On the problem of approximating the number of bases of a matroid

Y. Azar* A. Z. Broder* A. M. Frieze†

March 12, 1996

In this note we consider the problem of counting the number of bases of a matroid. The problem is of practical significance as it contains graph reliability as a special case. This is a #P-Hard problem and the main focus in recent research has been on trying to approximate the number of bases.

The main result of this paper is that it is impossible to get a good approximation in *deterministic* polynomial time if the matroid $\mathcal{M}$ is given to us by an independence or basis oracle. Thus our result has the same flavour as those of Elekes [5] and Bárányi and Füredi [1] on the problem of computing the volume of a convex body given by a membership oracle.

It should be noted that the main thrust of recent work on approximation for #P-Hard problems has been on randomized algorithms, in particular the Markov chain approach initiated by Broder [2]; see Dyer and Frieze [3], Féder and Mihail [6] for examples of this approach to counting matroid bases. It is to be hoped that randomisation can triumph in this case as it does for computing the volume of a convex body – Dyer, Frieze and Kannan [4] or Lovász and Simonovits [8].

We need very little from the theory of matroids, but see Welsh [12] or Oxley [9] for any of the basic definitions we use. Our computational model is that of Robinson and Welsh [11]. We assume that we have an oracle which

*DEC Systems Research Center, 130 Lytton Avenue, Palo Alto, CA 94301, USA. E-mail: azar@src.dec.com, broder@src.dec.com.

†Department of Mathematics, Carnegie Mellon University. A portion of this work was done while the author was visiting DEC Systems Research Center. Supported in part by NSF grant CCR9024935. E-mail: af1p@euler.math.cmu.edu.

answers questions about a specific matroid $\mathcal{M} = (E, \mathcal{B})$ where E denotes the groundset and $\mathcal{B}$ denotes the set of bases of $\mathcal{M}$. Specifically, if $S \subseteq E$ then one *probe* of the oracle will tell us if S is independent and if it is provide a basis $B \in \mathcal{B}$ containing S.

We consider algorithms whose only knowledge of $\mathcal{M}$ is obtained through probes. We call these *Matroid Oracle Algorithms.* Now it is well known that such algorithms can be used to optimise very efficiently and it is always the hope in a matroid problem that there is a Matroid Oracle Algorithm which can be used to solve it. Here we have a negative result, which we express in two ways. In the proof below we assume that n is sufficiently large to justify any of the inequalities used.

Theorem 1 *Let $|E| = n$ and let A be a deterministic oracle algorithm which outputs a number β approximating the number of bases of a given matroid $\mathcal{M} = (E, \mathcal{B})$. Suppose A makes $k = 2^{o(n)}$ probes. Then A can only guarantee that*

$$2^{-\Omega(n/(\log k)^2)}|\mathcal{B}| \le \beta \le 2^{\Omega(n/(\log k)^2)}|\mathcal{B}|.$$

In particular

(a) *If A makes only a polynomial number of probes then its estimate can only be guaranteed accurate to within $2^{\Omega(n/(\log n)^2)}$.*

(b) *Suppose that A always computes an α-approximation β to $|\mathcal{B}|$ where $0 < \alpha$ is a constant. Then in the worst case A requires $2^{\Omega(\sqrt{n})}$ probes.*

Proof: For a finite set X let $X^{(k)}$ denote the set of k-subsets of X.

Fact 1: Suppose $\mathcal{H} = \{H_1, H_2, \ldots, H_p\} \subseteq X^{(k)}$ and

$$|H_i \cap H_j| \le k - 2 \text{ for } 1 \le i < j \le p. \tag{1}$$

Then there exists a matroid with groundset X whose set of bases is precisely $X^{(k)} \setminus \mathcal{H}$. (It is in fact straightforward to check that this collection of sets satisfies the basis axioms of a matroid.). This was observed by Piff and Welsh [10] in their proof that the number of matroids on a ground set of size n is doubly exponential in n.

Fact 2: Using the same notation as Fact 1, if $|X| = m$ then there is a collection $\mathcal{H}$ satisfying (1) with $p \ge \binom{m}{\lfloor m/2 \rfloor}/(2m)$. This was shown by Knuth [7] in a paper that sharpened the lower bound of Piff and Welsh.

Now back to the proof proper. Let $s = \lceil \log_2 k + \frac{3}{2}\log_2\log_2 k + 10 \rceil$ and $r = \lfloor n/s \rfloor$. Partition $E = \{1, 2, \ldots, n\}$ into $E_1 \cup E_2 \cup \cdots \cup E_r$ where $s \le m_i = |E_i| \le s+1$ for $1 \le i \le r$. Let $\mathcal{M}$ be the partition matroid where a set $I \subseteq E$ is independent if and only if

$$|E_i \cap I| \le \lfloor m_i/2 \rfloor \text{ for } 1 \le i \le r. \tag{2}$$

We can now state the policy followed by the oracle: When given input I our oracle will answer NO if (2) fails to hold and otherwise will say YES and provide some $B \supseteq I$ satisfying (2) with equality. (B can be chosen arbitrarily.)

Suppose that our algorithm A makes k positive probes and learns bases $B_1, B_2, \ldots, B_k$. Let $D_{i,j} = E_i \cap B_j$ for $1 \le i \le r$, $1 \le j \le k$, and let $\mathcal{D}_i = \{D_{i,1}, D_{i,2}, \ldots D_{i,k}\}$. Using Fact 2 choose a set $\mathcal{H}_i = \{H_1, H_2, \ldots, H_p\} \subseteq E_i^{(\lfloor m_i/2 \rfloor)}$ with $p \ge \binom{m_i}{\lfloor m_i/2 \rfloor}/(2m_i)$ which satisfies (1). Let $\hat{\mathcal{H}}_i = \mathcal{H} \setminus \mathcal{D}_i$. Using Fact 1 we know that there is a matroid $\mathcal{M}_i$ with groundset E_i which has $E_i^{(\lfloor m_i/2 \rfloor)} \setminus \hat{\mathcal{H}}_i$ as its set of bases.

Now notice that A cannot distinguish between the two matroids $\mathcal{M}$ and the direct sum $\hat{\mathcal{M}} = \bigoplus_{i=1}^r \mathcal{M}_i$, since the oracle gave answers to A's probes consistent with either matroid. But $\mathcal{M}$ has

$$\mu = \prod_{i=1}^r \binom{m_i}{\lfloor m_i/2 \rfloor}$$

bases, and $\hat{\mathcal{M}}$ has at most

$$\hat{\mu} = \prod_{i=1}^r \left(\binom{m_i}{\lfloor m_i/2 \rfloor} \left(1 - \frac{1}{2m_i}\right) + k \right)$$

bases. Thus A can not guarantee to be more accurate than within a factor $\sqrt{\mu/\hat{\mu}}$ of the true number of bases. But

$$\frac{\hat{\mu}}{\mu} = \prod_{i=1}^r \left(1 - \frac{1}{2m_i} + \frac{k}{\binom{m_i}{\lfloor m_i/2 \rfloor}} \right). \tag{3}$$

Now our choice of s implies that $k \le \binom{m_i}{\lfloor m_i/2 \rfloor}/(4m_i)$ and so (3) implies

$$\frac{\hat{\mu}}{\mu} \le \prod_{i=1}^r \left(1 - \frac{1}{4m_i}\right) = e^{-\Omega(r/s)} = e^{-\Omega(n/(\log k)^2)}.$$

This completes our proof. $\square$

One might be tempted to think that one of the main difficulties that A faces is in finding the exact partition of E into $E_1 \cup E_2 \cup \cdots \cup E_r$. This is not so. One can decompose a direct sum of matroids into its components with a polynomial number of probes.

Observe also that at least for part (a) the number of different possible matroids $2^{O(n^{d+1})}$ that the oracle chooses from is "not much larger" than the number $2^{O(n^d)}$ that the algorithm could possibly distinguish between.

Acknowledgement. We are grateful to Dominic Welsh for his pointers to the relevant literature.

References

[1] I. Bárányi and Z. Füredi, *Computing the volume is difficult*, Proceedings of the 18'th Annual ACM Symposium on Theory of Computing (1986) 442-447.

[2] A. Z. Broder, *How hard is it to marry at random? (On the approximation of the permanent)*, Proceedings of the 18'th Annual ACM Symposium on Theory of Computing (1986) 50–58. Erratum in Proceedings of the 20'th Annual ACM Symposium on Theory of Computing (1988), p. 551.

[3] M. E. Dyer and A. M. Frieze, *Random walks, totally unimodular matrices and a randomised dual simplex algorithm*, Proceedings of second IPCO conference, Carnegie Mellon University, (1992) 72-84.

[4] M. E. Dyer, A. M. Frieze and R. Kannan, *A randomised polynomial time algorithm for approximating the volume of convex bodies*, Journal of the Association for Computing Machinery 38 (1991) 1-17.

[5] G. Elekes, *A geometric inequality and the complexity of computing volume*, Discrete and Computational Geometry 1 (1986) 289-292.

[6] T. Féder and M. Mihail, *Balanced matroids*, Proceedings of the 24'th Annual ACM Symposium on Theory of Computing (1992) 26-38.

[7] D. E. Knuth, *The asymptotic number of geometries*, Journal of Combinatorial Theory (A) 16, (1974) 398-400.

[8] L. Lovász and M. Simonovits, *Random walks in a convex body and an improved volume algorithm*, to appear in Random Structures and Algorithms.

[9] J. G. Oxley, Matroid Theory, Oxford University Press, 1992.

[10] M. J. Piff and D. J. A. Welsh, *On the number of combinatorial geometries*, Bulletin of the London Mathematical Society 3, (1971) 55-56.

[11] G. C. Robinson and D. J. A. Welsh, *The computational complexity of matroid algorithms*, Mathematical Proceedings of the Cambridge Philosophical Society 87, (1980) 29-45.

[12] D. J. A Welsh, *Matroid Theory*, Academic Press, 1976.