Electronic Journal of Linear Algebra ISSN 1081-3810
A publication of the International Linear Algebra Society
Volume 10, pp. 77-80, April 2003

ELA
http://math.technion.ac.il/iic/ela

A NOTE ON LINEAR DISCREPANCY*

GEIR DAHL†

Abstract. Close upper and lower bounds on the linear discrepancy of incidence matrices of directed graphs are determined. For such matrices this improves on a bound found in the work of Doerr [Linear discrepancy of basic totally unimodular matrices, The Electronic Journal of Combinatorics, 7:Research Paper 48, 4 pp., 2000].

Key words. Linear discrepancy, Combinatorial matrix theory, Incidence matrices.

AMS subject classifications. 05C50, 15A15.

1. Introduction. The *linear discrepancy* of an $m \times n$ matrix A is defined as

$$\operatorname{lindisc}(A) = \max_{p \in [0,1]^n} \min_{x \in \{0,1\}^n} \|A(p-x)\|_\infty.$$

See [3] for a treatment of linear discrepancy in connection with set-systems. A matrix is *totally unimodular* if the determinant of each square submatrix is -1, 0 or 1. The following theorem was proved in [2].

THEOREM 1.1. *Let A be a totally unimodular $m \times n$ matrix which has at most two nonzeros in every row. Then*

$$\operatorname{lindisc}(A) \le 1 - 1/(n+1).$$

A major subclass of the totally unimodular matrices, which frequently arises in applications, consists of the incidence matrices of digraphs. Consider a digraph $D = (V, E)$ with vertex set V and arc set E, and define $n = |V|$ and $m = |E|$. The associated *incidence matrix* A is the $m \times n$ matrix whose rows correspond to the arcs as follows: the row corresponding to the arc (i, j) has a 1 in column j, a -1 in column i, and zeros in the remaining columns. So, an incidence matrix has two nonzeros in every row. We consider the problem of bounding the linear discrepancy of incidence matrices of digraphs.

The incidence vector χ^S of a subset $S \subseteq V$ is the vector in $\mathbb{R}^V$ whose vth coordinate equals 1 if $v \in S$ and 0 otherwise. By the terms path and cycle we mean simple path and simple cycle. The length of a path or a cycle is its number of edges. We shall identify the spaces $\mathbb{R}^V$ and $\mathbb{R}^n$.

2. The result. Let A be the incidence matrix of a digraph $D = (V, E)$. If $z \in \mathbb{R}^n$ and (i, j) is an arc in D, then the corresponding component $(Az)_{(i,j)}$ of Az is equal to $z_j - z_i$. So $\|A(p-x)\|_\infty = \max_{(i,j)\in E} |(p_j - p_i) - (x_j - x_i)|$ and it follows that

*Received by the editors on 13 March 2003. Accepted for publication on 20 March 2003. Handling Editor: Richard A. Brualdi.

†University of Oslo, Dept. of Mathematics and Dept. of Informatics, P.O. Box 1080 Blindern, 0316 Oslo, Norway (geird@ifi.uio.no).

Electronic Journal of Linear Algebra ISSN 1081-3810
A publication of the International Linear Algebra Society
Volume 10, pp. 77-80, April 2003
ELA
http://math.technion.ac.il/iic/ela

lindisc(A) depends on D only via the associated undirected graph G (so the direction of each arc plays no role).

Note that a simple lower bound on lindisc(A) is 1/2. For, if i and j are adjacent vertices and we define $p_i = 0$ and $p_j = 1/2$, then $|(p_i - p_j) - (x_i - x_j)| \geq 1/2$ for each $x \in \{0,1\}^V$.

We now give our main result.

Theorem 2.1. *Let A be the incidence matrix of a (nontrivial) digraph D, and let G be the undirected graph corresponding to D. Let d (d') be the maximum length of a path (cycle) in G (if no cycle exits, we let $d' = 0$). Finally, define $k(A) = \max\{d-2, d'-1, 2\}$. Then*

$$1 - \frac{1}{k(A)} \leq \text{lindisc}(A) \leq 1 - \frac{1}{k(A)+2}.$$

Proof. We first prove the lower bound. If $d \leq 2$, then $k(A) = 2$ and we get the trivial bound lindisc$(A) \geq 1/2$. So assume that $d \geq 3$. Consider a path in G of length d, say $P : w, v_0, v_1, \ldots, v_{d-2}, w'$. Define $p_w = 1$, $p_{v_j} = j/(d-2)$ for $j = 0, 1, \ldots, d-2$, and $p_{w'} = 0$, and let the remaining components of the vector $p \in \mathbb{R}^V$ be arbitrary. We shall prove that $\min\{\|A(p-x)\|_\infty : x \in \{0,1\}^V\} \geq 1 - 1/(d-2)$. So let $x \in \{0,1\}^V$, and we may assume that $\|A(p-x)\|_\infty < 1$. This implies that $x_w = 1$, $x_{v_0} = 0$, $x_{v_{d-2}} = 1$ and $x_{w'} = 0$. So there must exist a $j \in \{0, 1, \ldots, d-3\}$ with $x_{v_j} = 0$ and $x_{v_{j+1}} = 1$. But $p_{v_{j+1}} - p_{v_j} = 1/(d-2)$ so

$$\|A(p-x)\|_\infty \geq |(p_{v_{j+1}} - p_{v_j}) - (x_{v_{j+1}} - x_{v_j})| = 1 - 1/(d-2)$$

as desired. Next, assume that G contains a cycle of length d', say with vertices $v_0, v_1, \ldots, v_{d'}$. We define p by letting $p_{v_j} = j/(d'-1)$ for $j = 0, 1, \ldots, d'-1$ and using arguments as above we see that $\|A(p-x)\|_\infty \geq 1 - 1/(d'-1)$. This discussion proves that lindisc$(A) \geq 1 - 1/k(A)$.

We now turn to the upper bound. Define $k' = k(A) + 2$. We first reformulate the problem combinatorially.

Claim 1: Let $p \in [0,1]^n$. If $S \subseteq V$ satisfies

$$\text{(2.1)} \quad \begin{array}{lll} \textit{(i)} & |p_j - p_i| \leq 1 - 1/k' & \textit{when } i \textit{ and } j \textit{ are adjacent vertices in the same} \\ & & \textit{set among } S \textit{ and } V \setminus S, \textit{ and} \\ \textit{(ii)} & p_i \leq p_j - 1/k' & \textit{when } i \in V \setminus S \textit{ and } j \in S \textit{ are adjacent,} \end{array}$$

then $\|A(p - \chi^S)\|_\infty \leq 1 - 1/k'$.

Proof of Claim 1: Let $x = \chi^S$ and consider $\Delta_{ij} := |(p_j - p_i) - (x_j - x_i)|$ where i and j are adjacent. If i and j belong to the same set among S and $V \setminus S$, then $\Delta_{ij} = |p_j - p_i| \leq 1 - 1/k'$ as desired. If i and j are in different sets among S and $V \setminus S$, we may assume that $i \in V \setminus S$ and $j \in S$. Then, as $p_i, p_j \in [0,1]$, $\Delta_{ij} = |p_j - p_i - 1| = -p_j + p_i + 1 \leq 1 - 1/k'$ by (ii), and the claim follows.

Electronic Journal of Linear Algebra ISSN 1081-3810
A publication of the International Linear Algebra Society
Volume 10, pp. 77-80, April 2003

Let now $p \in [0,1]^n$ be given. We shall define a set S with properties as in Claim 1, but we need some preparations first. Define

$$E' = \{[i,j] \in E(G) : |p_i - p_j| < 1/k'\}$$

and let $V_1, V_2, \ldots, V_t$ be the partition of V corresponding to the connected components of the subgraph (V, E') of G. From this partition we construct a digraph D^* with vertices $V_1, V_2, \ldots, V_t$ and, for each $r \neq s$, D^* contains an arc (V_r, V_s) if there are vertices $i \in V_r$ and $j \in V_s$ such that $[i,j] \in E$ and $p_i < p_j$. Note that we then have the stronger inequality $p_i \leq p_j - 1/k'$ as $[i,j] \in E(G) \setminus E'$.

We may find a reordering of the vertices of the digraph D^*, say, for notational simplicity,

$$V_1, \ldots, V_{r_1}, \;\; V_{r_1+1}, \ldots, V_{r_2}, \;\; \ldots, \;\; V_{r_{m-1}+1}, \ldots, V_{r_m}$$

so that (i) the strongly connected components of D^* are $S_i := \{V_{r_i+1}, \ldots, V_{r_{i+1}}\}$ $(i = 0, 1, \ldots, m-1)$ where we let $r_0 = 0$ and $r_m = t$, and (ii) each arc of D^* that do not join two vertices in the same strong component has the form (V_r, V_s) where $r < s$. We remark that with this reordering of the vertices the adjacency matrix of D^* is in the Frobenius normal form (see [1]).

An arc (V_r, V_s) in D^* is called *red* if there are adjacent vertices (in G) $i \in V_r$ and $j \in V_s$ with $p_j - p_i > 1 - 1/k'$. Note that, in this case, $p_i < 1/k'$ and $p_j > 1 - 1/k'$. We may now define our set $S \subseteq V$. If there is no red arc, simply let $S = V$. Otherwise, let q be smallest possible such that the strong component S_q contains a vertex V_s with an ingoing red arc, and let

$$S = \cup_{h=q}^{t} S_h.$$

More precisely, S is the union of the vertex sets V_s contained in some S_h $(h \geq q)$. Note that no arc in D^* leaves S due to the ordering of the sets V_r discussed above.

Claim 2: No strong component S_h contains both a red arc entering it and a red arc leaving it.

Proof of Claim 2: Assume, on the contrary, that S_h contains V_r and V_s such that there is a red arc entering V_s and a red arc leaving V_r. Therefore V_r contains a vertex i with $p_i < 1/k'$ and V_s contains a vertex j with $p_j > 1 - 1/k'$. Since V_s and V_r lie in the same strong component S_h of D^*, this graph contains a directed path P' from V_s to V_r. From P' we may construct a path P in G between j and i (so we select endvertices in V for each arc in P', and add paths in the corresponding connected components V_h). Consider the components of the vector p corresponding to the vertices of P. In the endvertices we have $p_j > 1 - 1/k'$ and $p_i < 1/k'$. Consider an edge $[u,v]$ in P that corresponds to an arc in P', where u is closer to j than v is (in P). Then, by the definition of the arcs of D^* (and with appropriate choice of u and v), $p_u < p_v$. For the remaining edges $[u,v]$ of P we have that $|p_u - p_v| < 1/k'$. From this it follows that P must contain at least $k' - 1$ edges. So, if we add to P the two edges corresponding to the two red arcs (these edges must be disjoint), we get a path in G of length at least $k' + 1$. Since $k' + 1 = k(A) + 3 \geq (d-2) + 3 = d + 1$,

Electronic Journal of Linear Algebra ISSN 1081-3810
A publication of the International Linear Algebra Society
Volume 10, pp. 77-80, April 2003

ELA
http://math.technion.ac.il/iic/ela

this contradicts that the maximum length of a path in G is d, and we have proved Claim 2.

Claim 3: $p_i \le p_j - 1/k'$ *when* $i \in V \setminus S$ *and* $j \in S$ *are adjacent in* G.

Proof of Claim 3: Let $i \in V \setminus S$ and $j \in S$ be adjacent in G. By the definition of S, we must have $i \in V_r$ and $j \in V_s$ for some $r < s$. Therefore (V_r, V_s) is an arc in D^* and $p_i < p_j$. But since $[i,j] \in E \setminus E'$, $|p_i - p_j| \ge 1/k'$, it follows that $p_i \le p_j - 1/k'$.

Claim 4: $|p_j - p_i| \le 1 - 1/k'$ *when* i *and* j *are adjacent vertices in the same set among* S *and* $V \setminus S$.

Proof of Claim 4: Let i and j be adjacent vertices in G that belong to the same set among S and $V \setminus S$. We need to consider two cases.

Case 1: $i, j \in V_r$ *for some* r. If $[i,j] \in E'$, then $|p_j - p_i| < 1/k' \le 1 - 1/k'$ as $k' \ge 2$. So, assume that $[i,j] \notin E'$. Since V_r is a connected component of (V, E'), there is a path P in V_r between i and j containing edges of E' only. Say that this path has l edges. Since $|p_r - p_s| < 1/k'$ holds for every edge $[r,s]$ in this path, we conclude that $|p_i - p_j| < l/k'$. But $l \le d' - 1$ as P and the edge $[i,j]$ make up a cycle and the maximum length of a cycle is d'. So $|p_i - p_j| < l/k' \le (d'-1)/k' = d'/k' - 1/k' < 1 - 1/k'$ as $d' < k'$.

Case 2: $i \in V_r$ *and* $j \in V_s$ *where* $r \ne s$. Assume that $|p_j - p_i| > 1 - 1/k'$. We may here assume that $p_j - p_i > 1 - 1/k'$, so (V_r, V_s) is a red arc. Therefore, by the definition of S, $V_s \subseteq S$. Moreover, due to Claim 2, there is no red arc entering the strong component that V_r belongs to. Therefore $V_r \subseteq V \setminus S$. So, $i \notin S$ and $j \in S$. But this contradicts that i and j lie in the same set among S and $V \setminus S$. Therefore $|p_j - p_i| \le 1 - 1/k'$ and Claim 4 follows.

Finally, due to Claim 3 and Claim 4, the vector $x = \chi^S$, satisfies the assumptions (2.1) in Claim 1 and the proof of the theorem is complete. $\square$

We give some concluding remarks:

1. Since $d' \le n = |V|$ and $d \le n-1$, we see that $1 - \frac{1}{k(A)+2} \le 1 - \frac{1}{n+1}$ when $n \ge 3$ (if $n = 2$ we have $\mathrm{lindisc}(A) = 1/2$). Thus, the upper bound on $\mathrm{lindisc}(A)$ in Theorem 2.1 is at least as good as the one given in Theorem 1.1. Clearly, the new bound may be much better.
2. Our proof also contains a polynomial-time algorithm which for given $p \in [0,1]^V$ finds a $(0,1)$-vector x with $\|A(p-x)\|_\infty \le 1 - \frac{1}{k(A)+2}$.
3. In connection with the bounds given in our theorem we also note that the problem of calculating $k(A)$ (with A as input) is *NP*-hard as it corresponds to finding the longest path in a graph.

REFERENCES

[1] R.A. Brualdi and H.J. Ryser. *Combinatorial matrix theory.* Encyclopedia of Mathematics. Cambridge University Press, 1991.
[2] B. Doerr. Linear discrepancy of basic totally unimodular matrices. *The Electronic Journal of Combinatorics*, 7:Research Paper 48, 4 pp., 2000.
[3] H. Peng and C.H. Yan. On the discrepancy of strongly unimodular matrices. *Discrete Mathematics*, 219:223–233, 2000.