On the Erdős Discrepancy Problem

Ronan Le Bras, Carla P. Gomes, and Bart Selman
Computer Science Department,
Cornell University, Ithaca, NY, 14850, USA

Abstract. According to the Erdős discrepancy conjecture, for any infinite ±1 sequence, there exists a homogeneous arithmetic progression of unbounded discrepancy. In other words, for any ±1 sequence \((x_1, x_2, \ldots)\) and a discrepancy \(C\), there exist integers \(m\) and \(d\) such that \(|\sum_{i=1}^{m} x_{id}| > C\). This is an 80-year-old open problem and recent development proved that this conjecture is true for discrepancies up to 2. Paul Erdős also conjectured that this property of unbounded discrepancy even holds for the restricted case of completely multiplicative sequences, namely sequences \((x_1, x_2, \ldots)\) where \(x_{ab} = x_a \cdot x_b\) for any \(a, b \geq 1\). The longest such sequence of discrepancy 2 has been proven to be of size 246. In this paper, we prove that any completely multiplicative sequence of size 127,646 or more has discrepancy at least 4, proving the Erdős discrepancy conjecture for discrepancy up to 3. In addition, we prove that this bound is tight and increases the size of the longest known sequence of discrepancy 3 from 17,000 to 127,645. Finally, we provide inductive construction rules as well as streamlining methods to improve the lower bounds for sequences of higher discrepancies.

Introduction

Much like Ramsey theory studies how order must emerge in combinatorial objects as their size increases, discrepancy theory investigates how deviations from uniformity necessarily occur. Namely, discrepancy theory addresses the problem of distributing points uniformly over some geometric object, and studies how irregularities ineluctably appear in these distributions. For example, this subfield of combinatorics aims to answer the following question: for a given set \(U\) of \(n\) elements, and a finite family \(S = \{S_1, S_2, \ldots, S_m\}\) of subsets of \(U\), is it possible to color the elements of \(U\) in red or blue, such that the difference between the number of blue elements and red elements in any subset \(S_i\) is small?

Important contributions in discrepancy theory include the Beck-Fiala theorem \([1]\) and Spencer’s Theorem \([2]\). The Beck-Fiala theorem guarantees that if each element appears at most \(t\) times in the sets of \(S\), the elements can be colored so that the imbalance, or discrepancy, is no more than \(2t - 1\). According to the Spencer’s theorem, the discrepancy of \(S\) grows at most as \(\Omega(\sqrt{n \log(2m/n)})\).

Nevertheless, some important questions remain open. According to Paul Erdős himself, two of his oldest conjectures relate to the discrepancy of homogeneous arithmetic progressions (HAPs) \([3]\). Namely, a HAP of length \(k\) and of common difference \(d\) corresponds to the sequence \((d, 2d, \ldots, kd)\). The first conjecture can be formulated as follows:
Conjecture 1. Let \((x_1, x_2, \ldots)\) be an arbitrary \pm 1 sequence. The discrepancy of \(x\) w.r.t. HAPs must be unbounded, i.e. for any integer \(C\) there is an integer \(m\) and an integer \(d\) such that \(|\sum_{i=1}^{m} x_{i \cdot d}| > C\).

This problem has been open for over eighty years, as is the weaker form according to which one can restrict oneself to completely multiplicative functions. Namely, \(f\) is a completely multiplicative function if \(f(a \cdot b) = f(a) \cdot f(b)\) for any \(a, b\). The second conjecture translates to:

Conjecture 2. Let \((x_1, x_2, \ldots)\) be an arbitrary completely multiplicative \pm 1 sequence. The discrepancy of \(x\) w.r.t. HAPs must be unbounded, i.e. for any integer \(C\) there is a \(m\) and a \(d\) such that \(|\sum_{i=1}^{m} x_{i \cdot d}| > C\).

Hereinafter, when non-ambiguous, we refer to the discrepancy of a sequence as its discrepancy with respect to homogeneous arithmetic progressions. Formally, we denote \(\text{disc}(x) = \max_{m, d} |\sum_{i=1}^{m} x_{i \cdot d}|\). We denote \(\mathcal{E}_1(C)\) the minimum length for which any sequence has discrepancy at least \(C + 1\), or equivalently, one plus the maximum length of a sequence of discrepancy \(C\). Similarly, we define \(\mathcal{E}_2(C)\) the minimum length for which any completely multiplicative sequence has discrepancy at least \(C + 1\).

| \(C\) | \(\mathcal{E}_1(C)\) | \(\mathcal{E}_2(C)\) |
|-----|----------------|----------------|
| 12  | 1,161          | \(127,646\)   |
| 10  | 247            | \(127,646\)   |

A proof or disproof of these conjectures would constitute a major advancement in combinatorial number theory \[4\]. To date, both conjectures have been proven to hold for the case \(C \leq 2\). As illustrated in Table[1] the values of \(\mathcal{E}_1(1), \mathcal{E}_2(1),\) and \(\mathcal{E}_2(2)\) have been long proven to be 12, 10, and 247 respectively, while recent development proved \(\mathcal{E}_1(2) = 1161\) \[5\]. Konev and Lisitsa \[5\] also provide a new lower bound for \(\mathcal{E}_1(3)\). After 3 days of computation, a SAT solver was able to find a satisfying assignment for a sequence of length 13,000. Yet, it would fail to find a solution of size 14,000 in over 2 weeks of computation. They also report a solution of length 17,000, the longest known sequence of discrepancy 3.

In this work, we explore streamlining for this problem, an effective combinatorial search strategy that exploits regularities in some problem solutions, beyond the structure of the combinatorial problem itself. Streamlining provides and exploits structural information about the problem, and we believe that a sine qua non condition for tackling huge sequences requires deep insights into the structure of the problem. In addition, streamlining provides a vision for a broader strategy for solving problems. Overall, we

\[1\] Note that, if Conjecture 1 (resp. Conjecture 2) were to be rejected, \(\mathcal{E}_1(C)\) (resp. \(\mathcal{E}_2(C)\)) would correspond to infinity.