Combinatorics, Probability and Computing (2000) 9, 349–354. Printed in the United Kingdom
© 2000 Cambridge University Press

Linear and Hereditary Discrepancy

BENJAMIN DOERR†

Mathematisches Seminar, Christian-Albrechts-Universität zu Kiel,
Ludewig-Meyn-Str. 4, 24098 Kiel, Germany
(e-mail: bed@numerik.uni-kiel.de)

Received 22 January 1999; revised 10 June 1999

Let A be an $m \times n$ matrix and $q := \lfloor \log_2(m) \rfloor + 1$. In this article we improve the well-known bound $\operatorname{lindisc}(A) \leqslant 2 \operatorname{herdisc}(A)$ and show that

$$\operatorname{lindisc}(A) \leqslant 2\left(1 - 2^{-q}\right) \operatorname{herdisc}(A) \quad \left(\leqslant 2\left(1 - \frac{1}{2m}\right) \operatorname{herdisc}(A)\right).$$

As with the previous proofs relating to this problem, ours is constructive. We will give an on-line algorithm and analyse it using game theory.

Introduction

Discrepancy theory deals with questions of the kind, 'How far does the optimal solution of a problem deviate from an ideal solution?' In combinatorial discrepancy theory our problem is to colour the points of a hypergraph with two colours in such a way that all hyperedges are about evenly coloured. So an ideal solution would be one where every hyperedge has the same number of points of each colour. Let us be more precise.

Let $\mathscr{H} = (X, \mathscr{E})$ denote a finite *hypergraph*, that is, X is a finite set (of '*vertices*') and $\mathscr{E}$ is a family of subsets of X (called *hyperedges*). A *colouring* of $\mathscr{H}$ is simply a mapping $\varepsilon : X \to \{-1, 1\}$ (we call -1 and 1 *colours*). For a hyperedge $E \in \mathscr{E}$ define $\varepsilon(E) = \sum_{x \in E} \varepsilon(x)$. The *discrepancy* of $\mathscr{H}$ is defined by

$$\operatorname{disc}(\mathscr{H}) = \min_{\varepsilon : X \to \{-1,1\}} \max_{E \in \mathscr{E}} |\varepsilon(E)|.$$

This concept may be generalized to matrices in a natural way. Let $A = (a_{ij})$ be any $m \times n$ matrix and set

$$\operatorname{disc}(A) := \min_{\varepsilon \in \{-1,1\}^n} \|A\varepsilon\|,$$

where the norm will always denote the maximum norm.

† Supported by the graduate school 'Effiziente Algorithmen und Multiskalenmethoden', Deutsche Forschungsgemeinschaft.

https://doi.org/10.1017/S0963548300004272 Published online by Cambridge University Press

If A is the incidence matrix of $\mathscr{H}$, we have $\operatorname{disc}(A) = \operatorname{disc}(\mathscr{H})$. Hence discrepancy of matrices is indeed a more general concept (if, of course, we restrict ourselves to zero-one matrices, both concepts are equivalent). For our specific problem the matrix concept provides no extra difficulties, so we shall continue in the language of matrices.

There are two related notions. The *linear discrepancy* of A is defined by

$$\operatorname{lindisc}(A) := \max_{p\in[-1,1]^n} \min_{\varepsilon\in\{-1,1\}^n} \|A(p-\varepsilon)\|.$$

Linear discrepancy can be seen as a measure of how well a fractional solution can be rounded to an integer solution. The *hereditary discrepancy* is

$$\operatorname{herdisc}(A) := \max_{J\subseteq[n]} \operatorname{disc}((a_{ij})_{i\in[m],j\in J}).$$

If A is the incidence matrix of a hypergraph, $\operatorname{herdisc}(A)$ is just the maximum discrepancy of the induced subgraphs. As a general reference we recommend Beck and Sós [1].

It is easy to see that we have both $\operatorname{disc}(A) \leqslant \operatorname{lindisc}(A)$ and $\operatorname{disc}(A) \leqslant \operatorname{herdisc}(A)$ for any A. More interesting is the relation between $\operatorname{lindisc}(A)$ and $\operatorname{herdisc}(A)$. Results of Beck and Spencer [2] and Lovász, Spencer and Vesztergombi [3] show that there is a nontrivial relation between the two concepts: for any A we have $\operatorname{lindisc}(A) \leqslant 2\operatorname{herdisc}(A)$. In [4] Spencer gives an improvement in terms of n: $\operatorname{lindisc}(A) \leqslant 2(1-2^{-2^n})\operatorname{herdisc}(A)$. He also gives an example of an A satisfying $\operatorname{lindisc}(A) = 2(1-\frac{1}{n+1})\operatorname{herdisc}(A)$ and introduces the problem of closing this gap.

Our results

The main result of this article is as follows.

Theorem 1. *Let A be any $m \times n$ matrix. Set $q := \lfloor \log_2(m) \rfloor + 1$. Then*

$$\operatorname{lindisc}(A) \leqslant 2\left(1-2^{-q}\right)\operatorname{herdisc}(A).$$

If A is the incidence matrix of a hypergraph (*i.e.*, $A \in \{0,1\}^{m\times n}$), then we may assume $m \leqslant 2^n - 1$, as any two rows can be assumed to be different, and different from $(0,0,\ldots,0)$. (In the language of hypergraphs this just means that no edge occurs twice and that the empty edge can be ignored.) This yields $2\left(1-2^{-q}\right) \leqslant 2\left(1-2^{-n}\right)$, giving the following result.

Corollary 2. *For $A \in \{0,1\}^{m\times n}$ we have*

$$\operatorname{lindisc}(A) \leqslant 2\left(1-2^{-n}\right)\operatorname{herdisc}(A).$$

So, for hypergraphs in particular, and matrices with $m \ll 2^{2^n}$ in general, our result is a considerable improvement on the known results, the improvement being better as the matrix becomes sparser. Our result is quite close to optimal: the example of [4] has

$$\operatorname{lindisc}(A) \quad = \quad 2\left(1-\frac{1}{m}\right)\operatorname{herdisc}(A),$$

https://doi.org/10.1017/S0963548300004272 Published online by Cambridge University Press

while the theorem gives

$$\operatorname{lindisc}(A) \leqslant 2\left(1-\frac{1}{m+1}\right)\operatorname{herdisc}(A)$$

for $m = 2^l - 1$.

The proof

Our proof uses the original proof of $\operatorname{lindisc}(A) \leqslant 2\operatorname{herdisc}(A)$, which we state here for convenience.

Original proof. Let $p \in [-1,1]^n$. We will construct an $\varepsilon \in \{-1,1\}^n$ such that $\|A(p-\varepsilon)\|$ is small. Set $a_j^{(0)} := \frac{1}{2}(p_j+1) \in [0,1]$ for all $j \in [n]$. As

$$p \mapsto \min_{\varepsilon \in \{-1,1\}^n} \|A(p-\varepsilon)\|$$

is a continuous function and $\{\sum_{i=1}^n x_i 2^{-i} \mid n \in \mathbb{N}, x \in \{0,1\}^n\}$ is dense in $[0,1]$, we may assume that there is a $k \in \mathbb{N}$ such that $a_j^{(0)} 2^k \in \mathbb{Z}$ for all $j \in [n]$. We are going to round the $a^{(0)}$ successively to a vector of shorter binary expansion until we have a 0–1 vector.

Suppose that, for $l \in \{0,\dots,k-1\}$, the $(a_j^{(l)})_{j=1,\dots,n}$ are already defined and satisfy $a_j^{(l)} 2^{k-l} \in \mathbb{Z}$ for all $j \in [n]$. Set $X := \{j \in [n] \mid a_j^{(l)} 2^{k-l} \text{ odd}\}$, the set of all j such that the binary expansion of $a_j^{(l)} 2^{k-l}$ ends in 1 (these are the components of $a^{(l)}$ that need to be rounded). Find $\varepsilon^{(l)} : X \to \{-1,1\}$ such that

$$d_i^{(l)} := \sum_{j \in X} \varepsilon_j^{(l)} a_{ij} \in [-\operatorname{herdisc}(A), \operatorname{herdisc}(A)]$$

for all $i \in [m]$.

Define

$$a_j^{(l+1)} := \begin{cases} a_j^{(l)} + 2^{-(k-l)}\varepsilon_j^{(l)}, & \text{if } j \in X, \\ a_j^{(l)}, & \text{otherwise.} \end{cases}$$

Then we have $a_j^{(l+1)} 2^{k-l-1} \in \mathbb{Z}$ for all $j \in [n]$, and

$$\sum_{j \in [n]} a_{ij}\left(a_j^{(l+1)} - a_j^{(l)}\right) = 2^{-(k-l)} d_i^{(l)}$$

for all $i \in [m]$. This means that we have rounded the $a^{(l)}$ to $a^{(l+1)}$ in such a way that $\|A(a^{(l+1)} - a^{(l)})\|$ is small.

Having defined $a_j^{(l)}$ for all $j \in [n], l \in \{0,\dots,k\}$, we set $\varepsilon_j := 2a_j^{(k)} - 1$ (this is in $\{-1,1\}$), and have

$$\begin{aligned} \|A(\varepsilon - p)\| &= 2\left\|A(a^{(k)} - a^{(0)})\right\| \\ &= 2\left\|\sum_{l \in [k]} A(a^{(l)} - a^{(l-1)})\right\| \\ &\leqslant \left\|\sum_{l \in [k]} 2^{-k+l} d^{(l)}\right\| \\ &\leqslant 2\operatorname{herdisc}(A). \end{aligned}$$

□

https://doi.org/10.1017/S0963548300004272 Published online by Cambridge University Press

The key idea now is based on the following simple observation: if we replace $\varepsilon^{(l)}$ by $-\varepsilon^{(l)}$, we get $-d^{(l)}$ instead of $d^{(l)}$. By choosing signs for the $\varepsilon^{(l)}$, $l \in [k]$ in a clever way and not using the triangle inequality, we improve the above result.

Note that, if we change the sign of one $\varepsilon^{(l)}$, this leads to a different $a^{(l+1)}$ and thus may change all the subsequently determined variables. So we have to decide the signs 'on-line'. This might be described best in the language of games. Consider the following two-player game.

The game. At the start of the game the vector $v \in \mathbb{R}^m$ is zero. One round of the game is: Player A gives a vector $w \in [-1,1]^m$ and Player B then chooses a sign $\delta \in \{-1,1\}$. The vector v is then updated to $v := 0.5v + \delta w$. The game is played for a fixed number k of rounds. Player A aims to maximize $\|v\|$ while Player B wants to keep the norm down. What is the maximum number c that Player A can reach?

This is the game we are playing (as Player B, against the algorithm as Player A) when deciding on the signs of the $\varepsilon^{(l)}$, $l \in \{0,\dots,k-1\}$. In the game we normalized the w to be in $[-1,1]^m$ (while above we have $d_i^{(l)} \in \{-\text{herdisc}(A),\dots,\text{herdisc}(A)\}$), but it is clear that this just changes c to $c\,\text{herdisc}(A)$ as an upper bound. So we have the following general result.

Lemma 3. *If c is the maximum value Player A can reach in the game described above, then* $\text{lindisc}(A) \leqslant c\,\text{herdisc}(A)$. □

We complete the proof by determining this constant c.

Lemma 4. *The maximum value Player A can reach is* $c = 2\left(1 - 2^{-q}\right)$.

Proof. We investigate the following strategy for Player B. Whatever vectors $w^{(1)},\dots,w^{(k-q)}$ Player A chooses in the first $k-q$ rounds, pick $\delta^{(1)},\dots,\delta^{(k-q)} := 1$ (any other choice would do, too). Set $w := \sum_{j=1}^{k-q} 2^{-k+j} w^{(j)}$. Choose the next sign $\delta^{(k-q+1)}$ in such a way that the number of components $i \in [m] =: X_1$ such that $\text{sgn}(w_i)$ and $\text{sgn}(\delta^{(k-q+1)} w_i^{(k-q+1)})$ are different is maximal. Set

$$X_2 := \left\{ i \in [m] \mid \text{sgn}(w_i) = \text{sgn}\left(\delta^{(k-q+1)} w_i^{(k-q+1)}\right) \right\}.$$

Next choose $\delta^{(k-q+2)} \in \{-1,1\}$ such that the number of components $i \in X_2$ such that $\text{sgn}(w_i)$ and $\text{sgn}(\delta^{(k-q+2)} w_i^{(k-q+2)})$ are different is maximal. Set

$$X_3 := \left\{ i \in X_2 \mid \text{sgn}(w_i) = \text{sgn}\left(\delta^{(k-q+2)} w_i^{(k-q+2)}\right) \right\}.$$

Continue in this fashion until $\delta^{(k)}$ and X_q are determined.

Note that

$$|X_{j+1}| \leqslant \left\lfloor \frac{|X_j|}{2} \right\rfloor$$

for all $j \in [q-1]$, which gives $|X_q| < 1$, that is, $X_q = \emptyset$. So, for every component i there is a $j \in \{k-q+1,\dots,k\}$ such that w_i and $\delta^{(q)} w_i^{(q)}$ have different signs. The worst case is

https://doi.org/10.1017/S0963548300004272 Published online by Cambridge University Press

the one where all $\delta^{(j)}w_i^{(j)}$, $j \in \{k-q+1,\ldots,k\}$ are 1 (or -1) and w_i is zero. This gives us

$$\begin{aligned}\left\|\sum_{j=1}^{k} 2^{-k+j}\delta^{(j)}w^{(j)}\right\| &= \left\|\sum_{j=k-q+1}^{k} 2^{-k+j}\delta^{(j)}w^{(j)} + w\right\| \\ &\leqslant \sum_{z=0}^{q-1} 2^{-z} \\ &= 2\left(1-2^{-q}\right).\end{aligned}$$

Player B is unable to do any better, as the following strategy for Player A reveals. Let r denote the biggest power of 2 that is less than or equal to m (so $r = 2^{q-1}$). Choose the first $k-q$ vectors as zero. Choose the last q vectors as follows: components with index greater than r are set to zero (for instance). For an index $i = 1 + \sum_{j=0}^{q-2} x_j 2^j \leqslant r$, $x : \{0,\ldots,q-2\} \to \{0,1\}$ and a $p \in \{0,\ldots,q-1\}$ set $w_i^{(k-p)} := 2x_p - 1$. Here is an example for $m = 5$:

$$\begin{aligned} w^{(j)} &= (0,0,0,0,0) \text{ for } i \in [k-q], \\ w^{(k-2)} &= (-1,-1,-1,-1,0), \\ w^{(k-1)} &= (-1,-1,+1,+1,0), \\ w^{(k)} &= (-1,+1,-1,+1,0). \end{aligned}$$

Whatever signs $\delta^{(j)}, j \in [k]$ are chosen, there will always be a component $i \in [r]$ such that $w_i^{(k-q+1)} = \cdots = w_i^{(k)}$, and thus

$$\sum_{j=1}^{k} 2^{-k+j}w_i^{(j)} = w_i^{(k)}\sum_{j=0}^{q-1} 2^{-j} = w_i^{(k)}2\left(1-2^{-q}\right).$$

This proves Lemma 4 and thus the theorem. □

Discussion

All of the above is constructive in the following sense. Let A and p be given as above. Assume that one has a colouring for every submatrix of discrepancy no greater than herdisc(A) (discrepancy is an NP-complete problem, so we cannot skip this assumption). Then we have an algorithm rounding p to ε in time at most $\mathcal{O}(k\max\{n,m\})$ such that

$$\|A(p-\varepsilon)\| \leqslant 2^{-k}\|A\| + 2\left(1-2^{-q}\right)\text{herdisc}(A),$$

where $\|A\| := \sup_{\|x\|=1}\|Ax\|$ and k is the maximum binary length we use to express the p_i in our algorithm.

There are some ideas which we did not know how to use or have not explored. Remember that the $d_i^{(l)}$ are discrepancies of hyperedges of induced subgraphs under an optimal colouring. We assumed each $d_i^{(l)}$ took the worst possible value in [−herdisc(A), herdisc(A)], but in general an optimal colouring just creates a few badly coloured hyperedges. Furthermore, in general not all induced subgraphs have herdisc(A) as discrepancy.

This last point might also be used in a different ε-choosing strategy: choose the $\varepsilon^{(l)}$ in such a way that the resulting $a^{(l+1)}$ represents a hypergraph with small discrepancy.

https://doi.org/10.1017/S0963548300004272 Published online by Cambridge University Press

Acknowledgements

I would like to thank J. Spencer for pointing out a mistake in Lemma 4, and for some discussion on this topic. It was also he who drew my attention to this problem at the Discrepancy Workshop in Kiel in the summer of 1998.

References

[1] Beck, J. and Sós, V. (1995) Discrepancy theory. In *Handbook of Combinatorics* (R. Graham, M. Grötschel and L. Lovász, eds), Elsevier, Amsterdam, Chapter 26.
[2] Beck, J. and Spencer, J. (1984) Integral approximation sequences. *Math. Programming* **30** 88–98.
[3] Lovász, L., Spencer, J. and Vesztergombi, K. (1986) Discrepancies of set-systems and matrices. *Europ. J. Combin.* **7** 151–160.
[4] Spencer, J. (1987) *Ten Lectures on the Probabilistic Method*, SIAM, Philadelphia.

https://doi.org/10.1017/S0963548300004272 Published online by Cambridge University Press