

Pristop AS^3D in problem zaporedij z nizkimi avtokorelacijami

Jana Herzog, Janez Brest, Borko Bošković

Laboratorij za računalniške arhitekture in jezike,
Inštitut za računalništvo,
Fakulteta za elektrotehniko, računalništvo in informatiko,
Univerza v Mariboru,
Koroška cesta 46, 2000 Maribor, Slovenija
E-pošta: jana.herzog1@um.si

The AS^3D Approach and Low Autocorrelation Binary Sequence Problem

This paper presents a thorough analysis of the stochastic algorithm with a self-avoiding walk on the optimization problem of the Low Autocorrelation Binary Sequences. A target approach with optimal values was utilized. Both, the problem and the algorithm are analysed by using the recently proposed approach AS^3D . This approach considers a statistical distribution of the observed variable. Based on the parameters of the identified statistical distribution, a predictive model is established. In all of the chosen dimensions, the exponential distribution was recognized. With the empirical validation of the predicted model, we show that the proposed approach is useful for our scenario.

1 Uvod

Neperiodično binarno zaporedje z nizkimi avtokorelacijami (angl. Low-Autocorrelation Binary Sequences – LABS) predstavlja enega izmed zahtevnejših problemov diskretne optimizacije. Njegova uporabnost sega na različna področja kot so fizika, kemija, inženirstvo, kriptografija in procesiranje signalov [2].

Binarno sekvenco $S_L = (s_0, s_1, \dots, s_{L-1})$, kjer L predstavlja dolžino binarnega zaporedja in s_i , $i \in \{0, 1, \dots, L-1\}$ zavzame vrednosti -1 ali 1 . Aperiodična avtokorelacijska funkcija zaporedja S_L je definirana v enačbi (1).

$$C_k(S) = \sum_{i=0}^{L-k-1} s_i \cdot s_{i+k} \quad (1)$$

Obravnavali bomo najvišji nivo stranskega režnja (angl. Peak Sidelobe Level – PSL), ki ga definiramo z enačbo (2).

$$PSL(S_L) = \max_{1 \leq k < L} |C_k(S_L)|. \quad (2)$$

Cilj predstavljenega problema je poiskati optimalno zaporedje S_L^* , kot prikazuje enačba (3). V tej enačbi B_L predstavlja prostor vseh zaporedij.

$$S_L^* = \arg \min_{S_L \in B_L} PSL(S_L). \quad (3)$$

V prispevku se bomo osredotočili na analizo stohastičnega algoritma samoizogibajočega se sprehoda (angl. self-avoiding walk – SAW) [6] na predstavljenem optimizacijskem problemu. Oboje bomo analizirali s pomočjo pristopa AS^3D [1] (angl. Analysis of Stochastic Solvers based on their Statistical Distributions).

Nadaljevanje članka je sledeče, in sicer sledi pregled sorodnih del v 2. poglavju. V 3. poglavju bomo opisali pristop AS^3D . Eksperiment in analiza rezultatov so opisani v 4. poglavju. Zaključek sledi v 5. poglavju.

2 Sorodna dela

Temeljita analiza stohastičnih algoritmov nam podaja odgovor na vprašanje, kako izbrati najprimernejši algoritem za določen optimizacijski problem. Analize stohastičnih algoritmov se raziskovalci lotevajo na različne načine. Najosnovnejša oblika je statistična analiza, ki pove ali obstaja med dvema ali večimi stohastičnimi algoritmi statistično signifikantna razlika. Pri tem se uporabljajo statistični testi [12], [9] in [13]. Zaradi njihove nenatančne in napačne uporabe ter zlorab so se pojavili alternativni pristopi analize, in sicer Bayesove statistike [14], poglobljena statistična analiza [15] in šahovski sistem rangiranja [16].

Da bi pridobili čimbolj poglobljeno znanje o algoritmih, njihovem delovanju in učinkovitosti, se jih analizira iz različnih vidikov. Pojavili so se pristopi, ki analizirajo algoritem iz vidika števila funkcijskih ovrednotenij oz. koliko časa algoritem porabi, da doseže ciljno rešitev (angl. fixed-target approach). Na določenih tekmovanjih [21], [20] se uporablja pristop fiksne števila ovrednotenij (angl. fixed budget approach). Pri tem pristopu je vnaprej podan zaustavitveni pogoj in opazuje se kvaliteta rešitve, ki jo je algoritem sposoben doseči [17]. Novejši pristopi analizirajo tudi lastnosti funkcij z namenom čimvečje raznovrstnosti problemov – ELA (angl. Exploratory Landscape Analysis) [18]. Raziskovalci se ukvarjajo tudi z ustvarjanjem novih testnih funkcij, in sicer se osredotočajo, kako izbrati primerno množico testnih funkcij [19].

Raziskovalci uporabljajo uveljavljene in nove metode, da bi določili, kateri algoritem prevlada nad danimi testnimi funkcijami. Tako si prizadevajo prepoznati prednosti in slabosti vsakega algoritma ter pridobiti obsežen vpogled v njihovo delovanje. V tem prispevku

bomo uporabili pred kratkim objavljen pristop AS^3D [1] in preverili njegovo delovanje na izbranem diskretnem problemu in algoritmu.

3 Pristop AS^3D

Stohastični algoritmi temeljijo na naključnosti, zato je težko odkriti izvor njihovega uspeha. Lahko jih zaustavimo v vsakem trenutku in dobili bomo rešitve, ki ne bodo nujno optimalne, ampak določene kvalitete [3]. Raziskovalci stohastične algoritme analizirajo iz dveh vidikov, in sicer iz vidika fiksnega pristopa [4] in ciljnega pristopa [5]. Pri ciljnem pristopu določimo kvaliteto rešitve, ki jo mora stohastični algoritem doseči v zagonu. Ob tem opazujemo število funkcijskih ovrednotenj ali čas, ki je potreben, da algoritem doseže želeno kvaliteto rešitev. V kolikor algoritem doseže ciljno rešitev, ta zagon definiramo kot uspešen. Pri fiksnem pristopu imamo na voljo vnaprej določen zaustavitveni pogoj (število funkcijskih ovrednotenj ali čas). Pri tem opazujemo kvaliteto rešitve, ki jo stohastični algoritem doseže pri določenem zaustavitvenem pogoju.

Pojavi se vprašanje, kako določiti primeren zaustavitveni pogoj pred začetkom eksperimenta, da bo stohastični algoritem dosegel želeno kvaliteto rešitve z določeno verjetnostjo?

Pri pristopu AS^3D , ki smo ga predstavili v [1], opazujemo stohastični algoritem na izbranem optimizacijskem problemu. Tukaj opazujemo doseganje zelene kvalitete rešitve na manjših dimenzijah in na podlagi teh napovedujemo zaustavitveni pogoj pri večjih dimenzijah. Pri tem lahko opazujemo dve spremenljivki, in sicer čas ali število ovrednotenj, potrebnih za doseg izbrane kvalitete rešitve. Sledi določanje statistične porazdelitve spremenljivk in njenih parametrov. Na podlagi parametrov za različne dimenzije optimizacijskega problema vzpostavimo napovedni model. Pristop AS^3D omogoča sledeče:

- Določanje zaustavitvenih pogojev glede na verjetnost doseganja izbrane kvalitete rešitve.
- Napovedovanje verjetnosti, da bo zagon uspešen glede na dani zaustavitveni pogoj in določeno kvaliteto rešitve.

V tem prispevku se bomo osredotočili na napovedovanje zaustavitvenega pogoja pri določeni verjetnosti. Pristop AS^3D ima določene omejitve, in sicer:

- Uspešnost števila zagonov na izbranem optimizacijskem problemu z določenim stohastičnim algoritmom mora biti 100 %. To pomeni, da mora izbran stohastični algoritem v vsakem izmed zagonov doseči kvaliteto rešitve.
- Uporaba ciljnega pristopa. Izbira kvalitete rešitve je odvisna od uporabnika, ali želi optimalne rešitve ali rešitve določene kvalitete. Prav tako je izbira kvalitete rešitve odvisna od učinkovitosti/sposobnosti stohastičnega algoritma in težavnosti optimizacijskega problema.

- Napoved bo natančnejša za dimenzije, ki so bližje tistim, na katerih je bil vzpostavljen napovedni model. Z oddaljevanjem od izbranih dimenzij, bo tudi natančnost modela manjša.
- Parametri statistične porazdelitve morajo slediti nekemu trendu.

Pristop AS^3D omogoča uporabniku, da opazuje obnašanje algoritma na izbranem optimizacijskem problemu in tako pridobi čimveč informacij o algoritmu in tudi o problemu. V primerjavi s preostalimi pristopi, nam pristop AS^3D omogoča napoved zaustavitvenih pogojev za katerokoli verjetnost uspešnega zagona. Prednost pristopa je tudi v tem, da omogoča vpogled v celotno statistično porazdelitev opazovane spremenljivke. Ne osredotoča se zgolj na eno značilko, kot je npr. povprečje ali mediana [7], [9]. Do sedaj so raziskovalci napovedovali zgolj povprečne vrednosti [6]. Parametrični ali neparametrični statistični testi pokažejo, ali obstaja signifikantna razlika med dvema ali večimi stohastičnimi algoritmi. Velikokrat se zgodi, da niso pravilno uporabljeni, saj niso izpolnjene predhodne predpostavke o statistični porazdelitvi. S pristopom AS^3D omogočimo podrobnejši vpogled v statistično porazdelitev opazovane spremenljivke in napovedovanjem njenega obnašanja.

4 Eksperiment in rezultati

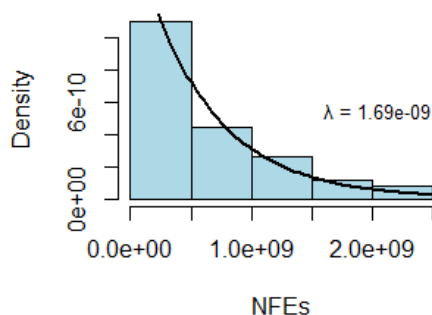
V tem prispevku smo se osredotočili na analizo diskretne optimizacijskega problema LABS v okviru kriterija PSL s stohastičnim algoritmom samoizogibajočega se sprehoda.

Uporabili smo ciljni pristop z optimalnimi rešitvami. Pri tem smo opazovali, kolikšno število funkcijskih ovrednotenj je potrebnih za doseg optimalne rešitve. V eksperimentu smo obravnavali dimenzije od 29 do 43. Za vsako dimenzijo je bilo izvedenih 100 neodvisnih zagonov. Število zagonov je bilo izbrano z namenom lažje določitve statistične porazdelitve.

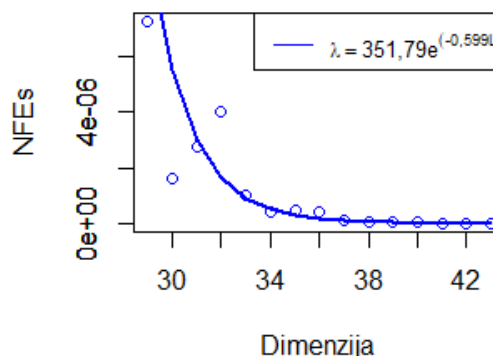
Statistično porazdelitev smo preverili na vsaki izmed dimenzij, in sicer s pomočjo histograma. Uporabili smo tudi Anderson-Darlingov test, ki se uporablja za preverjanje, ali dani podatki pripadajo določeni statistični porazdelitvi [10]. Glede na obe metriki se izkaže, da je porazdelitev števila funkcijskih ovrednotenj eksponentna. Primer eksponentne porazdelitve za dimenzijo $L = 43$ prikažemo na sliki 1. Parameter eksponentne statistične porazdelitve predstavlja λ . Ta se izračuna v enačbi (4).

$$\lambda = \frac{1}{E(X)} \quad (4)$$

V tej enačbi, $E(X)$ predstavlja povprečje vseh zagonov, medtem ko X predstavlja opazovano naključno spremenljivko. V našem primeru X predstavlja število funkcijskih ovrednotenj ($NFEs$). Za vsako izbrano dimenzijo problema od 29 do 43 smo izračunali vrednosti λ . Na podlagi izračunanih λ smo vzpostavili napovedni model. Pogledali smo, katera krivulja se najbolje prilega danim podatkom. Dobro prileganje modela točkam pokažemo z



Slika 1: Eksponentna porazdelitev za dimenzijo $L = 43$ in pripadajoča vrednost λ .



Slika 2: Napovedni model λ za dimenzije od $L = 29$ do 43 za problem PSL ($R^2 = 0,80$).

determinacijskim koeficientom R kvadrat (R^2) [11]. Višji kot je determinacijski koeficient, boljše je prilaganje modela podatkom. Če pogledamo sliko 1, lahko vidimo, da vrednosti λ sledijo eksponentnemu trendu. Splošen zapis tega trenda prikazuje enačba (5).

$$\lambda_{solver}(L) = a \cdot e^{-b \cdot L} \quad (5)$$

Napovedni model na podlagi eksponentne porazdelitve prikazuje enačba (6).

$$\lambda_{SAW}(L) = 351,79 \cdot e^{-0,599 \cdot L} \quad (6)$$

V enačbi (6) in na sliki 1 prikažemo napovedni model za λ glede na dane dimenzije problema. Na podlagi tega modela bomo napovedali vrednost zaustavitvenega pogoja, da bo optimalna rešitev dosežena pri 95 % verjetnosti. Iz enačbe (7) za kumulativno porazdelitveno funkcijo eksponentne porazdelitve sledi izpeljava za izračun zaustavitvenega pogoja pri 95 % verjetnosti uspešnega zagona.

$$P(x, \lambda) = \begin{cases} 1 - e^{-\lambda \cdot x} & , x \geq 0 \\ 0 & , x < 0 \end{cases} \quad (7)$$

Spremenljivka x predstavlja v našem primeru število funkcijskih ovrednotenj in v ta namen bomo uporabljali zapis $NFEs$. Sledi ponazoritev za $L = 44$ v enačbah (8) in (9).

$$P_{SAW}(NFEs, 44) = \begin{cases} 1 - e^{1,26 \cdot 10^{-9} \cdot NFEs} & , NFEs \geq 0 \\ 0 & , NFEs < 0 \end{cases} \quad (8)$$

$$0,95 = 1 - e^{1,26 \cdot 10^{-9} \cdot NFEs} \quad (9)$$

$$NFEs = \frac{\ln 0,05}{-1,26 \cdot 10^{-9}} = 2,377 \cdot 10^9 \quad (9)$$

Da bi empirično validirali naš model, smo za dimenzijo 44 naredili 100 neodvisnih zagonov. Ker smo napovedovali vrednost $NFEs$ pri 95 % verjetnosti uspešnih zagonov, nas je zanimala empirična vrednost

$NFEs$, da dosežemo 95 % uspešnih zagonov. Izkaže se, da samoizogibajoči se sprehod potrebuje $2,2 \cdot 10^9$ ovrednotenj, da doseže optimalne rešitve v 95 zagonih od skupno 100 zagonov. V tabeli 1 prikažemo še napovedi, kdaj bo optimalna rešitev dosežena z verjetnostjo 90 %, 95 % in 99 %. V tabeli 1 je prikazano tudi empirično število funkcijskih ovrednotenj, ki jih je samoizogibajoči se sprehod potreboval, da je dosegel 90 %, 95 % in 99 % uspešnih zagonov. Vidimo, da je razlika med napovedano in empirično vrednostjo majhna, kar kaže na uporabnost napovednega modela. Opazimo lahko, da je napovedana vrednost v tabeli 1 zmeraj nekoliko višja od empirične vrednosti $NFEs$.

Tabela 1: Vrednosti napovedanega in izmerjenega $NFEs$ za verjetnosti 90 %, 95 % in 99 %.

Verjetnost	Napoved	Empirična vrednost	Razlika [%]
90 %	$1,8274 \cdot 10^9$	$1,6435 \cdot 10^9$	11
95 %	$2,3775 \cdot 10^9$	$2,1551 \cdot 10^9$	9
99 %	$3,6548 \cdot 10^9$	$3,2665 \cdot 10^9$	11

V tabeli 2 prikazujemo empirično vrednost in napovedano vrednost λ ter razliko med njima. Vidimo, da je vzpostavljen model uporaben, kajti njegova natančnost je večja od 87 % za vse izbrane velikosti sekvenc.

5 Zaključek

V tem prispevku smo se osredotočili na analizo algoritma samoizogibajočega se sprehoda pri reševanju problema binarnih sekvenc z nizkimi avtokorelacijami, ki upoštevajo kriterij najvišjega nivoja stranskega režnja. Analizo smo opravili s pomočjo pristopa AS^3D . Dobili smo model za napoved zaustavitvenega pogoja oz. števila ovrednotenj, da dosežemo optimalno rešitev s

Tabela 2: Vrednosti napovedane in izmerjene λ za dimenzije problema 44, 45 in 46.

Dimenzija	Napoved λ	Empirična vrednost λ	Razlika [%]
44	$1,26 \cdot 10^{-9}$	$1,37 \cdot 10^{-9}$	9
45	$6,91 \cdot 10^{-10}$	$6,02 \cdot 10^{-10}$	13
46	$3,79 \cdot 10^{-10}$	$3,06 \cdot 10^{-10}$	11

pomočjo določene verjetnosti. Ta model smo empirično validirali. Ugotovili smo, da je model dokaj natančen. Razlika med napovedano in empirično vrednostjo zaustavitvenega pogoja je znašala manj kot 13 %. Tako smo pokazali, da lahko s pomočjo pristopa AS^3D uspešno določimo zaustavitveni pogoj stohastičnega algoritma in pričakujemo določeno kvaliteto dosežene rešitve.

Zahvala

Priznavamo financiranje prispevka s strani Javne agencije za raziskovalno dejavnost Republike Slovenija, raziskovalni program P2-0041 – Računalniški sistemi, metodologije in inteligentne storitve.

Literatura

- [1] J. Herzog, J. Brest in B. Boškovič. Analysis based on statistical distributions: A practical approach for stochastic solvers using discrete and continuous problems. *Information Sciences*, 633, strani 469-490, 2023.
- [2] J. Brest in B. Boškovič. Low autocorrelation binary sequences: Best-known peak sidelobe level values. *IEEE Access*, 9, strani 67713-67723, 2021.
- [3] M. Buzdalov, B. Doerr, C. Doerr in D. Vinokurov. Fixed-target runtime analysis. *Algorithmica*, strani 1-32, 2021.
- [4] T. Jansen. Analysing stochastic search heuristics operating on a fixed budget. *Theory Of Evolutionary Computation: Recent Developments In Discrete Optimization*, strani 249-270, 2020.
- [5] N. Hansen, A. Auger, R. Ros, O. Mersmann, T. Tušar in D. Brockhoff. COCO: A platform for comparing continuous optimizers in a black-box setting. *Optimization Methods And Software*, 36, strani 114-144, 2021.
- [6] B. Boškovič, F. Brglez in J. Brest. Low-autocorrelation binary sequences: On improved merit factors and runtime predictions to achieve them. *Applied Soft Computing*, 56, strani 262-285, 2017.
- [7] J. Derrac, S. García, S. Hui, P. N. Suganthan in F. Herrera, Analyzing Convergence Performance of Evolutionary Algorithms: A Statistical Approach. *Information Sciences*, 289, strani 41-58, 2014.
- [8] S. García, A. Fernández, J. Luengo in F. Herrera, Advanced Nonparametric Tests for Multiple Comparisons in the Design of Experiments in Computational Intelligence and Data Mining: Experimental Analysis of Power. *Information sciences*, 180 (10), strani 2044-2064, 2010.
- [9] J. Derrac, S. García, D. Molina in F. Herrera, A Practical Tutorial on the use of Nonparametric Statistical Tests as a Methodology for Comparing Evolutionary and Swarm Intelligence Algorithms. *Swarm and Evolutionary Computation*, 1 (1), strani 3-18, 2011.
- [10] M. A. Stephens, EDF statistics for goodness of fit and some comparisons. *Journal Of The American Statistical Association*, 69(347), strani 730-737, 1974.
- [11] A. Cameron in F. Windmeijer. An R-squared measure of goodness of fit for some common nonlinear regression models. *Journal Of Econometrics*, 77, strani 329-342, 1997.
- [12] J. Carrasco, S. García, M. Rueda, S. Das in F. Herrera. Recent trends in the use of statistical tests for comparing swarm and evolutionary computing algorithms: Practical guidelines and a critical review. *Swarm And Evolutionary Computation*, 54, 2020.
- [13] S. Garcia, D. Molina, M. Lozano in F. Herrera. A study on the use of non-parametric tests for analyzing the evolutionary algorithms' behaviour: a case study on the CEC'2005 special session on real parameter optimization. *Journal Of Heuristics*, 15, strani 617-644, 2009.
- [14] A. Benavoli, G. Corani, J. Demšar in M. Zaffalon. Time for a change: a tutorial for comparing multiple classifiers through Bayesian analysis. *The Journal Of Machine Learning Research*, 18, strani 2653-2688, 2017.
- [15] T. Eftimov, P. Korošec in B. Seljak. A novel approach to statistical comparison of meta-heuristic stochastic optimization algorithms using deep statistics. *Information Sciences*, 417, strani 186-215, 2017.
- [16] N. Veček, M. Mernik in M. Črepinšek. A chess rating system for evolutionary algorithms: A new method for the comparison and ranking of evolutionary algorithms. *Information Sciences*, 277, strani 656-679, 2014.
- [17] T. Tušar, N. Hansen in D. Brockhoff. Anytime benchmarking of budget-dependent algorithms with the COCO platform. IS 2017-International Multiconference Information Society, strani 1-4, 2017.
- [18] A. Nikolikj, R. Lang, P. Korošec in T. Eftimov. Explaining Differential Evolution Performance Through Problem Landscape Characteristics. *Bioinspired Optimization Methods And Their Applications: 10th International Conference, BIOMA 2022, Maribor, Slovenia*, strani 99-113, november 2022.
- [19] G. Cenikj, R. Lang, A. Engelbrecht, C. Doerr, P. Korošec in T. Eftimov. Selector: selecting a representative benchmark suite for reproducible statistical comparison. *Proceedings Of The Genetic And Evolutionary Computation Conference*, strani 620-629, 2022.
- [20] A. W. Mohamed, A. A. Hadi, A. K. Mohamed, P. Agrawal, A. Kumar in P. N. Suganthan Problem Definitions and Evaluation Criteria for the CEC 2021 Special Session and Competition on Single Objective Bound Constrained Numerical Optimization. (Nanyang Technological University, Singapore). <https://github.com/P-N-Suganthan/2021-SO-BCO>
- [21] A. W. Mohamed, A. A. Hadi, A. K. Mohamed, P. Agrawal, A. Kumar in P. N. Suganthan Problem Definitions and Evaluation Criteria for the CEC 2022 Special Session and Competition on Single Objective Bound Constrained Numerical Optimization. (Nanyang Technological University, Singapore). <https://github.com/P-N-Suganthan/2022-SO-BO>