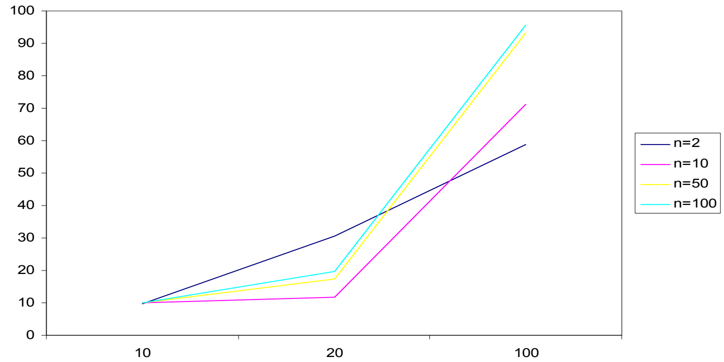
čia T_1 yra laikas, kurį sugaišta nuoseklusis algoritmas, spręsdamas uždavinį su vienu procesoriumi, o T_p - laikas, per kurį lygiagretusis algoritmas išsprendžia uždavinį su p procesorių.

Algoritmo efektyvumas skaičiuojamas tokiu būdu:

$$E_p = \frac{S_p}{p}.$$

Lentelė 6.2. SPSAL lygiagretaus algoritmo spartinimo ir efektyvinimo koeficientai

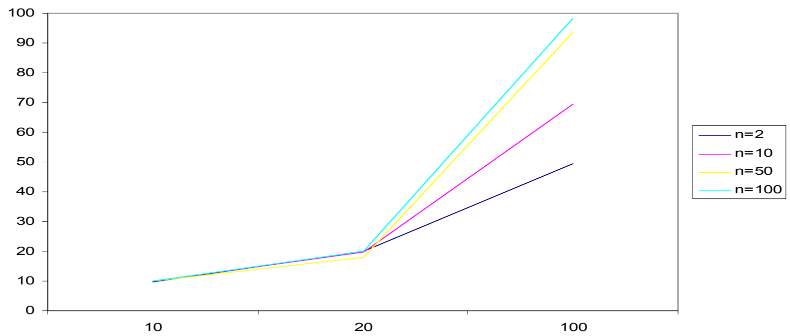
n=100	p10	p20	p100
Spartinimo koeficientas	9,9373	19,727	95,6765
Efektyvumo koeficientas	0,9937	0,9863	0,9567



Pav. 6.1. SPSAL lygiagretaus algoritmo spartinimo koeficientas

Lentelė 6.3. Laikas (*sekundėmis*), per kurį išsprendžia uždavinį SPSAU lygiagretusis algoritmas naudojant p procesorių

n \ p	1	10	20	100
2	5,2868	0,5392	0,2655	0,1069
10	12,930	1,3077	0,6529	0,1860
50	73,58	7,3946	4,0875	0,7864
100	196,6	19,799	9,9019	2,0006



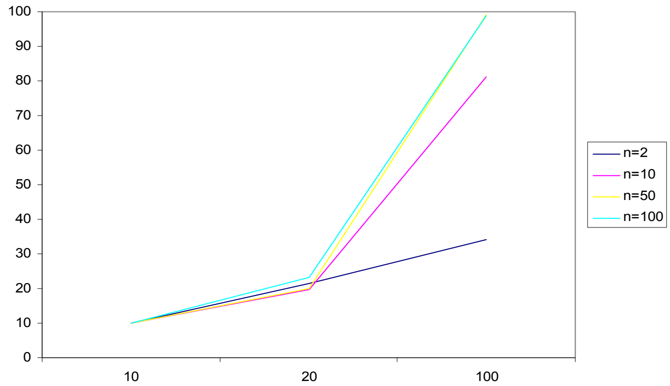
Pav. 6.2. SPSAU lygiagretaus algoritmo spartinimo koeficientas

Lentelė 6.4. SPSAU lygiagretaus algoritmo spartinimo ir efektyvinimo koeficientai

n=100	p10	p20	p100
Spartinimo koeficientas	9,9298	19,855	98,2705
Efektyvumo koeficientas	0,9929	0,9927	0,9827

Lentelė 6.5. Laikas (*sekundėmis*), per kurį išsprendžia uždavinį FDSA lygiagretusis algoritmas naudojant p procesorių

n \ p	1	10	20	100
2	7,914	0,8032	0,3688	0,2319
10	21,179	2,1409	1,0758	0,2606
50	133,90	13,455	6,7251	1,3529
100	365,03	36,822	15,809	3,6925



Pav. 6.3. FDSA lygiagretaus algoritmo spartinimo koeficientas

Lentelė 6.6. FDSA lygiagretaus algoritmo spartinimo ir efektyvinimo koeficientai

n=100	p10	p20	p100
Spartinimo koeficientas	9,9134	23,090	98,8571
Efektyvumo koeficientas	0,9913	1,1545	0,9886

Lygiagrečiųjų skaičiavimų kompiuterinio modeliavimo būdu gauti rezultatai, pateikti 3 ir 4 skyrių lentelėse *3.3*, *4.1* bei *4.3*, leido patvirtinti teorines stochastinės aproksimacijos algoritmų konvergavimo sąlygas, kai sprendžiami didelio matavimo optimizavimo uždaviniai (funkcija su aštriu minimumu, n=100) bei patvirtinti nepriklausomų vienodai pasiskirsčiusių atsitiktinių dydžių ekstremaliųjų reikšmių teorijos tinkamumą tikslo funkcijos minimaliai reikšmei įvertinti. Be to MPICC kalba sukurti programiniai kodai gali būti lengvai realizuojami su bet kokia lygiagrečiųjų skaičiavimų architektūra. O sukurtos lygiagrečiųjų SPSAL, SPSAU ir FDSA algoritmų programos gali būti pritaikytos taikomiesiems uždaviniams spręsti.

6.2. TAIKOMŲJŲ UŽDAVINIŲ SPRENDIMAS

6.2.1. SPSA algoritmo taikymas Heston'o stochastinio volatiliškumo modeliui kalibruoti

Nagrinėsime sukurto SPSA algoritmo praktinio taikymo galimybes finansinio opciono kainai įvertinti. Finansinė duomenų analizė rinkos tyrimuose ir rizikos analizė taip pat kaip ir valdymas labai dažnai susiję su numanomu ir realizuotu volatiliškumu (*implied* ir *realized volatility*). Pademonstruosime SPSA algoritmo tinkamumą vidutinei absoliučiajai opciono kainos paklaidai minimizuoti įvertinant Heston'o stochastinio volatiliškumo modelio parametrus (Heston (1993)). Šis uždavinys buvo suformuluotas mokslinės stažuotės tarptautiniame fizikos centre ICTP (2004, Triestas, Italija) metu vykusiame darbiniame seminare "*Matematinis modeliavimas pramonės ir gamtos moksle*" bei gauti rezultatai pristatyti tarptautinėse konferencijose: *Kompiuterininkų dienos-2005* (KU, Klaipėda, 2005), *INYS workshop Optimal Design of Technological Processes*, (Vilnius, 2006), *taikomosios statistikos XXVI tarptautinėje konferencijoje* (Killarnei, Airija, 2006), o taip pat publikuoti moksliniuose leidiniuose: Bartkutė & Sakalauskas (2006a), Bartkutė & Sakalauskas (2005a), Bartkutė (2006c).

Hestono modeliu paskaičiuota opciono kaina yra lyginama su rinkoje turimomis opciono kainomis. Tuo tikslu apibrėšime vidutinę absoliutinę opciono kainos paklaidos funkciją:

$$MAE(\kappa, \sigma, \rho, \mathrm{r}, \mathrm{q}, \theta) = \frac{1}{n}\sum_{i=1}^{n}\left|C_i^H(\kappa, \sigma, \rho, \mathrm{r}, \mathrm{q}, \theta) - C_i\right|, \tag{6.1}$$

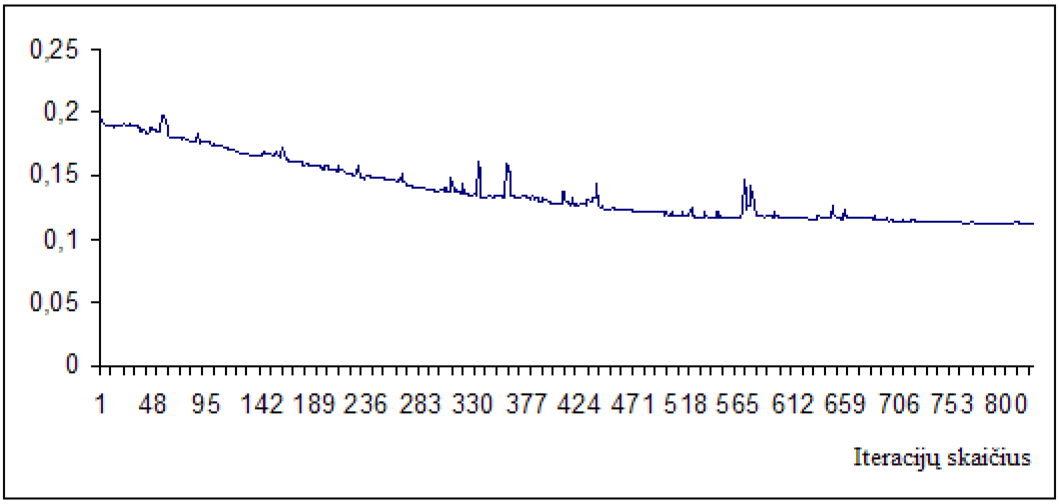
čia n yra visų turimų opcionų skaičius, C_i- opciono kaina rinkoje ir C_i^H- Heston'o modeliu apskaičiuota opciono kaina, $(\kappa, \sigma, \rho, r, q, \theta)$ - įvertinami Heston'o modelio parametrai (κ-reversijos greitis (*reverting speed*), σ kintamumo volatiliškumas ("*volatility of volatility*"), ρ koreliacijos koeficientas tarp aktyvo grąžos ir volatiliškumo (*correlation coefficient between the asset return and its volatility*), r, q yra palūkanų ir dividendų normos atitinkamai, θ ilgalaikio trendo vidurkis (*the long run mean level*)) (Heston (1993)).

Norėdami apskaičiuoti opciono kainą pagal Hestono'o modelį, turime įvertinti šio modelio parametrus, kurių negalime nustatyti turėdami rinkos duomenis. Parametrams įvertinti sudarysime tikslo funkciją (6.1):

$$MAE(\kappa, \sigma, \rho, \mathrm{r}, \mathrm{q}, \theta) \to \min . \tag{6.2}$$

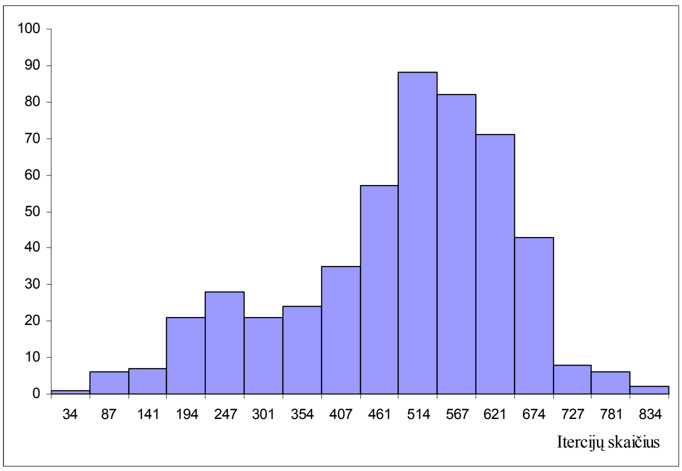
Teoriniu modeliu aspkaičiuojama opciono kaina C_i^H, yra išreiškiama kaip $C_i^H = S \cdot e^{-q \cdot \tau} P_1 - K \cdot e^{-r \cdot \tau} \cdot P_2$, kur $P_j, j = 1, 2$, yra dvi tikimybinės funkcijos, ir S finansinio aktyvo kaina, K opciono įvykdymo kaina. Detalus šios išraiškos išvedimas yra pateiktas (Heston (1993)).

Heston'o modelis buvo realizuotas su Vokietijos akcijų SPX (29 Gegužės, 2002) pirkėjo opcionų (*Call option*) kursais. Heston'o stochastinio volatiliškumo modelis paaiškina daugelį efektų vertybinių popierių rinkose, tačiau šio modelio pritaikymas reikalauja gana sudėtingo modelio parametrų įvertinimo panaudojant optimizavimo algoritmus. Tuo tikslu buvo pritaikytas SPSAL algoritmas.



Pav. 6.4. Vidutinės absoliutinės opciono kainos paklaidos priklausomybė nuo iteracijų skaičiaus, stabdant algoritmą, kai $\varepsilon = 0{,}0005$.

Pav. 6.4 pateikta vidutinės absoliutinės opciono kainos paklaidos priklausomybė nuo iteracijų skaičiaus, kai algoritmas stabdomas esant stabdymo sąlygai $\varepsilon = 0{,}0005$. Ši priklausomybė iliustruoja taikomo metodo konvergavimą, tačiau norint pasiekti optimalią reikšmę reikia mažinti pasikliautinojo intervalo plotį, tuo pačiu didinant iteracijų skaičių. *Pav. 6.5* pateiktoje histogramoje matome algoritmo žingsnių skaičiaus sklaidą priklausomai nuo algoritmui stabdyti nustatyto pasikliautinojo intervalo pločio. Vidutinis iteracijų skaičius, kai tikslo funkcijos minimalios reikšmės pasikliautinojo intervalo plotis yra 0,02 lygus 463,308 iteracijų.



Pav.6.5. Iteracijų skaičius reikalingas algoritmui stabdyti, kai $\varepsilon = 0{,}02$, $n = 6$, M=500.

6.2.2. Stochastinės aproksimacijos taikymas tankerio pertvarų svoriui optimizuoti

Optimizavimo uždaviniai su netiesiniais ribojimais labai dažnai sutinkami inžineriniame projektavime. Panagrinėsime stochastinės aproksimacijos taikymą optimizavimo uždaviniams su ribojimais spręsti baudos funkcijos metodu. Šio taikomojo uždavinio rezultatai paskelbti tarptautinėse konferencijose: *INYS workshop Optimal Design of Technological Processes*, (Vilnius, 2006), *22nd European Conference on Operational Research* (EURO XXII) (Praha, 2007), o taip pat publikuoti moksliniame leidinyje Bartkutė & Sakalauskas (2006a).

Tarkime turime netiesinio programavimo uždavinį:

$$f_0(x) \to \min,$$

$$f_i(x) \le 0 \ , i = 1, \ldots, m. \tag{6.3}$$

Šį uždavinį spręsime taikant baudos funkcijas. Tada baudos funkcija šiam uždaviniui yra sudaroma tokiu būdu:

$$F(x) = f_0(x) + \sum_{i=1}^{m} \mu_i \cdot \max(0, \ f_i(x)), \tag{6.4}$$

čia $\mu_i \ge 0$, $i = 1, \ldots, m$, yra baudos koeficientai.

Pademonstruosime disertacijoje pasiūlyto metodo sprendimams apie gauto sprendinio optimalumą priimti taikymo galimybę sprendžiant tankerio pertvarų svorio optimizavimo uždavinį, panaudojant baudos funkcijų metodą.

Projektuojant tankerį yra labai svarbu parinkti tokius pertvarų dydžius (plotį, gylį, ilgį, storį), kad šių pertvarų svoris būtų minimalus. Praleidus tam tikras techninio projektavimo detales, tankerio pertvarų minimizavimas gali būti formuluojamas kaip netiesinio programavimo uždavinys (Ravindran et al. (2006)):

$$f(x) = \frac{5{,}885 \cdot x_4 (x_1 + x_3)}{x_1 + \sqrt{(x_3^2 - x_2^2)}} \to \min,$$

kai

$$g_1(x) = x_2 x_4 \left(0{,}4 \cdot x_1 + \frac{1}{6} x_3 \right) - 8{,}94 \cdot \left(x_1 + \sqrt{(x_3^2 - x_2^2)} \right) \ge 0,$$

$$g_2(x) = x_2^2 x_4 \left(0{,}2 \cdot x_1 + \frac{1}{12} x_3 \right) - 2{,}2 \cdot \left(8{,}94 \cdot \left(x_1 + \sqrt{(x_3^2 - x_2^2)} \right) \right)^{\frac{4}{3}} \ge 0,$$

$$g_3(x) = x_4 - 0{,}0156 x_1 - 0{,}15 \ge 0,$$

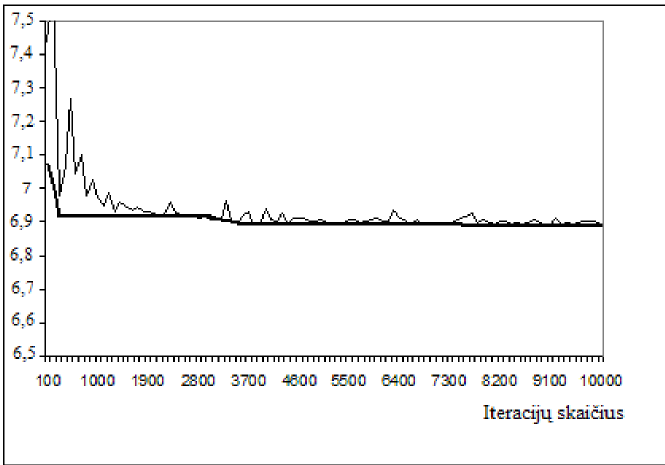
$$g_4(x) = x_4 - 0{,}0156 x_3 - 0{,}15 \ge 0,$$

$$g_5(x) = x_4 - 1{,}05 \ge 0,$$

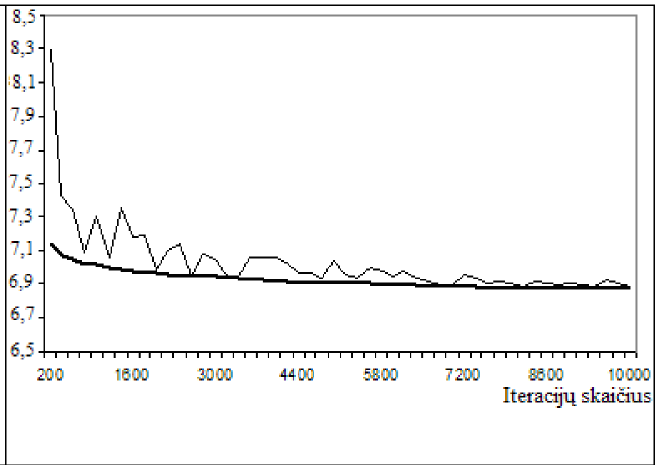
$$g_6(x) = x_3 - x_2 \ge 0,$$

čia x_1 - plotis, x_2 -gylis, x_3 - ilgis, x_4 - storis.

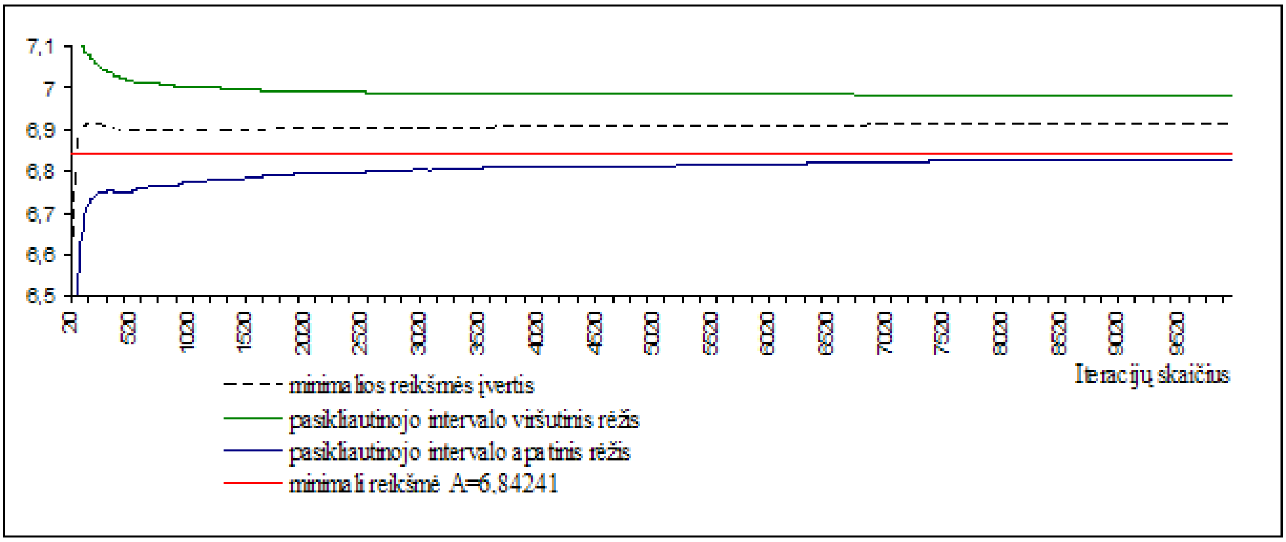
Pav. 6.6 ir *6.7* pateiktos baudos funkcijos ir geriausių pasiektų tikslo funkcijos reikšmių priklausomybės nuo iteracijų skaičiaus, taikant SPSAL ir FDSA metodus. *Pav. 6.8* pateikti tikslo funkcijos minimalios reikšmės pasikliautinojo intervalo įverčių vidurkiai. Palyginimui pateikta tikslo funkcijos minimali reikšmė A=6,84241 (Reklaitis et al. (2006)). Iš pateiktų priklausomybių matome, kad disertacijoje pasiūlytas metodas sprendimams apie gauto sprendinio optimalumą priimti, leidžia įvertinti tikslo funkcijos minimalios reikšmės pasikliautinąjį intervalą su reikiama pasikliovimo tikimybe. Gautos priklausomybės iliustruoja sukurto metodo SPSAL pritaikomumą matematinio programavimo uždaviniams su nediferencijuojamomis baudos funkcijomis spręsti.



Pav. 6.6. —baudos funkcija, —geriausios pasiektos tikslo funkcijos reikšmės (SPSAL metodas)



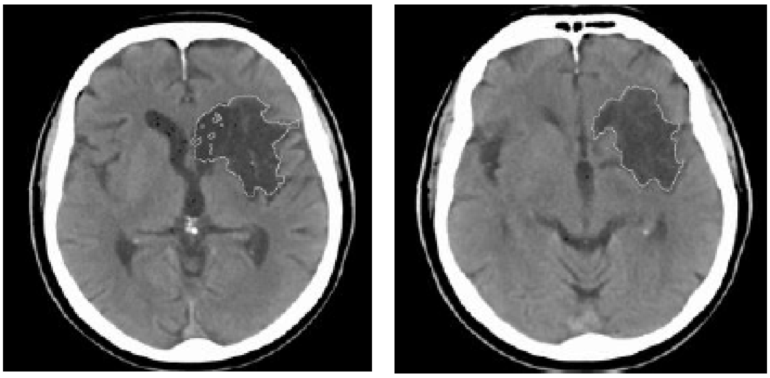
Pav. 6.7. —baudos funkcija, — geriausios pasiektos tikslo funkcijos reikšmės (FDSA metodas)



Pav.6.8*.* Tikslo funkcijos minimalios reikšmės pasikliautinasis intervalas*,*
(SPSAU, A=6,84241, M=100, N=1000, k=20, δ=0,95)

6.2.3. Išeminio insulto sričių kompiuterinės tomografijos vaizduose automatinio atpažinimo sistemos optimizavimas

Sudarant vaizdų atpažinimo algoritmus tenka spręsti algoritmų vidinių parametrų parinkimo uždavinį, įtakojantį vaizdų atpažinimo kokybę. Ši problema taip pat iškyla sprendžiant išeminio insulto sričių kompiuterinės tomografijos vaizduose automatinio atpažinimo uždavinį (Goktur et al. (2001), Grigaitis et al. (2004)). Insulto sričių atsitiktinis išsidėstymas, nevienodi sričių parametrai ir informacijos perteklius kompiuterinės tomografijos vaizduose lemia išeminio insulto sričių nustatymo automatiniu režimu sudėtingumą. Apibrėžiant insulto sritis, viena iš sudėtingiausių problemų yra galvos smegenų vingių bei kitų tamsių vaizdo sričių panaikinimas (Mikelaitis (2002), Ušinskas (2003)). Tamsios insulto sritys tomogramose dažniausiai yra gana didelės, didesnės nei tamsios galvos smegenų vingių sritys. Tuo pasinaudojant kompiuterinės tomografijos vaizduose galima nustatyti ir pašalinti galvos smegenų vingių sritis. *Pav. 6.9* pateiktos automatiniu rėžimu nustatytos galvos smegenų išeminio insulto sritys (Novelline & Squire (1987)).



Pav. 6.9. Automatiniu režimu nustatytos galvos smegenų išeminio insulto sritys

Insulto srities nustatymo kokybę galima išskirti į dvi dalis: insulto srities vietos nustatymo kokybę; insulto srities formos nustatymo kokybę. Atskirai šie kriterijai sunkiau įvertinami, todėl automatinio atpažinimo sričių sutapimas su etaloninėmis buvo vertinamas pagal formulę:

$$f(q) = \frac{S_c(q) \cap S_r(q)}{S_c(q) + S_r(q)}, \tag{6.5}$$

čia S_c – automatiškai atpažintos insulto srities plotas; S_r – eksperto suformuotas srities plotas, f – sutapimo kriterijaus reikšmė, q – apdorojamo vaizdo eilės numeris.

Insulto sritims atpažinti automatiniu režimu kompiuterinės tomografijos vaizduose yra sprendžiamas stochastinės aproksimacijos optimizavimo uždavinys. Atpažinimo algoritmas reguliuojamas penkiais įėjimo parametrais, kurie tiesiogiai įtakoja atpažinimo rezultatus. Parametru x_1 reguliuojamas triukšmo mažinimo filtras, siekiant nustatyti tokį bendrą triukšmo lygį kompiuterinės tomografijos vaizduose, kad insulto sritis būtų nustatoma geriausiai. Parametras x_2 susijęs su galvos smegenų vingių sričių atpažinimu, parametrais x_3 ir x_4 reguliuojami vaizdo skaisčio bei simetrijos lygiai atitinkamai. Parametru x_5 reguliuojamas pajėgumas atrinkti išeminio insulto sritis. Atpažinimo kokybė matuojama etaloninių ir atpažintų vaizdų sankirtos ploto ir šių vaizdų plotų sumos santykiu. Tokiu būdu vaizdo kokybė gali būti apibrėžiama glodžia iškila skaičiuojama su triukšmu funkcija $f(x)$, čia $x = (x_1, x_2, x_3, x_4, x_5)$. Kadangi parametrai yra reguliuojami tam tikru tikslumu todėl jų reikšmės buvo diskretizuojamos. Šiai tikslo funkcijai optimizuoti ir parametrams x_1, x_2, x_3, x_4, x_5 parinkti naudojamas darbe sukurtas SPSA algoritmas, kuriame įvestas Metropolio patvirtinimo kriterijus (Spall (1992), Yang (2000), Bartkutė et al. (2006), Grigaitis et al. (2008)). Sukurto algoritmo savybės buvo įvertintos kompiuterinio modeliavimo būdu, taikant Monte-Karlo metodą.

Algoritmu 6.1 aprašomas stochastinės aproksimacijos optimizavimo uždavinio insulto sritims kompiuterinės tomografijos vaizduose atpažinti sprendimas.

Algoritmas 6.1: **Stochastinės aproksimacijos insulto sričių kompiuterinės tomografijos vaizduose automatinio atpažinimo algoritmas**

Tikslas: sudaryti tikslo funkcijos stochastinės aproksimacijos seką.

Pradinės sąlygos (*preconditions*)**:** tikslo funkcijos apskaičiavimo algoritmas, stochastinio gradiento įvertinimo algoritmas, formalizuotas stabdymo taisyklės aprašymas, optimizavimo žingsnio ilgio ρ bei suglodinimo parametro σ apskaičiavimo funkcijos, pradinis taškas x^0.

Galinės sąlygos (*postconditions*)**:** optimizavimo taškų seka, tikslo funkcijos reikšmių tuose taškuose seka.

1. Inicializuoti pradinį tašką x_0, optimizavimo žingsnio ilgį ρ_0, suglodinimo parametrą σ_0, $t = 0$.
2. **While** neišpildyta stabdymo taisyklė (pvz.: $t = t_{\max}$) **do**
3. Generuoti vektorių $\Delta^t = \left(\Delta_1^t, \Delta_2^t, ..., \Delta_n^t\right)$, kurio komponentės yra atsitiktiniai dydžiai, su tikimybe $p = 0{,}5$ įgyjantys reikšmes 1 arba -1.
4. Skaičiuoti vektorius $y^{1,t} = \left(y_1^{1,t}, y_2^{1,t}, ..., y_n^{1,t}\right)$ ir $y^{2,t} = \left(y_1^{2,t}, y_2^{2,t}, ..., y_n^{2,t}\right)$:

$$y_i^{1,t} = \min\left(\max\left(x_i^t + \Delta_i^t \cdot C_t, x_{i\,\min}\right), x_{i\,\max}\right),$$

$$y_i^{2,t} = \max\left(\min\left(x_i^t - \Delta_i^t \cdot C_t, x_{i\,\max}\right), x_{i\,\min}\right),$$

čia $C_t = \max(1, [\,\sigma_t + 0{,}5\,])$, $[\cdot]$ žymi skaičiaus sveikąją dalį.

{*Komentaras:* vektorių $y_i^{1,t}$, $y_i^{2,t}$ reikšmės gali būti sveikieji skaičiai}

5. Skaičiuoti tikslo funkcijos reikšmes: $z^{1,t} = f\left(y^{1,t}\right)$, $z^{2,t} = f\left(y^{2,t}\right)$.
6. Apskaičiuoti tašką $v^t = \left(v_1^t, v_2^t, ..., v_n^t\right)$:

$$v_i^t = x_i^t + sign\left(\cdot \frac{z^{1,t} - z^{2,t}}{y_i^{1,t} - y_i^{2,t}} \right) \cdot W_t, \quad i = 1, 2, ..., n,$$

čia $W_t = \max\left(1, \left[\rho_t \cdot \left| \frac{z^{1,t} - z^{2,t}}{y_i^{1,t} - y_i^{2,t}} \right| \right]\right), \quad i = 1, 2, ..., n$, $[\cdot]$ žymi skaičiaus sveikąją dalį.

6. Skaičiuoti vektorius $y^{1,t} = \left(y_1^{1,t}, y_2^{1,t}, ..., y_n^{1,t}\right)$ ir $y^{2,t} = \left(y_1^{2,t}, y_2^{2,t}, ..., y_n^{2,t}\right)$:

$$y_i^{1,t} = \min\left(\max\left(v_i^t + \Delta_i^t \cdot C_t, x_{i\,\min}\right), x_{i\,\max}\right),$$

$$y_i^{2,t} = \max\left(\min\left(v_i^t - \Delta_i^t \cdot C_t, x_{i\,\max}\right), x_{i\,\min}\right),$$

čia $C_t = \max(1, [\sigma_t + 0{,}5])$, $[\cdot]$ žymi skaičiaus sveikąją dalį.

7. Skaičiuoti funkcijos reikšmes: $z^{1,t} = f\left(y^{1,t}\right)$, $z^{2,t} = f\left(y^{2,t}\right)$. Taip pat apskaičiuojame $z^t = \max\left(z^{1,t}, z^{2,t}\right)$.

Taikyti Metropolio kriterijų taškui x^{t+1} patvirtinti.

8. Generuoti atsitiktinį skaičių κ, tolygiai pasiskirsčiusį intervale [0,1], ir skaičiuoti naują tašką $x^{t+1} = \left(x_1^{t+1}, x_2^{t+1}, ..., x_n^{t+1}\right)$:

$$x_i^{t+1} = \begin{cases} v_i^t, \ \ jei \ \ \kappa \le e^{\frac{z^t - z^{t-1}}{T}}, \\ x_i^t, \ \ kitu \ \ atveju. \end{cases} \quad i = 1, 2, ..., n$$

čia $T = 1$, $t = t + 1$.

9. **done**. ◊

{*Komentaras:* x_i^{t+1} reikšmės gali būti sveikieji skaičiai}

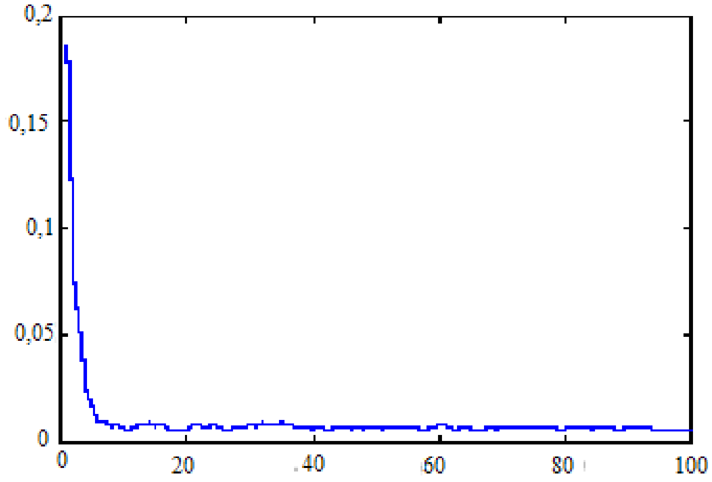
Uždavinio sprendiniu gali būti imamos geriausios pasiektos $z^{1,t}$, $z^{2,t}$ reikšmės. Šiam algoritmui stabdyti taip pat galime taikyti stabdymo taisykles, sudarytas taikant pozicines statistikas (Жиглявский & Жилинскас (1991), Bartkutė & Sakalauskas (2004a)).

Empiriniai tyrimai rodo, kad nagrinėjama tikslo funkcija (6.5) gali įgyti minimumą ant leistinosios srities krašto, čia leistinoji sritis dažniausiai yra nusakoma stačiakampiais ribojimais. Todėl pasiūlytas metodas buvo testuojamas minimizuojant testinę funkciją leistinoje srityje su stačiakampiais ribojimais $0 \le x_i \le 1$, $i = 1, ..., n$:

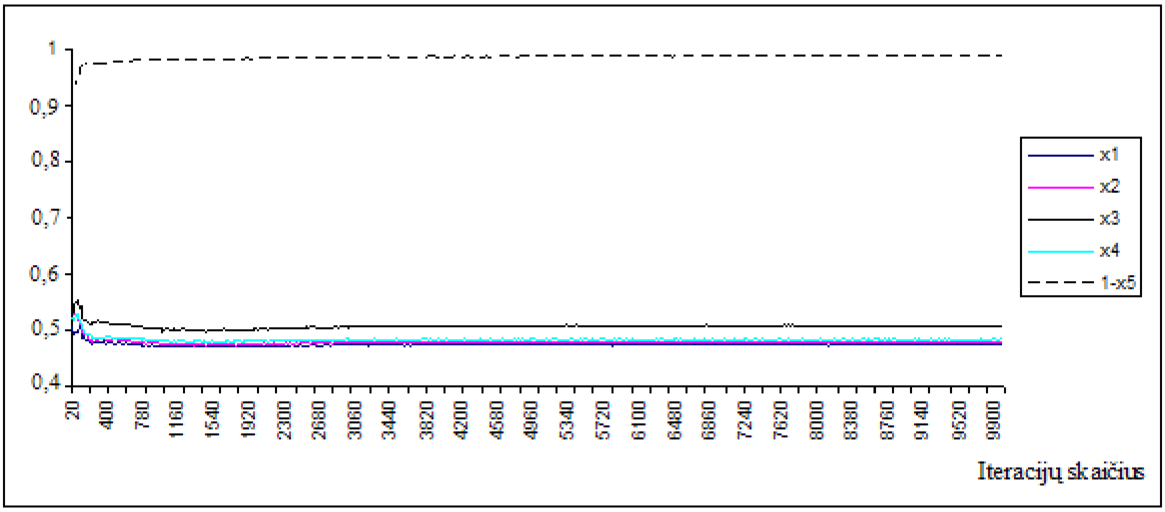
$$f(x) = \sum_{j=1}^{n} a_j \cdot \left(\left(x_i - x_i^*\right)^2 + 0.1 \cdot \left(1 - \cos\left(18 \cdot \left(x_i - x_i^*\right)\right)\right) \right),$$

čia a_j atsitiktiniai realūs skaičiai, tolygiai pasiskirstę intervale [μ, K], $K > \mu > 0$, $n = 5$.

Testinė funkcija buvo minimizuojama M=500 kartų, $\mu = 2$, K=5, $x_1^* = x_2^* = x_3^* = x_4^* = 0{,}5$, $x_5^* = 1$. Optimizavimo sekos (2.14) koeficientai buvo parenkami tokiu būdu: $\rho_t = \min\left(0{,}1; \frac{1}{t}\right)$, $\sigma_t = \min\left(0{,}1; \frac{1}{t^{\varphi}}\right)$, $\varphi = 0{,}25$ (Grigaitis et. al (2008)).



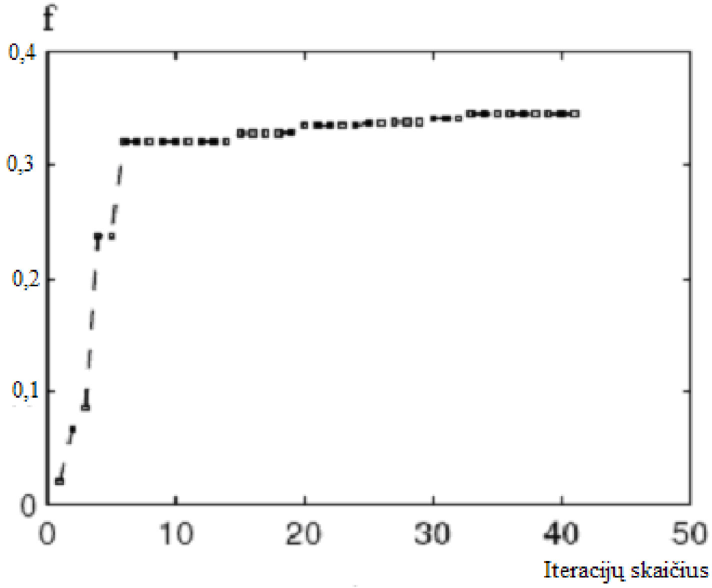
Pav.6.10. Tikslo funkcijos vidurkis priklausomai nuo iteracijų skaičiaus



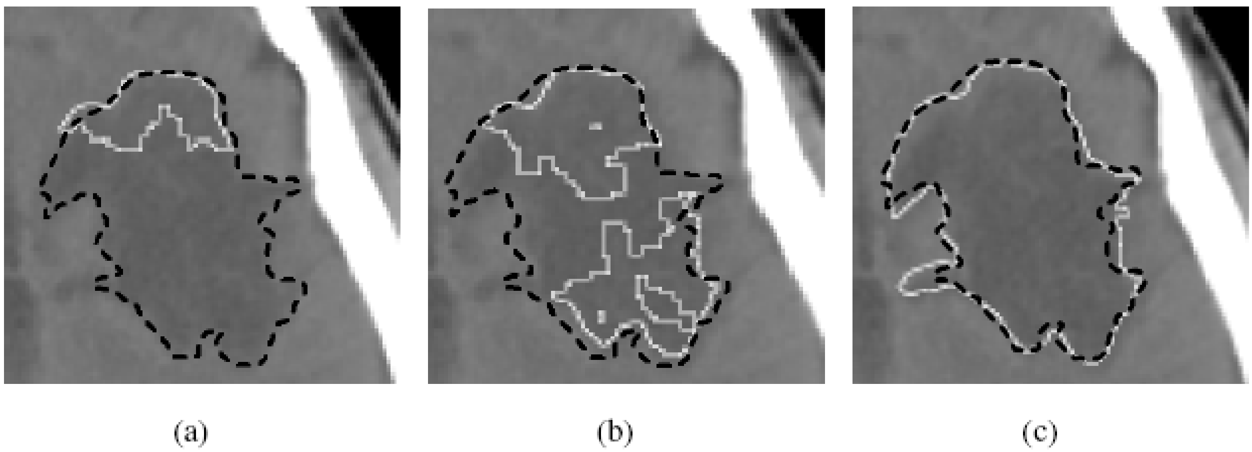
Pav.6.11. Algoritmo parametrų vidurkių reikšmės priklausomai nuo iteracijų skaičiaus

Pav. 6.10 pateiktas tikslo funkcijos vidurkis priklausomai nuo iteracijų skaičiaus. *Pav. 6.11* pateiktos algoritmo parametrų x_1^t, x_2^t, x_3^t, x_4^t, x_5^t vidurkių priklausomybės nuo iteracijų skaičiaus, iliustruojančios algoritmu gautos sekos x^t konvergavimą į optimalų sprendinį x^*.

Įsitikinus testavimo metu sukurto algoritmo tinkamumu spręsti optimizavimo uždavinius su stačiakampiais ribojimais, algoritmas buvo pritaikytas insulto srities atpažinimo uždaviniams spręsti, kai tikslo funkcija aprašoma (6.5) būdu. *Pav. 6.12 ir 6.13* pateikti insulto tomografinių vaizdų atpažinimo, taikant sukurtą SPSA metodą, rezultatai. *Pav. 6.12* pateikta geriausios pasiektos tikslo funkcijos reikšmės priklausomybė nuo iteracijų skaičiaus parodo, kad sukurtas algoritmas greitai randa optimumo zoną, kurioje pradeda konverguoti. *Pav. 6.13* parodytas insulto srities atpažinimo pavyzdys. pateiktos išeminio insulto sritys atpažintos optimizavimo metu. Iš šiame paveiksle pateiktų rezultatų matome, kad jau po 8 iteracijų algoritmo atpažinta insulto sritis beveik visiškai sutampa su etaloniniu vaizdu. Paveiksluose pateikta balta kreive apvesta sritis atitinka automatiniu režimu atpažintą išeminio insulto sritį, juoda kreive apvesta sritis atitinka eksperto nustatytą išeminio insulto sritį.



Pav.6.12. Geriausios pasiektos tikslo funkcijos reikšmės priklausomybė nuo iteracijų skaičiaus



Pav 6.13. Išeminio insulto sritys atpažintos optimizavimo metu: insulto sritis optimizavimo proceso pradžioje, kai f=0,096 (a); insulto sritis po 5 iteracijų, kai f=0,24 (b); insulto sritis po 8 iteracijų, kai f=0,35 (c)

6.3. REZULTATAI IR IŠVADOS

1. Lygiagrečiųjų skaičiavimų kompiuterinio modeliavimo rezultatai leido:
   - patvirtinti teorines stochastinės aproksimacijos algoritmų konvergavimo sąlygas, kai optimizavimo uždavinių dimensija iki 100;
   - patvirtinti nepriklausomų vienodai pasiskirsčiusių atsitiktinių dydžių ekstremaliųjų reikšmių teorijos tinkamumą tikslo funkcijos minimaliai reikšmei įvertinti;

2. MPICC kalba sukurti programiniai kodai gali būt realizuojami su bet kokia lygiagrečiųjų skaičiavimų architektūra;

3. Sukurto SPSAL algoritmo taikymo rezultatai Heston'o stochastinio volatiliškumo modeliui kalibruoti parodė, kad šis algoritmas gali būti taikomas finansinių duomenų rinkos ir rizikos analizei;

4. Sukurto SPSAL algoritmo taikymo rezultatai tankerio pertvarų svoriui optimizuoti parodė, kad sukurtas algoritmas gali būti taikomas optimizavimo uždaviniams su ribojimais spręsti, taikant nediferencijuojamų baudos funkcijų metodą;

5. Sukurto SPSAL algoritmo taikymo rezultatai išeminio insulto sričių kompiuterinės tomografijos vaizdų automatinio atpažinimo sistemoms optimizuoti parodė, kad šis metodas gali būti taikomas diskretaus optimizavimo uždaviniams su stačiakampiais ribojimais spręsti, kai tikslo funkcija yra apskaičiuojama su paklaida.

7. REZULTATAI IR IŠVADOS

Markovo tipo stochastiniai algoritmai yra gerai žinomi ir plačiai taikomi determinuotojo ir stochastinio programavimo uždaviniams spręsti. Tačiau aktuali teoriniu ir praktiniu požiūriu šių algoritmų stabdymo problema beveik nebuvo tiriama. Markovo tipo algoritmų stabdymo taisyklės sudaromos konverguojančioms sekoms, t.y. jų sudarymas bei optimalumo tyrimas remiasi algoritmo konvergavimo tyrimo rezultatais. Optimaliems sprendimams priimti optimizavimo algoritmuose gali būti naudojamos geriausios pasiektos tikslo funkcijos reikšmės, kurių teikiama informacija yra pritaikoma minimaliai tikslo funkcijos reikšmei bei jos pasikliautinajam intervalui įvertinti.

Geriausių tikslo funkcijos reikšmių taikymas optimaliems sprendimams priimti remiasi prielaida, kad, pozicinės statistikos, gautos optimizavimo metu, yra pasiskirsčiusios pagal ekstremaliųjų reikšmių skirstinį. Nors šis skirstinys yra labiau išnagrinėtas nepriklausomų vienodai pasiskirsčiusių dydžių atveju, juo paremti modeliai gali būti taikomi ir tais atvejais, kai nebūtinai yra išpildytos nepriklausomumo bei homogeniškumo prielaidos. Juolab, galima įrodyti, kad funkcijos reikšmių seka, gauta daugelyje stochastinės paieškos algoritmų, asimptotiškai konverguoja į nepriklausomų Gauso dydžių su skirtingomis dipersijomis seką. Tuo būdu teorinių išvadų, prieštaraujančių ekstremaliųjų reikšmių skirstinio taikymui optimaliems sprendimams priimti, nėra žinoma. Kadangi šių reikšmių skirstiniai, reikalingi tokiems įverčiams gauti, teoriškai dar nėra ištirti, darbe atlikti kompiuterinio modeliavimo eksperimentai leido ištirti ekstremaliųjų reikšmių teorijoje žinomų įverčių tinkamumą optimizavimo algoritmų stabdymo taisyklei sudaryti bei optimaliems sprendimams priimti. Sukurta metodika ištirta pastaruoju metu intensyviai tiriamų stochastinio optimizavimo metodų pavyzdžiu.

Sprendžiant darbe suformuluotus uždavinius, gauti šie rezultatai:

5. Lipšico funkcijų klasėje įvestas suglodinimo operatorius su Lipšico tankio funkcija, leidžiantis gauti du kartus diferencijuojamas suglodintas funkcijas bei sukurtas SPSA algoritmas Lipšico funkcijoms, suglodintoms su šiuo operatoriumi, optimizuoti;
6. Sukurtas metodas ir sudarytas algoritmas tikslo funkcijos minimaliai reikšmei bei jos pasikliautinajam intervalui įvertinti, naudojant optimizavimo sekų, gautų Markovo tipo stochastiniais algoritmais, pozicines statistikas. Šis algoritmas pritaikytas tolydaus ir diskretaus optimizavimo uždaviniams spręsti stochastinės aproksimacijos bei modeliuojamojo atkaitinimo metodais;

7. Sudarytas algoritmas Veibulo skirstinio parametrams (formos, padėties ir skalės) įvertinti maksimalaus tikėtinumo metodu, naudojant necenzūruotos imties pozicines statistikas, suformuluotos ir įrodytos maksimalaus tikėtinumo įverčio egzistavimo sąlygos;
8. Sukurtas analitinis metodas, apibrėžtas bet kokiam pozicinių statistikų rinkiniui, ir sudarytas jį realizuojantis algoritmas Veibulo skirstinio parametrams (formos, padėties ir skalės) įvertinti, taikant tikslo funkcijos minimalios reikšmės tiesinį įvertį ir naudojant necenzūruotos imties pozicines statistikas, suformuluotos ir įrodytos analitinio įverčio egzistavimo sąlygos bei įrodytas analitinio algoritmo konvergavimas į Veibulo skirstinio formos parametro įvertį tiesiniu greičiu;
9. Sudaryti algoritmai bei programinė įranga nediferencijuojamo optimizavimo uždaviniams spręsti SPSAL, SPSAU, FDSA, SA metodais bei sukurta testinių uždavinių biblioteka stochastinės paieškos algoritmams testuoti;
10. Sudaryti algoritmai bei programinė įranga programinių sistemų MathCad, C++, PASCAL, MPICC priemonėmis nuoseklių ir lygiagrečių skaičiavimų kompiuteriams, stochastinių Markovo tipo optimizavimo algoritmų konvergavimui bei optimalių sprendimų priėmimo taisyklėms tirti Monte - Karlo metodu;
11. Sudaryti algoritmai bei programinė įranga stochastinių Markovo tipo optimizavimo algoritmų stabdymo taisyklei realizuoti;
12. Sudaryti algoritmai bei programinė įranga Veibulo skirstinio trims parametrams įvertinti;
13. Sudaryta programinė įranga ir algoritmai buvo pritaikyti taikomiesiems uždaviniams spręsti:
    - Heston'o stochastinio volatiliškumo modeliui kalibruoti;
    - tankerio pertvarų svoriui optimizuoti;
    - išeminio insulto sričių kompiuterinės tomografijos vaizduose automatinio atpažinimo sistemoms optimizuoti.

Gauti rezultatai ir atlikti tyrimai leidžia daryti tokias išvadas.

1. Sudarytas SPSAL algoritmas konverguoja Lipšico funkcijų klasėje, esant gana bendroms sąlygoms, nustatytoms teoremoje 3.1. Šis algoritmas konverguoja greičiu $O\left(\frac{1}{k^{\gamma}}\right)$, $1<\gamma<2$, pusiauglodžių funkcijų klasėje ir konvergavimo greičio eilė nepriklauso nuo uždavinio dimensijos (*teorema 3.2*). Nustatytos konstruktyvios konvergavimo greičio konstantos, išreiškiamos per tikslo funkcijos skaitmenines charakteristikas bei metodo parametrus. Atitinkamai derinant šiuos parametrus, galima gauti konverguojančias sekas;

2. Skaičiuojamojo eksperimento Monte-Karlo metodu rezultatai su įvairiomis testinėmis funkcijomis parodė, kad eksperimento metu nustatyta sukurto SPSAL algoritmo konvergavimo greičio eilė, atitinka darbe gautas teorines išvadas ($O\left(\frac{1}{k^{\gamma}}\right)$, $1<\gamma<2$, kintamųjų skaičius $n \leq 100$) ;
3. Sudaryti metodai gali būti taikomi sprendimams apie minimalią tikslo funkcijos reikšmę bei jos pasikliautinąjį intervalą priimti, pasinaudojant pozicinėmis statistikomis, gautomis optimizavimo Markovo tipo stochastiniais algoritmais metu, su iš anksto nustatyta pasikliovimo tikimybe ($\delta = 0{,}9,....,0{,}99$), kai kintamųjų skaičius $n \leq 100$;
4. Minimalios reikšmės ir jos pasikliautinojo intervalo tiesinių įverčių koeficientai gali būti skaičiuojami, pritaikius ekstremaliųjų reikšmių teorijos rezultatus nepriklausomiems vienodai pasiskirsčiusiusiems atsitiktiniams dydžiams, kai ekstremaliųjų reikšmių skirstinio formos parametras nustatomas pagal tikslo funkcijos homogeniškumo rodikį minimumo aplinkoje arba įvertinamas statistiniais metodais;
5. Kompiuterinio modeliavimo rezultatai Monte-Karlo metodu parodė sukurto analitinio metodo pranašumą lyginant jį su žinomu literatūroje analitiniu Veibulo skirstinio formos parametro įverčiu bei patvirtino gautų įverčių pagrįstumą ir tinkamumą Veibulo skirstinio parametrams vertinti;
6. Tikslo funkcijos minimalią reikšmę galima įvertinti pasinaudojus tiesiniais įverčiais, kai ekstremaliųjų reikšmių parametras vertinamas pagerintu analitiniu arba maksimalaus tikėtinumo metodais, o pozicinių statistikų skaičius ir algoritmo iteracijų skaičius yra pakankamai dideli;
7. Sudaryta optimizavimo algoritmo stabdymo taisyklė, naudojant optimizavimo metu pasiektas geriausias tikslo funkcijos reikšmes - algoritmą stabdyti kai pasikliautinojo intervalo plotis tampa mažesnis už iš anksto užsiduotą reikšmę $\varepsilon > 0$;
8. Sukurtas SPSAL algoritmas bei pasiūlyta stabdymo taisyklė pasižymi paprastumu bei universalumu ir gali būti plačiai taikomi įvairiems tolydaus arba diskretaus programavimo uždaviniams spręsti.

LITERATŪRA

1. Aiex R.M., Resende M.G.C., Ribeiro C.C. (2002). Probability distribution of solution time in GRASP: An experimental investigation. // *Journal of Heuristics*, vol. 8, p. 343–373.
2. Anderson R.L. (1953). Recent advances in finding best operating conditions. *// Journal of the American Statistical Association*, vol. 48(264), p. 789 - 798.
3. Bagirov A. M., Ugon J. (2006). Piecewise Partially Separable Functions and a Derivative-free Algorithm for Large Scale Nonsmooth Optimization. // *Journal of Global Optimization*, vol. 35, p. 163-195.
4. Bartkutė V., Sakalauskas L. (2004a). Order statistics for testing optimality in stochastic optimization*. // Proceedings of the 7th International Conference "Computer data analysis and Modelling"*, Minsk, p. 128-131.
5. Barkutė V., Sakalauskas L. (2004b) Vienalaikio trikdžio stochastinės aproksimacijos konvergavimas Lipšico funkcijų klasėje. // *Liet. matemat. rink.* T. 44, spec. nr., p. 603-608.
6. Bartkutė V., Sakalauskas L. (2005a). Pozicinių statistikų taikymas tiriant modeliuojamojo pokyčio stochastinės aproksimacijos algoritmo optimalumą. // *Informacijos mokslai*. T. 34, p. 136-141.
7. Bartkutė V., Sakalauskas L. (2005b). Pozicinių statistikų taikymas optimalumo tyrimui. // *Matematika ir matematinis modeliavimas*. T. 1, p. 15–19.
8. Bartkutė V., Sakalauskas L. (2005c). Modeliuojamojo pokyčio stochastinės aproksimacijos algoritmo konvergavimo greičio tyrimas. // *Konferencijos „Informacinės tecnnologijos 2005" pranešimų medžiaga,* Kaunas, p. 415-425.
9. Bartkutė V., Sakalauskas L. (2006a). Application of Stochastic Approximation in Technical Design. // *Series on Computers and Operations Research*, vol. 7. Computer Aided Methods in Optimal Design and Operations, p. 29-38.
10. Bartkutė V., Sakalauskas L., Felinskas G. (2006). Optimality testing in stochastic and heuristic algorithms. // *Technological and economic development of economy*, vol. 12, Nr. 1, p. 4-10.
11. Bartkutė V., Sakalauskas L. (2006b). Pozicinių statistikų taikymas stochastinės aproksimacijos algoritmų optimalumui tirti. // *Jaunųjų mokslininkų darbai,* Nr. 4 (11), p. 202-210.
12. Bartkutė V. (2006c). Application of Order Statistics to Terminate the Heston Model Calibration Procedure. *// Proceedings of the 26th Conference on Applied Statistics in Ireland,* May 17-19, Killarney, Ireland, p. 69-70.

13. Bartkutė V., Sakalauskas L. (2007a). Simultaneous perturbation stochastic approximation for nonsmooth functions. // *European Journal on Operational research,* vol. 181, No. 3, p. 1174-1188.

14. Bartkutė V., Sakalauskas L. (2007b). Three parameter estimation of the Weibull distribution by order statistics. Recent Advances on Stochastic Modeling and Data Analysis, p. 91-101.

15. Bartolucci A. A., Singh K. P., Bartolucci A. D., Bae S. (1999). Applying medical survival data to estimate the three-parameter Weibull distribution by the method of probability-weighted moments. // *Mathematics and Computers in Simulation*, vol. 48, No. 4-6, p. 385 – 392.

16. Bazaraa M.S., Sherali H.D., Shetty C.M. (1992). Nonlinear Programming: Theory and Algorithms. Wiley.

17. Bhatnagar S., Kowshik H.J. (2005). A discrete parameter stochastic approximation algorithm for simulation optimization. // *Simulation*, vol. 81, No. 11, p. 757-772.

18. Blischke W. R. (1974). On nonregular estimation. II. Estimation of the Location Parameter of the Gamma and Weibull Distributions. // *Communications in Statistics*, vol. 3, p. 1109-1129.

19. Большев Л. Н., Смирнов Н.В. (1983). Таблицы математической статистики. Москва: Наука.

20. Brooks S.H. (1958). A discussion of random methods for seeking maxima. // *Operations research*, vol. 6(2), p. 244–251.

21. Chen H. (1996). Estimation of the Location of the Maximum of a Regression Function Using Extreme Order Statistics. // *Journal of multivariate analysis,* vol. 57, p. 191 – 214.

22. Chen H. F., Duncan T. E., Pasik-Duncan B. (1999). A Kiefer-Wolfowitz algorithm with randomized differences. // *IEEE Trans. Automat. Contr.*, vol. 44, p. 442–453.

23. Chin D.C. (1997). Comparative Study of Stochastic Algorithms for System Optimization Based on Gradient Approximations. // *IEEE Trans. Systems, Man, and Cybernetics- Part B: Cybernetics,* vol. 27, p. 244-249.

24. Chung K. L. (1954). On a Stochastic Approximation Method. *// Annals of Mathematical Statistics*, vol. 25, p. 463-483

25. Clarke F. H. (1975). Generalized gradients and applications. // *Trans. Amer. Math. Soc.,* vol. 205, p. 247 – 262

26. Clarke F. H. (1983). Optimization and Nonsmooth Analysis. John Wiley, New York.

27. Collins N.E., Eglese R.W., Golden B.L. (1988). Simulated annealing-An annotated bibliography. // *American Journal of Mathematical and Management Sciences,* vol. 8, p. 209-307.

28. Čiegis R. (2001). Lygiagretieji algoritmai. Vilnius: Technika.

29. Dantzig G. B. (1963). Linear Programming and Extensions. Princeton: Princeton U P.

30. David H. A., Nagaraja H. N. (2003). Order statistics. John Wiley & Sons Inc. N.Y.

31. Dennis J. E., Shnabel R. B. (1996). Numerical Methods for Unconstrained Optimization and Nonlinear Equations. SIAM, Philadelphia.

32. Dičiūnas V., Skersys G. (2003). Diskrečioji matematika. Mokymo priemonė. VU.

33. Dieudonné J. (1960). Foundations of modern analysis. Academic Press, N.Y., London.

34. Dorea C.C.Y. (1997). On the efficiency of a continuous version of the simulated annealing algorithm. // *Statistics & probability letters*, vol. 31, p. 247-253.

35. Dosinas G., Janušauskaitė N., Navickas Z. (1999). Taikomoji matematika. Technologija-Kaunas.

36. Dress H., Ferreira A., Haan de L. (2004). On maximum likelihood estimation of the extreme value index. // *Ann. Appl. Probab*., vol. 14, No. 3, p. 1179-1201.

37. Dupac V. (1988). Stochastic approximation. In Handbook of Statistics. Nonparametric methods. Eds. Krishnaiah P.R., Sen P.K. Nord Holand, N.Y.

38. Dvoretzky A. (1956). On stochastic approximation. // *Proc. 3rd Berkeley Symposium of Mathematical Statistics and Probability*, J.Neumann(eds.), University of California Press, Berkeley, vol. 1, p. 39-55.

39. Ермольев Ю. Н. (1976). Методы стохастического программирования. Москва: Наука.

40. Ermoliev Yu. M., Norkin V. I., Wets R. J-B. (1995). The minimization of semicontinuous functions: mollifier subgradients. *// Control and optimization,* vol. 33, No 1, p. 149-167.

41. Fu M. C., Hill S. D. (1997). Optimization of discrete event systems via simultaneous perturbation stochastic approximation. // *IIE Transactions,* vol. 29, No. 3, p. 233-243.

42. Галамбош Я. (1984). Асимптотичсекая теория экстремальных порядковых статистик. Москва: Наука.

43. Garey M. R., Johnson D. S. (1979). Computers and Intractability: A Guide to the Theory of NP-Completeness, W. H. Freeman & Co. New York, NY.

44. Gerencsér L. (1999). Rate of convergence of moments for a simultaneous perturbation stochastic approximation method for function minimization. // *IEEE Trans. Auto. Cont*., vol. 44, p. 894-906.

45. Gidas B. (1985). Non stationary Markov chains and convergence of Simulated Annealing algorithms. // *J. Statist. Phys*., vol. 39, p. 73-131.

46. Goktur S. B., Tomasi C., Girod B., Beuleur C. (2001). Medical Image compression based on region of interest, with application to colon CT images. // *IEEE Proceedings of the 23rd Annual International Conference, Engineering in Medicine and Biology Society*, vol. 3, p. 2453-2456.

47. Граничин О. Н., Поляк Б. Т. (2003). Рандомизированные алгоритмы оценивания и оптимизации при почти произвольных помехах. Москва: Наука.

48. Granville V., Krivanek M., Rasson J. P. (1994). Simulated Annealing: A Proof of Convergence. // *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 16, No. 6, p. 652-656.

49. Grigaitis D., Kirvaitis R., Dobrovolskis R. A. (2004). Galvos smegenų vingių bei pavienių tamsesnių sričių segmentavimas kompiuterinės tomografijos vaizduose. // *Elektronika ir elektrotechnika*. Nr. 6(55), p. 29 - 33.

50. Grigaitis D., Bartkutė V., Sakalauskas L. (2008). An Optimization of System for Automatic Recognition of Ischemic Stroke Areas in Computed Tomography Images. // *Informatica,* No.1 (priimtas)

51. Gupal A. M., Norkin V. I. (1977). An algorithm for minimization of discontinuous functions. // *Kibernetika,* p. 73-75.

52. Haan L. (1981). Estimation of the Minimum of a Function Using Order Statistics. // *Journal of the American Statistical Association*, vol. 76, No. 374, p. 467-469.

53. Hall P. (1982). On estimating the endpoint of a distribution. // *Annals of Statistic,* vol. 10, p. 556-568.

54. Heston S. L. (1993). A closed-form solution for options with stochastic volatility with applications to bond and currency options. // *The Review of Financial Studies,* vol. 6, issue 2, p. 327-343.

55. Heuckroth M. W, Gady J. L, Gaines L. D. (1976). An examination of the adaptive random search technique. // *AIChE Journal,* vol. 22(4), p. 744-50.

56. Hirose H. (1991). Percentile point estimation in the three parameter Weibull distribution by the extended maximum likelihood estimate. // *Computational Statistics & Data Analysis*, vol. 11, p. 309-331.

57. Horst R., Pardalos P.M., Thoai Van N. (2000). Introduction to Global Optimization (Nonconvex Optimization and its Applications), vol. 48, Kluwer Academic Publishers.

58. Hutchison D. W., Spall J.C. (2005). A Method for Stopping Nonconvergent Stochstic Approximation Processes. // *Proceedings of the Joint IEEE Conference on Decision and Control and European Control Conference*, 12−15 December 2005, Seville, Spain, p. 6620−6625.

59. Johnson N. L, Kotz S. (1970). Distributions in statistics. Continuous Univariate Distributions -2. New York: Wiley.

60. Yang R. L. (2000). Convergence of the Simulated Annealing Algorithm for Continuous Global Optimization. // *JOTA,* vol. 104, No. 3, p. 691-716.

61. Karnopp D. C. (1963). Random search techniques for optimization problems. // *Automatica*, vol. 1(1), p. 111–121.

62. Kiefer J., Wolfowitz, J. (1952). A stochastic estimation of the maximum of a regression function. // *Annals of Mathematical Statistics*, vol. 23(3), p. 462-466.

63. Kirkpatrick S., Gelatt C.D., Vecchi M.P. (1983). Optimization by Simulated Annealing. // *Science*, vol. 220, No. 4598, p. 671-680.

64. Kleene C. (1952). Introduction to Metamathematics. Tenth Edition 1991, North-Holland Publishing Company.

65. Knuth D. E. (1973). The Art of Computer Programming. Second Edition, vol. 1. Fundamental Algorithms, Addison-Wesley.

66. Kouvelis P. G. Yu. (1997). Robust Discrete Optimization and Its Applications. Kluwer Academic Publishers, Amsterdam, The Netherlands

67. Korn G. A., Korn T. M. (2003). Mathematical Handbook for Scientists and Engineers: Definitions, Theorems, and Formulas for Reference and Review. Dover publications, INC, Mineola, New York.

68. Kuhn H. W., Tucker A. W. (1951). Nonlinear Programming. Proceedings of the Second Berkeley Symposium on Mathematical Statistics and Probability, Ed. Jerzy Neyman, University of California Press, Berkeley, California, p. 481 -492.

69. Kushner H.J., Yin G. G. (2003). Stochastic Approximation and Recursive Algorithms and Applications. *Springer*: N.Y., Heidelberg, Berlin.

70. Kuratowski K. (1968). Topology. Vol. II. N.Y.: Academic Pres.

71. Lemon G. H. (1975). Maximum Likelihood Estimation for the Three Parameter Weibull Distribution based on censored samples. // *Technometrics,* vol. 17, No. 2, p. 247-254.

72. Locatelly M. (2000a). Convergence of Simulated Annealing algorithm for continuous global optimization. // *Journal of global optimization*, vol. 18, p. 219-234.

73. Locatelly M. (2000b). Simulated Annealing algorithms for continuous global optimization: convergence conditions. // *Journal of optimization theory and applications*, vol. 104, No.1, p. 121-133.

74. Luksan L., Vlcek J. (1999). Sparse and Partially Separable Test Problems for Unconstrained and Equality Constrained Optimization, (http://www.cs.cas.cz/~luksan/test.html).

75. Metropolis N., Rosenbluth A., Rosenbluth M., Teller A., Teller E. (1953). Equation of State Calculations by Fast Computing Machines. // *J. Chem. Phys.* vol. 21, No. 6, p. 1087-1092.

76. Mifflin R. (1977). Semismooth and semiconvex functions in constrained optimization. // *SIAM J. Control and Optim.,* vol. 15, p. 957-972.

77. Михалевич В. С., Гупал А. М., Норкин В. И. (1987). Методы невыпуклой оптимизации. Москва: Наука.

78. Mikelaitis V. (2002). Medicininių atvaizdų apdorojimas –nereikalingų vaizdo dalių pašalinimas // *Penktoji Lietuvos jaunųjų mokslininkų konferencija „Lietuva be mokslo –Lietuva be ateities*“. Vilnius: Technika, p. 13–17.

79. Моцкус Й. Б. (1976). Многоэкстремальные задачи в проектирование. Москва: Наука.

80. Murthy D. N. P, Xie M., Hang R. (2004). Weibull Models. Wiley Series in Probability and Statistics. John Wiley & Sons, Inc. New Jersey, USA.

81. Nelder J. A., Mead R. (1965). A Simplex Method for Function Minimization. // *Comput. J.*, vol. 7, p. 308-313.

82. Nesterov Y. (2004). Inroductory lectures on convex optimization. Kluwer Academic Publishers.

83. Novelline R. A., Squire L. F. (1987). Living anatomy. A working atlas using computed tomography magnetic resonance and angiography images. Hanley & Belfus, Philadelphia, PA 19107, p. 117.

84. Nurminski E. A. (1979). Numerical methods for solving deterministic and stochastic minimax problems. Naukova Dumka: Kiev (rusų kalba).

85. Пизо Ш. Заманский М. (1971). Курс математики. Алгебра и анализ. Москва: Наука.

86. Poliak B.T. (1987). Introduction to Optimization. Translations Series in Mathematics and Engineering. Optimization Software, Inc., Publications Division, New York.

87. Растригин Л.А. (1968). Статистические методы поиска. Москва: Наука.

88. Rardin R. L. (1998). Optimization in Operations Research. Prentice Hall, New Jersey.

89. Ravindran A., Ragsdell K. M., Reklaitis G. V. (2006). Engineering Optimization: Methods and Applications. John Wiley & Sons.

90. Reklaitis G. V., Ravindran A., Ragsdell K. M. (1986). Engineering Optimization. Methods and Applications, Moscow.

91. Robins H., Monro S. (1951). A stochastic approximation method. // *Annals of Mathematical Statistics*, vol. 22, No. 3, p. 400-407.

92. Rockafellar R. T. (1979). Directionally Lipschitzian functions and subdifferential calculus. // *Proc. London Math. Soc.*, vol. 39, p. 331-355.

93. Rockette H., Antle Ch., Klimko L. A. (1974). Maximum Likelihood Estimation with the Weibull Model. // *Journal of the American Statistical Association*, vol. 69, No. 345, p. 246-249.

94. Rubinstein R., Shapiro A. (1993). Discrete Event Systems: Sensitivity Analysis and Stochastic Optimization by the Score Function Method. New York: Wiley.

95. Sadegh P. (1997). Constrained optimization via stochastic approximation with a simultaneous perturbation gradient approximation. // *Automatica*, vol. 33, p. 889-892.

96. Sakalauskas L. (2002). Nonlinear stochastic programming by Monte-Carlo estimators. // *Informatica*, vol. 137, p. 558-573.

97. Schumer M. A, Steiglitz K. (1972). Adaptive step size random search. // *IEEE Trans Automat Control.*, vol. 17, p. 1004-1007.

98. Smith R. L. (1987). Estimating tails of probability distribution. // *Ann. Statist.*, vol. 15, No. 3, p. 1174-1207.

99. Spall J. C. (1992). Multivariate stochastic approximation using a simultaneous perturbation gradient approximation. // *IEEE Transactions on Automatic Control,* vol. 37. p. 332-341.

100. Spall J. C. (2003). Introduction to Stochastic Search and Optimization: Estimation, Simulation, and Control, New York: Wiley.

101. Ušinskas A. (2003). Medicinoje gaunamų atvaizdų analizė ir klasifikavimas. Daktaro disertacija. Vilniaus Gedimino technikos universitetas.

102. Wasan M. T. (1969). Stochastic Approximation. // *Transactions in Mathematics and Mathematical Physics.* Cambridge University Press*,* Cambridge.

103. Weibull W. (1951). A statistical distribution functions with wide applicability. // *Appl Mech*, vol. 18 p. 293–297.

104. Weiss L. (1971). Asymptotic inference about a density functions at end of its range. Naval Rs. Logistics Quart, vol. 18, No 1, p. 111-115.

105. Zanakis S. H., Kyparisis J. (1986). A Review of Maximum Likelihood Estimation Methods for the Three Parameter Weibull Distribution. // *Journal of Statistical Computation and Simulation*, vol. 25, p. 53-73.

106. Жиглявский А. А., Жилинскас А. Г. (1991). Методы поиска экстремума. Москва: Наука.

107. Жиглявский А. А. (1985). Математическая теория глобального случайного поиска. Л: ЛГУ.

108. Zhigljavsky A. (1990). Branch and probability bound methods for global optimization. // *Informatica*, vol. 1, No. 1, p. 125-140.

PRIEDAI

P.1. TOLYDINIS IR PUSIAU TOLYDINIS ATVAIZDAVIMAS EUKLIDO ERDVĖJE

P.1.1. Euklidinė erdvė. Norma Euklidinėje erdvėje

Viena iš pagrindinių vektorių algebros teorijos sąvokų yra algebrinio vektoriaus, Euklidinės erdvės sąvoka.

<u>P1. Apibrėžimas</u>. *Skaičių rinkinį su jame fiksuota skaičių užrašymo tvarka* $(x_1, x_2, ..., x_n)$, *čia* $x_k \in R,\ k = \overline{1, n}$, *vadinsime n-mačiu vektoriumi. Skaičių* x_k *- vadinsime vektoriaus k-tąja komponente.*

Vektorių aibėje apibrėšime dvi operacijas:

a) dviejų *n*-mačių vektorių sumą:
$$x + y = (x_1, x_2, ..., x_n) + (y_1, y_2, ..., y_n) = (x_1 + y_1, x_2 + y_2, ..., x_n + y_n)$$

b) vieno *n*-mačio vektoriaus sandaugą iš skaliaro:
$$\alpha \cdot x = (\alpha \cdot x_1, \alpha \cdot x_2, ..., \alpha \cdot x_n),\ \alpha \in R$$

<u>P2. Apibrėžimas.</u> *Vektorių aibę su joje apibrėžtomis sudėties ir daugybos iš skaliaro operacijomis žymėsime* $\langle R^n ;+| R\rangle$ *ir vadinsime vektorine tiesine erdve arba tiesiog erdve.*

<u>P3. Apibrėžimas</u>. *Baigtinis iškilas funkcionalas p, apibrėžtas vektorinėje erdvėje, vadinamas norma, jei yra išpildomos šios savybės:*

1) $p(x) \geq 0,$ *kai* $x = 0,$ *tai* $p(x) = 0$,

2) $p(\alpha x) = |\alpha| p(x)$ *su visais* α,

3) $p(x + y) \leq p(x) + p(y)$.

<u>P4. Apibrėžimas</u>. *Tiesinė erdvė, kurioje egzistuoja norma yra vadinama normuota tiesine erdve. Normą žymėsime* $\|x\|$.

<u>P5. Apibrėžimas</u> *Normuota erdvė yra vadinama metrine erdve, jei joje yra įvedamas atstumas*

$$\rho(x, y) = \|x - y\|$$

Pavyzdžiai: 1. Skaičių tiesė R^1 yra vadinama normuota erdve, jei kiekvienam $x \in R^1$ galima įvesti $\|x\| = |x|$.

2. n-matėje erdvėje R^{n} galima įvesti normą $\|x\| = \sqrt{\sum_{k=1}^{n} x_k^2}$. Toje pačioje ervėje galima įvesti ir tokias normas $\|x\|_1 = \sum_{k=1}^{n} |x_k|$, $\|x\|_\alpha = \sqrt[\alpha]{\sum_{k=1}^{n} |x_k|^\alpha}$ arba $\|x\|_0 = \max_{1 \le k \le n} |x_k|$.

Duotoje vektorinėje tiesinėje erdvėje apibrėšime dar vieną operaciją – skaliarinę sandaugą.

<u>P6. Apibrėžimas.</u> Jeigu kiekvienai vektorių $x, y \in R^n$ porai vieninteliu būdu yra priskiriamas realusis skaičius, žymimas simboliu (x, y) ir tenkina sąlygas:

1. $(x, y) = (y, x)$ *- komutatyvumo sąlyga;*
2. $(x + y, z) = (x, z) + (y, z)$*- distributyvumo sąlyga;*
3. $(\lambda x, y) = \lambda (x, y)$, *kai* $\lambda \in R$.
4. $(x, x) > 0,$ *kai* $x \neq 0$ *ir* $(x, x) = 0$, *kai* $x = 0$,

tai sakoma, kad vektorinėje tiesinėje erdvėje $\langle R^n ;+ | R \rangle$ *yra apibrėžta skaliarinė sandauga* (x, y).

<u>P7. Apibrėžimas.</u> Vektorinę tiesinę erdvę, kurioje yra apibrėžta skaliarinė sandauga vadinsime ***Euklidine erdve.***

<u>P8. Apibrėžimas.</u> Vektoriaus x norma Euklidinėje erdvėje vadinsime skaičių $\|x\| = \sqrt{(x, x)} = \left(\sum_{k=1}^{n} |x_k|^2 \right)^{\frac{1}{2}}$, *tenkinantį šias savybes:*

1. $\|x\|$ neneigiamas skaičius apibrėžiamas kiekvienam $x \in R^n$. Jei $x = 0$, tai $\|x\| = 0$, kai $x_k = 0$ su visais k. Ir atvirkščiai: jei $\|x\| = 0$, tai $\|x\|^2 = 0$ ir $\sum_{k=1}^{n} |x_k|^2 = 0$.Tai reiškia, kad $|x_k|^2 = 0$ arba suma neneigiamų skaičių yra lygi nuliui tiktai tada, kai visi tie skaičiai yra lygūs nuliui. Vadinasi, $|x_k| = 0$ ir $x_k = 0$ su visais k.

2. Jei $\alpha \in R$, tai $\|\alpha \cdot x\| = \left(\sum_{k=1}^{n} |\alpha \cdot x_k|^2 \right)^{\frac{1}{2}}$, kai $\alpha \cdot x = (\alpha \cdot x_1, \alpha \cdot x_2, ..., \alpha \cdot x_n)$.

Iš $|\alpha \cdot x_k|^2 = |\alpha|^2 |x_k|^2$, $\sum_{k=1}^{n} |\alpha \cdot x_k|^2 = |\alpha|^2 \sum_{k=1}^{n} |x_k|^2$ seka $\left(|\alpha|^2 \sum_{k=1}^{n} |x_k|^2 \right)^{\frac{1}{2}} = |\alpha| \left(\sum_{k=1}^{n} |x_k|^2 \right)^{\frac{1}{2}}$. Taigi $\|\alpha \cdot x\| = |\alpha| \cdot \|x\|$ su visais $x \in R^n$ ir $\alpha \in R$.

3. Tegu $x=(x_1, x_2, ..., x_n)$ ir $y=(y_1, y_2, ..., y_n)$. Tai $\|x+y\| \le \|x\|+\|y\|$.

<u>P9. Apibrėžimas.</u> *Atstumu* $\rho(x,y)$ *tarp vektorių* x *ir* y *Euklidinėje erdvėje vadiname jų skirtumo normą:*

$$\rho(x,y)=\|x-y\|=\sqrt{(x-y,x-y)}=\left(\sum_{k=1}^{n}|x_k-y_k|^2\right)^{\frac{1}{2}}.$$

<u>P1. Teorema.</u> *Su visais* $x, y \in R^n$ *Euklidinėje erdvėje teisinga nelygybė* $(x,y)^2 \le \|x\|^2 \cdot \|y\|^2$, *kuri vadinama Koši ir Buniakovskio nelygybe.*

Iš Koši ir Buniakovskio nelygybės išplaukia trikampio (Minkovskio) nelygybė: vektorių sumos norma ne didesnė už jos narių normų sumą:

$$\|x+y\| \le \|x\|+\|y\|, \text{ kai } x, y \in R^n.$$

P.1.2. Sekos konvergavimas

Tegu aibė E – apibrėžimo sritis, o aibė F- reikšmių sritis. Veiksmas, kuriuo kiekvienam elementui $x \in E$ priskiria elementą $y \in F$, vadinamas atvaizdavimu aibės E į aibę F. Toks atvaizdavimas yra vadinamas funkcija $y=f(x)$.

<u>P10. Apibrėžimas</u>. *Jei turime aibę E, tai seka aibėje E yra vadinamas atvaizdavimas natūraliųjų skaičių aibės N į aibę E, t.y. funkcija, kurios kintamasis yra* $n \in N$, *o reikšmė – elementas iš aibės E (funkcijos reikšmes žymėsime* x_n*).*

<u>P11. Apibrėžimas.</u> *Seka* $\{x_n\}$ *konverguoja į* x_0, *kai* $n \to \infty$, *jei seka* $\{x_n - x_0\} \to 0$, *t.y. bet kokiam* $\varepsilon > 0$ *egzistuoja toks* $p(\varepsilon)$, *kad su visais* $n > p(\varepsilon)$ *yra išpildoma nelygybė* $|x_n - x_0| < \varepsilon$.

Įvesime pažymėjimą: $\lim\limits_{n\to\infty} x_n = x_0$, kur x_0 yra vadinamas sekos $\{x_n\}$ riba.

Savybės:

1. Jei seka $\{x_n\}$ konverguoja, tai ji turi vienintelę ribą x_0.

$\triangleright$ Tarkime, kad seka $\{x_n\}$ konverguoja į $x_0' \ne x_0$. Tarkime $\varepsilon = \dfrac{|x_0 - x_0'|}{3}$. Kadangi $\{x_n\}$ konverguoja į x_0, tai $|x_n - x_0| < \varepsilon$ su visais $n \ge p$. O jei $\{x_n\}$ konverguoja ir į x_0', tai

$|x_n - x_0'| < \varepsilon$ su visais $n \geq p'$.Tegu *P*- didžiausias iš *p* ir *p'*. Tada su visais $n \geq P$ tuo pačiu metu $|x_n - x_0| < \varepsilon$ ir $|x_n - x_0'| < \varepsilon$. Taigi

$$|x_0 - x_0'| = |x_0 - x_n + x_n - x_0'| \leq |x_0 - x_n| + |x_n - x_0'| = \varepsilon + \varepsilon = 2\varepsilon .$$

O kadangi $\varepsilon = \dfrac{|x_0 - x_0'|}{3}$, tai gauname, kad $3\varepsilon \leq 2\varepsilon$, o taip negali būti. ◁

2. Jei seka $\{x_n\}$konverguoja, tai $|x_n - x_0| < \varepsilon$ išskyrus baigtinį indeksų skaičių. Tarkime turime du sekos narius x_p ir x_q, tai $|x_p - x_q| = |x_p - x_0 + x_0 - x_q| \leq |x_p - x_0| + |x_0 - x_q| = \varepsilon + \varepsilon = 2\varepsilon$ išskyrus baigtinį skaičių reikšmių *p* ir *q*. Kitais žodžiais tariant, koks bebūtų $\varepsilon > 0$ egzistuoja toks sveikasis skaičius *P*, kad jei $p \geq P$ ir $q \geq P$, tai $|x_p - x_q| < \varepsilon$.

3. Jeigu seka $\{x_n\}$ konverguoja į x_0, tai seka $\{|x_n|\}$ konverguoja į $|x_0|$, nes

$$\big||x_n| - |x_0|\big| \leq |x_n - x_0| .$$

<u>*Koši sąlyga.*</u> *Sąlyga* $\lim\limits_{p,q\to\infty} |x_p - x_q| = 0$ *yra būtina ir pakankama realiųjų skaičių sekos konvergavimui.*

<u>*P12. Apibrėžimas.*</u> *Seka* $\{x_n\}$*vadinama Koši seka* (*fundamentaliąja*), *jei ji tenkina Koši sąlygą* $\lim\limits_{p,q\to\infty} |x_p - x_q| = 0$.

<u>*P13. Apibrėžimas.*</u> *Tegu* $\{f_n\}$*yra seka skaitinių funkcijų, apibrėžtų realiųjų skaičių intervale* I. *Ši seka tolygiai konverguoja į skaitinę funkciją f, apibrėžtą intervale I, jei bet kokiam* $\varepsilon > 0$*galima rasti tokį sveikąjį skaičių* $P(\varepsilon)$, *nepriklausantį nuo* $x \in I$, *kad bet kokiam* $n > P$ *yra teisinga nelygybė* $|f_n(x) - f(x)| < \varepsilon$.

<u>*P14. Apibrėžimas.*</u> *Pilna normuota erdvė vadinama Banacho erdve.*

<u>*P15. Apibrėžimas*</u>. *Metrinė erdvė R vadinama kompaktu, jei ji yra pilna ir pilnai aprėžta.*

P.1.3. Monotoninės aprėžtos sekos konvergavimas

<u>*Teiginys 1*</u>. Kiekvienoje begalinėje sekoje galima rasti konverguojantį posekį.

<u>*Teiginys 2*</u>. Bet kokia monotoninė aprėžta seka konverguoja: $\lim\limits_{k\to\infty} x_k = x'$ ir $x_k - x' \geq 0$ visiems *k*.

▷ Turime aprėžtą monotoninę seką $0 \leq ... \leq x_{k+1} \leq x_k \leq ...$. Tegul $\{x_p\}$ yra konverguojantis į x' šios sekos posekis (Teiginys 1), t.y. kiekvienam $\varepsilon > 0$ egzistuoja toks $P(\varepsilon)$, kad su visais $p > P(\varepsilon)$

yra išpildoma nelygybė $\left|x_p - x'\right| < \varepsilon$. Kadangi kiekvienam indeksui k galima rasti tokį indeksą p, kad kai $x_{p+1} \le x_k \le x_p$, gauname, $-\varepsilon \le x_{p+1} - x' \le x_k - x' \le x_p - x' \le \varepsilon$. Iš to seka, kad $\left|x_k - x'\right| < \varepsilon$ visiems $k > P(\varepsilon)$, t.y. $\lim\limits_{k\to\infty} x_k = x'$.

Tarkime, kad nelygybė $x_k - x' \ge 0$ visiems *k* yra neteisinga. Vadinasi galima rasti tokį indeksą p, kad $x_p - x' < 0$, t.y. egzistuoja $\varepsilon' > 0$, kad $x_p - x' \le -\varepsilon'$. Esant monotoninei sekai, turime, kad $x_k - x' \le x_p - x' \le -\varepsilon'$ visiems $k > p$. Iš čia gauname, kad $\left|x_k - x'\right| \ge \varepsilon'$ visiems $k > p$. Jei $\varepsilon < \varepsilon'$, tai $\left|x_k - x'\right| \ge \varepsilon' > \varepsilon$, o tai prieštarauja ribos apibrėžimui. ⊲

P.1.4. Viršutinė ir apatinė sekos riba

Realiųjų taškų aibė E vadinama mažoruojančia arba aprėžta iš viršaus (minoruojančia arba aprėžta iš apačios), jei egzistuoja toks taškas *a*, kad su visais $x \in E$ yra teisinga nelygybė $x \le a$ (atitinkamai $x \ge a$).

Realiųjų taškų aibė *E* vadinama aprėžta, jei ji yra mažoruojanti ir minoruojanti.

Jei *E* yra aprėžta, tai egzistuoja tokie skaičiai *a* ir *b*, kad su visais $x \in E$, $a \le x \le b$.

P16. Apibrėžimas. Mažoruojančios aibės E mažorante vadinamas bet koks skaičius a, su kuriuo $x \le a$, *esant bet kokiam* $x \in E$, *o minoruojančios aibės minorante vadinamas bet koks skaičius b, su kuriuo* $x \ge b$, *esant bet kokiam* $x \in E$.

Jei *E* yra mažoruojanti ir *a* – jos mažorantė, tai bet koks skaičius didesnis už *a*, taip pat bus mažorante.

Pvz. Aibę E sudaro visi neneigiami skaičiai ir $0 \le x < 1$. Ši aibė nėra minoruojanti, bet ji yra mažoruojanti ir bet koks skaičius ≥ 1 yra šios aibės mažorante.

P2. Teorema. Aibėje mažorančių mažoruojančios aibės E egzistuoja mažiausias elementas.

⊳ 1. Jei *E* yra baigtinė, tai mažiausia mažorantė yra laikomas didžiausias aibės *E* elementas.

2. Jei aibėje *E* egzistuoja elementas, kuris yra laikomas mažorante visiems aibės E elementams, tai šis elementas ir yra laikomas mažiausiu iš visų mažorančių.

3. Tegu a yra aibės *E* mažorantė, nepriklausanti tai aibei, o *x*- tam tikras aibės E elementas. Tai jei *x* nėra aibės *E* mažorantė, tai galima rasti tokį a_1, kuris yra intervalo $[x, a]$ vidurys. Jei $a_1 \in E$ ir yra aibės *E* mažorantė, tai ši mažorantė yra mažiausia aibės *E* mažorantė. Priešingu atveju viename iš intervalų $(x, a_1]$, $[a_1, a)$ yra aibės *E* taškų. Jei abejuose intervaluose yra aibės E taškų, tai pasirenkamas dešinysis intervalas. Tęsdami šį veiksmą arba mes po baigtinio žingsnių skaičiaus

gausime aibės E tašką , kuris yra aibės E mažorantė, arba gausime intervalų seką $I_n = [a_n, b_n]$, kur a_n ir b_n turi bendrą ribą b, priklausančią visiems intervalams iš sekos I_n. Iš to seka, kad b yra aibės E mažorantė. $\lhd$

<u>P17. Apibrėžimas.</u> *Viršutiniu rėžiu* (*atitinkamai apatiniu*) *mažoruojančios* (*atitinkamai minoruojančios*) *aibės E yra vadinama mažiausia* (*atitinkamai didžiausia*) *iš aibės E mažorančių* (*atitinkamai minorančių*).

Viršutinis aibės E rėžis yra žymimas $\sup E$, o apatinis - $\inf E$. Iš teoremos įrodymo seka, kad jei E yra uždara mažoruojanti aibė, tai ji turi viršutinį rėžį.

Kaip jau buvo minėta, seka $\{x_n\}$yra funkcijų seka, kurių kintamasis yra $n \geq 1$, o reikšmės užrašomos x_n. Seka $\{x_n\}$konverguoja į x_0, jei bet kokiame atvirame intervale, kuriame galima rasti x_0, yra seka $\{x_n\}$, išskyrus nebent baigtinį jų skaičių.

Įvesime viršutinės ir apatinės ribų apibrėžimus. Tam įvesime ribinių aibės E taškų aibę.

<u>P18. Apibrėžimas</u>. *Viršutine riba vadiname mažoruojančios sekos* $\{x_n\}$ *viršutinį rėžį iš aibės.*

<u>P19. Apibrėžimas.</u> *Apatine riba vadiname minoruojančios sekos* $\{x_n\}$ *apatinį rėžį iš aibės.*

Apatinę ir viršutinę ribas žymėsime atitinkamai: $\overline{\lim_{n\to\infty}} x_n$ ir $\underline{\lim_{n\to\infty}} x_n$.

Tam kad skaičius a būtų sekos $\{x_n\}$ *viršutine riba būtina ir pakankama sąlyga yra ta, kad esant bet kokiam* $\varepsilon > 0$, *galima būtų rasti tokį sveiką skaičių p, kad su visais* $n \geq p$ *būtų teisinga nelygybė* $x_n \leq a + \varepsilon$ *ir esant begalinei reikšmių n aibei būtų teisinga nelygybė* $x_n \geq a - \varepsilon$.

Tam kad skaičius a būtų sekos $\{x_n\}$ *apatine riba būtina ir pakankama sąlyga yra ta, kad esant bet kokiam* $\varepsilon > 0$, *galima būtų rasti tokį sveiką skaičių p, kad su visais* $n \geq p$ *būtų teisinga nelygybė* $x_n \geq a - \varepsilon$ *ir esant begalinei reikšmių n aibei būtų teisinga nelygybė* $x_n \leq a + \varepsilon$.

Galima sakyti, kad a yra sekos $\{x_n\}$ viršutinė riba, jei bet kokiam $\varepsilon > 0$, skaičius indeksų *n,* kuriems $x_n \geq a + \varepsilon$ yra baigtinis, o kuriems $x_n \geq a - \varepsilon$, tai indeksų n skaičius yra begalinis.

Taip pat a yra sekos $\{x_n\}$apatinė riba, jei bet kokiam $\varepsilon > 0$, skaičius indeksų *n,* kuriems $x_n \leq a - \varepsilon$ yra baigtinis, o kuriems $x_n \leq a + \varepsilon$, tai indeksų n skaičius yra begalinis.

<u>P20. Apibrėžimas</u>. *Aibių* A_n, *kai* $n = 1,2,...$, *sekos viršutinė (apatinė) riba, tai aibė* $\overline{A} = \overline{\lim_{n\to\infty}} A_n$ *iš tokių elementų x, kurie priklauso begaliniam aibių* A_n *skaičiui. Atitinkamai aibė* $\underline{A} = \underline{\lim_{n\to\infty}} A_n$ *iš tokių*

elementų x, kurie priklauso visoms aibėms A_n, pradedant nuo tam tikro numerio $n = n(x)$. Akivaizdu, kad $\underline{A} \subset \overline{A}$.

P.1.5. Tolydinis ir pusiautolydinis atvaizdavimas

<u>*P21. Apibrėžimas.*</u> *Tegul f – realiųjų skaičių intervalo I atvaizdis į aibę R^n. Tada f yra vadinamas tolydžiuoju atvaizdavimu (funkcija) taške $x_0 \in I$, jeigu $|f(x) - f(x_0)| \to 0$, kai $x \to x_0$.*

<u>*P22. Apibrėžimas.*</u> *Atvaizdavimas f erdvės X į erdvę Y vadinamas tolydžiuoju atvaizdavimu taške x_0, jei bet kokiai taško $y_0 = f(x_0)$ aplinkai U_{y0} galima rasti tokią taško x_0 aplinką V_{x0}, kad $f(V_{x0}) \subset U_{y0}$.*

Atvaizdavimas *f*: $X \to Y$ vadinamas tolydžiuoju, jei jis tolydus kiekviename taške $x \in X$ .Tegu $\{f_n\}$ seka atvaizdavimų intervalo *I* į aibę R^n. Jei visos funkcijos f_n yra tolydinės intervale *I* ir jei seka $\{f_n\}$ tolygiai konverguoja, tai ribinė funkcija *f* taip pat tolydi intervale *I*.

Tegu *f* yra apibrėžta taško $x_0 \in R^n$ aplinkoje. Ši funkcija yra tolydi tame taške, jei bet kokiam $\varepsilon > 0$ galima rasti tokį $\alpha(x, \varepsilon)$, kad iš sąlygos $\|x - x'\| < \alpha$ seka nelygybė $\|f(x) - f(x')\| < \varepsilon$.

<u>*P23. Apibrėžimas.*</u> *Atvaizdavimas $f(x)$ vadinamas tolydžiu taške x_0, jei bet kuriai sekai $\{x_n\}$ konverguojančiai į x_0, seka $\{y_n = f(x_n)\}$ konverguoja į $y_0 = f(x_0)$.*

<u>*P24. Apibrėžimas.*</u> *Funkcija $f(x)$ vadinama pusiautolydžia iš viršaus (iš apačios) taške x_0, jei bet kokiam $\varepsilon > 0$ egzistuoja tokia taško x_0 aplinka, kurioje $f(x) > f(x_0) - \varepsilon$ (atitinkamai, $f(x) < f(x_0) + \varepsilon$).*

<u>*P25. Apibrėžimas.*</u> *Pusiautolydinis atvaizdavimas iš viršaus (iš apačios)- tai atvaizdavimas f vienos topologinės erdvės X į kitą erdvę Y, toks, kad iš*

$$\lim_{n\to\infty} x_n = x$$

seka

$$\overline{\lim_{n\to\infty}} f(x_n) \subseteq f(x) \ \left(atitinkamai \ \underline{\lim_{n\to\infty}} f(x_n) \supseteq f(x) \right).$$

<u>*P26. Apibrėžimas.*</u> *Atvaizdavimą $G(x)$ vadiname pusiautolydiniu atvaizdavimu iš viršaus taške $x \in X$, jei*

$$\lim_{y\to x} \sup_{g\in G(x)} \inf_{p\in G(y)} \|g - p\| = 0.$$

P1. Lema. Tegul turime seką $x_1, x_2, ..., x_n$*, konverguojančią į x', ir tegu* $f(x)$ *yra tolydinė funkcija. Jei sekos taškai sutvarkyti taip, kad funkcijos reikšmių seka būtų monotoniška:* $f(x_{(1)}) \le f(x_{(2)}) \le ... \le f(x_{(n)})$*, tai tuomet bet kokiems* ε *ir k galima rasti tokį* n_0*, kad* $\|x_i - x'\| \le \varepsilon$*, kai* $i \in [n-i, n]$*,* $n \ge n_0$.

▷ *Įrodymas.* Jei seka $\{x_n\}$ konverguoja į x', tai galima rasti tokį skaičių n_0, kad kai $n \ge n_0$, $q \ge n_0$ $\|x_n - x_q\| < \varepsilon$, kur $\frac{\|x_n - x_q\|}{3} = \varepsilon$, $\|x_n - x_q\| = 3\varepsilon$.

Tarkime priešingai, $\|x_n - x'\| > \varepsilon$. Vadinasi egzistuoja toks x_q, kad $\|x_n - x_q\| = \|x_n - x' + x' - x_q\| \le \|x_n - x'\| + \|x' - x_q\| = 2\varepsilon$. Bet $\|x_n - x_q\| = 3\varepsilon$, iš čia gauname prieštaravimą $3\varepsilon \le 2\varepsilon$. *Lema įrodyta.* ◁

PI.6. Lipšico funkcijų apibendrintas diferencijavimas

Įvesime neiškilų neglodžių funkcijų klasę, vadinamą apibendrintai diferencijuojamomis funkcijomis. Šių funkcijų klasę galima laikyti kaip apibendrinimą tolydžiai diferencijuojamų ir iškilų funkcijų klasės.

P27. Apibrėžimas. Funkcija $f(x) : \Re^n \to \Re^1$ vadinama lokaliai Lipšico, jei bet kokiam kompaktui $K \subset \Re^n$ egzistuoja Lipšico konstanta K tokia, kad

$$|f(x) - f(y)| \le K\|x - y\|, \text{ kiekvienam } x, y \in K .$$

P2. Lema. (Михалевич et al (1987), Lema 2.1). Lokaliai Lipšico funkcijos dalinė išvestinė $\frac{\bar{\partial} f(x)}{\partial x_i}$, $i = 1,...,n$ egzistuoja visur, išskyrus nulinio mato aibes.

P3. Teorema. (*Rademacherio teorema*, žr. Михалевич et al (1987), teorema 2.1). Lokaliai Lipšico funkcija yra diferencijuojama beveik visur, t.y. $\nabla f(x)$ egzistuoja visur, išskyrus nulinio mato aibes.

Lokaliai Lipšico funkcijos gali neturėti išvestinės pagal kryptį:

$$f'(x, e) = \lim_{\lambda \to 0} \frac{f(x + \lambda e) - f(x)}{\lambda} .$$

Tegul $f(x)$ yra lokaliai Lipšico funkcija. Šiai funkcijai galima įvesti apibendrintą gradientą, t.y. Klarko subdiferencialą (Clarke (1975)).

P28. Apibrėžimas. (Михалевич et al (1987)). Apibendrintą Lipšico funkcijos $f(x)$ išvestinę taške x pagal kryptį e, žymėsime $f^0(x, e)$, apibrėžiamą taip:

$$f^0(x,e) = \lim_{\delta\to 0} \sup_{\substack{\|v\|\le\delta \\ 0<\lambda\le\delta}} \frac{f(x+v+\lambda e)-f(x+v)}{\lambda}.$$

P3. Lema. (Михалевич et al (1987), Lema 2.2). Funkcija $e \to f^0(x;e)$ yra iškila.

P30. Apibrėžimas. (Михалевич et al (1987), Clarke (1975)). Apibendrintas funkcijos $f(x)$ gradientas taške *x*, žymimas $\partial f(x)$, yra pusiautolydus iš viršaus daugiareikšmis atvaizdavimas, kurio reikšmės yra netuščios aprėžtos iškilos uždaros aibės, kuriom teisinga

$$f^0(x,e) \ge (y,e) \text{ su kiekvienu } e, \text{ jei } y \in \partial f(x).$$

P4. Teorema. (Михалевич et al (1987), teorema 2.4). Aibė $\partial f(x)$ yra iškilasis uždarinys visų aibės taškų

$$G = \lim_{k\to\infty} \nabla f(x^k),$$

čia $\{x^k\}$- seka, konverguojanti į *x*, ir tokia, kad funkcija $f(x)$ yra diferencijuojama kiekviename taške x^k.

P5. Teorema. (Михалевич et al (1987), teorema 2.5). Tegu

$$g(x^k) \in \partial f(x^k), \quad x^k \to x, \quad g(x^k) \to g.$$

Tada $g \in \partial f(x)$.

P31. Apibrėžimas. (Mifflin (1977)). Funkcija yra vadinama pusiau glodžia taške *x*, jei tam tikroje taško *x* aplinkoje egzistuoja išvestinė pagal kryptį (P28 Apibrėžimas) ir $r(\bar{x},e) = f'(\bar{x}+e,e) - f'(\bar{x},e) \to 0$, kai $h \to 0$ visiem $\bar{x}$ priklausantiems kuriai nors taško *x* aplinkai.

Galima įrodyti, kad bet kuriai sekai $y^t \to x$, tokiai, kad $\frac{y^t - x}{\|y^t - x\|} \to e$ ir $g^t \in \partial f(y^t)$, egzistuoja riba $\lim_{t\to\infty} \langle g^t, e\rangle = f'(x,e)$(Михалевич et al (1987), teorema 1.14).

Žymėkime $\{\Theta_k\}_{k=0}^{\infty}$ σ-algebrų srautą, sugeneruotą sekos $\{x^k\}_{k=0}^{\infty}$.

P4. Lema. Jei $\{\varphi_k\}$ yra tokia atsitiktinių dydžių seka, suderinta su $\{\Theta_k\}_{k=0}^{\infty}$, $\sum_{k=0}^{\infty} E\varphi_k^2 < \infty$, tai $\sum_{k=0}^{\infty} \varphi_k - E(\varphi_k | \Theta_{k-1})$ b.t. konverguoja.

Lemos įrodymą galima rasti (Wasan (1969), Lema 1, priedo 2 skyrius 4).

Visi aukščiau pateikti apibrėžimai, jei nenurodyti kiti autoriai, suformuluoti remiantis Пизо & Заманский (1971) ir Dosinas et al. (1999) pateikta teorija.