A SAT Attack on the Erdős Discrepancy Conjecture

Boris Konev and Alexei Lisitsa

Department of Computer Science
University of Liverpool, United Kingdom

Abstract In 1930s Paul Erdős conjectured that for any positive integer C in any infinite ± 1 sequence (x_n) there exists a subsequence $x_d, x_{2d}, x_{3d}, \ldots, x_{kd}$, for some positive integers k and d, such that $| \sum_{i=1}^{k} x_{id} | > C$. The conjecture has been referred to as one of the major open problems in combinatorial number theory and discrepancy theory. For the particular case of $C = 1$ a human proof of the conjecture exists; for $C = 2$ a bespoke computer program had generated sequences of length 1124 of discrepancy 2, but the status of the conjecture remained open even for such a small bound. We show that by encoding the problem into Boolean satisfiability and applying the state of the art SAT solver, one can obtain a discrepancy 2 sequence of length 1160 and a *proof* of the Erdős discrepancy conjecture for $C = 2$, claiming that no discrepancy 2 sequence of length 1161, or more, exists. We also present our partial results for the case of $C = 3$.

1 Introduction

Discrepancy theory is a branch of mathematics dealing with irregularities of distributions of points in some space in combinatorial, measure-theoretic and geometric settings [5,9,19,4]. The paradigmatic combinatorial discrepancy theory setting can be described in terms of a hypergraph $\mathcal{H} = (U, S)$, that is, a set U and a family of its subsets $S \subseteq 2^U$. Consider a colouring $c : U \to \{+1, -1\}$ of the elements of U in *blue* $(+1)$ and *red* (-1) colours. Then one may ask whether there exists a colouring of the elements of U such that in every element of S colours are distributed uniformly or a discrepancy of colours is always inevitable. Formally, the discrepancy (deviation from a uniform distribution) of a hypergraph $\mathcal{H}$ is defined as $\min_c(\max_{s \in S} | \sum_{e \in s} c(e)|)$. Discrepancy theory also has practical applications in computational complexity [9], complexity of communication [1] and differential privacy [20].

One of the oldest problems of discrepancy theory is the discrepancy of hypergraphs over the set of natural numbers with the subsets (hyperedges) forming arithmetical progressions over this set [18]. Roth's theorem [22], one of the main results in the area, states that for the hypergraph formed by the arithmetic progressions in $\{1, \ldots, l\}$, that is $\mathcal{H}_l = (U_l, S_l)$, where $U_l = \{1, 2, \ldots, l\}$ and elements of S_l being of the form $(ai+b)$ for arbitrary a, b, the discrepancy grows at least as $\frac{1}{20} l^{1/4}$.

Surprisingly, for the more restricted case of *homogeneous* arithmetic progressions of the form (ai), the question of the discrepancy bounds is open for more than eighty years. In 1930s Paul Erdős conjectured [10] that the discrepancy is unbounded. This conjecture became known as the Erdős discrepancy problem (EDP) and its proving or disproving has been referred to as one of the major open problems in combinatorial number theory and discrepancy theory [5,4,21].

The problem can be naturally described in terms of sequences of $+1$ and -1 (and this is how Erdős himself introduced it). Then Erdős's conjecture states that for any $C > 0$ in any infinite ± 1 sequence (x_n) there exists a subsequence $x_d, x_{2d}, x_{3d}, \ldots, x_{kd}$, for some positive integers k and d, such that $|\sum_{i=1}^{k} x_{id}| > C$. The general definition of discrepancy given above can be specialised as follows. The discrepancy of a finite ± 1 sequence $\bar{x} = x_1, \ldots, x_l$ of length l can be defined as $\max_{d=1,\ldots,l}(|\sum_{i=1}^{\lfloor \frac{l}{d} \rfloor} x_{id}|)$. For an infinite sequence (x_n) its discrepancy is the supremum of discrepancies of all its initial finite fragments.

For random ± 1 sequences of length l the discrepancy grows as $l^{1/2+o(1)}$ and the explicit constructions of a sequence with slowly growing discrepancy at the rate of $\log_3 l$ have been demonstrated [14,8]. It is known [17] that discrepancy of any infinite ± 1 sequence can not be bounded by 1, that is, Erdős's conjecture holds for the particular case $C = 1$. For all other values of C the status of the conjecture remained unknown. Although widely believed not to be the case, there was still a possibility that an infinite sequence of discrepancy 2 existed.

The EDP has attracted renewed interest in 2009-2010 as it became a topic of the Polymath project [11], a widely publicised endeavour in collective math initiated by T. Gowers [13]. As part of this activity (see discussion in [11]) an attempt has been made to attack the problem using computers. A purposely written computer program had successfully found ± 1 sequences of length 1124 having discrepancy 2; however, it failed to produce a discrepancy 2 sequence of a larger length and it has been claimed that "given how long a finite sequence can be, it seems unlikely that we could answer this question just by a clever search of all possibilities on a computer" [11].

In this paper we settle the status of the EDP for $C = 2$. We show that by encoding the problem into Boolean satisfiability and applying the state of the art SAT solvers, one can obtain a sequence of length 1160 of discrepancy 2 and a proof of the Erdős discrepancy conjecture for $C = 2$, claiming that no sequence of length 1161 and discrepancy 2 exists. We also present our partial results for the case of $C = 3$ and demonstrate the existence of a sequence of length 13 000 of discrepancy 3.

2   SAT Encoding

Checking that a ± 1 sequence of length l has discrepancy C is quite straightforward and so for the existence claims the specific encoding details are of limited interest and could be left as an exercise to the reader. The negative results (that is, our claim that no infinite discrepancy 2 sequence exists), however, require us to give a short description of our SAT encoding of the EDP. The encoding in full for all cases discussed in this paper and the program generating the encoding of the EDP for arbitrary given values of C and l can be found in [16].

We employ the automata based approach similar to the encoding of temporal formulae for bounded model checking [6]. In Figure 1 we give an automaton that accepts a ± 1 word if, and only if, the word represents a ± 1 sequence whose discrepancy is larger than C. Notice that if a subsequence $x_d, x_{2d}, \ldots, x_{kd}$ of $\bar{x} = x_1, \ldots, x_l$ contains less than C elements, this subsequence does not contribute to the discrepancy of $\bar{x}$. It should be clear then that if for every $d : 1 \leq d \leq \lfloor \frac{l}{C+1} \rfloor$ the automaton $\mathcal{A}_C$ *does not*
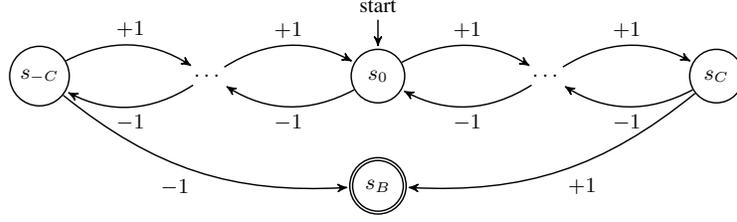
Figure 1. Automaton $\mathcal{A}_C$ accepting a sequence of discrepancy exceeding C.

accept the subsequence $x_d, x_{2d}, \ldots, x_{kd}$, where $k = \lfloor \frac{l}{d} \rfloor$, then the discrepancy of the sequence $\bar{x}$ does not exceed C.

The trace of the automaton $\mathcal{A}_C$ on the subsequence $x_d, x_{2d}, \ldots, x_{kd}$ can be encoded by a Boolean formula in the obvious way. To explain representation details, first consider

$$\phi_{(l,C,d)} = s_0^{(1,d)} \bigwedge_{i=1}^{\lfloor \frac{l}{d} \rfloor} \left[\bigwedge_{-C \leq j < C} \left(s_j^{(i,d)} \wedge p_{id} \rightarrow s_{j+1}^{(i+1,d)} \right) \wedge \right.$$
$$\bigwedge_{-C < j \leq C} \left(s_j^{(i,d)} \wedge \neg p_{id} \rightarrow s_{j-1}^{(i+1,d)} \right) \wedge$$
$$\left(s_C^{(i,d)} \wedge p_{id} \rightarrow B \right) \wedge$$
$$\left. \left(s_{-C}^{(i,d)} \wedge \neg p_{id} \rightarrow B \right) \right], \tag{1}$$

where the intended meaning is that proposition $s_j^{(i,d)}$ is true if, and only if, the automaton $\mathcal{A}_C$ is in the state s_j having read first $(i-1)$ symbols of the input word, and proposition p_i is true if, and only if, the i-th symbol of the input word is $+1$.

Let

$$\phi_{(l,C)} = \neg B \wedge \bigwedge_{d=1}^{\lfloor \frac{l}{C+1} \rfloor} \phi_{(l,C,d)} \wedge \mathsf{frame}_{(l,C)},$$

where $\mathsf{frame}_{(l,C)}$ is a Boolean formula encoding that the automaton state is correctly defined, that is, exactly one proposition from each of the sets $\{s_j^{(i,d)} \mid -C \leq j \leq C\}$, for $d = 1, \ldots, \lfloor \frac{l}{C+1} \rfloor$ and $1 \leq i \leq \lfloor \frac{l}{d} \rfloor$, is true in every model of $\phi_{(l,C)}$.

The following statement can be easily proved by an investigation of models of $\phi_{(l,C)}$ and the traces of $\mathcal{A}_C$. Notice that although $\phi_{(l,C)}$ encodes the traces of $\mathcal{A}_C$ on all subsequences of $\bar{x}$ they all share the same proposition B—as soon as the automaton accepts any of these subsequences, the entire sequence should be rejected.

Proposition 1. *The formula $\phi_{(l,C)}$ is satisfiable if, and only if, there exists a ± 1 sequence $\bar{x} = x_1, \ldots, x_l$ of length l of discrepancy C. Moreover, if $\phi_{(l,C)}$ is satisfiable, the sequence $\bar{x} = x_1, \ldots, x_l$ of discrepancy C is uniquely identified by the assignment of truth values to propositions $p_1, \ldots p_l$.*

The encoding described above, albeit very natural, is quite wasteful: the size of formula $\mathsf{frame}_{(l,C)}$ is quadratic in the number of states. To reduce the size, in our implementation we use a slightly different encoding of the traces of $\mathcal{A}_C$. Namely, we replace in (1) every occurrence of $s_j^{(i,d)}$ with a conjunction of propositions representing the numerical value of j in binary, where the most significant bit encodes the sign of j and the other bits encode an unsigned number $0 \ldots C$ in the usual way. We denote the resulting formula $\phi_{(l,C,d)}^b$.

For example, for $C = 2$ the values $-C \ldots C$ can be represented in binary by 3 bits. Then $\phi_{(l,C,d)}^b$ contains, for example,

$$(\neg b_2^{(i,d)} \wedge \neg b_1^{(i,d)} \wedge \neg b_0^{(i,d)}) \wedge p_{id} \rightarrow \neg b_2^{(i+1,d)} \wedge \neg b_1^{(i+1,d)} \wedge b_0^{(i+1,d)}$$

encoding the transition from s_0 to s_1 having read $+1$.

We also exclude by a formula $\mathsf{frame}_{(l,C)}^b$ all combinations of bits that do not correspond to any states of $\mathcal{A}_C$. For example, for $C = 2$ we have

$$\mathsf{frame}_{(l,C)}^b = \bigwedge_{d=1}^{\lfloor \frac{l}{C+1} \rfloor} \bigwedge_{i=1}^{\lfloor \frac{l}{d} \rfloor + 1} \left[\neg(b_2^{(i,d)} \wedge \neg b_1^{(i,d)} \wedge \neg b_0^{(i,d)}) \wedge \\ \neg(\neg b_2^{(i,d)} \wedge b_1^{(i,d)} \wedge b_0^{(i,d)}) \wedge \\ \neg(b_2^{(i,d)} \wedge b_1^{(i,d)} \wedge b_0^{(i,d)}) \right].$$

The first conjunct disallows the binary value `100`, a 'negated zero', the other two encode that $\mathcal{A}_C$, for $C = 2$, does not have neither s_3 nor s_{-3}. The following statement is a direct consequence of Proposition 1.

Proposition 2. *The formula* $\phi_{(l,C)}^b = \neg B \wedge \bigwedge_{d=1}^{\lfloor \frac{l}{C+1} \rfloor} \phi_{(l,C,d)}^b \wedge \mathsf{frame}_{(l,C)}^b$ *is satisfiable if, and only if, there exists a* ± 1 *sequence* $\bar{x} = x_1, \ldots, x_l$ *of length* l *of discrepancy* C. *Moreover, if* $\phi_{(l,C)}$ *is satisfiable, the sequence* $\bar{x} = x_1, \ldots, x_l$ *of discrepancy* C *is uniquely identified by the assignment of truth values to propositions* $p_1, \ldots p_l$.

3 Results

In our experiments we used the Lingeling SAT solver [7] version ats, the winner of the *SAT-UNSAT* category of the SAT'13 competition [3] and the Glucose solver [2] version 3.0, the winner of the *certified UNSAT* category of the SAT'13 competition [3]. All experiments were conducted on PCs equipped with an Intel Core i5-2500K CPU running at 3.30GHz and 16GB of RAM.

By iteratively increasing the length of the sequence, we establish precisely that the maximal length of a ± 1 sequence of discrepancy 2 is 1160. On our system it took Plingeling, the parallel version of the Lingeling solver, about 800 seconds[1] to find a satisfying assignment. One of the sequences of length 1160 of discrepancy 2 can be found in Appendix A for reader's amusement.

---

[1] The time taken by the solver varies significantly from experiment to experiment; in one rerun it took the solver just 166.8 seconds to find a satisfying assignment.

Proposition 3. *There exists a sequence of length* 1160 *of discrepancy* 2.

When we increased the length of the sequence to 1161, Plingeling reported unsatisfiability. In order to corroborate this statement, we also used Glucose. It took the solver about 21 500 seconds to compute a Delete Reverse Unit Propagation (DRUP) certificate of unsatisfiability, which is a compact representation of the resolution refutation of the given formula [12]. The correctness of the unsatisfiability certificate has been independently verified with the drup-trim tool [15]. The size of the certificate is about 13 GB, and the time needed to verify the certificate was comparable with the time needed to generate it. Combined with Proposition 2, we obtain a computer proof of the following statement.

Theorem 1. *No sequence of length* 1161 *has discrepancy* 2.

As there is no finite sequence of discrepancy 2, there is no infinite such sequence. So we conclude the following.

Corollary 1. *The Erdős discrepancy conjecture holds true for* $C = 2$.

In an attempt to better understand this result, we looked at the smaller unsatisfiable subset of $\phi_{(l,C)}^{b}$ identified by the drup-trim tool. It turned out that the encoding of some automata traces is not present in the subset. A further manual minimisation showed that, although $\lfloor \frac{1161}{3} \rfloor$ is 387, to show unsatisfiability it suffices to consider subsequences of $x_1, \ldots, x_{1161}$ of the form $x_d, \ldots, x_{kd}$ for the values of d ranging from 1 to 358. It remains to be seen whether or not this observation can be helpful for a human proof of the conjecture

We also applied our methodology to identify sequences of discrepancy 3, however, we did not manage to prove the conjecture. Having spent 3 days, 7 hours and 30 minutes (or 286247.9 seconds total), on the encoding of the problem using 356 048 variables and 4 342 612 clauses Plingeling has successfully identified a sequence of length 13 000 with discrepancy 3. The encoding and the generated sequence can be found in [16].

Proposition 4. *There exists a sequence of length* 13 000 *of discrepancy* 3.

Unfortunately, our attempts to improve this result did not succeed: Plingeling did not return any answer on the encodings of discrepancy 3 problems for sequences of length 14 000 and 16 000 even within 1 550 000 and 2 280 000 seconds, respectively; the computations are still going on and the problem is still open.

4 Discussion

We have demonstrated that SAT-based methods can be used to tackle the longstanding mathematical question on the discrepancy of ± 1 sequences. For EDP with $C = 2$ we have identified the exact boundary between satisfiability and unsatisfiability, that is, we found the longest discrepancy 2 sequence and proved that no larger sequence of discrepancy 2 exists. There is, however, a noticeable asymmetry between these findings. The fact that a sequence of length 1160 has discrepancy 2 can be easily checked either by a straightforward computer program or even manually. The negative witness, that is,

the DRUP unsatisfiability certificate, is probably one of longest proofs of a non-trivial mathematical result ever produced. Its gigantic size is comparable, for example, with the size of the whole Wikipedia, so one may have doubts about to which degree this can be accepted as a proof of a mathematical statement.

But this is the best we can get for the moment. Essentially, the unsatisfiability proof corresponds to the verification that the search in a huge search space has been done correctly and completed without finding a satisfying assignment. It is a challenging problem to produce a compact proof more amenable for human comprehension.

Finally notice that apart from the obtained results the proposed methodology can be used to further experimentally explore variants of the Erdős problem as well as more general discrepancy theory problems.

References

1. Alon, N.: Transmitting in the *n*-dimensional cube. Discrete Applied Mathematics 37/38, 9–11 (1992)
2. Audemard, G., Simon, L.: Glucose 2.3 in the SAT 2013 Competition. In: Proceedings of SAT Competition 2013. pp. 42–43. University of Helsinki (2013)
3. Balint, A., Belov, A., Heule, M.J.H., Järvisalo, M. (eds.): Proceedings of SAT competition 2013. University of Helsinki (2013)
4. Beck, J., Sós, V.T.: Discrepancy theory. In: Graham, R.L., Grötschel, M., Lovász, L. (eds.) Handbook of combinatorics, vol. 2, pp. 1405–1446. Elsivier (1995)
5. Beck, J., Chen, W.W.L.: Irregularities of Distribution. Cambridge University Press (1987)
6. Biere, A.: Bounded model checking. In: Handbook of Satisfiability, Frontiers in Artificial Intelligence and Applications, vol. 185, pp. 457–481. IOS Press (2009)
7. Biere, A.: Lingeling, Plingeling and Treengeling entering the SAT Competition 2013. In: Proceedings of SAT Competition 2013. pp. 51–52. University of Helsinki (2013)
8. Borwein, P., Choi, S.K.K., Coons, M.: Completely multiplicative functions taking values in $\{1, -1\}$. Transactions of the American Mathematical Society 362(12), 6279–6291 (2010)
9. Chazelle, B.: The Discrepancy Method: Randomness and Complexity. New York: Cambridge University Press (2000)
10. Erdős, P.: Some unsolved problems. The Michigan Mathematical Journal 4(3), 291–300 (1957)
11. Erdős discrepancy problem: Polymath wiki. `http://michaelnielsen.org/polymath1/index.php?title=The_Erd%C5%91s_discrepancy_problem`, accessed 29 January 2014
12. Goldberg, E.I., Novikov, Y.: Verification of proofs of unsatisfiability for CNF formulas. In: Proceedings of Design, Automation and Test in Europe Conference and Exposition (DATE 2003), 3-7 March 2003, Munich, Germany. pp. 10886–10891 (2003)
13. Gowers, T.: Is massively collaborative mathematics possible? http://gowers.wordpress.com/2009/01/27/is-massively-collaborative-mathematics-possible/, accessed 29 January 2014
14. Gowers, T.: Erdős and arithmetic progressoins. In: Erdős Centennial conference (2013), `http://www.renyi.hu/conferences/erdos100/program.html`, accessed 29 January 2014
15. Heule, M.J.H.: DRUP checker. `http://www.cs.utexas.edu/˜marijn/drup/`, accessed 29 January 2014
16. Konev, B., Lisitsa, A.: Addendum to: A SAT attack on the Erdős discrepancy conjecture. `http://www.csc.liv.ac.uk/˜konev/SAT14`

17. Mathias, A.R.D.: On a conjecture of Erdős and Čudakov. Combinatorics, geometry and probability (1993)
18. Matousek, J., Spencer, J.: Discrepancy in arithmetic progressions. Journal of the American Mathematical Society 9(1), 195–204 (1996)
19. Matousek, J.: Geometric Discrepancy: An Illustrated Guide, Algorithms and combinatorics, vol. 18. Springer (1999)
20. Muthukrishnan, S., Nikolov, A.: Optimal private halfspace counting via discrepancy. In: Proceedings of the 44th Symposium on Theory of Computing. pp. 1285–1292. STOC '12, ACM, New York, NY, USA (2012)
21. Nikolov, A., Talwar, K.: On the hereditary discrepancy of homogeneous arithmetic progressions. CoRR abs/1309.6034v1 (2013)
22. Roth, K.F.: Remark concerning integer sequence. Acta Arithmetica 9, 257–260 (1964)

A One of the sequences of length 1160 having discrepancy 2

We give a graphical representation of one of the sequences of length 1160 obtained from the satisfying assignment computed with the Plingeling solver. Here + stands for +1 and − for −1, respectively.

```
− + + − + − − + + − + + − + − − + − − + + − + − − + − − +
+ − + − − + + − + + − + − + + − − + + − + − − − + − + + −
+ − − + − − + + + + − − + − − + + − + − − + + − + + − − −
− + + − + + − + − + + − − + + − + − + − − − + + − + − − +
+ − + + − + − − + + − + − − + − − − + − + + − + − − + + −
+ + − + − − + − − + + − + + − + − − + + − + − − + + + − +
− + − − − + + + − + − − + − − + + + − − − + + − + + − + −
− − + − − + + + − − + − + − + − − + − + + + − + + − + − −
+ − − + + − + − − + + − + + − + − − − + − − + + − + + + −
− − + + + − + − − − + + − + − − + + − − + − + − − + − + +
+ − + − − + + − + + − + − − + + − + − − + − − + + − + − −
+ + − − + − + + − + − + − − + − + − + + − + − − + + − + −
− + − − + + − + − + − + + − + − + − + + − − − + − + − − +
+ + + − − + − − − + + − + − + + − + − − + + − + − − + − −
+ + − + − − + + + + − − + − − − + − + + + − − + − − + + +
− + + − + − − + + − + − − + − − + + − + − − + + − + + − +
− − + + − + − − + − − + + − + + − + − − + + + − + − + − −
+ + − − + + + − − − + − + + − + − − + − + + + − − − + − +
− + + − + − − + − − + + − + − + + + − + − − + − − + − − +
+ + + − − + − − − + + + − − − + + − + + − + − − + + − − +
+ − − + − − + + − + + − + − − + − − + − − + + + + − + + −
− + − − + + − − + − + + − + + − + − − + − − + − − + + + −
+ − + − − + + + − − − + + − + − − + + − + + − + − − + + +
− − + + − + + − − − − + + + − − + − + + − + − − + − − + +
− + − − + + − + + − + − + + − − + + − − + + − − − + + + +
− + + − − + + − − − − + + − + + + − − + + − − − + + + − −
− − + − + − + + − + + − + + − + − + − + − − − + + + − − +
− + − − + + − + + − + − − + − − + − − + + − + − − + + − +
```

```
+ − + − − + + − − + − + − − + − + − + − + + + + − − − + −
+ − + + − − + − − + − + − + − + + − + − + + + − − + − + −
− + − − + − + + + − − + − + + + − − − + + − + − − + − − +
+ − + + − − + + − − − + + − + − + + − − + + − + − − − + −
+ + − + − − + − + + − − + + − + − − + + − + − − + − + + +
− + − − + + − − + − + − + + + − − + − + − − + + − + + − +
− − + − − + − + + − − − + − + + − + − + + − − + + − + − −
+ + + − + − − − − + + − − + − + + − + − + + − − + + − + −
− + + − + − + + − − + + − + − − − + − + + − + − − + + + −
− − − + − + − + + − − + + − + − − + + − + − + + + − + − −
+ − − + − − + + + + − − − + + − − − + − + − + + − + − + +
+ − − + − + + − − + − + − − + − + − + + − − − + + + − + +
```