On the Power of Additive Combinatorial Search Model

Vladimir Grebinski

INRIA-Lorrainne, 615, rue du Jardin Botanique,
BP 101, 54602 Villers-lès-Nancy, France
Vladimir.Grebinski@inria.fr

Abstract. We consider two generic problems of combinatorial search under the additive model. The first one is the problem of reconstructing bounded–weight vectors. We establish an optimal upper bound and observe that it unifies many known results for coin–weighing problems. The developed technique provides a basis for the graph reconstruction problem. Optimal upper bound is proved for the class of k–degenerate graphs.

1 Introduction

In many practical situations, one needs to obtain some information indirectly available through some physical device. Sometimes this implies costly or lengthy experiments so that the viability of the method crucially depends on the total number of them. Such problems are studied in the field of combinatorics called *combinatorial search*. We refer to monographs [2, 6] for a detailed account of modern methods and results in this area.

Informally, a general combinatorial search problem is described by three parameters: a universe of objects, a set of queries to the oracle and a set of possible answers. Objects are accessible only by the oracle. As every query to the oracle yields some information about the object, we repeat the process until we have enough information in order to uniquely identify the object. Our goal is to minimize the number of queries to the oracle.

One can distinguish two major classes of combinatorial search problems, namely the adaptive and non-adaptive ones. The latter class contains all algorithms which make all queries in advance, before any answer is known. In contrast, an adaptive algorithm takes into account outcomes of previous queries in order to form a next one. The non-adaptive algorithms form a subclass of adaptive ones and they are generally weaker. Surprisingly, in many cases non-adaptive algorithms achieve the power of adaptive ones. This will be the case for our problems.

In this paper we concentrate on two sets of objects. The first one is the set of d-bounded weight vectors $\Omega(n, d)$, which consists of all n–dimensional, non-negative integer–valued vectors of the total weight (sum of components) at most d. The second class is the set $\mathcal{G}_{n,k}$ of k-degenerate graphs on n vertices $v_1, \ldots, v_n$.

Wen-Lian Hsu, Ming-Yang Kao (Eds.): COCOON'98, LNCS 1449, pp. 194–203, 1998.

The definition of k-degenerate graphs is given below. Terms "d-bounded weight vector reconstruction problem" and "k-degenerate graph reconstruction problem" will refer to these two sets respectively.

The set of allowed queries and the set of oracle's answers are crucial for the complexity of the combinatorial search problem. For the set $\Omega(n, d)$, an allowed query is a subset $S \subset \{1, \ldots, n\}$ of vector positions. The answer to such a query S is the sum of entries corresponding to indices in S and will be denoted $\mu_{\boldsymbol{v}}(S)$. That is, if the unknown vector is $\boldsymbol{v} = (a_1, \ldots, a_n)$, then $\mu_{\boldsymbol{v}}(S) = \sum_{i \in S} a_i$. For $\mathcal{G}_{n,k}$, an allowed query is a subset of vertices $Q \subset \{v_1, \ldots, v_n\}$. For a graph $G = (V, E) \in \mathcal{G}_{n,k}$, the answer to the query $Q \subset V$ is the number of edges with both endpoints in Q, we denote $\mu_G(Q) = |(Q \times Q) \cap E|$. Such a choice of queries and answers corresponds to the *additive* or *quantitative model* of combinatorial search.

Historically, the additive model takes roots in a coin-weighing problem, posed by Södenberg and Shapiro in 1963 (see [2]). In this problem there is a finite number of coins, defective and authentic ones. The goal is to find the set of defective coins by possibly minimal number of weighings (or experiments). Each experiment consists in weighing an arbitrary subset of coins which reveals the number of defective ones. The problem was solved by B. Lindström [11], who gave an explicit optimal construction for the set of $\frac{2n}{\log_2 n}$ queries. A probabilistic proof can be found in [7]. This result was extended in several ways. In [10] Lindström obtained an explicit construction of a d-detecting matrix, which provides an optimal reconstruction algorithm for vectors with each entry bounded by d. This construction can be shown to be optimal for the class of non-adaptive algorithms (see [9]). Paper [9] studies the coin–weighing problem where the number of defective coins is bounded by a constant d_0. The upper bound of $4\frac{d_0}{\log d_0} \log n$ was established for the non-adaptive version of this problem. The naive information–theoretic lower bound for non-adaptive algorithms was improved in [3] to $2\frac{d_0}{\log d_0 - c} \log n$ for all $d_0 < n$ and some constant c. Again, this class of objects is a proper subclass of d_0–bounded weight vectors.

To introduce main results of this paper, we point out a connection between coin–weighing and vector reconstruction problems. Namely, assuming all counterfeit coins are heavier, we can associate with every coin its "degree of falsity", that is the difference between the coin weight and the weight of an authentic one. Our goal is to reconstruct the degree of falsity of every coin, i.e. the vector of coin overweights. A weighing of a subset of coins reveals the total overweight which is equal to the sum of corresponding entries of the coin overweights vector. This establishes correspondence between coin–weighing and vector reconstruction problems.

In the first part of this paper we extend previous results in the following direction: we show that an optimal algorithm exists for the problem when only the total overweight is known and the overweight of each individual coin is not bounded. Furthermore, the optimal upper bound can be achieved by a non-adaptive algorithm. This bound is of the same order as for the classical coin-weighing problem where degrees of falsity are restricted to $(0, 1)$ only. Thus, we gain a uniform viewpoint to all previously mentioned results.

In the second part of the paper, we apply the results for bounded–weight vectors to reconstruction of graphs. Reconstruction of graphs covers a broad class of combinatorial search problems. Note that the problem of graph reconstruction is different from that of *verifying* a graph property [2].

In [8, 9] optimal algorithms were proposed for some classes of graphs. For example, it was shown that d-bounded degree graphs have reconstruction complexity $O(dn)$ which can be reached by a non-adaptive algorithm. Another example is provided by general graphs, where the universe of objects is the set of all labeled graphs on n vertices. This class has complexity $O(\frac{n^2}{\log n})$ matched by a non-adaptive algorithm. The same problem was already considered in [1] in a slightly different setting.

While these results already cover many classes of graphs, they all assume some local restriction (except for the extremal case of the class of all graphs). In particular, the maximum degree of a vertex turns out to be the main parameter in complexity bounds. We get rid of this restriction, but require a graph to be k-degenerate (see Definition 2). We prove that for this graph reconstruction problem, the lower and upper bounds asymptotically coincide up to a multiplicative factor. Furthermore, this can be achieved by a non-adaptive algorithm.

Definitions and Conventions

The following notation will be used throughout the paper. We assume implicitly that all graphs are labeled and simple, i.e. without loops or multiple edges. The *weight* of a vector is the sum of its entries, $\mathbf{wt}(\boldsymbol{v}) = \sum_{i=1}^{n} v_i$ if $\boldsymbol{v} = \{v_1, \ldots, v_n\}$. The non-zero positions of a vector represent its *support*, $\mathbf{sp}(\boldsymbol{v}) = \{i | v_i \neq 0\}$. All logarithms are natural unless the base is indicated. Finally, all considered matrices are $(0, 1)$-matrices over the ring of integers.

Throughout the paper we make several assumptions about the range of parameters. In the first part of the paper, we consider only n–dimensional vectors, whose weight is bounded by a $n^{1+\epsilon}$, for an $\epsilon > 0$. This choice excludes the range of values where a trivial construction can be applied. In the second part of this paper we consider only k–degenerate graphs with $k \leq n^{\alpha}$, with $\alpha < 1$, the choice is motivated by similar considerations.

2 Non-adaptive Vector Reconstruction Problem

In this section we give a lower and upper bounds for the complexity of reconstruction of bounded-weight vectors by a non-adaptive algorithm. Recall that a d-bounded weight vector is a vector $\boldsymbol{v} = (v_1, \ldots, v_n)$, with non-negative integer components $v_i \in \{0\} \cup \mathbb{N}$ and $\sum v_i \leq d$. The set of all such vectors will be denoted by $\Omega(n, d)$ or Ω. An algorithm tries to reconstruct a vector from Ω by asking for a sum of entries with indices in a set $S \subset \{1, \ldots, n\}$ which it is free to choose. The complexity measure of the algorithm is the number of queries and will be denoted by $k(n, d)$.

2.1 Separating Matrices and Bounded Weight Vectors

The notion of separating matrix plays a central role in the study of non-adaptive algorithms for coin-weighing problems.

Definition 1. *A matrix $M \in (0,1)^{k\times n}$ with n columns and k rows is called* separating *for a set of vectors V iff the function $\boldsymbol{v} \to M \cdot \boldsymbol{v}$ is injective on V.*

The importance of this notion is due to the following simple observation:

Proposition 1. *Constructing a non-adaptive algorithm for a coin-weighing problem with n coins under the additive model is equivalent to constructing a separating matrix with n columns.*

Indeed, let V be the set of all possible input vectors. Each query can be represented as an incidence $(0,1)$–vector of the objects that are put in the query. Consider the matrix M, whose rows correspond to queries and columns to objects. A crucial observation is that the vector of answers for configuration $\boldsymbol{v}$ coincides with the vector $M \cdot \boldsymbol{v}$ (in the additive model). Since the algorithm must distinguish between different vectors $\boldsymbol{v}_1 \neq \boldsymbol{v}_2$ we have $M \cdot \boldsymbol{v}_1 \neq M \cdot \boldsymbol{v}_2$. Thus, M is a separating matrix for V. On the other hand, given a separating matrix M for a set of vectors V we obtain a non-adaptive algorithm, by treating rows of M as incidence vectors of queries. □

2.2 Lower Bound

In this section we obtain a lower bound using the second moment method [4]. This lower bound is the factor of two away from the upper bound which will be obtained later. The idea of the proof is to consider the set of all vectors of the weight d as a uniform probabilistic space. Then, an estimation of a certain variance will show that the image $M \cdot \boldsymbol{w}$ of at least a half of vectors $\boldsymbol{w} \in \Omega$ belong to a sphere of small radius if $M \in (0,1)^{k\times n}$ is a separating matrix for Ω. We then obtain an estimation of the dimension of the matrix.

Let $\Omega = \Omega(n,d) = \{(d_1,\ldots,d_n) | \sum_{i=1}^{n} d_i = d\}$ be a probabilistic space with uniform distribution (here we consider only vectors of weight exactly d.) The $\mathbb{P}[d_1 = i] = \binom{n+d-2-i}{n-2} / \binom{n+d-1}{n-1}$, and a simple calculation shows that $E[d_i] = \frac{d}{n}$ and $Var[d_i] = \frac{n-1}{n+1} \cdot \frac{(n+d)d}{n^2}$. Consider a random vector $\boldsymbol{w} = (d_1,\ldots,d_n) \in \Omega$, and let $\boldsymbol{v} = M \cdot \boldsymbol{w}$, where $\boldsymbol{v} = (v_1,\ldots,v_k)$. The first step is to estimate $Var[v_i]$. Suppose there are exactly m non-zero entries in i-th line of the matrix M. The symmetric structure of Ω imposes that $Var[v_i] = Var[d_{i_1} + \cdots + d_{i_m}] = Var[d_1 + \cdots + d_m]$. A direct calculation shows that

$$Var[d_1 + \cdots + d_m] = \frac{d(n+d)}{n^2(n+1)} \cdot m \cdot (n-m) \leq \frac{d(n+d)}{n^2(n+1)} \cdot \frac{n^2}{4} \tag{1}$$

Together with the linearity of expectation this gives:

$$E_{\boldsymbol{w}\in\Omega}\left[\sum_{i=1}^{k} (v_i - E[v_i])^2\right] \leq k\frac{d(n+d)}{4(n+1)} \tag{2}$$

From Markov inequality it follows that:

$$\mathbb{P}\left[\sum_{i=1}^{k}(v_i - E[v_i])^2 \leq k\frac{d(n+d)}{2(n+1)}\right] \geq \frac{1}{2} \quad (3)$$

Hence, at least $\frac{1}{2}\binom{n+d-1}{n-1}$ vectors $\boldsymbol{v}$ belong to a k-dimensional sphere of radius $\sqrt{\frac{k\cdot d(n+d)}{2(n+1)}}$. The volume of k-dimensional sphere is known to be $\left(\frac{2c_1\cdot R^2}{k}\right)^{k/2}$, for a constant c_1. Therefore, by volume argument,

$$\left(\frac{c_1 d(n+d)}{(n+1)}\right)^{k/2} \geq \frac{1}{2}\binom{n+d-1}{n-1} \quad (4)$$

From this we obtain:

$$k \geq 2\frac{\min(n-1,d)\log\left(1+\frac{\max(n-1,d)}{\min(n-1,d)}\right)}{\log d + \log(1+\frac{d}{n+1}) + \log c_1} \quad (5)$$

Considering two cases of $d < n-1$ and $d \geq n-1$ and taking into account that $d \leq n^{1+\epsilon}$, we can further simplify the last expression and formulate the result in the following theorem:

Theorem 1. *There exists an absolute constant c, such that for all $n \to \infty$ and $d \leq n^{1+\epsilon}$:*

$$k(n+1,d) \geq 2\frac{\min(n,d)\log\left(1+\frac{\max(n,d)}{\min(n,d)}\right)}{(1+2\epsilon)\log\min(n,d) + c} \quad (6)$$

2.3 Upper Bound for the Vector Reconstruction Problem

In this section we apply the probabilistic method [7, 4] to obtain an upper bound on the dimension of a separating matrix M for the set $\Omega(n,d)$ of d-bounded weight vectors. The general idea is to consider a set of "bad" events, defined by *critical pairs*, and estimate the probability for a uniformly drawn matrix that any of them takes place. When this probability is strictly below 1 there is a matrix where no "bad" events occurs. Thus, we will estimate the dimension of the matrix M.

For two different vectors $\boldsymbol{v}_1, \boldsymbol{v}_2 \in V$ and a matrix M, we define a characteristic function $\chi(\boldsymbol{v}_1, \boldsymbol{v}_2, M)$:

$$\chi(\boldsymbol{v}_1, \boldsymbol{v}_2, M) = \begin{cases} 1 \text{ if } M\boldsymbol{v}_1 = M\boldsymbol{v}_2, \\ 0 \text{ otherwise.} \end{cases}$$

For a matrix M which is *not* a separating matrix for Ω we can find two witness vectors $\boldsymbol{a} = (a_1, \ldots, a_n)$, $\boldsymbol{b} = (b_1, \ldots, b_n)$ that enjoy two additional properties:

1. $\mathbf{sp}(\boldsymbol{a}) \cap \mathbf{sp}(\boldsymbol{b}) = \emptyset$. Otherwise, consider $(\boldsymbol{a}', \boldsymbol{b}')$, where $\boldsymbol{a}' = (a'_1, \ldots, a'_n)$, $\boldsymbol{b}' = (b'_1, \ldots, b'_n)$, $a'_i = a_i - \min(a_i, b_i)$, $b'_i = b_i - \min(a_i, b_i)$. Obviously, $\mathbf{wt}(\boldsymbol{a}') \leq \mathbf{wt}(\boldsymbol{a})$, $\mathbf{wt}(\boldsymbol{b}') \leq \mathbf{wt}(\boldsymbol{b})$ and $M\boldsymbol{a}' = M\boldsymbol{b}'$ when $M\boldsymbol{a} = M\boldsymbol{b}$.

2. $\mathbf{wt}(\boldsymbol{a}) = \mathbf{wt}(\boldsymbol{b})$. This can be insured by adding to M an additional row with all entries equal to 1. We implicitly assume this row is always present in the matrix M.

An ordered pair of vectors $\boldsymbol{v}_1, \boldsymbol{v}_2 \in \Omega(n, d)$ satisfying the two properties above is said to be a *critical* pair. Let $\mathcal{C} = \mathcal{C}(\Omega)$ be the set of all critical pairs. We have

$$\mathbb{P}\left[M \text{ is not separating for } \Omega\right] = \mathbb{P}\left[\bigvee_{(\boldsymbol{v}_1,\boldsymbol{v}_2)\in\mathcal{C}(\Omega)} (\chi(\boldsymbol{v}_1, \boldsymbol{v}_2, M) = 1)\right] \tag{7}$$

We estimate this probability from above:

$$\mathbb{P}\left[\bigvee_{(\boldsymbol{v}_1,\boldsymbol{v}_2)\in\mathcal{C}} (\chi(\boldsymbol{v}_1, \boldsymbol{v}_2, M) = 1)\right] \leq \frac{1}{2} \sum_{(\boldsymbol{v}_1,\boldsymbol{v}_2)\in\mathcal{C}} \mathbb{P}\left[\chi(\boldsymbol{v}_1, \boldsymbol{v}_2, M) = 1\right] \tag{8}$$

From now on we assume the uniform distribution over $k \times n$ matrices M, except for the implicit row of all 1's mentioned above. The idea of obtaining an upper bound is to find the smallest k which makes the above sum smaller than 1. The first step is to obtain an upper bound for $\mathbb{P}\left[\chi(\boldsymbol{v}_1, \boldsymbol{v}_2, M) = 1\right]$.

Lemma 1. *Given a critical pair $(\boldsymbol{v}_1, \boldsymbol{v}_2)$ and M uniformly distributed over $(0, 1)^{k\times n}$*

$$\mathbb{P}_M\left[\chi(\boldsymbol{v}_1, \boldsymbol{v}_2, M) = 1\right] \leq \left(\frac{8}{9 \cdot |\mathbf{sp}(\boldsymbol{v}_1)|}\right)^{k/4} \cdot \left(\frac{8}{9 \cdot |\mathbf{sp}(\boldsymbol{v}_2)|}\right)^{k/4} \tag{9}$$

Proof. Let $\xi_1, \ldots, \xi_n$ be a set of independent random variables with $\mathbb{P}[\xi_i = 0] = \mathbb{P}[\xi_i = 1] = 1/2$. The event $M\boldsymbol{v}_1 = M\boldsymbol{v}_2$ is equivalent to k independent events corresponding to the equality in each row. Therefore,

$$\mathbb{P}\left[M\boldsymbol{v}_1 = M\boldsymbol{v}_2\right] = \mathbb{P}\left[\langle \boldsymbol{s}, \boldsymbol{v}_1\rangle = \langle \boldsymbol{s}, \boldsymbol{v}_2\rangle\right]^k, \tag{10}$$

where $\boldsymbol{s} = (\xi_1, \ldots, \xi_n)$, and $\langle \boldsymbol{s}, \boldsymbol{v}_i\rangle$ is the inner product of $\boldsymbol{s}$ and $\boldsymbol{v}_i$. Since $\mathbf{sp}(\boldsymbol{v}_1) \cap \mathbf{sp}(\boldsymbol{v}_2) = \emptyset$, then $\langle \boldsymbol{s}, \boldsymbol{v}_1\rangle$ and $\langle \boldsymbol{s}, \boldsymbol{v}_2\rangle$ are independent and

$$\mathbb{P}\left[\langle \boldsymbol{s}, \boldsymbol{v}_1\rangle = \langle \boldsymbol{s}, \boldsymbol{v}_2\rangle\right] = \sum_i \mathbb{P}\left[\langle \boldsymbol{s}, \boldsymbol{v}_1\rangle = i\right] \cdot \mathbb{P}\left[\langle \boldsymbol{s}, \boldsymbol{v}_2\rangle = i\right] \leq \tag{11}$$

$$\sqrt{\sum_i \mathbb{P}\left[\langle \boldsymbol{s}, \boldsymbol{v}_1\rangle = i\right]^2} \cdot \sqrt{\sum_i \mathbb{P}\left[\langle \boldsymbol{s}, \boldsymbol{v}_2\rangle = i\right]^2} \tag{12}$$

The sum $\sum_i \mathbb{P}\left[\langle \boldsymbol{s}, \boldsymbol{v}_j\rangle = i\right]^2$, $j = 1, 2$, can be bounded from above by $\max_i \mathbb{P}\left[\langle \boldsymbol{s}, \boldsymbol{v}_j\rangle = i\right]$. Indeed, consider an arbitrary integer-valued random variable ξ and let $p_{max}(\xi) = \max_{i\in\mathbb{Z}} \mathbb{P}[\xi = i]$. Then $\sum_i \mathbb{P}[\xi = i]^2 \leq p_{max}(\xi) \sum_i \mathbb{P}[\xi = i] = p_{max}(\xi)$. Therefore, we can weaken (12) to

$$\mathbb{P}\left[\langle \boldsymbol{s}, \boldsymbol{v}_1\rangle = \langle \boldsymbol{s}, \boldsymbol{v}_2\rangle\right] \leq \sqrt{p_{max}(\langle \boldsymbol{s}, \boldsymbol{v}_1\rangle)} \cdot \sqrt{p_{max}(\langle \boldsymbol{s}, \boldsymbol{v}_2\rangle)} \tag{13}$$

To estimate $p_{max}(\langle \boldsymbol{s}, \boldsymbol{v}\rangle)$ we need the following technical proposition.

Proposition 2. *Let t be a natural number, $a_1, \ldots, a_t > 0$, and $\xi_1, \ldots, \xi_t$ be independent random variables with $\mathbb{P}[\xi_i = 0] = \mathbb{P}[\xi_i = 1] = 1/2$. Then*

1. $2^{-t}\binom{t}{\lfloor t/2 \rfloor} \leq \sqrt{\frac{8}{9t}}$, *for all* $t \geq 1$,
2. $p_{max}(\xi_1 + \cdots + \xi_t) = 2^{-t}\binom{t}{\lfloor t/2 \rfloor}$,
3. $p_{max}(a_1\xi_1 + \cdots + a_t\xi_t) \leq p_{max}(\xi_1 + \cdots + \xi_t)$,

Proof. 1. For big t the inequality easily follows from Stirling formula. The constant was chosen to satisfy the inequality for *all* $t \geq 1$.
2. This is obvious since $\mathbb{P}[\xi_1 + \cdots + \xi_t = i] = 2^{-t}\binom{t}{i} \leq 2^{-t}\binom{t}{\lfloor t/2 \rfloor}$.
3. Let $P^{\max} = p_{max}(a_1\xi_1 + \cdots + a_t\xi_t)$. By definition, there is a value s such that $\mathbb{P}[a_1\xi_1 + \cdots + a_t\xi_t = s] = P^{\max}$. Consider the family $\mathcal{F} = \{A | \sum_{i \in A} a_i = s\}$. Clearly, $card(\mathcal{F}) \cdot 2^{-t} = P^{\max}$. Since $a_i > 0$, $\mathcal{F}$ is a Sperner family of sets, that is there is no two sets $A, B \in \mathcal{F}$ such that $A \subset B$. By Sperner's theorem [4], $card(\mathcal{F}) \leq \binom{t}{\lfloor t/2 \rfloor}$.

We return to the proof of Lemma 1. To bound $p_{max}(\langle \boldsymbol{s}, \boldsymbol{v}_j \rangle)$, $j = 1, 2$, we apply Proposition 2 with $t = \mathbf{sp}(\boldsymbol{v}_j)$. We have

$$p_{max}(\langle \boldsymbol{s}, \boldsymbol{v}_j \rangle) \leq 2^{-t}\begin{pmatrix} t \\ \lfloor t/2 \rfloor \end{pmatrix} \leq \left(\frac{8}{9 \cdot |\mathbf{sp}(\boldsymbol{v}_j)|}\right)^{1/2}$$

By (10), (13), the Lemma follows. □

Let $\mathcal{C}_w = \{(\boldsymbol{v}_1, \boldsymbol{v}_2) | \mathbf{wt}(\boldsymbol{v}_1) = \mathbf{wt}(\boldsymbol{v}_2) = w$ and $\mathbf{sp}(\boldsymbol{v}_1) \cap \mathbf{sp}(\boldsymbol{v}_2) = \emptyset\}$ and rewrite the right-hand side of (8) as

$$\sum_{(\boldsymbol{v}_1, \boldsymbol{v}_2) \in \mathcal{C}} \mathbb{P}[\chi(\boldsymbol{v}_1, \boldsymbol{v}_2, M) = 1] = \sum_{w=1}^{d} \sum_{\mathcal{C}_w} \mathbb{P}[\chi(\boldsymbol{v}_1, \boldsymbol{v}_2, M) = 1] \tag{14}$$

Using Lemma 1, we bound the inner sum for some fixed w.

$$\textstyle\sum_{\mathcal{C}_w} \mathbb{P}[\chi(\boldsymbol{v}_1, \boldsymbol{v}_2, M) = 1] \leq \sum_{\mathcal{C}_w} \left(\frac{8}{9 \cdot |\mathbf{sp}(\boldsymbol{v}_1)|}\right)^{\frac{k}{4}} \cdot \left(\frac{8}{9 \cdot |\mathbf{sp}(\boldsymbol{v}_2)|}\right)^{\frac{k}{4}} \leq \tag{15}$$

$$\left(\sum_{\substack{\boldsymbol{v}_1, \\ \mathbf{wt}(\boldsymbol{v}_1) = w}} \left(\frac{8}{9 \cdot |\mathbf{sp}(\boldsymbol{v}_1)|}\right)^{\frac{k}{4}}\right)^2 = \left(\sum_{s=1}^{w} \sum_{\substack{\boldsymbol{v}_1, \\ \mathbf{wt}(\boldsymbol{v}_1) = w, \\ |\mathbf{sp}(\boldsymbol{v}_1)| = s}} \left(\frac{8}{9s}\right)^{\frac{k}{4}}\right)^2 = \left(\sum_{s=1}^{w} \binom{n}{s}\binom{w-1}{s-1}\left(\frac{8}{9s}\right)^{\frac{k}{4}}\right)^2 \tag{16}$$

The last inequality is obtained by dropping the condition $\mathbf{sp}(\boldsymbol{v}_1) \cap \mathbf{sp}(\boldsymbol{v}_2) = \emptyset$. Next we used the fact that there are $\binom{n}{s}\binom{w-1}{s-1}$ vectors $\boldsymbol{v}_1$ of weight w with $|\mathbf{sp}(\boldsymbol{v}_1)| = s$, which follows from simple combinatorial considerations.

Now we are left with the technical problem of finding a possibly minimal k which makes (16) smaller than $\frac{2}{d}$. This will make (14) smaller than 1 and achieve our goal. Finding such k requires some routine calculations that we omit. The following proposition gives the final result.

Theorem 2. *There exist absolute constants* C_1, C_2, C_3 *such that for all* n, d *there exists a* $k \times n$ *separating matrix for the set of* d*-bounded weight vectors with* $k(n, d)$ *bounded as*

$$k(n, d) \leq \frac{4\min(n, d)\log(C_1 \cdot \max(n, d)/\min(n, d))}{\log\min(n, d) + C_2} + C_3 \log d. \tag{17}$$

Comparing (17) with lower bound (6), we conclude that upper bound (17) is within the factor of $2(1 + 2\epsilon)$ from the lower bound provided that $d < n^{1+\epsilon}$ for our fixed parameter $\epsilon > 0$.

3 Non-adaptive Reconstruction of k-Degenerate Graphs

In this section we study the complexity of non-adaptive algorithms which reconstruct the class of k-degenerate graphs. This class of graphs is large enough to contain k-bounded degree graphs, sums of $k/2$ trees and other interesting structures.

Definition 2. *A graph $G = (V, E)$ is called k-degenerate if there exists an ordering of vertices $V = \{v_1, v_2, \ldots, v_n\}$ such that for every i we have $deg(v_i) \leq k$ in the subgraph induced by the vertices $\{v_i, v_{i+1}, \ldots, v_n\}$.*

The class of k-degenerate graphs on n vertices will be denoted $\mathcal{G}_{n,k}$. For example, every tree is 1-degenerate, planar graphs are 5-degenerate (see [5]). Note that our definition is equivalent to the one in [5]. We mention that k-degenerate graphs are $k+1$-colorable and have at most $n \cdot k - \binom{k+1}{2}$ edges. For other properties of k-degenerate graphs see [5].

Let $\mu_G(X)$ be the query function, i.e. the number of edges of the graph G with endpoints in X. The complexity $c(\mathcal{G})$ of graph reconstruction for a class of graphs $\mathcal{G}$ is the number of queries sufficient to uniquely identify every graph in $\mathcal{G}$.

Theorem 3. *For any constant $\alpha < 1$ there are two constants b_α and c_α such that for all $k \leq n^\alpha$*

$$b_\alpha \leq \frac{c(\mathcal{G}_{n,k})}{nk} \leq c_\alpha \tag{18}$$

We start the proof by establishing the lower bound. Next we reformulate our problem in terms of bipartite graphs and finally apply the techniques developed for bounded weight vectors.

Proof of the lower bound: To establish the information–theoretic lower bound we need to estimate from below the number $N(n, k)$ of k-degenerate graphs with n vertices. To obtain a k-degenerate graph with $m+1$ vertices one can take a k-degenerate graph with m vertices and choose any k vertices to be adjacent to the new vertex v_{m+1}. Since this can be done in $\binom{m}{k}$ ways, we obtain the following estimation

$$N(n+1, k) \geq \prod_{i=k+1}^{n} \binom{i}{k} \geq \prod_{i=1}^{n} \left(\frac{i}{k}\right)^k = \frac{(n!)^k}{k^{nk}} \tag{19}$$

As it was mentioned above, the number of edges in a k-degenerate graph is at most $kn - k(k+1)/2$. From (19), our assumption $k \leq n^\alpha$ and asymptotic $n! \approx (n/e)^n$ we obtain the following information–theoretic lower bound:

$$\log_{k(n+1-\frac{k+1}{2})} N(n+1, k) \geq \log_{nk} \left(\frac{n}{ke}\right)^{nk} = \frac{nk(\log n - \log k - 1)}{\log n + \log k} \geq \frac{1-\alpha}{1+\alpha} nk + o(nk)$$

Therefore, we can set $b_\alpha = \frac{1-\alpha}{1+\alpha}$. □

Proof of the upper bound: In order to prove the upper bound, we reduce our problem to a problem of reconstructing a bipartite graph of special form.

Specifically, we reduce the graph $G = (V, E)$ and query function $\mu(X)$ to a bipartite graph $G' = (V', V'', E')$ and a new query function μ'. Here G' is the bipartite representation of G, i.e. V' and V'' are copies of V, and there is an edge between $v' \in V'$ and $v'' \in V''$ iff $(v', v'') \in E$. The query function $\mu'(X, Y)$ for $X \subset V'$ and $Y \subset V''$ is *defined* to be $\mu'(X, Y) = |E' \cap (X \times Y)|$, the number of edges between X and Y.

Lemma 2. *One query $\mu'(\cdot, \cdot)$ can be evaluated by five queries $\mu(\cdot)$.*

Proof. In [9] it was shown that for arbitrary $X \subset V'$, $Y \subset V''$ one query μ' can be simulated by five queries μ:

$$\mu'(X, Y) = \mu((X \setminus Y) \cup (Y \setminus X)) - 2(\mu(X \setminus Y) + \mu(Y \setminus X)) + \mu(X) + \mu(Y). \quad \square$$

We are going to explicitly describe a family of queries $\mu'_{G'}(X_i, Y_i)$ that reconstruct G' uniquely provided that G' corresponds to a k-degenerate graph G as above. Let $\{Q_j\}_{j=1}^m$ be a family of sets corresponding to rows of a matrix that is separating for the set of k-bounded weight vectors. Theorem 2 states that $m = O(k\frac{\log n}{\log k})$ as $n \to \infty$. Recall that for a given k–bounded weight vector $\boldsymbol{v} = (v_1, \ldots, v_n)$, values $s_j = \sum_{i \in Q_j} v_i$ uniquely define $\boldsymbol{v}$. Let $\{P_i\}_{i=1}^l$ be a family of sets corresponding to rows of a matrix which is separating for the set of $(2nk)$–bounded weight vectors. Theorem 2 implies that $l = O(n\frac{\log k}{\log n})$.

Lemma 3. *Values $\{\mu'_{G'}(P_i, Q_j)\}_{i=1..l}^{j=1..m}$ uniquely identify graph G'.*

Proof. The proof relies on the following essential properties of reconstruction of bounded-weight vectors:
1. For fixed j, the value of $\mu'(\{v_r\}, Q_j)$ can be uniquely reconstructed for all $r = 1 \ldots n$. Indeed, $\sum_{r=1}^n \mu'(\{v_r\}, Q_j) \leq \sum_{r=1}^n \mu'(\{v_r\}, V'') = \mu'(V', V'') = 2\mu_G(V) \leq 2nk$. Consider a vector $\boldsymbol{w} = (w_1, \ldots, w_n)$, where $w_r = \mu'(\{v_r\}, Q_j)$. By the choice of $\{P_i\}$, vector $\boldsymbol{w}$ is uniquely defined by values of the sum $\sum_{r \in P_i} w_r$ for $i = 1 \ldots l$, which are known, since by definition of μ', $\sum_{r \in P_i} w_r = \mu'(P_i, Q_j)$.
2. Fix an order on vertices of $V' = \{v_1, v_2, \ldots, v_n\}$, which is compatible with the definition of k-degenerate graph. Thus $\mu'(\{v_i\}, \{v_{i+1}, \ldots, v_n\}) \leq k$.
3. Consider a vertex $v_1 \in V'$ and vector $\boldsymbol{e} = (e_1, \ldots, e_n)$, where $e_i = \mu'(\{v_1\}, \{v''_i\})$, the incidence vector of v_1 in G'. If one reconstructs $\boldsymbol{e}$ one will find all vertices adjacent to v_1. By Step 3, v_1 has at most k adjacent vertices in V'', so the values $\sum_{k \in Q_j} e_k = \mu'(\{v_1\}, Q_j)$ $(j = 1 \ldots m)$ uniquely define $\boldsymbol{e}$ by the property of $\{Q_j\}$. According to Step 3, the values $s_j = \mu'(\{v_1\}, Q_j)$ can be reconstructed for all $j = 1 \ldots m$, which proves that vector $\boldsymbol{e}$ can be reconstructed and all vertices adjacent to v_1 can be found.
4. To proceed to vertex v_2, we "exclude" vertex v_1 from graph G and update $\mu'(P_i, Q_j)$. This can be done without additional queries due to the additive nature of μ'. Namely, given an edge (v_1, w), we subtract 2 from $\mu'(P_i, Q_j)$ if both v_1 and w belong to P_i and Q_j, we subtract 1 if exactly one of v_1 or w belongs to P_i and the other to Q_j, and we do not change the value if $\{(v_1, w) \cup (w, v_1)\} \cap (P_i \times Q_j) = \emptyset$.
5. We repeat the process for $v_2, v_3, \ldots v_{n-1}$.

6. It is possible that there are several orders on vertices compatible with the definition of k-degenerate graphs. The uniqueness of reconstruction follows from the fact that at the i-th step we reconstruct *exactly* those edges which are adjacent to v_i in the graph. This implies that different graphs have different values $\{\mu'(P_i, Q_j)\}$. □

The total number of queries μ' is $m \cdot l = O(nk)$. The reduction between μ' and μ gives a factor of 5, according to Lemma 2. Thus, Theorem 3 follows. □

4 Open Problems

A plausible conjecture is that the result of Theorem 3 holds for the graphs with a specified number of edges (i.e. $|E| = nk$), but we are unable to prove it with our technique.

5 Acknowledgments

I am grateful to Gregory Kucherov for numerous valuable discussions and remarks about these topics.

References

[1] Alok Aggarwal, Don Coppersmith, and Dan Kleitman. A generalized model for understanding evasiveness. *Information Processing Letters*, 30:205–208, 1989. 196
[2] Martin Aigner. *Combinatorial Search*. John Wiley & Sons, 1988. 194, 195, 196
[3] Noga Alon. Separating matrices. private communication, May 1997. 195
[4] Noga Alon and Joel Spencer. *The Probabilistic Method*. John Wiley & Sons, 1992. 197, 198, 200
[5] Béla Bollobás. *Extremal Graph Theory*. Academic Press, 1978. 201
[6] Ding-Zhu Du and Frank K. Hwang. *Combinatorial Group Testing and its applications*, volume 3 of *Series on applied mathematics*. World Scientific, 1993. 194
[7] Paul Erdős and Joel Spencer. *Probabilistic Methods in Combinatorics*. 1974. 195, 198
[8] Vladimir Grebinski and Gregory Kucherov. Optimal query bounds for reconstructing a hamiltonian cycle in complete graphs. In *Proceedings of the 5th Israeli Symposium on Theory of Computing and Systems*, pages 166–173. IEEE Press, 1997. 196
[9] Vladimir Grebinski and Gregory Kucherov. Optimal reconstruction of graphs under the additive model. In Rainer Burkard and Gerhard Woeginger, editors, *Algorithms – ESA'97*, volume 1284 of *LNCS*, pages 246–258. Springer, 1997. 195, 196, 202
[10] Bernt Lindström. On Möbius functions and a problem in combinatorial number theory. *Canad. Math. Bull.*, 14(4):513–516, 1971. 195
[11] Bernt Lindström. Determining subsets by unramified experiments. In J.N. Srivastava, editor, *A Survey of Statistical Designs and Linear Models*, pages 407–418. North Holland, Amsterdam, 1975. 195