**HAMBURGER BEITRÄGE
ZUR MATHEMATIK**

Heft 470

The Parity Search Problem.

Christian Reiher

Version März 2013

THE PARITY SEARCH PROBLEM

CHRISTIAN REIHER

ABSTRACT. We prove that for any positive integers n and d there exists a collection consisting of $f = d \log n + O(1)$ subsets $A_1, A_2, \ldots, A_f$ of $[n]$ such that for any two distinct subsets X and Y of $[n]$ whose size is at most d there is an index $i \in [f]$ for which $|A_i \cap X|$ and $|A_i \cap Y|$ have different parity. Here we think of d as fixed whereas n is thought of as tending to infinity, and the base of the logarithm is 2.

Translated into the language of combinatorial search theory, this tells us that $d \log n + O(1)$ queries suffice to identify up to d marked items from a totality of n items if the answers one gets are just whether an even or an odd number of marked elements has been queried, even if the search is performed non–adaptively. Since the entropy method easily yields a matching lower bound for the adaptive version of this problem, our result is asymptotically best possible.

This answers a question posed by DÁNIEL GERBNER and BALÁZS PATKÓS in GYULA O.H. KATONA's Search Theory Seminar at the Rényi institute.

§1. INTRODUCTION.

In a typical problem from combinatorial search theory a finite number of entities is given to you some of which are considered to be *marked* or *defective* and your task is to find out which of them these are. For example, many recreational problems involving coins a few of which are forged as well as a scale that may be used to expose them belong to this area.
For a thorough introduction to combinatorial search theory, the reader is referred to the excellent and comprehensive survey article [3].

Recently GERBNER and PATKÓS ([4]) started to consider the following search problem: One gets confronted with n items – the set of which may for convenience be identified with the set $[n] = \{1, 2, \ldots, n\}$ – and one knows in advance that at most d of these items are marked, where $0 \leqslant d \leqslant n$. To identify them, one may make a sequence of queries, i.e. specify a sequence of subsets of $[n]$, and each time one makes such a query one is told the parity of the number of marked elements in ones query set. So for instance by querying a set containing

2010 *Mathematics Subject Classification.* 05D05, 11T99, 90B40.

Key words and phrases. Combinatorial Search Theory, Parity search, finite fields.

only one element one learns whether this element is marked or not and hence the problem is solvable in principle. As usual, however, it is more interesting to think about the least number of queries one needs to perform this task. More precisely, Gerbner and Patkós asked what the asymptotic behaviour of this minimal number is if d is fixed whilst n tends to infinity.

Now actually there are two version of this question. In the first of these, called the *adaptive problem*, one allows ones query sets to depend on the answers one has gotten to all previous queries. In the second, *non–adaptive* version, one has to specify all query sets before getting the first answer. The main result of this paper asserts that for both of these versions $d \log n + O(1)$ queries are necessary and sufficient.

It is quite standard to obtain a lower bound of the form $d \log n + O(1)$ to the adaptive problem, where the base of the logarithm is 2. For if $2^m < \sum_{i=0}^{i=d} \binom{n}{i} = \Theta(n^d)$, then it may happen that each of the first m answers reduces the number of outcomes still possible by no more than a factor of two, for which reason m queries cannot be enough. It is also clear that the non–adaptive problem requires no less queries than the adaptive one for one may pretend to search adaptively while in fact not caring about the answers. Thus it suffices to prove an upper bound of the form $d \log n + O(1)$ to the non–adaptive problem.

It seems worth while to observe that the non–adaptive problem may also be viewed as a question from extremal set theory. Specifically, one is interested in the number $f(n, d)$ defined as follows:

Definition 1.1. Given two positive integers n and d, let $f(n, d)$ be the least integer f such that there exist f subsets $A_1, A_2, \ldots, A_f$ of $[n]$ with the following property: For any two distinct subsets X and Y of $[n]$ the size of which is at most d, there exists an index $i \in [f]$ such that the cardinalities of $A_i \cap X$ and $A_i \cap Y$ have different parity.

I would like to record here that Gerbner and Patkós showed that choosing these sets A_i uniformly at random one can get $f(n, d) \leqslant 2d \log n + O(1)$. Their proof uses the first moment method. It may be observed that a routine application of the symmetric version of Lovász's Local Lemma (see [2] or Corollary 5.1.2 from [1]) allows you to improve this to $f(n, d) \leqslant (2d - 1) \log n + O(1)$. But in fact we shall prove $f(n, d) = d \log n + O(1)$ below. Somewhat more explicitly, we shall get

Theorem 1.2. *Suppose that d, m and n denote three positive integers such that $md \leqslant n < 2^m$. Then $f(n,d) \leqslant md$.*

The proof will be given in the next section.

§2. The proof of Theorem 1.2.

The actual proof of Theorem 1.2 is prepared by a sequence of three Lemmas most of which are of an algebraic nature. Throughout we denote the finite field with q elements by $\mathbb{F}_q$ and refer to the multiplicative group of its non–zero elements by $\mathbb{F}_q^{\times}$. If F is a field we write F^n for the n–dimensional vector space over F. Finally we would like to remind the reader that the number of ones appearing in a vector from $\mathbb{F}_2^n$ is sometimes called its *weight*.

The basic strategy of our proof is as follows: one interprets the problem as a statement about vector spaces over $\mathbb{F}_2$ and applies a change of basis to see that all one needs to do is proving the assertion contained in Lemma 2.3. Using a direct sum decomposition this task can be reduced to showing Lemma 2.2, which in turn is accomplished by means of an explicit construction based on the following algebraic fact exploiting the multiplicative structure of fields having characteristic 2.

Lemma 2.1. *If $A \not\subseteq \{0\}$ denotes a finite subset of a field of characteristic 2, then for some odd positive integer $k \leqslant |A|$ one has $\sum_{x\in A} x^k \neq 0$.*

Proof. Pick any $a \in A - \{0\}$ and observe that

$$\sum_{x\in A} x \prod_{b\in A-\{a\}} (x-b) = a \prod_{b\in A-\{a\}} (a-b) \neq 0.$$

Expanding the product appearing under the sum of the left hand side and rearranging we get

$$\sum_{i=1}^{i=|A|} \alpha_i \sum_{x\in A} x^i \neq 0$$

with certain irrelevant coefficients $\alpha_1, \dots, \alpha_{|A|}$ from our base field. Thus there exists some positive integer $k' \leqslant |A|$ such that $\sum_{x\in A} x^{k'} \neq 0$. Now if k denotes the least such k', then k automatically has to be odd, for otherwise we could use the equation

$$\sum_{x\in A} x^k = \left(\sum_{x\in A} x^{k/2}\right)^2$$

to obtain a contradiction. Thereby our Lemma is proved. □

Lemma 2.2. *For any two positive integers d and m, the vector space $\mathbb{F}_2^{dm}$ has a generating subset B of size at least $2^m - 1$ such that each vector x admits at most one representation as the sum of at most d distinct members of B.*

Proof. Plainly it suffices to exhibit a set consisting of $2^m - 1$ vectors from $\mathbb{F}_2^{dm}$ possessing the unique representability property. For once we have found such a set B we may look at a direct sum decomposition $\mathbb{F}_2^{dm} = \langle B \rangle \oplus U$ with some vector space U and extend B by a basis of U to achieve both goals.
For the purpose of finding such a set B, we may evidently replace the vector space $\mathbb{F}_2^{dm}$ appearing in this statement by $\mathbb{F}_{2^m}^d$. Corresponding to each number $\xi \in \mathbb{F}_{2^m}$ we define v_ξ to be the vector $(\xi, \xi^3, \dots, \xi^{2d-1})$ from the latter space, and then we claim that

$$B = \{v_\xi \mid \xi \in \mathbb{F}_{2^m}^\times\}$$

is as desired. To see this, suppose that some vector x admitted two distinct representations as the sum of at most d elements from B. Adding these representations up and cancelling terms appearing twice, we obtain a non–empty subset A of $\mathbb{F}_{2^m}^\times$ whose size is at most $2d$ such that $\sum_{\xi \in A} v_\xi = 0$. So in particular for all odd $k \leqslant |A|$ we have $\sum_{\xi \in A} \xi^k = 0$, contrary to our previous Lemma. □

Lemma 2.3. *Let d, m and n denote three positive integers such that $md \leqslant n < 2^m$. Then there is some vector subspace of $\mathbb{F}_2^n$ of codimension dm containing no non–zero vector whose weight is at most $2d$.*

Proof. It is convenient to think of $\mathbb{F}_2^n$ as being the space $V = \mathbb{F}_2^{dm} \oplus \mathbb{F}_2^{n-dm}$. It has $W = \{0\} \oplus \mathbb{F}_2^{n-dm}$ as a subspace of codimension dm. By our foregoing lemma there exist n distinct vectors $b_1, b_2, \dots, b_n$ from $\mathbb{F}_2^{dm}$ such that each vector from this space is expressible in at most one way as the sum of at most d distinct vectors from this sequence, and such that $b_1, b_2, \dots, b_{dm}$ is a basis of this space. Now define $v_1, v_2, \dots, v_{dm}$ to be the zero vector of $\mathbb{F}_2^{n-dm}$ and let $v_{dm+1}, v_{dm+2}, \dots v_n$ be any basis of this space. Clearly the set

$$L = \{(b_i, v_i) \mid i = 1, 2, \dots, n\}$$

forms a basis of V and by our construction it is not possible that a non–empty sum comprised of at most $2d$ distinct terms from L gives a vector from W. Applying an automorphism of V transforming L to the standard basis to this whole situation, W is mapped to a vector subspace W' whose codimension is still dm and that does not contain any vector whose weight is at most $2d$. So W' is as desired. □

We are now ready to prove Theorem 1.2. To do so we identify the power set of $[n]$ with the vector space $\mathbb{F}_2^n$ via characteristic functions. It is well known that the parity of the size of the intersection of two sets corresponds in this way to the standard scalar product. Our task now consists in exhibiting dm vectors $v_1, v_2, \ldots, v_{dm}$ such that for any two distinct vectors x and y the weight of which is at most d there is some $i \in [dm]$ such that $x \cdot v_i \neq y \cdot v_i$. This may be achieved by taking W to be a vector subspace of $\mathbb{F}_2^n$ as obtained in our third lemma and then choosing the vectors $v_1, v_2, \ldots, v_{dm}$ so as to span its orthogonal complement. Given any two distinct vectors x and y from $\mathbb{F}_2^n$ whose weights are at most d, one easily sees that their difference is non–zero and has weight at most $2d$, wherefore it cannot belong to W, which in turn means that there is indeed some $i \in [dm]$ satisfying $(x - y) \cdot v_i \neq 0$. This completes the proof of Theorem 1.2 and hence of our main result.

ACKNOWLEDGEMENT. I would heartily like to thank GYULA O.H. KATONA for inviting me to the Alfred Rényi Institute for one week, where this piece of research was carried out, and DÁNIEL GERBNER and BALÁZS PATKÓS for telling me about the problem solved here.

REFERENCES

[1] Noga Alon, Joel H. Spencer: *The probabilistic method*, Third edition, Wiley–Interscience Series in Discrete Mathematics and Optimization 2008.

[2] Paul Erdős and László Lovász: *Problems and results on 3–chromatic hypergraphs and some related questions*, in: Infinite and Finite Sets (1975), ed. by A. Hajnal, R. Rado and Vera T. Sós, North–Holland Publishing Company, Amsterdam, pp. 609–628.

[3] Gyula O.H. Katona: *Combinatorial Search Problems*, in: A survey of combinatorial theory (1973), ed. by J. N. Srivastava et al., North–Holland Publishing Company, Amsterdam, pp. 285–308.

[4] Balázs Patkós: Talk given in Katona's research seminar on combinatorial search theory at the Rényi institute in Feb. 2012

MATHEMATISCHES SEMINAR DER UNIVERSITÄT HAMBURG, BUNDESSTR. 55, D–20146 HAMBURG, GERMANY

E-mail address: Christian.Reiher@uni-hamburg.de