LABS – binarna zaporedja z nizkimi avtokorelacijami

Janez Brest[1], **Borko Bošković**[1]

[1] ***Laboratorij za računalniške arhitekture in jezike, Inštitut za računalništvo,***
Fakulteta za elektrotehniko, računalništvo in informatiko, Univerza v Mariboru,
Koroška cesta 46, 2000 Maribor, Slovenija,
janez.brest@um.si, borko.boskovic@um.si

LABS – Low autocorrelated binary sequences

Abstract. *Low autocorrelated binary sequences (*labs*) problem is a difficult combinatorial problem. It has various applications in many communications and radar systems. The merit factor of a sequence is a well-known measure for the ratio of the energy in the sequence's aperiodic autocorrelation peak to the total energy in its sidelobes.*

In this paper we give an overview of the labs *problem, some discussions of state-of-the-art algorithms for solving this combinatorial problem and their characteristics. We also present the best-known solution values for this problem and an interesting observation regarding these solutions, since most of them are skew-symmetric.*

1 Uvod

V članku opisujemo binarna zaporedja z nizkimi avtokorelacijami (ang. "low autocorrelated binary sequences", krajše labs), ki imajo mnogo praktičnih aplikacij v radarskih in komunikacijskih sistemih ter statistični mehaniki [3, 13]. Iskanje takih zaporedij predstavlja izjemno težak optimizacijski problem [4, 15, 17, 2, 16, 14].

Ločimo dve vrsti avtokorelacij in sicer:

- periodične in
- aperiodične ali neperiodične.

V tem članku se bomo omejili le na aperiodični problem labs.

Preostanek prispevka je organiziran takole. V drugem poglavju predstavimo problem labs. Tretje poglavje opisuje algoritme, ki so primerni za reševanje omenjenega problema, in podajamo pregled stanja na tem področju. Sledi še zaključno poglavje z idejami za nadaljnje raziskave.

2 Problem labs

Opis problema labs povzemamo po [15]. Za dano *binarno zaporedje* dolžine L

$$S = (s_1, s_2, \cdots, s_L), \quad s_i \in \{-1, 1\} \tag{1}$$

izračunamo njeno aperiodično avtokorelacijo

$$C_k(S) = \sum_{i=1}^{L-k} s_i s_{i+k}, \quad k = 0, 1, \cdots, L-1. \tag{2}$$

Energija zaporedja S je definirana kot

$$E(S) = \sum_{k=1}^{L-1} C_k^2. \tag{3}$$

Faktor F ("merit factor") danega zaporedja, kot ga je definiral Golay [9], izračunamo

$$F(S) = \frac{L^2}{2\sum_{k=1}^{L-1} C_k^2} = \frac{L^2}{2E(S)}. \tag{4}$$

Želimo, da imajo binarna zaporedja čim večji $F(S)$, kar z drugimi besedami pomeni čim manjšo energijo $E(S)$.

Zgled: Za lažje razumevanje prikažimo kratko binarno zaporedje s petimi elementi: $s = (s_1, s_2, s_3, s_4, s_5)$. S pomočjo enačbe (2) izračunamo:

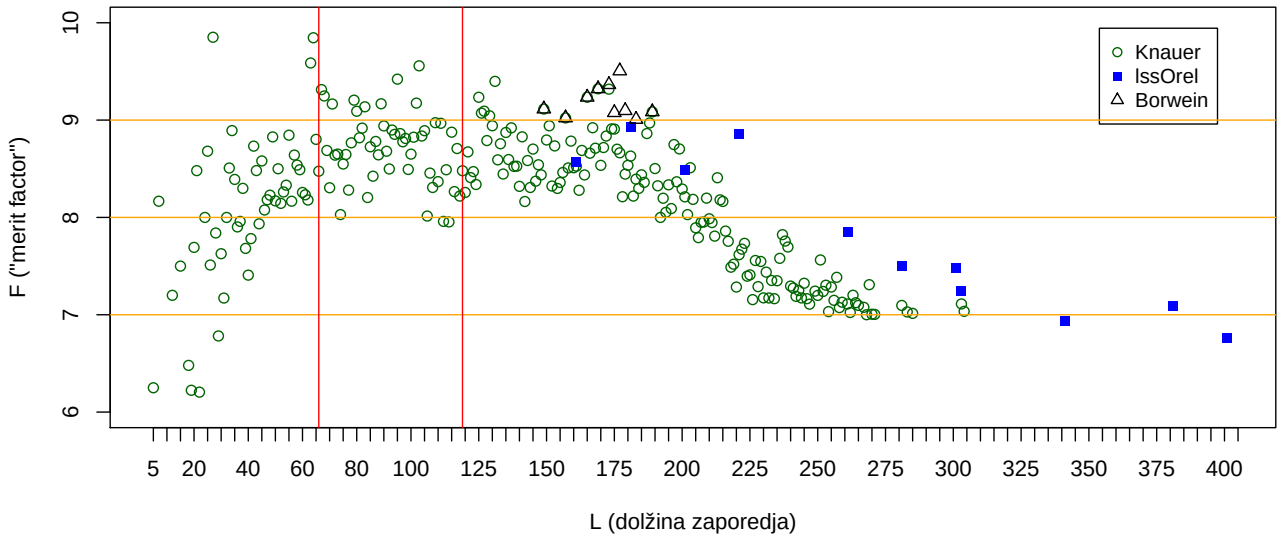
$$\begin{aligned} C_1 &= s_1 s_2 + s_2 s_3 + s_3 s_4 + s_4 s_5, \\ C_2 &= s_1 s_3 + s_2 s_4 + s_3 s_5, \\ C_3 &= s_1 s_4 + s_2 s_5, \\ C_4 &= s_1 s_5. \end{aligned}$$

Člen C_0 smo izpustili, saj ni vključen v enačbo (3), kjer izračunamo energijo. Omenimo, da ima vsak C_k ($k = 1, 2, 3, 4$) $L - k$ elementov in $L = 5$ je dolžina binarnega zaporedja.

Prikažimo še nekaj primerov zaporedij za $L = 5$ in izračunajmo vrednosti E in F :

$$\begin{aligned} &S = (-1, -1, -1, -1, -1), && E = 30, && F = 0.4167; \\ &S = (1, 1, -1, 1, -1), && E = 6, && F = 2.0833; \\ &S = (1, 1, 1, -1, 1), && E = 2, && F = 6.25; \end{aligned}$$

Vrednost $F = 6.25$ je optimalna za $L = 5$.



Slika 1: Najboljši znani rezultati za problem labs: Knauerjeva zbirka [12], ter rezultati dveh stohastičnih algoritmov (Borwein [5], in Bošković [6], ki ima oznako lssOrel). Do $L \leq 66$ so optimalne rešitve, do $L \leq 119$ so optimalne rešitve za lihe dolžine s simetrijo 'skew', ostalo so pa doslej najboljše najdene rešitve.

V splošni obliki problem labs lahko zapišemo:

$$\lim_{L\to\infty} F(S). \tag{5}$$

Problem labs (vrednost limite v enačbi (5)) še dandanes ostaja nerešen. Golay [10] je zapisal domnevno asimptotično vrednost $F = 12.3248$ za zelo dolga zaporedja.

3 Eksaktni in stohastični algoritmi

Zaporedja pri majhnih vrednostih L lahko optimalni (maksimalni) F izračunamo z *eksaktnim algoritmom*, s katerim izračunamo vrednost F za vse možne razvrstitve -1 in 1 v danem zaporedju. Vseh možnostih je 2^L. Tako problem labs spada med probleme z eksponentno zahtevnostjo.

Iskalni prostor problema labs je zelo skokovit in nazobčan ter ima mnogo lokalnih optimumov. Če naredimo majhno spremembo znotraj zaporedja S, ko npr. zamenjamo en predznak $s_i = -s_i$, energija (glej enačbo (3)) močno spremeni svojo vrednost, kar posredno vpliva tudi na F.

Z eksaktnim algoritmom ne pridemo prav daleč (približno do vrednosti $L = 35$ na današnjem osebnem računalniku). Doslej so s pomočjo zmogljivih paralelnih računalnikov eksaktno izračunali zaporedja do velikosti $L \leq 66$ [14]. Kot zanimivost povejmo, da so leta 1996 bili znani rezultati za $L \leq 60$ [13] in potrebni sta bili dve desetletji, da se je meja premaknila iz 60 na 66. Pri eksaktnem računanju se ponavadi uporabljajo algoritmi razveji in omeji ("branch-and-bound").

Pri problemu labs obstajajo tudi *simetrije*. Na primer, če v zaporedju zamenjamo vse -1 z 1 in obratno, dobimo prvo simetrijo. O ostalih simetrijah pa lahko bralec razlago s primeri najde v [6]. Povejmo, da imajo zaporedja lihih dolžin vsaj 4 simetrične optimalne rešitve, zaporedja sodih dolžin pa imajo vsaj 8 simetričnih rešitev [6]. Z uporabo simetrij se iskalni prostor le nekoliko zmanjša, a z večanjem L iskalni prostor raste eksponentno.

Samo za zaporedja lihih dolžin, $L = 2k - 1$, je definirana posebna simetrija 'skew-symmetric' (v slovarju smo našli prevod 'popačena simetrija'):

$$s_{k+1} = (-1)^i s_{k-i},\ i = 1, 2, \ldots k-1. \tag{6}$$

Ta oblika simetrije občutno zmanjša efektivno velikost iskalnega prostora – gre za zmanjšanje približno za faktor 2 (dimenzija iskalnega prostora se prepolovi), vendar dobljene rešitve pri nekaterih vrednostih niso nujno optimalne. Optimalne rešitve s simetrijo 'skew' so znane za $L \leq 119$ [14]. Pri dolgih zaporedjih ($L > 200$) se rešitve s to simetrijo pravzaprav izkažejo za kar precej dobre rešitve, kar bomo lahko natančneje opazili pri rezultatih, ki jih podajamo v nadaljevanju prispevka.

Druga možnost je uporaba *stohastičnih algoritmov*, ki lahko najdejo dobre rešitve tudi za daljša zaporedja, za katere pa žal ne moremo reči, ali so optimalne. V zadnjem času so se pojavili stohastični algoritmi, ki so uspeli poiskati najboljša znana zaporedja.

Na kratko opišimo nekatere nedavno nastale stohastične algoritme:

- S. Halim [11]: algoritem uporablja lokalno iskanje in iskanje s tabuji ("tabu search"). Primeren je za reševanje celotnih (brez simetrije 'skew') zaporedij.
- P. Borwein [5]: usmerjeni stohastični algoritem je uspel najti dobre rešitve do velikosti približno $L = 200$.
- J. Gallardo et al. [8]: memetski algoritem, ki tudi uporablja iskanje s tabuji. Je zelo hiter in učinkovit.
- B. Bošković et al. [6]: algoritem lssOrel, ki temelji na preiskovanju, ki se izogiba ponavljanju ("self-avoiding walk" - SAW). Z njim so avtorji uspeli najti precej doslej najboljših rešitev (glej sliko 1 in tabelo 1).

Pri reševanju problema labs algoritmi ponavadi težijo k pogostemu ponavaljanju že obiskanih rešitev. Kot je pozoren bralec lahko razbral iz opisov algoritmov, imajo vsi mehanizem za preprečitev večkratnega obiska iste rešitve, kar lahko vodi k ciklanju. Da ciklanje preprečimo, je težavno opravilo. Še ena lastnost tega kombinatoričnega problema, ki otežuje iskanje, je, da ponavadi obstaja mnogo rešitev z isto vrednostjo E oziroma F. Vsem opisanim algoritmom je skupen ponoven naključni start, ki ga algoritmi opravljajo precej pogosto.

Slika 1 prikazuje rezultate najboljših znanih algoritmov in Knauerjeve zbirke rezultatov [12] iz leta 2004. Opazimo, da je algoritem lssOrel (Bošković et al. [6]) uspel izboljšati rezultate iz [12], in je našel boljše rešitve kot algoritem avtorjev Gallardo et al. [8] (glej tabelo 1).

Dolžine zaporedij, ki jih najdemo pri Borwein [5] in so prikazane v tabeli 2, ne najdemo v [8] in [6], zato ne moremo nič več povedati o rezultatih v tabeli 2, saj v [5] ni bila narejena detaljna analiza algoritma, kot je na primer prikazana na sliki 2. Na sliki je prikazano, da je izračunan model

Tabela 1: Najboljše znane vrednosti za F.

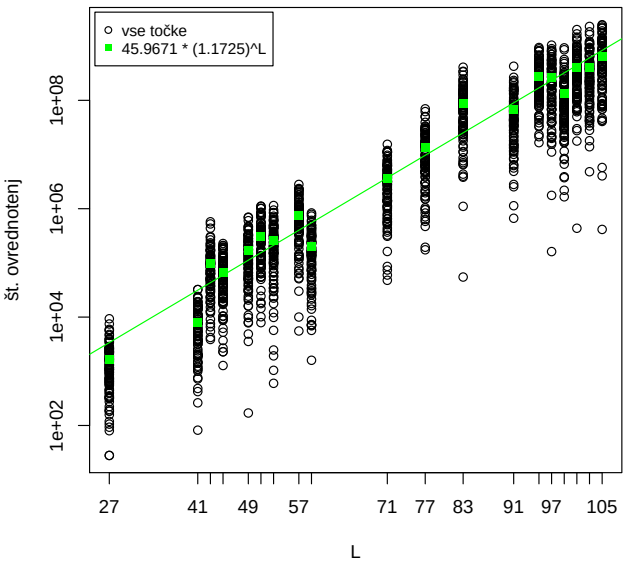
L	Knauer [12]	Gallardo [8]	Bošković [6]
119	**8.4796**	**8.4796**	**8.4796**
121	**8.6736**	**8.6736**	**8.6736**
141	**8.8282**	**8.8282**	**8.8282**
161	8.5266	**8.5718**	**8.5718**
181	8.6304	7.7194	**8.9316**
201	8.2116	7.6633	**8.4876**
215	8.1641	-	**8.5888**
221	7.6171	-	**8.8544**
241	7.2747	-	**8.0668**
249	7.2431	-	**8.1323**
259	7.1287	-	**8.0918**

Tabela 2: Zbrani rezultati za vrednosti F usmerjenega stohastičnega iskalnega algoritma (Borwein [5]).

L	Borwein [5]
126	**9.0720**
149	**9.1137**
157	**9.0223**
165	**9.2351**
169	**9.3215**
172	**9.0526**
173	**9.3645**
175	**9.0768**
177	**9.5052**
178	**9.2915**
179	**9.0974**
183	**9.0073**
189	**9.0874**

naslednji: $a * b^L = 45.9671 * (1.1725)^L$. S pomočjo asimptotične analize (glej [6, 7]) lahko primerjamo zmogljivosti več algoritmov.

Ob sliki 1 se poraja vprašanje, ali vrednost F pada pri zaporedjih z $L > 190$ ali pa da algoritmi le niso uspeli najti boljših rešitev v izjemno velikem iskalnem prostoru. Na primer, algoritem lssOrel se je izvajal na računalniški gruči [1]. Odgovor na zastavljeno vprašanje je, da zaradi izjemno velikega iskalnega prostora algoritmi niso uspeli najti boljših rešitev!



Slika 2: Asimptotična krivulja za dani algoritem, ki je izračunana s pomočjo povprečnih vrednosti (zelena barva). Za vsako dolžino zaporedja L je bilo opravljenih 100 zagonov (Pri danem L imajo dobljeni rezultati eksponentno porazdelitev.). Y-os ima logaritemsko skalo, kar kaže na eksponentno zahtevnost problema labs.

V tabeli 1 so zbrane vrednosti F za lažjo primerjavo. Vse prikazane vrednosti, ki so v stolpcu Bošković [6], predstavljajo rešitve s simetrijo 'skew'. Prav tako so vse rešitve pri lihih dolžinah L v tabeli 2 rešitve s simetrijo 'skew'. S tem želimo povedati, da če primerjamo rezultate Knauerjeve zbirke iz leta 2004 in rezultate, ki so bili najdeni s pomočjo algoritmov [5] in [6], lahko opazimo, da so nove najdene najboljše rešitve v veliki večini s simetrijo 'skew'.

Pri primerjavi tabel 1 in 2 moramo biti posebej pozorni tudi na dolžine L, saj med tabelama ni preseka in ne smemo pozabiti na eksponentno zahtevnost problema labs.

4 Zaključek

V članku smo predstavili binarna zaporedja z nizkimi avtokorelacijami – problem labs, ki spada med najzahtevnejše kombinatorične probleme. Predstavili smo algoritme za reševanje problema labs, na kratko opisali njihove lastnosti in opravili pregled doslej najdenih najboljših rešitev.

Algoritmi imajo težave pri reševanju problema labs zaradi velikosti iskalnega prostora, ki se povečuje eksponentno, možnosti ciklanja pri pregledovanju rešitev in lastnosti velike spremembe energije že ob majhni spremembi zaporedja.

Nadaljenje raziskave na tem področju pa lahko potekajo v smeri opravljanja eksperimentov za lihe dolžine $149 \leq L \leq 191$, da bomo lahko opravili še boljšo primerjavo predstavljenih algoritmov.

Zahvala

J. Brest in B. Bošković priznavata financiranje prispevka s strani Javne agencije za raziskovalno dejavnost Republike Slovenije, raziskovalni program P2-0041 – Računalniški sistemi, metodologije in inteligentne storitve.

Literatura

[1] SLING - Slovenian Initiative for National Grid. `http://www.sling.si/`, February 2017.

[2] J.M. Baden. Efficient optimization of the merit factor of long binary sequences. *Information Theory, IEEE Transactions on*, 57(12):8084–8094, Dec 2011.

[3] J. Bernasconi. Low autocorrelation binary sequences: statistical mechanics and configuration space analysis. *J. Physsique*, 48:559–567, April 1987.

[4] P. Borwein, K.-K.S. Choi, and J. Jedwab. Binary sequences with merit factor greater than 6.34. *IEEE Transactions on Information Theory*, 50(12):3234–3249, Dec 2004.

[5] Peter Borwein, Ron Ferguson, and Joshua Knauer. The merit factor problem. *London Mathematical Society Lecture Note Series*, 352:52, 2008.

[6] B. Bošković, F. Brglez, and J. Brest. Low-Autocorrelation Binary Sequences: On Improved Merit Factors and Runtime Predictions to Achieve Them. *Applied Soft Computing*, 56:262–285, 2017.

[7] F. Brglez, X. Y. Li, M. Stallmann, and B. Militzer. Reliable Cost Predictions for Finding Optimal Solutions to LABS Problem: Evolutionary and Alternative Algorithms. In *Proc. of The Fifth Int. Workshop on Frontiers in Evolutionary Algorithms (FEA2003), Cary, NC, USA*, September 2003. Also available at http://militzer.berkeley.edu/papers-/2003-FEA-Brglez-posted.pdf.

[8] José E. Gallardo, Carlos Cotta, and Antonio J. Fernández. Finding low autocorrelation binary sequences with memetic algorithms. *Appl. Soft Comput.*, 9(4):1252–1262, September 2009.

[9] M. J. E. Golay. Sieves for low autocorrelation binary sequences. *IEEE: Transactions on Information Theory*, 23:43–51, 1977.

[10] M. J. E. Golay. The merit factor of long low autocorrelation binary sequences. *IEEE: Transactions on Information Theory*, 28:543–549, 1982.

[11] Steven Halim, Roland H. Yap, and Felix Halim. Engineering stochastic local search for the low autocorrelation binary sequence problem. In *Proceedings of the 14th international conference on Principles and Practice of Constraint Programming*, CP '08, pages 640–645, Berlin, Heidelberg, 2008. Springer-Verlag.

[12] Joshua Knauer. A table of 'Merit Factor Records' for Low Autocorrelation Binary Sequences. http://signalslab.marstu.net/?page_id=4270, http://labraj.uni-mb.si/xBed/xProj/B.labs/xResults/2004-records-asymp-all-Knauer.html.

[13] S. Mertens. Exhaustive search for low-autocorrelation binary sequences. *Journal of Physics A: Mathematical and General*, 29:473–481, 1996. http://www-e.uni-magdeburg.de/mertens/research/labs/open.dat.

[14] Tom Packebusch and Stephan Mertens. Low autocorrelation binary sequences. *Journal of Physics A: Mathematical and Theoretical*, 49(16):165001, 2016.

[15] Abhisek Ukil. Low autocorrelation binary sequences: Number theory-based analysis for minimum energy level, barker codes. *Digit. Signal Process.*, 20(2):483–495, March 2010.

[16] Abhisek Ukil. On asymptotic merit factor of low autocorrelation binary sequences. In *Industrial Electronics Society, IECON 2015-41st Annual Conference of the IEEE*, pages 004738–004741. IEEE, 2015.

[17] Toby Walsh. CSPLib problem 005: Low autocorrelation binary sequences. `http://www.csplib.org/Problems/prob005`.