Simultaneous linear discrepancy for unions of intervals*

Ron Holzman and Nitzan Tur
Department of Mathematics
Technion-Israel Institute of Technology
32000 Haifa, Israel
holzman@tx.technion.ac.il, nitzan.tur@gmail.com

July 30, 2017

Abstract

Lovász proved (see [7]) that given real numbers $p_1, \dots, p_n$, one can round them up or down to integers $\epsilon_1, \dots, \epsilon_n$, in such a way that the total rounding error over every interval (i.e., sum of consecutive p_i's) is at most $1 - \frac{1}{n+1}$. Here we show that the rounding can be done so that for all $d = 1, \dots, \lfloor \frac{n+1}{2} \rfloor$, the total rounding error over every union of d intervals is at most $(1 - \frac{d}{n+1})d$. This answers a question of Bohman and Holzman [1], who showed that such rounding is possible for each value of d separately.

1 Introduction

Let $[n] = \{1, \dots, n\}$. The *linear discrepancy* of a hypergraph $\mathcal{H} \subseteq 2^{[n]}$ is defined by

$$\text{lindisc}(\mathcal{H}) \quad = \quad \max_{p_1,\dots,p_n \in [0,1]} \; \min_{\epsilon_1,\dots,\epsilon_n \in \{0,1\}} \; \max_{X \in \mathcal{H}} \left| \sum_{i \in X} (\epsilon_i - p_i) \right| .$$

Thus, given any assignment of real numbers $p_1, \dots, p_n$ to the vertices of $\mathcal{H}$, the goal is to round them up or down to integers $\epsilon_1, \dots, \epsilon_n$ so that the total rounding error over any edge of $\mathcal{H}$ will be as small as possible. This concept was introduced by Lovász, Spencer and Vesztergombi [5], who studied its relationship to several other notions of hypergraph discrepancy. Additional investigations of linear discrepancy include [7, 4, 6, 2, 3, 1].

A natural example for studying linear discrepancy is the *interval hypergraph* $\mathcal{H}_n$ on the vertex set $[n]$, having as edges all the integer intervals, i.e., sets

*Running head: Simultaneous linear discrepancy. MSC: 05C65, 11K38.

of consecutive elements of $[n]$. Spencer [7] gave a short argument (a 'gem' attributed to Lovász) that $\text{lindisc}(\mathcal{H}_n) = 1 - \frac{1}{n+1}$. An example of an assignment of $p_1, \ldots, p_n$ which forces a rounding error of at least $1 - \frac{1}{n+1}$ over some interval is $p_1 = \cdots = p_n = \frac{1}{n+1}$.

More generally, one may consider the *d-interval hypergraph* $\mathcal{H}_n^{(d)}$, where a subset of $[n]$ is an edge if it is the union of at most d intervals. The relevant values of d are $1, \ldots, \lfloor \frac{n+1}{2} \rfloor$, with $\mathcal{H}_n^{(1)} = \mathcal{H}_n$ and $\mathcal{H}_n^{(\lfloor \frac{n+1}{2} \rfloor)} = 2^{[n]}$. It is straightforward to deduce from $\text{lindisc}(\mathcal{H}_n) = 1 - \frac{1}{n+1}$ that $\text{lindisc}(\mathcal{H}_n^{(d)}) \leq (1 - \frac{1}{n+1})d$. Bohman and Holzman [1] improved this, showing that $\text{lindisc}(\mathcal{H}_n^{(d)}) = (1 - \frac{d}{n+1})d$ for every $d \in \{1, \ldots, \lfloor \frac{n+1}{2} \rfloor\}$. But the rounding used to establish this was devised for each value of d separately. The question whether the same rounding can work simultaneously for all $d = 1, \ldots, \lfloor \frac{n+1}{2} \rfloor$ was left open in [1]. Here we answer this affirmatively:

Theorem 1. *For any $p_1, \ldots, p_n \in [0, 1]$ there exist $\epsilon_1, \ldots, \epsilon_n \in \{0, 1\}$ such that the following holds true:*

For all $d = 1, ..., \lfloor \frac{n+1}{2} \rfloor$ and for any $2d$ integers $0 \leq a_1 < b_1 < \cdots < a_d < b_d \leq n$ we have

$$\sum_{t=1}^{d} \left| \sum_{i=a_t+1}^{b_t} (\epsilon_i - p_i) \right| \leq \left(1 - \frac{d}{n+1}\right) d.$$

Note that $|\sum_{t=1}^{d} \sum_{i=a_t+1}^{b_t} (\epsilon_i - p_i)| \leq \sum_{t=1}^{d} |\sum_{i=a_t+1}^{b_t} (\epsilon_i - p_i)|$, so the form that appears in Theorem 1 is stronger than in the definition of linear discrepancy. Yet, as shown in [1] using the assignment $p_1 = \cdots = p_n = \frac{d}{n+1}$, the upper bound $(1 - \frac{d}{n+1})d$ is sharp even when taking the absolute value of the total rounding error over the entire union of d intervals.

The proof of Theorem 1 is based on an adaptation of the above-mentioned argument of Lovász, and on an auxiliary result which is interesting in its own right, about partitions of a circle. Consider a circle of length one, partitioned into arcs $J_0, J_1, \ldots, J_n$ in cyclic order. (Some of these arcs may have length zero. Indices of arcs are taken modulo $n + 1$.) For each J_k, we look at its length $|J_k|$, the 2-length around J_k defined as $2|J_k| + |J_{k-1}| + |J_{k+1}|$, and in general the d-length around J_k defined as $d|J_k| + \sum_{i=1}^{d-1} (d - i) (|J_{k-i}| + |J_{k+i}|)$, for $d = 1, \ldots, \lfloor \frac{n+1}{2} \rfloor$. Note that the average over all k of the d-length around J_k equals $\frac{d^2}{n+1}$. Hence for each d there is some J_k around which the d-length is at least this average. The nontrivial fact that we shall prove is that there is always a J_k around which *all* d-lengths are at least the respective averages:

Theorem 2. *Let $J_0, J_1, \ldots, J_n$ be a cyclically ordered partition of a circle of length one into arcs. Then there exists k such that for all $d = 1, \ldots, \lfloor \frac{n+1}{2} \rfloor$ we have*

$$d\,|J_k| + \sum_{i=1}^{d-1} (d-i)\,(|J_{k-i}| + |J_{k+i}|) \geq \frac{d^2}{n+1}.$$

In Section 2 we shall derive Theorem 1 from Theorem 2 (this is essentially the adaptation by Bohman and Holzman of the argument of Lovász, but is repeated here for completeness). In Section 3 we shall prove Theorem 2.

2 Proof of Theorem 1

Given the real numbers $p_1, \ldots, p_n \in [0,1]$, we consider a string of length $\sum_{j=1}^n p_j$, with $n+1$ marked points, namely the points at distance $0, p_1, p_1+p_2, \ldots, \sum_{j=1}^n p_j$ from the left endpoint of the string. Now we wrap this string around a circle of length one, and the marked points appear on the circle as the points $\sum_{j=1}^i p_j$ modulo 1, $i = 0, 1, \ldots, n$. These points partition the circle into $n+1$ arcs (connected components), which we denote $J_0, J_1, \ldots, J_n$ in cyclic order (marked points may coincide on the circle, so we allow arcs of length zero). Applying Theorem 2, we find an arc J_k around which the d-length is at least $\frac{d^2}{n+1}$, for all $d = 1, \ldots, \lfloor \frac{n+1}{2} \rfloor$.

Note that for each p_i there is a corresponding piece of the string, that we denote P_i, which has length p_i and lies between the marked points $\sum_{j=1}^{i-1} p_j$ and $\sum_{j=1}^{i} p_j$. After wrapping around the circle, P_i becomes the union of some cyclically consecutive arcs among $J_0, J_1, \ldots, J_n$. We set $\epsilon_i = 1$ if J_k (found above) is one of the consecutive arcs forming P_i, and $\epsilon_i = 0$ otherwise.

We verify that this rounding scheme satisfies the statement of Theorem 1. For an integer interval $\{a_t+1, \ldots, b_t\}$, observe that $|\sum_{i=a_t+1}^{b_t} (\epsilon_i - p_i)|$ equals the difference (in absolute value) between the length of the piece of string $\bigcup_{i=a_t+1}^{b_t} P_i$ and the number of times it wraps around J_k. This difference equals the length of the circular arc between the two endpoints of $\bigcup_{i=a_t+1}^{b_t} P_i$ that *does not* contain J_k. The length of this circular arc is at most $1 - |J_k|$. When we consider d such intervals $\{a_t+1, \ldots, b_t\}$, $t = 1, \ldots, d$, with $0 \leq a_1 < b_1 < \cdots < a_d < b_d \leq n$, the $2d$ endpoints of $\bigcup_{i=a_t+1}^{b_t} P_i$, $t = 1, \ldots, d$, occupy $2d$ distinct marked points on the circle. Thus, J_k is contained in none of the corresponding d circular arcs, $J_{k\pm 1}$ are each contained in at most one of them, $J_{k\pm 2}$ are each contained in at most two of them, etc. Hence the total length of these d circular arcs is at most $d - d|J_k| - \sum_{i=1}^{d-1} (d-i)\,(|J_{k-i}| + |J_{k+i}|)$, which by our choice of k is at most $d - \frac{d^2}{n+1} = (1 - \frac{d}{n+1})d$, as required.

3 Proof of Theorem 2

We first restate Theorem 2 in an equivalent but more convenient form. Instead of working with the lengths $|J_i|$, we work with the excess lengths (compared to the average length), namely

$$e_i = |J_i| - \frac{1}{n+1}, \quad i = 0, 1, \ldots, n.$$

Clearly, the excess lengths satisfy

$$\sum_{i=0}^{n} e_i = 0,$$

and we need to prove that there exists k such that for all $d = 1, \ldots, \left\lfloor \frac{n+1}{2} \right\rfloor$ we have

$$de_k + \sum_{i=1}^{d-1} (d-i)\,(e_{k-i} + e_{k+i}) \geq 0.$$

We recall that the entries $e_0, e_1, \ldots, e_n$ are cyclically ordered, and their indices are taken modulo $n+1$. The circular distance between two indices i and j is denoted by $\|i-j\|$, that is, for $0 \leq i, j \leq n$ we have

$$\|i-j\| = \min\big(|i-j|, n+1-|i-j|\big).$$

Let us assume, for the sake of contradiction, that for each k there exists d_k such that

$$d_k e_k + \sum_{i=1}^{d_k-1} (d_k - i)\,(e_{k-i} + e_{k+i}) < 0.$$

Consider the $(n+1)\times(n+1)$ matrix A, with entries $\big(A_{ij}\big)_{i=0,\ldots,n,\, j=0,\ldots,n}$ defined by

$$A_{ij} = \begin{cases} d_i - \|i-j\| & \text{if } \|i-j\| < d_i, \\ 0 & \text{otherwise.} \end{cases}$$

Our assumption is equivalent to

$$A \begin{pmatrix} e_0 \\ e_1 \\ \cdot \\ \cdot \\ e_n \end{pmatrix} < \begin{pmatrix} 0 \\ 0 \\ \cdot \\ \cdot \\ 0 \end{pmatrix}. \tag{1}$$

Consider also the $(n+1)\times(n+1)$ matrix B, with entries $\big(B_{ij}\big)_{i=0,\ldots,n,\, j=0,\ldots,n}$ defined by

$$B_{ij} = \begin{cases} 2 & \text{if } i = j, \\ -1 & \text{if } \|i-j\| = 1, \\ 0 & \text{otherwise.} \end{cases}$$

We claim that

$$\begin{pmatrix} e_0 \\ e_1 \\ \cdot \\ \cdot \\ e_n \end{pmatrix} \in \operatorname{Im} B. \tag{2}$$

As B is symmetric, its kernel is the subspace orthogonal to its image. We know that $\sum_{i=0}^{n} e_i = 0$, hence it suffices to prove that

$$\operatorname{Ker} B = \operatorname{Span}\left\{ \begin{pmatrix} 1 \\ 1 \\ \cdot \\ \cdot \\ 1 \end{pmatrix} \right\}.$$

Indeed, a vector $\vec{x}$ in $\operatorname{Ker} B$ satisfies $x_i - x_{i-1} = x_{i+1} - x_i$. So its entries form an arithmetic progression, and as $x_{n+1} = x_0$ they must all be equal.

By (2), there exists $\vec{v}$ such that

$$B\vec{v} = \begin{pmatrix} e_0 \\ e_1 \\ \cdot \\ \cdot \\ e_n \end{pmatrix}.$$

Substituting in (1), we get

$$AB\vec{v} < \begin{pmatrix} 0 \\ 0 \\ \cdot \\ \cdot \\ 0 \end{pmatrix}. \tag{3}$$

We proceed to compute the matrix AB. Noting that $A_{ij} = d_i - \min\big(\|i-j\|, d_i\big)$, we have

$$\begin{aligned} \big(AB\big)_{ij} &= -2\min\big(\|i-j\|, d_i\big) + \min\big(\|i-(j-1)\|, d_i\big) + \min\big(\|i-(j+1)\|, d_i\big) \\ &= \begin{cases} 2 & \text{if } i = j, \\ -1 & \text{if } \|i-j\| = d_i < \frac{n+1}{2}, \\ -2 & \text{if } \|i-j\| = d_i = \frac{n+1}{2}, \\ 0 & \text{otherwise.} \end{cases} \end{aligned}$$

Thus, (3) requires that

$$2v_i - v_{i-d_i} - v_{i+d_i} < 0, \quad i = 0, 1, \ldots, n.$$

But this does not hold for $i_0 = \mathrm{argmax}_i v_i$, because

$$2v_{i_0} - v_{i_0 - d_{i_0}} - v_{i_0 + d_{i_0}} \geq 2v_{i_0} - v_{i_0} - v_{i_0} = 0.$$

This contradiction proves Theorem 2.

References

[1] T. Bohman and R. Holzman, Linear versus hereditary discrepancy, *Combinatorica* **25** (2005), 39-47.

[2] B. Doerr, Linear and hereditary discrepancy, *Combinatorics, Probability and Computing* **9** (2000), 349-354.

[3] B. Doerr, Linear discrepancy of totally unimodular matrices, *Combinatorica* **24** (2004), 117-125.

[4] D. E. Knuth, Two-way rounding, *SIAM Journal on Discrete Mathematics* **8** (1995), 281-290.

[5] L. Lovász, J. Spencer and K. Vesztergombi, Discrepancy of set systems and matrices, *European Journal of Combinatorics* **7** (1986), 151-160.

[6] J. Matoušek, On the linear and hereditary discrepancies, *European Journal of Combinatorics* **21** (2000), 519-521.

[7] J. Spencer, *Ten Lectures on the Probabilistic Method*, 2nd edition, SIAM, 1994.