On the number of matroids

Nikhil Bansal**, Rudi A. Pendavingh, and Jorn G. van der Pol†

Abstract

We consider the problem of determining m_n, the number of matroids on n elements. The best known lower bound on m_n is due to Knuth (1974) who showed that $\log\log m_n$ is at least $n - \frac{3}{2}\log n - O(1)$. On the other hand, Piff (1973) showed that $\log\log m_n \leq n - \log n + \log\log n + O(1)$, and it has been conjectured since that the right answer is perhaps closer to Knuth's bound.

We show that this is indeed the case, and prove an upper bound on $\log\log m_n$ that is within an additive $1 + o(1)$ term of Knuth's lower bound. Our proof is based on using some structural properties of non-bases in a matroid together with some properties of independent sets in the Johnson graph to give a compressed representation of matroids.

1 Introduction

Matroids, introduced by Whitney in his seminal paper [25], are fundamental combinatorial objects and have been extensively studied due to their very close connection to combinatorial optimization, see e.g. [23], and their ability to abstract core notions from areas such as graph theory and linear algebra [15, 20].

There are several ways to define a matroid. Perhaps the most natural one is using the notion of independence. A matroid M is a pair $(E, \mathcal{I})$, where E is the ground set of elements, and $\mathcal{I}$ is a nonempty collection of subsets of E called the independent sets with the following properties:

*Partially supported by the Netherlands Organisation for Scientific Research (NWO) grant 639.022.211.

†Eindhoven University of Technology, P.O. box 513, 5600 MB, Eindhoven, The Netherlands. E-mail: bansal@gmail.com, R.A.Pendavingh@tue.nl, jornvanderpol@gmail.com.

1. Subset property: $A \in \mathcal{I}$ implies $A' \in \mathcal{I}$ for all $A' \subset A$, and

2. Exchange property: If $A, B \in \mathcal{I}$ with $|A| > |B|$, then there exists an element x in $A \setminus B$, such that $B \cup \{x\} \in \mathcal{I}$.

A basic question is: how many distinct matroids can there be on a ground set of n elements? We denote this number by m_n. Clearly, there are 2^n subsets of E and hence at most 2^{2^n} ways to choose $\mathcal{I}$, which gives the trivial upper bound $\log\log m_n \leq n$. Here, and throughout the paper, log denotes the logarithm to the base 2.

This bound is easily improved to $\log\log m_n \leq n - \frac{1}{2}\log n + O(1)$ by focussing on matroids of a fixed *rank*. In a matroid, the maximal independent sets are called *bases*, and by the exchange property all bases of a matroid have the same cardinality. This common cardinality is the rank of the matroid. Let $m_{n,r}$ be the number of matroids of rank r, and note that $m_n = \sum_{r=0}^{n} m_{n,r}$. By the subset property, any matroid of rank r is completely determined by specifying its bases. As there are at most $\binom{n}{r} \leq \binom{n}{\lfloor n/2 \rfloor} = O(2^n/\sqrt{n})$ (call this ℓ) such bases, this gives $m_{n,r} \leq 2^\ell$ and thus

$$\log\log m_n \leq \log\log\left(\sum_{r=0}^{n} m_{n,r}\right) \leq \log\log((n+1)2^\ell) = n - \frac{1}{2}\log n + O(1).$$

In 1973, Piff [21] improved this bound further to $\log\log m_n \leq n - \log n + \log\log n + O(1)$, by observing that a matroid is also completely determined by the closures of its circuits, and using a counting argument to show that there "only" $O(2^n/n)$ such closures (we describe Piff's proof in section 2.5). This is the best upper bound known to date.

In the other direction, the best known lower bound is due to Knuth [14] from 1974, who showed that $\log\log m_n \geq n - \frac{3}{2}\log n - 1$. Knuth's bound is based on an elegant construction of matroids whose *non-bases*[1] satisfy a particular property. Specifically, he constructs a large family of so-called *sparse paving matroids*. These are matroids of rank r, where any two non-bases of size r intersect in at most $r-2$ elements (i.e. their incidence vectors have Hamming distance 4 or more). Such sets of non-bases are precisely the independent sets in the so-called *Johnson graph* $J(n,r)$. This is the graph with vertex set $\binom{[n]}{r}$, in which two vertices are adjacent if and only if their intersection contains $r-1$ elements.

Knuth's bound follows by taking collection of $k = \frac{1}{n}\binom{n}{\lfloor n/2 \rfloor}$ such non-bases, equivalently, an independent set in the graph $J(n, n/2)$ of this size (section 2.4 has an explicit description of this set) and considering the family of size 2^k of sparse paving matroids obtained by taking each possible subset of this family. Thus $m_n \geq s_n \geq 2^k$, where s_n is the number of sparse paving matroids on n elements. This gives the lower bound

$$\log\log m_n \geq \log\log s_n \geq \log k = n - \frac{3}{2}\log n + \frac{1}{2}\log\frac{2}{\pi} - o(1). \qquad (1)$$

[1]For a matroid of rank r, a non-base is a r-subset of the ground set that is dependent.

We explain Knuth's bound in more detail in section 2.4.

Historically, the interest in paving matroids seems to be a response to the publication of the catalog of matroids on at most 8 elements by Blackburn, Crapo, and Higgs [4] in the early 1970's. With reference to such numerical evidence, Crapo and Rota consider it probable that paving matroids "would actually predominate in any asymptotic enumeration of geometries" [8, p.3.17]. In his book "Matroid Theory", Welsh also notes that paving matroids predominate among the small matroids, and puts the question whether this pattern extends to matroids in general as an exercise [24, p.41]. An earlier lower bound on the number of matroids due to Piff and Welsh [22] was also based on a bound on the number of (sparse) paving matroids. Mayhew and Royle recently confirmed that the predominance of sparse paving matroids extends to the matroids on 9 elements [17].

In recent years, (sparse) paving matroids have received attention in relation to a wide variety of matroid topics [12, 9, 19, 5]. These authors all suggest that the class of sparse paving matroids is probably a very substantial subset of all matroids, pointing out Knuth's argument for the lower bound.

Mayhew, Newman, Welsh and Whittle [16] present a very nice collection of conjectures on the asymptotic behavior of matroids. In particular, they conjecture that asymptotically almost every matroid is sparse paving:

Conjecture 1 (Mayhew, Newman, Welsh and Whittle [16]). $\lim_{n\to\infty} \frac{s_n}{m_n} = 1$.

If true, this would imply

Conjecture 2. $\log\log m_n = \log\log s_n + o(1)$.

Note that this is in fact a much weaker statement as $\log\log(\cdot)$ is a very "forgiving" function, e.g. if $m_n = \Omega(ns_n)$ or even if $m_n = \Omega(2^{2^{\sqrt{n}}} s_n)$, then $\frac{s_n}{m_n} \to 0$, while still $\log\log m_n = \log\log s_n + o(1)$.

1.1 Our results

Our main result is a substantial strengthening of the upper bound on m_n. Specifically, we show that

Theorem 1. *The number of matroids m_n on n elements satisfies*

$$\log\log m_n \leq n - \frac{3}{2}\log n + \frac{1}{2}\log\frac{2}{\pi} + 1 + o(1).$$

Combining theorem 1 with Knuth's lower bound (1) on the number of sparse paving matroids s_n, this gives

Corollary 2. $\log\log m_n \leq \log\log s_n + 1 + o(1)$.

Thus, this result comes quite close to conjecture 2, except for the additive $+1$

term. In particular, it implies that the number of matroids is indeed much closer to Knuth's lower bound, and perhaps also lends support to the conjecture that most matroids are indeed sparse paving.

1.2 Our Techniques

The proof of theorem 1 is based on a combination of the following:

1. Techniques for proving refined upper bounds on the total number of independent sets in a graph.
2. Defining a notion of a *local cover* of a matroid, which serves as a short certificate to identify the bases in the neighborhood of an r-set. Combining the local covers for a carefully chosen set of r-sets then serves as a compressed representation of any matroid.

To see the connection to the total number of independent sets, note that any upper bound on m_n is also an upper bound on s_n. As $s_n = s_{n,0} + s_{n,1} + \ldots + s_{n,n}$, where $s_{n,r}$ denotes the number of sparse paving matroids of rank r, and $s_{n,r}$ is precisely the total number of independent sets in the Johnson graph $J(n, r)$, any method to upper bound m_n must also bound the number of such sets.

We first give an overview of each of these two ideas, and then describe how these are combined to prove theorem 1. These ideas are already useful by themselves to improve the currently known bounds on s_n and m_n. In section 3 we show how local covers can be used in a very simple way to obtain the bound

Theorem 3. $\log\log m_n \leq n - \frac{3}{2}\log n + 2\log\log n + O(1)$.

While this bound is weaker than the one in theorem 1, it already improves Piff's upper bound substantially, and matches Knuth's lower bound up to the additive $O(\log\log n)$ term.

Similarly, in section 5 we show how the refined counting technique for independent sets implies

Theorem 4. $\log\log s_n \leq n - \frac{3}{2}\log n + \frac{1}{2}\log\frac{2}{\pi} + 1 + o(1)$.

Previously, the best known upper bound on s_n seems to be $\log\log s_n \leq n - \frac{3}{2}\log n + O(\log\log n)$ [18] (we sketch an argument below).

Finally, we prove theorem 1 in section 6.

Upper-bounding m_n via local covers:

Let $m_{n,r}$ denote the number of matroids of rank r on n elements. As $m_n = m_{n,0} + \ldots + m_{n,n}$, it suffices to bound each $m_{n,r}$ separately. For a matroid of rank r, let us call a collection of *flats* a *flat cover* if it completely describes the matroid by certifying for each r-set whether it is a basis or not.

A related notion is that of a *local cover*: a collection of flats that allows us to identify the bases in the neighborhood of some fixed r-set. Our main observation

is that given any matroid, for every r-set, one can associate to it a local cover consisting of at most r flats. This implies that if we pick any dominating set D in the Johnson graph and list all the local covers for the vertices in D, then this gives a valid flat cover consisting of at most $|D|r$ flats . Together with standard arguments about the existence of small dominating sets in any regular graph, this implies that each matroid $M \in \mathbb{M}_{n,r}$ can be described by a "small" flat cover, which gives the bound in theorem 3.

Upper-bounding s_n via independent sets:

As $s_n = s_{n,0} + s_{n,1} + \ldots + s_{n,n}$ it suffices to bound each of these terms separately and we focus on the case of $r = n/2$, as this term has the largest contribution to s_n. For a graph G, let $i(G)$ denote the number of independent sets in G, and recall that $s_{n,r} = i(J(n,r))$. While it is hard to obtain any reasonable estimate of $i(G)$ for general graphs, it was shown in [18] that

$$\log\log s_n \leq n - \frac{3}{2}\log n + \log\log n + O(1) \tag{2}$$

One may argue this as follows. Let $G = (V, E)$ be a d-regular graph and $-\lambda$ denote the smallest eigenvalue of its adjacency matrix. Then the size of maximum independent set of G is at most $|V|\lambda/(d+\lambda)$ by Hoffmann's bound (see e.g. theorem 3.5.2 of [7], or our corollary 16). Let us denote $\alpha = \lambda/(d+\lambda)$. This implies that

$$i(G) \leq \sum_{j=0}^{\alpha|V|} \binom{|V|}{j}. \tag{3}$$

For the graph $J(n, n/2)$ it is known that $\alpha = (2+o(1))/n$, which implies that the maximum independent set has size at most αN where $N = \binom{n}{n/2}$. Note that this bound is quite good and is within a factor $2 + o(1)$ of the size of the explicit independent set used in Knuth's lower bound. Applying (3) to $J(n, n/2)$ then gives $s_{n,n/2} = \exp(2 + o(1))N \log n/n$, which implies the bound (2). We note that the proof of (2) in [18] is similar, except that there the same bound on the maximal size of an independent set of $J(n, n/2)$ was shown by a combinatorial argument.

It turns out however that counting all the subsets in (3) is rather wasteful and that this bound can be improved. In particular, we show that

Theorem 5. *If G is a d-regular graph on N vertices with smallest eigenvalue $-\lambda$. Then*

$$i(G) \leq \lceil \sigma N \rceil \binom{N}{\lceil \sigma N \rceil} 2^{\alpha N}$$

where $\alpha = \frac{\lambda}{d+\lambda}$ and $\sigma = \frac{\ln(d+1)}{d+\lambda}$.

For the graph $J(n, n/2)$, $\sigma \leq \frac{8 \ln n}{n^2}$ and hence this gives the stronger bound $i(G) \leq 2^{(2+o(1))N/n}$. As $\alpha N = (2+o(1))N/n$ was our bound on the size of the maximum independent set, this bound on $i(G)$ roughly implies that most independent

sets occur as subsets of a few large independent sets of size αN. Using standard bound on the binomial coefficients, this directly implies theorem 4.

Our proof of theorem 5 is based on a procedure for encoding independent sets in a graph that appears in several places in the literature. We remain very close to the description of the procedure as given in Alon, Balogh, Morris, and Samotij [2], see also this paper for detailed references on the earlier uses of the procedure. Compared to [2], we have given a somewhat improved analysis (specifically lemma 14) to obtain a sufficient bound in the parameter range that is of interest to us.

The improved upper bound on m_n:

To obtain the bound in theorem 1, we combine the two ideas above. The main observation is that given a matroid M, if X is a dependent r-set (i.e. a non-basis) in M, then X has a local cover consisting of at most 2 flats (as opposed to up to r flats if X was an arbitrary r-set). Thus if we could construct a flat cover using few such local covers, then we would obtain a much smaller description of a matroid. To this end, we generalize the procedure of Alon et al. [2] for encoding independent graphs to more generally encode flat covers of the kind described above using a few number of bits. This gives the improved bound on $m_{n,r}$ and hence on m_n.

Finally, we remark that the $+1$ additive gap in our upper bound on m_n arises only because of the factor $2 + o(1)$ gap between the known upper and lower bounds on the size of the maximum independent set in the graphs $J(n, r)$ for $r \approx n/2$. It is likely that reducing this gap could lead to improved bounds for m_n. In section 7, we elaborate on this issue a bit further.

2 Preliminaries

2.1 Matroids

As mentioned previously, a matroid M is specified by $M = (E, \mathcal{I})$, where the sets in the collection $\mathcal{I}$ satisfy the independence axioms. The elements of $\mathcal{I}$ are *independent*, the remaining elements of $2^E \setminus \mathcal{I}$ are *dependent*. The set E is the *ground set*, and we say that M is a matroid *on* E. There are various setsystems and functions defined on M that each allow one to distinguish between dependent and independent sets, such as the set of bases, the rank function, the circuits, the closure operator, etc. We define these notions and state some of their basic properties here, but for a detailed account of their interrelations and for proofs we refer to Oxley [20].

A *basis* of M is an inclusionwise maximal independent set of M. By the independence axioms, each basis has the same cardinality. In this paper, we will present matroids as $M = (E, \mathcal{B})$, where $\mathcal{B}$ is the set of bases of M. The following is an alternate characterization of matroids in terms of the basis axioms, which we shall need later. A set $\mathcal{B} \subseteq 2^E$ is the set of bases of a matroid on E if and only if $\mathcal{B} \neq \emptyset$ and $\mathcal{B}$ satisfies the *basis exchange axiom*

$$\forall B, B' \in \mathcal{B}, e \in B \setminus B' \quad \exists f \in B' \setminus B : \quad B - e + f \in \mathcal{B}. \tag{4}$$

Here, we write $X + y := X \cup \{y\}$ and $X - y := X \setminus \{y\}$.

The *rank* of a set $X \subseteq E$ is $r_M(X) := \max\{|I| \mid I \subseteq X,\ I \in \mathcal{I}\}$, i.e. the cardinality of any maximal independent set in X. The rank function is *submodular*:

$$r_M(X \cap Y) + r_M(X \cup Y) \leq r_M(X) + r_M(Y).$$

We write $r(M) := r_M(E)$. Then $r(M)$ is the common cardinality of all bases, the *rank of* M. We say that an r-set X is a *non-basis* if $r_M(X) < r$. Clearly, a matroid of rank r with set of bases $\mathcal{B}$ is also uniquely defined by its set of *non-bases*, $\binom{E}{r} \setminus \mathcal{B}$.

A *circuit* of M is an inclusionwise minimal dependent set of M. We denote the set of circuits of M by $\mathcal{C}(M)$. By definition, each dependent set contains some circuit. We will use that if X is an r-set with $r_M(X) = r(M) - 1$, then it contains a unique circuit $C \subseteq X$.

In M, the *closure* of a set $X \subseteq E$ is the set $\mathrm{cl}_M(X) := \{e \in E \mid r_M(X + e) = r_M(X)\}$. We will often use that $r_M(\mathrm{cl}_M(X)) = r_M(X)$ for any set X, which follows easily from induction and the submodularity of the rank function. A set $F \subseteq E$ is called a *flat* of M if $\mathrm{cl}_M(F) = F$, and $\mathcal{F}(M)$ denotes the set of all flats of M. As $\mathrm{cl}_M(\mathrm{cl}_M(X)) = \mathrm{cl}_M(X)$ for any set X, every closure $\mathrm{cl}_M(X)$ is a flat.

The following simple property of flats will be crucially used in our construction of flat covers: A set $X \subseteq E$ is dependent if and only if there exists a flat F such that $|X \cap F| > r_M(F)$. In other words, F acts as witness that X contains a dependency when restricted to F.

The *dual* of M is the matroid M^* whose bases are $\mathcal{B}^* = \{E \setminus B \mid B \in \mathcal{B}\}$. The bases, circuits, rank, and closure of sets in M^* are called the cobases, cocircuits, corank, and coclosure of sets in M, and we write $r^*_M(X) := r_{M^*}(X)$, $\mathcal{C}^*(M) := \mathcal{C}(M^*)$, $\mathrm{cl}^*_M := \mathrm{cl}_{M^*}$, etc.

The rank and corank functions of M are related by

$$r^*_M(X) = r_M(E \setminus X) - r(M) + |X|. \tag{5}$$

We write

$$\mathbb{M}_n := \{M \text{ a matroid} \mid E(M) = \{1, \ldots, n\}\},\ \mathbb{M}_{n,r} := \{M \in \mathbb{M}_n \mid\ r(M) = r\}.$$

Also, we put $m_n := |\mathbb{M}_n|, \quad m_{n,r} := |\mathbb{M}_{n,r}|$.

A matroid M is *paving* if $|C| \geq r(M)$ for each circuit C of M (or equivalently if there is no dependent set of size $< r(M)$), and *sparse* if M^* is paving. M is said to be *sparse paving* if it is both sparse and paving. We write

$$s_n := |\{M \in \mathbb{M}_n \mid M \text{ is sparse paving}\}|, \qquad s_{n,r} := |\{M \in \mathbb{M}_{n,r} \mid M \text{ is sparse paving}\}|.$$

2.2 Bounds on binomial coefficients

We will frequently use the following standard bounds.

$$\binom{n}{r} \leq \left(\frac{\mathrm{e}n}{r}\right)^r \tag{6}$$

$$\frac{2^n}{\sqrt{n}}\left(\sqrt{\frac{2}{\pi}} - o(1)\right) \leq \binom{n}{\lfloor \frac{n}{2} \rfloor} \leq \frac{2^n}{\sqrt{n}}\sqrt{\frac{2}{\pi}} \tag{7}$$

We will also use the following bounds on the sum of binomial coefficients.

$$\sum_{i=0}^{k} \binom{n}{i} \leq \frac{n-(k-1)}{n-(2k-1)} \binom{n}{k}. \tag{8}$$

In particular if $k = o(n)$, then

$$\sum_{i=0}^{k} \binom{n}{i} \leq (1+o(1)) \binom{n}{k}. \tag{9}$$

2.3 The Johnson graph

If E is a finite set and $r \leq |E|$, then we write

$$\binom{E}{r} := \{X \subseteq E \mid |X| = r\}$$

for the collection of r-subsets of E. We say that $X, Y \in \binom{E}{r}$ are adjacent (notation: $X \sim Y$) if they have Hamming distance $|X \triangle Y| = 2$ (or equivalently: $|X \cap Y| = r-1$). The *Johnson graph* $J(E,r)$ is defined as the graph with vertex set $\binom{E}{r}$, in which two vertices X and Y are adjacent if and only if $X \sim Y$. We abbreviate $J(n,r) := J([n],r)$. For any r-set $X \in \binom{E}{r}$, we write $N(X) := \{Y \in \binom{E}{r} \mid X \sim Y\}$ for the neighborhood of X in $J(E,r)$. Obviously, $J(E,r) \cong J(n,r)$ for any n-set E.

The following lemma points out the connection between the Johnson graph and sparse paving matroids. It was essentially shown by Piff and Welsh [22] in proving an earlier lower bound on s_n.

Lemma 6. *For $0 < r < n$, sparse paving matroids $M \in \mathbb{M}_{n,r}$ correspond one-to-one to independent sets in $J(n,r)$.*

Proof. Let $E = [n]$. We first show that the non-bases of a rank-r sparse paving matroid M on E form an independent set in $J(E,r)$. Suppose that there are non-bases $X, Y \in \binom{E}{r} \setminus \mathcal{B}(M)$ such that $X \sim Y$, then we would have

$$r_M(X \cap Y) + r_M(X \cup Y) \leq r_M(X) + r_M(Y) < 2r - 1,$$

so that either $r_M(X \cap Y) < r - 1 = |X \cap Y|$ or $r_M(X \cup Y) < r$. In the former case, $X \cap Y$ is a dependent set of size $< r(M)$, which contradicts that M is paving. In the latter case, it follows from (5) that

$$\begin{aligned} r^*_M(E \setminus (X \cup Y)) &= r_M(X \cup Y) - r(M) + |E \setminus (X \cup Y)| \\ &< r - r + |E \setminus (X \cup Y)| = n - r - 1 = r^*(M) - 1, \end{aligned}$$

so that $E \setminus (X \cup Y)$ is a dependent set of M^* of size $< r(M^*)$, which contradicts that M^* is paving.

Next, suppose that I is an independent set in $J(E,r)$. We will show that $\mathcal{B} := \binom{E}{r} \setminus I$ forms a valid collection of bases for some matroid on E.

First, it cannot be that $\mathcal{B} = \emptyset$ as this would imply that $I = \binom{E}{r}$ and hence that $J(E,r)$ has no edges. So the only way $\mathcal{B}$ may fail to be a basis is if it fails the basis exchange axiom (4). That is, there are distinct $B, B' \in \mathcal{B}$ and an $e \in B \setminus B'$ such that $B - e + f \notin \mathcal{B}$ for all $f \in B' \setminus B$. Now, it must be that $|B' \setminus B| > 1$, for otherwise it holds that $B - e + f = B' \in \mathcal{B}$ for the only $f \in B' \setminus B$. So, let f, f' be distinct elements of $B' \setminus B$, and consider $N = B - e + f$ and $N' = B - e + f'$. Since the base exchange axiom fails both $N, N' \in I$. On the other hand, $|N \triangle N'| = |\{f, f'\}| = 2$, i.e. $N \sim N'$, contradicting independence of I. ∎

2.4 Knuth's lower bound

In [14], Knuth argues that if $J(n,r)$ has an independent set I of size k, then $J(n,r)$ has at least 2^k independent sets, as each subset of I is itself independent. Knuth constructed an independent set of size $k = \frac{1}{2n}\binom{n}{r}$, but theorem 1 in [10] shows the existence of an independent set of size at least $k = \frac{1}{n}\binom{n}{r}$.

We sketch the construction in [10]. Identifying the vertices of $J(n,r)$ with their incidence vectors, we view them as $\{0,1\}$ vectors $(x_1, \ldots, x_n)$ with exactly r 1's. It is easily verified that the functional $\{0,1\}^n \to \mathbb{Z}/n\mathbb{Z}$, defined by

$$(x_1, x_2, \ldots, x_n) \mapsto \sum_{i=1}^{n} i x_i \mod n$$

gives a valid n-vertex-coloring of $J(n,r)$. As there are n color classes, at least one of them should contain at least $\frac{1}{n}\binom{n}{r}$ vertices.

Picking such an independent set gives $\log(s_{n,r}) \geq \frac{1}{n}\binom{n}{r}$, and in particular $\log(s_n) \geq \log(s_{n,\lfloor n/2 \rfloor}) \geq \frac{2^n}{n\sqrt{n}}\left(\sqrt{\frac{2}{\pi}} - o(1)\right)$ by (7). Therefore,

$$\log\log s_n \geq n - \frac{3}{2}\log n + \frac{1}{2}\log\frac{2}{\pi} - o(1). \tag{10}$$

2.5 Piff's upper bound

To prove his upper bound on m_n, Piff uses that any matroid M is characterized by the set of all closures of circuits and their ranks, i.e. by the collection

$$\mathcal{K}(M) := \{(\mathrm{cl}_M(C), r_M(C)) \mid C \text{ a circuit of } M\}. \tag{11}$$

This completely defines M as a set $X \subseteq E(M)$ is dependent in M if and only if $|X \cap \mathrm{cl}_M(C)| > r_M(C)$ for some circuit C of M. He then uses the following counting argument to bound the size of $\mathcal{K}(M)$.

Lemma 7. *If $M \in \mathbb{M}_n$, then $|\mathcal{K}(M)| \leq \frac{1}{n+1}2^{n+1}$.*

Proof. Fix an $i < n$. Let $C \in \mathcal{C}(M)$ be a circuit such that $|C| = i+1$. Then for each $e \in C$, we have $\mathrm{cl}_M(C) = \mathrm{cl}_M(C-e)$, i.e. there are $i+1$ sets $C - e \in \binom{E}{i}$ that map to $\mathrm{cl}_M(C)$. It follows that

$$|\{(\mathrm{cl}_M(C), r) \in \mathcal{K}(M) \mid r = i\}| \leq \frac{1}{i+1}\binom{n}{i} = \frac{1}{n+1}\binom{n+1}{i}.$$

Summing these upper bounds over all i, we get

$$|\mathcal{K}(M)| = \sum_{i=0}^{n-1} |\{(\mathrm{cl}_M(C), r) \in \mathcal{K}(M) \mid r = i\}| \leq \sum_{i=0}^{n-1} \frac{1}{n+1}\binom{n+1}{i} \leq \frac{1}{n+1}2^{n+1}.$$

It follows that the number of matroids on a set E of n elements is at most the number of subsets $\mathcal{K} \subseteq 2^E \times \{0, \ldots, n\}$ with $|\mathcal{K}| \leq \frac{1}{n+1}2^{n+1}$, so $m_n \leq \sum_{i=0}^{\frac{1}{n+1}2^{n+1}} \binom{2^n(n+1)}{i}$. Taking logarithms and using first (9) and then (6), we obtain

$$\log m_n \leq \log\left((1+o(1))\binom{2^n(n+1)}{\frac{1}{n+1}2^{n+1}}\right) \leq \frac{1}{n+1}2^{n+1}\log\frac{\mathrm{e}(n+1)^2}{2} + o(1)$$

and hence $\log\log m_n \leq n - \log n + \log\log n + O(1)$.

3 A weaker upper bound on the number of matroids

In this section, we introduce the notion of flat covers and local covers and use them to show that each matroid in $\mathbb{M}_{n,r}$ has a concise description. Using this, we then bound $m_{n,r}$.

Definition 8 (Flat cover). *Let $M = (E, \mathcal{B})$ be a matroid with n elements, of rank r. For a set $X \subseteq E$, we say that a flat $F \in \mathcal{F}(M)$* covers *X if $|F \cap X| > r_M(F)$. We say that a set of flats $\mathcal{Z}$ is a* flat cover *of M if each non-base $X \in \binom{E}{r} \setminus \mathcal{B}$ is covered by some $F \in \mathcal{Z}$.*

Note that if $\mathcal{Z}$ covers M, then M is characterized by E, r and the collection

$$\{(F, r_M(F)) \mid F \in \mathcal{Z}\},$$

since by definition of a cover, we have $\mathcal{B} = \{X \in \binom{E}{r} \mid |X \cap F| \leq r_M(F) \text{ for all } F \in \mathcal{Z}\}$.

Definition 9 (Local cover). *For a r-set $X \in \binom{E}{r}$, we say that a collection of flats $\mathcal{Z}_X \subseteq \mathcal{F}(M)$ is a* local cover *at X if $\mathcal{Z}_X$ covers all the non-bases $Y \in (N(X) \cup \{X\}) \setminus \mathcal{B}$.*

Lemma 10. *Let $M \in \mathbb{M}_{n,r}$. For each r-set $X \in \binom{E}{r}$, there is a local cover $\mathcal{Z}_X$ such that $|\mathcal{Z}_X| \leq r$.*

Proof. Let X be some fixed r-set. Take $\mathcal{Z}_X := \{\mathrm{cl}_M(X-x) \mid x \in X\}$. Then clearly $|\mathcal{Z}_X| \leq r$. We consider a $Y \in N(X) \cup \{X\}$.

If $Y = X$ and X is dependent, then $X \subseteq \mathrm{cl}_M(X - x_0)$ for some $x_0 \in X$. Then $\mathrm{cl}_M(X - x_0)$ covers X, as

$$|X \cap \mathrm{cl}_M(X - x_0)| = |X| = r > r_M(X) \geq r_M(X - x_0) = r_M(\mathrm{cl}_M(X - x_0)).$$

If $Y \in N(X)$, then $Y = X - x + y$ for some $x \in X$ and $y \in E \backslash X$. If $\mathrm{cl}_M(X-x)$ covers Y, we are done. Otherwise,

$$r - 1 = |X - x| \leq |\mathrm{cl}_M(X - x) \cap Y| \leq r_M(\mathrm{cl}_M(X - x)) \leq r - 1$$

so that equality holds throughout, and in particular $r_M(X - x) = r - 1$ and $y \notin \mathrm{cl}_M(X - x)$. It follows that $r_M(Y) = r_M(X - x + y) = r$, so that Y is a basis and it is not required to cover Y. ∎

If $G = (V, E)$ is a graph, then a set $D \subseteq V$ is *dominating* if $D \cup N(D) = V$. The point of introducing local covers is that one can construct a small flat cover from a collection of local covers at the vertices in some small dominating set, as every non-basis in the matroid will be covered by this collection. By standard probabilistic arguments (see theorem 1.2.2 of [3]), one has:

Lemma 11. *$J(n, r)$ has a dominating set of cardinality $\frac{\ln(r(n-r)+1)+1}{r(n-r)+1}\binom{n}{r}$.*

Corollary 12. *Let $M \in \mathbb{M}_{n,r}$. Then M has a flat cover $\mathcal{Z}$ with $|\mathcal{Z}| \leq r\frac{\ln(r(n-r)+1)+1}{r(n-r)+1}\binom{n}{r}$.*

Proof. By lemma 11, $J(n, r)$ has a dominating set D with $|D| \leq \frac{\ln(r(n-r)+1)+1}{r(n-r)+1}\binom{n}{r}$. For each $X \in D$, let $\mathcal{Z}_X$ be a local cover of M as in lemma 10. Then $|Z_X| \leq r$ for each $X \in D$. Take $\mathcal{Z} := \bigcup_{X \in D} \mathcal{Z}_X$. Then $\mathcal{Z}$ is a cover of M, and $|\mathcal{Z}| \leq r|D|$. ∎

Theorem 3. $\log\log m_n \leq n - \frac{3}{2}\log n + 2\log\log n + O(1)$.

Proof. Denote the upper bound in corollary 12 by $k_{n,r} := r\frac{\ln(r(n-r)+1)+1}{r(n-r)+1}\binom{n}{r}$. As each matroid in $\mathbb{M}_{n,r}$ is uniquely determined by the set $\{(F, r_M(F)) \mid F \in \mathcal{Z}\} \subseteq 2^E \times \{0, \ldots, n\}$, where $\mathcal{Z}$ is a cover of size bounded by $k_{n,r}$, the number of matroids in $\mathbb{M}_{n,r}$ is bounded by the number of subsets of a set of size $2^n(n+1)$ of cardinality at most $k_{n,r}$, so by (9) we have $m_{n,r} \leq \binom{2^n(n+1)}{k_{n,r}}(1 + o(1))$.

Now, for any $r \leq n/2$,

$$k_{n,r} \leq k_{n,\lfloor n/2 \rfloor} \leq \frac{4\ln n}{n}\binom{n}{\lfloor n/2 \rfloor}(1 + o(1)).$$

So, by (6) and (7), it follows that for $r \leq n/2$

$$\log m_{n,r} \leq k_{n,\lfloor n/2 \rfloor} \log \frac{\mathrm{e}2^n(n+1)}{k_{n,\lfloor n/2 \rfloor}} \leq \frac{4\ln n}{n}\binom{n}{\lfloor n/2 \rfloor}\frac{5 + o(1)}{2}\log n.$$

The same applies if $r > n/2$, as $m_{n,r} = m_{n,n-r}$ due to matroid duality. As $m_n = \sum_r m_{n,r}$, we have $\log\log m_n \leq n - \frac{3}{2}\log n + 2\log\log n + O(1)$ as required. □

The difference between this upper bound and the lower bound of Knuth is $2\log\log n + O(1)$. A better bound on the cardinality of a dominating set of the Johnson graph could improve this gap to $\log\log n + O(1)$ at best, but applied to bound the number of sparse paving matroids, the above proof is inherently as wasteful as using (3). Note that a minimal cover of a sparse paving matroid just lists the non-bases. We proceed by describing a better technique for bounding the number of sparse paving matroids.

4 A procedure for encoding vertex sets

4.1 The procedure

We describe the procedure by Alon, Balogh, Morris and Samotij [2], for which they refer to Kleitman and Winston [13] as the original source. The procedure was originally developed to encode an independent set I as a pair $(S, I \setminus S)$, such that $S \subseteq I$ and the number of possibilities for both S and $I \setminus S$ can be controlled. We will use it for that purpose in this section as well, but to prepare for other uses in this paper we generalize the procedure so that it takes a general vertex set K and produces a pair (S, A), satisfying

$$S \subseteq K \subseteq S \cup N(S) \cup A. \tag{12}$$

We stress that the encoding is not one to one, and several sets K may produce the same pair (S, A). We will later describe why such a pair (S, A) is useful.

Throughout this section, G is a d-regular graph on N vertices, and the smallest eigenvalue of its adjacency matrix is $-\lambda$. We denote $\alpha := \frac{\lambda}{d+\lambda}$. For a subset $A \subseteq V$, let $G[A]$ denote the subgraph of G induced by A. Let $e(A)$ denote the number of edges with both end points in A, i.e. the number of edges in $G[A]$. Let us assume there is some fixed linear ordering $\leq_V$ of the vertices of G (say according to their indices $1, \ldots, N$). By the canonical ordering of $A \subseteq V$, we refer to the following procedure to order the set A linearly. Let v be the vertex with maximum degree in $G[A]$; if there are multiple such v, take the one that is smallest with respect to $\leq_V$. Call v the first vertex in the canonical ordering, and apply the procedure iteratively to $A \setminus \{v\}$.

The procedure to produce (S, A) (see Figure 4.1) maintains two disjoint sets of vertices: S for selected and A for available. Initially, no vertices are selected ($S = \emptyset$) and all vertices are available ($A = V$). During the procedure, the set S will expand and the set A will shrink, until $|A| \leq \alpha N$. Throughout we will maintain (12) as an invariant.

Algorithm 4.1 The encoding procedure.
Set $A \leftarrow V$ and $S \leftarrow \emptyset$
While $|A| > \alpha N$ **do**
Pick the first vertex v in the canonical ordering of A

If $v \in K$ **do**
set $S \leftarrow S \cup \{v\}$ and set $A \leftarrow A \setminus (N(v) \cup \{v\})$
Else
set $A \leftarrow A \setminus \{v\}$
Output (S, A).

The following is a simple but subtle and crucial observation from [2].

Lemma 13. *Upon the termination of the algorithm, the set A is completely determined by S (irrespective of the set K).*

This follows as at any step in the algorithm, the vertices chosen thus far in S completely determine the remaining vertices and their ordering. In particular, given S one can recover A as follows: Initialize $X = V$ and $T = S$. Repeating the following steps until $|X| \leq \alpha N$ (the resulting set X when the algorithm terminates will be A). (i) Consider the canonical ordering of X, and let v be the first vertex in this ordering. (ii) If $v \in T$, discard v from T and $\{v\} \cup N(v)$ from X and go to step 1. Otherwise, discard v from X and go to step 1.

4.2 Application to counting independent sets

Later we will show that

Lemma 14. *The number of vertices selected into S is at most $\lceil \frac{\ln(d+1)}{d+\lambda} N \rceil$.*

Let us first see how this implies the following upper bound on $i(G)$, the number of independent sets in G.

Theorem 5. *If G is a d-regular graph on N vertices with smallest eigenvalue $-\lambda$. Then*

$$i(G) \leq \lceil \sigma N \rceil \binom{N}{\lceil \sigma N \rceil} 2^{\alpha N}$$

where $\alpha = \frac{\lambda}{d+\lambda}$ and $\sigma = \frac{\ln(d+1)}{d+\lambda}$.

Proof. Let K be any independent set. Running the procedure yields a pair S, A with $|A| \leq \alpha N$, such that (i) A is completely determined by S (by lemma 13) and (ii) $S \subseteq K \subset S \cup N(S) \cup A$. Now, since K is an independent and $S \subseteq K$, we have $N(S) \cap K = \emptyset$. Together with (ii) above, this implies that $(S \cup N(S)) \cap K = S$. Thus, $K = S \cup (K \cap A)$ and hence K is completely determined by S and $K \cap A$.

As A is completely determined by S, for a fixed S, there are at most $2^{\alpha N}$ possibilities for $K \cap A$. Moreover, as $|S| \leq \lceil \sigma N \rceil$, the number of ways of choosing S is at most $\sum_{k=0}^{\lceil \sigma N \rceil} \binom{N}{k} \leq \lceil \sigma N \rceil \binom{N}{\lceil \sigma N \rceil}$. ∎

4.3 Analysis

We now prove lemma 14. We first need the following lemma that was proved by Alon and Chung in [1], and earlier by Haemers (theorem 2.1.4 (i) of [11]). We use the version of the lemma stated in [2].

Lemma 15. *For all $A \subseteq V(G)$, we have $2e(A) \geq |A| \left(\frac{d}{N}|A| - \lambda \frac{N-|A|}{N} \right)$.*

Proof. Let x denote incidence vector of set A, and let B denote the adjacency matrix of G. Then the number of edges is given by $(1/2)x^T Bx$. Let v be the all 1's vector scaled by $|A|/N$, then $v \cdot (x - v) = 0$, and v is an eigenvector of B with eigenvalue d. As B is symmetric its eigenvectors are orthogonal and hence $v \perp (x - v)$ implies that $v \perp B(x - v)$. Thus

$$\begin{aligned} 2e(A) &= x^T Bx = (x - v + v)^T B(x - v + v) \\ &= (x - v)^T B(x - v) + v^T Bv \\ &\geq -\lambda \|x - v\|^2 + d\|v\|^2 \\ &= -\lambda \left(\frac{(N - |A|)|A|^2}{N^2} + \frac{|A|(N - |A|)^2}{N^2} \right) + d\frac{|A|^2}{N} \\ &= |A| \left(\frac{d|A|}{N} - \lambda \frac{(N - |A|)}{N} \right). \end{aligned}$$

□

Corollary 16. *For any $\varepsilon > 0$, if $|A| = (\alpha + \varepsilon)N$, then $G[A]$ contains a vertex of degree at least $\varepsilon(d + \lambda)$.*

Proof. Let A be a vertex set of size $(\alpha + \varepsilon)N$. By lemma 15, $2e(A) \geq |A|(d + \lambda)\varepsilon$. Hence the average degree in $G[A]$ is at least $\varepsilon(d + \lambda)$, and so there must be some vertex in $G[A]$ of degree $\geq \varepsilon(d+\lambda)$. □ In particular, it follows that an independent

set $A \subseteq V(G)$ has size at most αN.

Proof. [Proof of lemma 14] Say that the procedure is in phase j if the current A satisfies

$$\frac{|A|}{N} \in \left(\alpha + \frac{j-1}{d+\lambda}, \alpha + \frac{j}{d+\lambda} \right], \qquad (j = d, d-1, \ldots, 1).$$

Then each phase sees the removal of at most $N/(d+\lambda)$ vertices from A. By corollary 16, any vertex that gets selected into S during phase j has degree $> j - 1$, hence removes at least $j+1$ vertices from A (the vertex selected into S, and its neighbors). Let $S(j)$ be the set of vertices that get selected into S during phase j, then the above argument shows that $|S(j)| \leq \frac{N}{(d+\lambda)(j+1)} + \frac{u_j - u_{j+1}}{j+1}$, where u_j is the fractional

number of vertices that get removed in phase $j-1$ due to the insertion of a vertex in $S(j)$. It follows that

$$|S| = \sum_{j=1}^{d} |S(j)| \leq \frac{N}{d+\lambda} \sum_{j=1}^{d} \frac{1}{j+1} + \sum_{j=1}^{d} \frac{u_j - u_{j+1}}{j+1} < \frac{\ln(d+1)}{d+\lambda} N + 1,$$

as $0 \leq u_j < j+1$, $u_0 = 0$ and $\sum_{i=1}^{k} \frac{1}{k} \leq 1 + \ln k$. The lemma follows. ∎

5 An upper bound on the number of sparse paving matroids

As mentioned previously, sparse paving matroids of rank r on groundset $[n]$ correspond one-to-one with independent sets in the Johnson graph $J(n,r)$. Thus $s_{n,r} = i(J(n,r))$, and we may apply theorem 5 to bound the number of sparse paving matroids. We first investigate the parameters that occur in this application of the theorem.

Recall that $J(n,r)$ is $r(n-r)$-regular and has $\binom{n}{r}$ vertices. The eigenvalues of the adjancency matrix of $J(n,r)$ are $r(n-r) - i(n+1-i)$ for $i = 0, 1, \ldots, r$ (see [6]). This identifies the smallest eigenvalue as $-\lambda_{n,r}$, where

$$\lambda_{n,r} = \begin{cases} \left(\frac{n+1}{2}\right)^2 - r(n-r) & \text{if } n \text{ is odd,} \\ \frac{n}{2}\left(\frac{n}{2}+1\right) - r(n-r) & \text{if } n \text{ is even.} \end{cases} \tag{13}$$

Define

$$\alpha_{n,r} := \frac{\lambda_{n,r}}{r(n-r) + \lambda_{n,r}}.$$

If ε is such that $r = \frac{n}{2}(1+\varepsilon)$, then a straightforward calculation shows that

$$\alpha_{n,r} \leq \frac{2}{n} + \varepsilon^2 = \frac{2}{n} + (\frac{2r}{n} - 1)^2.$$

Lemma 17. $\alpha_{n,r}\binom{n}{r} \leq (\frac{2}{n} + \frac{1}{n^2})\binom{n}{\lfloor n/2 \rfloor}$.

Proof. (Sketch) It can be shown that the maximum of the function $\varepsilon \mapsto (\frac{2}{n} + \varepsilon^2)\binom{n}{\frac{n}{2}(1+\varepsilon)}$ is achieved at $\varepsilon = 0$. In particular, the two terms in the function behave as $\left(\frac{2}{n} + \varepsilon^2\right) \leq \frac{2}{n}(e^{\varepsilon^2 n/2})$ and

$$\binom{n}{n(1+\epsilon)/2} \leq \left(\frac{1}{1+\varepsilon}\right)^{\varepsilon n/2} \binom{n}{n/2} \approx e^{-\epsilon^2 n/2}\binom{n}{n/2},$$

and hence the term $e^{\epsilon^2 n/2}$ essentially cancels out. We omit the details. ∎

This gives us sufficient control over the parameters to prove theorem 4 from theorem 5.

Theorem 4. $\log\log s_n \le n - \frac{3}{2}\log n + \frac{1}{2}\log\frac{2}{\pi} + 1 + o(1)$.

Proof. By theorem 5, we have

$$s_{n,r} = i(J(n,r)) \le \lceil \sigma_{n,r} N \rceil \binom{N}{\lceil \sigma_{n,r} N \rceil} 2^{\alpha_{n,r} N}$$

where $N = \binom{n}{r}$ and $\sigma_{n,r} = \frac{\ln(r(n-r)+1)}{r(n-r)}$. Taking logarithms and applying (6) to the factor $\binom{N}{\lceil \sigma_{n,r} N \rceil}$, we obtain $\log s_{n,r} \le \log N + \lceil \sigma_{n,r} N \rceil \log(n^2) + \alpha_{n,r} N$.

As $\lceil \sigma_{n,r} N \rceil \le \frac{4\ln(n^2)}{n^2}\binom{n}{\lfloor n/2 \rfloor}$, an application of lemma 17 shows that

$$\log s_{n,r} \le \log \binom{n}{\lfloor n/2 \rfloor} + \frac{4\ln(n^2)}{n^2}\binom{n}{\lfloor n/2 \rfloor}\log(n^2) + \left(\frac{2}{n} + \frac{1}{n^2}\right)\binom{n}{\lfloor n/2 \rfloor}.$$

As the latter term is substantially larger than the first two, we have $\log s_{n,r} \le \frac{2}{n}\binom{n}{\lfloor n/2 \rfloor}(1 + o(1))$. As $s_n = \sum_{r=1}^{n} s_{n,r} \le (n+1)\max_r s_{n,r}$, we also have $\log s_n \le \frac{2}{n}\binom{n}{\lfloor n/2 \rfloor}(1 + o(1))$. Taking logarithms and applying (7) to bound the binomial coefficient, the result follows. □

6 An upper bound on the number of matroids

We will now show the upper bound on m_n claimed in theorem 1. To do this, we first show that substantially better local covers at X exist if X is a non-basis. Later, we combine this fact with the encoding procedure in section 4 to find a very concise encoding of a matroid $M \in \mathbb{M}_{n,r}$.

6.1 Improved Local Covers

Lemma 18. *Let $M \in \mathbb{M}_{n,r}$. For each r-set $X \in \binom{E}{r}$ that is dependent in M, there exists a set $\mathcal{Z}_X \subseteq \mathcal{F}(M)$ such that each non-basis $Y \in N(X) \cup \{X\}$ is covered by some $F \in \mathcal{Z}_X$, and $|\mathcal{Z}_X| \le 2$.*

Proof. Let X be some fixed r-set. If $r_M(X) < r - 1$, take $\mathcal{Z}_X := \{\mathrm{cl}_M(X)\}$. Then if $Y \in N(X)$ or $Y = X$, we have

$$|\mathrm{cl}_M(X) \cap Y| \ge |X \cap Y| \ge r - 1 > r_M(X) = r_M(\mathrm{cl}_M(X)).$$

If $r_M(X) = r - 1$, then X contains a unique circuit C of M. Take $\mathcal{Z}_X := \{\mathrm{cl}_M(C), \mathrm{cl}_M(X)\}$. If $Y \in N(X)$ is not a basis, then by submodularity

$$r_M(X \cup Y) + r_M(X \cap Y) \le r_M(X) + r_M(Y) < 2r - 1$$

so that $r_M(X \cup Y) < r$ or $r_M(X \cap Y) < r - 1$. In the former case, we have $Y \subseteq \mathrm{cl}_M(X)$, hence

$$|\mathrm{cl}_M(X) \cap Y| = r > r_M(X \cup Y) \ge r_M(X) = r_M(\mathrm{cl}(X)).$$

In the latter case, $X \cap Y$ is dependent and hence must contain a circuit C', and as C is the unique circuit contained X, we must have $C' = C$. Then $|\mathrm{cl}_M(C) \cap Y| \geq |C| > r_M(C) = r_M(\mathrm{cl}_M(C))$. □

6.2 Matroid Encoding

The crucial difference from lemma 10 is the assumption that X is a dependent set in M. This allows us to obtain a much smaller bound on the size of a cover of M, if we can identify a small collection of non-bases of M such that their neighborhood contains all non-bases in a large fraction of the r-sets. This is exactly what the encoding algorithm will accomplish. We now give the details.

Theorem 1. *The number of matroids m_n on n elements satisfies*

$$\log\log m_n \leq n - \frac{3}{2}\log n + \frac{1}{2}\log\frac{2}{\pi} + 1 + o(1).$$

Proof. Consider a matroid $M \in \mathbb{M}_{n,r}$, and let $K := \binom{E}{r} \setminus \mathcal{B}(M)$ be the set of its non-bases. Then K is a set of vertices of the graph $G = J(n,r)$. As before, let $N = \binom{n}{r}$ denote the number of vertices of G, let $d = r(n-r)$ be its degree, let $\alpha = \frac{\lambda}{d+\lambda}$ and $\sigma = \frac{\ln(d+1)}{d+\lambda}$, where $-\lambda$ is the smallest eigenvalue of G.

We describe how to obtain a concise description of M. Apply the encoding procedure to K and obtain sets S, A such that $|A| \leq \alpha N$, $|S| \leq \sigma N$, A is determined by S, and $S \subseteq K \subseteq S \cup N(S) \cup A$.

By lemma 10, there exists a local cover $\mathcal{Z}_X$ of $(\{X\} \cup N(X)) \setminus \mathcal{B}$ with $|\mathcal{Z}_X| \leq 2$ for each $X \in S$, noting that each such X is a dependent set of M. Then $\mathcal{Z} := \bigcup_{X \in S} \mathcal{Z}_X$ covers all $Y \in (S \cup N(S)) \setminus \mathcal{B}$, and $|\mathcal{Z}| \leq 2|S|$. As all members of $K \setminus A$ lie in $S \cup N(S)$, the set $K \setminus A$ is fully determined by $\{(F, r_M(F)) \mid F \in \mathcal{Z}\}$. For the remaining non-bases in $K \cap A$, we can simply list them. Thus, $(\{(F, r_M(F)) \mid F \in \mathcal{Z}\}, K \cap A)$ gives a complete and concise description of the non-bases in a matroid.

This bounds the number of matroids in $\mathbb{M}_{n,r}$ by the number of ways of choosing S from an N-set, times the number of ways of choosing the collection $\{(F, r_M(F)) \mid F \in \mathcal{Z}\}$ from a set of size $2^n(n+1)$, times the number of possible subsets from A. As $|A| \leq \alpha N$, $|S| \leq \lceil \sigma N \rceil$ and $|\mathcal{Z}| \leq 2|S|$, this yields

$$m_{n,r} \leq \lceil \sigma N \rceil \binom{N}{\lceil \sigma N \rceil} \binom{2^n(n+1)}{2\lceil \sigma N \rceil} 2^{\alpha N}.$$

We have $\lceil \sigma N \rceil \leq \frac{4 \ln n}{n^2} \binom{n}{\lfloor n/2 \rfloor}$, and $\alpha N \leq \frac{2}{n}\binom{n}{\lfloor n/2 \rfloor}(1 + o(1))$ by lemma 17. So the bound on αN dominates in

$$\log m_{n,r} \leq \log(\lceil \sigma N \rceil) + 2\lceil \sigma N \rceil \log \frac{\mathrm{e} 2^n (n+1)}{2\lceil \sigma N \rceil} + \alpha N \leq \frac{2}{n}\binom{n}{\lfloor n/2 \rfloor}(1 + o(1)).$$

Since $m_n = \sum_{r=0}^{n} m_{n,r} \leq (n+1) \max_r m_{n,r}$, we also have

$$\log m_n \leq \frac{2}{n}\binom{n}{\lfloor n/2 \rfloor}(1 + o(1)). \tag{14}$$

Taking logarithms and applying (7) to bound the binomial coefficient, the theorem follows. □

Combining the above theorem with Knuth's lower bound on s_n (10), we obtain:

Corollary 2. $\log\log m_n \leq \log\log s_n + 1 + o(1)$.

7 Further directions

7.1 The maximal independent sets of the Johnson graph

By Knuths lower bound and (14), we have

$$\frac{1}{n}\binom{n}{\lfloor n/2\rfloor} \leq \log s_n \leq \log m_n \leq \alpha_n \binom{n}{\lfloor n/2\rfloor}(1+o(1)),$$

where $\alpha_n \approx \frac{2}{n}$. So asymptotically, there is a factor 2 between lower and upper bound, which turns up as the additive $+1$ term in corollary 2. The lower bound is the size of an independent set in $J(n, \lfloor n/2\rfloor)$ as constructed by Graham and Sloane [10]. As far as we know, the best general upper bound on the size of such an independent set is $\alpha_n\binom{n}{\lfloor n/2\rfloor}$, as a consequence of corollary 16.

It seems that a better understanding of the maximum size of an independent set in $J(n,r)$ could lead to better bounds for m_n. If it could be shown that $J(n, \lfloor n/2\rfloor)$ actually has an independent set of size $\alpha_n\binom{n}{\lfloor n/2\rfloor}$, then the gap of $+1$ in corollary 2 would disappear. On the other hand, if the maximum size of an independent set is at most $\frac{1}{n}\binom{n}{\lfloor n/2\rfloor}$, then a technique to show such an upper bound could potentially be useful for bounding m_n.

7.2 The cover complexity of a matroid

For a matroid M, we define the *cover complexity* as

$$\kappa(M) := \min\{|\mathcal{Z}| \mid \mathcal{Z} \subseteq \mathcal{F}(M), \mathcal{Z} \text{ is a flat cover of } M\}.$$

In [16, Conj. 1.7] it is conjectured that if N is any sparse paving matroid, then

$$\lim_{n\to\infty} \frac{|\{M \in \mathbb{M}_n \mid M \text{ does not have an } N\text{-minor}\}|}{m_n} = 0.$$

In a forthcoming paper, we will show that the conjecture holds for $N = U_{2,k}$ and $N = U_{3,6}$, by deriving bounds on the cover complexity of matroids not having such a minor N. We pose the challenge of bounding

$$\max\{\kappa(M) \mid M \in \mathbb{M}_n, M \text{ does not have an } M(K_4)\text{-minor }\}.$$

Acknowledgements

We thank Andries Brouwer for several useful comments and Dominic Welsh for his help in tracing the origins of the conjecture that 'most matroids are paving'.

Bibliography

[1] N. Alon and F. R. K. Chung. Explicit construction of linear sized tolerant networks. *Discrete Math.*, 72(1-3):15–19, 1988.

[2] Noga Alon, József Balogh, Robert Morris, and Wojciech Samotij. Counting sum-free subsets in abelian groups. (Preprint, arXiv:1201.6654), 2012.

[3] Noga Alon and Joel H. Spencer. *The probabilistic method*. Wiley-Interscience Series in Discrete Mathematics and Optimization. John Wiley & Sons Inc., Hoboken, NJ, third edition, 2008. With an appendix on the life and work of Paul Erdős.

[4] John E. Blackburn, Henry H. Crapo, and Denis A. Higgs. A catalogue of combinatorial geometries. *Math. Comp.*, 27:155–166; addendum, ibid. 27 (1973), no. 121, loose microfiche suppl. A12–G12, 1973.

[5] Joseph E. Bonin. Sparse paving matroids, basis-exchange properties, and cyclic flats. arXiv:1011.1010v1, 2011.

[6] Andries E. Brouwer, Arjeh M. Cohen, and Arnold Neumaier. *Distance-regular graphs*. Springer, 1989.

[7] Andries E. Brouwer and Willem H. Haemers. *Spectra of graphs*. Universitext. Springer, New York, 2012.

[8] Henry H. Crapo and Gian-Carlo Rota. *On the foundations of combinatorial theory: Combinatorial geometries*. The M.I.T. Press, Cambridge, Mass.-London, preliminary edition, 1970.

[9] Jim Geelen and Peter J. Humphries. Rota's basis conjecture for paving matroids. *SIAM J. Discrete Math.*, 20(4):1042–1045 (electronic), 2006.

[10] R. L. Graham and N. J. A. Sloane. Lower bounds for constant weight codes. *IEEE Trans. Inform. Theory*, 26(1):37–43, 1980.

[11] Wilhelmus Hubertus Haemers. *Eigenvalue techniques in design and graph theory*, volume 121 of *Mathematical Centre Tracts*. Mathematisch Centrum, Amsterdam, 1980. Dissertation, Technische Hogeschool Eindhoven, Eindhoven, 1979.

[12] Mark Jerrum. Two remarks concerning balanced matroids. *Combinatorica*, 26(6):733–742, 2006.

[13] Daniel J. Kleitman and Kenneth J. Winston. On the number of graphs without 4-cycles. *Discrete Math.*, 41(2):167–172, 1982.

[14] Donald E. Knuth. The asymptotic number of geometries. *J. Combinatorial Theory Ser. A*, 16:398–400, 1974.

[15] Joseph P. S. Kung. Matroids. In *Handbook of algebra, Vol. 1*, pages 157–184. North-Holland, Amsterdam, 1996.

[16] Dillon Mayhew, Mike Newman, Dominic Welsh, and Geoff Whittle. On the asymptotic proportion of connected matroids. *European J. Combin.*, 32(6):882–890, 2011.

[17] Dillon Mayhew and Gordon F. Royle. Matroids with nine elements. *J. Combin. Theory Ser. B*, 98(2):415–431, 2008.

[18] Dillon Mayhew and Dominic Welsh. On the number of sparse paving matroids. http://homepages.ecs.vuw.ac.nz/ mayhew/Publications/MW.pdf, 2010.

[19] Criel Merino, Steven D. Noble, Marcelino Ramírez-Ibañez, and Rafael Villarroel. On the structure of the h-vector of a paving matroid. arXiv:1008.2031v2, 2010.

[20] James Oxley. *Matroid theory*, volume 21 of *Oxford Graduate Texts in Mathematics*. Oxford University Press, Oxford, second edition, 2011.

[21] M. J. Piff. An upper bound for the number of matroids. *J. Combinatorial Theory Ser. B*, 14:241–245, 1973.

[22] M. J. Piff and D. J. A. Welsh. The number of combinatorial geometries. *Bull. London Math. Soc.*, 3:55–56, 1971.

[23] Alexander Schrijver. *Combinatorial optimization. Polyhedra and efficiency. Vol. B*, volume 24 of *Algorithms and Combinatorics*. Springer-Verlag, Berlin, 2003. Matroids, trees, stable sets, Chapters 39–69.

[24] D. J. A. Welsh. *Matroid theory*. Academic Press [Harcourt Brace Jovanovich Publishers], London, 1976. L. M. S. Monographs, No. 8.

[25] Hassler Whitney. On the Abstract Properties of Linear Dependence. *Amer. J. Math.*, 57(3):509–533, 1935.