The size of the exceptional set of an asymptotic basis

Alina Firicel

December 18, 2006

Abstract

Problems related to the study of the asymptotic basis sets of integers belong to the realm of "additive number theory". In this note we survey a representative problem concerning the number of the elements of a basis with the property that, if we extract one of them from the initial set, the remaining set is not a basis anymore. It summarizes the article written by G. Grekos and B. Deschamps: "Estimations du nombre d'exceptions à ce qu'un ensemble de base privé d'un point reste un ensemble de base".

We prove that if A denotes a basis of order $h \geq 2$ and if A^* denotes the set of elements a such that $A \backslash \{a\}$ remains a basis, then $\#(A \backslash A^*) = O(\sqrt{\frac{h}{\log h}})$. Moreover, thanks to an example, we show that this result is not so far from the truth.

1. Some definitions and examples

If A is a set of nonnegative integers and $h \in \mathbb{N}^* = \{1, 2, \ldots\}$, we note

$$hA = \{x_1 + x_2 + \cdots + x_h; \quad x_1, x_2, \ldots, x_h \in A\}$$

This paper deals with additive **(asymptotic) bases**, that is, with sets of nonnegative integers A such that there exists an integer h satisfying $hA \sim \mathbb{N}$ where the notation $A \sim B$ denotes the fact that the symmetric difference of A and B is finite.

In other words, a set A of nonnegative integers is said to be a basis of order h if every sufficiently large integer can be expressed as a sum of exactly h integers taken from A (where repetition is allowed) and h is minimal. If A is a basis set, we denote by A^* the set of elements a in A having the property that $A \backslash \{a\}$ is also a basis set.

For a basis set A of order h, the study of the order of $A \backslash \{a\}$ (if this is a basis) was undertaken in [Nas] and [Gre1], where it was shown that the order of $A \backslash \{a\}$ is less than $\frac{h^2+3h}{2}$ and that, for any h, there exist a basis A_h of order h and an element a belonging to A_h^* such that the order of $A_h \backslash \{a\}$ is larger then $\frac{h^2-3h}{3}$.

Examples :

• If $A = \{1,2\} \cup 6\mathbb{N}$ then order(A)=4, order$(A \backslash \{6\}) = 4$, order$(A \backslash \{2\}) = 6$ but $A \backslash \{1\}$ is not a basis set.

• If $A = \{n^2; n \in \mathbb{N}\}$ then order (A)=4 because every nonnegative number can be represented as sum of 4 perfect squares, as Lagrange's theorem states.

If A is a basis and $a \in A$, [EG] demonstrates that $A \backslash \{a\}$ remains a basis (see lemma 1) if and only if

$$g.c.d.\{b - b'; b, b' \in A \backslash \{a\}\} = 1.$$

It is exceptional for an element a in A not to satisfy this last relation. One can see that the set $A \backslash \{a\}$ is a basis except for a *finite* number of a belonging to A (see lemma 2). This led to the study of the maximal cardinality of this set of "exceptional" elements : those $a \in A$ such that $A \backslash \{a\}$ is not an asymptotic basis.

Theorem 1.[DG] *If A denotes a basis set of order $h \geq 2$, then*:

$$\#\left(A \backslash A^*\right) \leq C\sqrt{\frac{h}{\log h}} \qquad with \quad C = \sqrt{4e^{2.0769} + e^{-1}}.$$

Furthermore, thanks to an example, it is shown that this estimate in $O(\sqrt{\frac{h}{\log h}})$ is best possible.

Theorem 2.[DG] *There exist $K > 0$ and a sequence of basis sets $(A_n)_n$ of order $h_n \geq 2$ such that:* $\lim_{n\to\infty} h_n = \infty$ *and* $\quad \#\left(A_n \backslash A_n^*\right) \geq K\sqrt{\frac{h_n}{\log h_n}}$.

2. Proof of theorem 1

Our first goal will be to prove the first theorem. We have to prove that the number of the elements a of a basis A satisfying the condition: "$A \backslash \{a\}$ *does not remain a basis set anymore*" is $O(\sqrt{\frac{h}{\log h}})$, where h is the order of A. We will need a sequence of lemmas to accomplish this.

Lemma 1. *Let A be a basis set and $a \in A$. Then the following statements are equivalent*:

i) $A \backslash \{a\}$ *is a basis set*;

ii) $g.c.d.\{b - b'; b, b' \in A \backslash \{a\}\}$=1.

Proof. Necessity. Suppose $g.c.d.\{b - b'; b, b' \in A \backslash \{a\}\} = d > 1$. Let $\Phi_d : \mathbb{N} \to \mathbb{Z}_d$ be the canonic surjection. Then $\Phi_d(A \backslash \{a\}) = \{\alpha\}$ and consequently

$\Phi_d(h(A\backslash\{a\})) = \{h\alpha\} \neq \mathbb{Z}_d$, for any h, which implies that it is impossible to have $h(A\backslash\{a\}) \sim \mathbb{N}$ hence $A\backslash\{a\}$ is not a basis set. Contradiction!

Sufficiency. The main idea of this proof comes from [EG].

Let A=$\{a_0 < a_1 < \ldots\}$ and $A_1 = A - a = \{x - a; \quad x \in A\}$. Then A_1 is a subset of $\mathbb{Z}$ which contains 0. Since $hA \sim \mathbb{N}$, it follows that $hA_1 \sim \mathbb{N}$ (because $hA_1 = hA - ha$). We denote $B = A\backslash\{a\}$ and $B_1 = A_1\backslash\{0\}$. Clearly, B is a basis set if and only if B_1 is a basis (note that $B_1 = B - a$). Moreover

$$g.c.d.\{b - b'; b, b' \in B_1\} = g.c.d.\{b - b'; b, b' \in B\} = 1$$

Let consider $B_1 = \{b_0 < b_1 < \ldots\}$; we remark that

$$g.c.d.\{b - b'; b, b' \in B_1\} = g.c.d.\{b_{k+1} - b_k; k \in \mathbb{N}\}.$$

As $g.c.d.\{b_{k+1} - b_k; \quad k \in \mathbb{N}\} = 1$, there exists a positive integer t such that $g.c.d.\{b_{k+1} - b_k; \quad k \leq t\} = 1$. Indeed, if we define $f(n) = g.c.d.\{b_{k+1} - b_k; k \leq n\}$, it is a decreasing function and $\lim_{n\to\infty} f(n) = 1$. Therefore the sequence $f(n)$ is a stationary one and the result follows. Applying Bézout's theorem, for suitable integers c_k we have:

$$\sum_{k=0}^{t} c_k\,(b_{k+1} - b_k) = 1$$

Define p_k and q_k as follows:

$$p_k = \begin{cases} b_{k+1} & \text{if} \quad c_k \geq 0 \\ b_k & \text{if} \quad c_k < 0 \end{cases} \qquad q_k = \begin{cases} b_k & \text{if} \quad c_k \geq 0 \\ b_{k+1} & \text{if} \quad c_k < 0 \end{cases}$$

The above equality can be rewritten as $\sum_{k=0}^{t} |c_k|\,(p_k - q_k) = 1$, i.e. $\sum_{k=0}^{t} |c_k| p_k = 1 + \sum_{k=0}^{t} |c_k| q_k$.

Now let consider $q \in B_1 \cap \mathbb{N}$ and $M = \sum_{k=0}^{t} q|c_k|p_k = q + \sum_{k=0}^{t} q|c_k|q_k$. We get $M \in nB_1 \cap (n+1)B_1$ where $n = q\sum_{k=0}^{t} |c_k|$.

It follows immediately that $2M \in 2nB_1 \cap (2n+1)B_1 \cap (2n+2)B_1$ and, more generally, that for any $w \geq 1$:

$$wM \in \bigcap_{k=0}^{w} (wn + k)B_1.$$

Setting $w = h-1$, we get $wM \in \bigcap_{k=0}^{h-1}((h-1)n+k)B_1$. For any $r = 1, \ldots, h$, $0 \leq h - r \leq h - 1$, hence $(h-1)M \in [(h-1)n + (h-r)]B_1$. This shows that

$$(h-1)M + rB_1 \subset [(h-1)n + (h-r) + r]B_1 = [(h-1)n + h]B_1$$

and

$$\bigcup_{r=1}^{h} [(h-1)M + rB_1] \subset [(h-1)n + h]B_1.$$

Since $hA_1 \sim \mathbb{N}$ and $0 \in A_1$, we obtain that $hA_1 = \{0\} \cup (\bigcup_{r=1}^{h} rB_1) \sim \mathbb{N}$. On the other hand, $\bigcup_{r=1}^{h}[(h-1)M + rB_1] = (h-1)M + \bigcup_{r=1}^{h} rB_1$. Consequently $\bigcup_{r=1}^{h}[(h-1)M + rB_1] \sim \mathbb{N}$ and so B_1 is a basis set .
Remark. Actually, we can see that the order of B_1 is less than $(h-1)n+h$.

Lemma 2. *$A \backslash A^*$ is a finite subset of $\mathbb{Z}$.*

Proof. Let us enumerate the elements of A : $x_0 \leq x_1 \leq x_2 \leq \ldots$. Let us consider the function $f : \mathbb{N} \to \mathbb{N}$ as $f(n) = g.c.d.\{x_i - x_j; i, j \leq n\}$. Obviously, f is decreasing and $\lim_{n\to\infty} f(n) = 1$. In particular, there exists $n_0 \in \mathbb{N}$ such that $f(n_0) = 1$. We denote A_1 the set of elements of A which are less than x_{n_0}. For any $a \in A \backslash A_1$, $g.c.d.\{b - b'; b, b' \in A \backslash \{a\}\} = 1$ because $A_1 \subset A \backslash \{a\}$. Applying the first lemma, $A \backslash \{a\}$ is a basis set. Hence $A^* \supset A \backslash A_1$ and $A \backslash A^* \subset A_1$, so $A \backslash A^*$ is a finite set.

As $A \backslash A^*$ is a finite set, throughout this section we write $A \backslash A^* = \{x_1, x_2, \ldots, x_s\}$ and for any $i = 1, 2, \ldots, s$: $d_i = g.c.d.\{b - b'; \ \ b, b' \in A \backslash \{x_i\}\}$.

Lemma 3. *For any $i = 1, 2, \ldots, s$, $2 \leq d_i \leq h$.*

Proof. The lower bound is easily deduced from lemma 1. We will prove the upper bound. As $d_i = g.c.d.\{b - b'; b, b' \in A \backslash \{x_i\}\}$ we deduce that $\Phi_{d_i}(A \backslash \{x_i\}) = \{\alpha\}$. Let a_i be a representant of α in $\mathbb{N}$. Then $\Phi_{d_i}(a_i) \neq \Phi_{d_i}(x_i)$; otherwise $\Phi_{di}(A)$ contains only one element and A is no longer a basis. Moreover, $\Phi_{d_i}(hA) = h\Phi_{d_i}(A)$. As h is the order of A, we get $h\Phi_{d_i}(A) = \mathbb{Z}_{d_i}$. We denote $w_1 = \Phi_{d_i}(a_i)$ and $w_2 = \Phi_{d_i}(x_i)$. Then

$$h\Phi_{d_i}(A) = \{iw_1 + jw_2; i + j = h\} = \mathbb{Z}_{d_i}$$

Since i takes values from 0 to h, in the left set there are $(h+1)$ elements, not necessarily all different. But $\mathbb{Z}_{d_i}$ has d_i elements. Consequently: $d_i \leq h + 1$.

If $d_i = h + 1$, then for all $k = 1, \ldots, h$ we have $kw_1 + (h-k)w_2 \neq hw_2$ (because the elements of the set are pairwise distinct). Let n be a sufficiently large integer such that $n \in hA$ and $n \equiv hx_i\,[d_i]$. There exists a $k \leq h$ such that $n = kx_i + \alpha_{k+1} + \ldots + \alpha_h$, where $\alpha_i \in A \backslash \{x_i\}$. We have $\Phi_{d_i}(n) = hw_2 = (h-k)w_1 + kw_2$. This is possible only if $k = h$. So, if we choose $n > hx_i$ and $n \equiv hx_i\,[d_i]$, then $n \notin hA$. This yields a contradiction since $hA \sim \mathbb{N}$.

Lemma 4. *For any i and j, $i \neq j$, we have $g.c.d.(d_i, d_j) = 1$.*

Proof. A proof by contradiction. Assume that the *g.c.d.* is greater than 1 and find that A is not a basis set anymore (like in lemma 3).

Lemma 5. *For any $n \geq 2$, $n \in \mathbb{N} : n! \geq n^n e^{-n}$.*
Proof. Use the fact that $e^x \geq \frac{x^n}{n!}$ for all $x \geq 0$ and, then, take $x = n$.

Lemma 6. *Let n and k be two positive integers. The number $A_n(k)$ of $n-$uples $(\alpha_1, \alpha_2, \ldots, \alpha_n) \in \mathbb{N}^n$ such that $\sum_{i=1}^{n} \alpha_i = k$ is:*

$$A_n(k) = C_{k+n-1}^{k} = \frac{(k+1)(k+2)\ldots(k+n-1)}{(n-1)!}$$

Proof. By induction on n.

Lemma 7. *For any $n \geq 2$,* we have : $p_1 \ldots p_n \geq n^n (\log^n n) \alpha^{-n}$, *where* $\alpha = e^{1.0769}$ and p_n denotes the $n-$th prime number.

Proof. This is equivalent to prove $\theta(p_n) \geq n(\log n + \log\log n - 1.0769)$, where $\theta(x) = \sum_{p \leq x} \log p$ and is called the Chebyshev function. The proof of this lemma, which can be found in [Rob], requires some other lemmas.

Remark. The constant which appears in theorem 1 satisfies $C^2 - 4e\alpha = e^{-1}$.

Lemma 8. *We have $s^2 \log s \leq 2e\alpha h$.*

Proof. By *lemma 1*, $\#\Phi_{d_i}(A \backslash \{x_1, x_2, \ldots, x_s\}) = 1$. By *lemma 4*, $(d_i, d_j) = 1$ and by the Chinese remainder theorem $\#\Phi_q(A \backslash \{x_1, x_2, \ldots, x_s\}) = 1$, where $q = d_1 d_2 \ldots d_s$. Denote $\lambda = \Phi_q(A \backslash \{x_1, x_2, \ldots, x_s\})$, $\lambda_i = \Phi_q(x_i)$, $i = 1, \ldots, s$. We have $\Phi_q(hA) = h\Phi_q(A)$ and since h is the order of A, we get $h\Phi_q(A) = \mathbb{Z}_q$. Consequently $\mathbb{Z}_q = \{\alpha_0 \lambda + \alpha_1 \lambda_1 + \ldots + \alpha_s \lambda_s; \quad \alpha_i \in \mathbb{N}, \quad \sum_{i=0}^{s} \alpha_i = h\}$. By *lemma 6*, the last set contains at most C_{h+s}^{h} elements. Hence

$$q = d_1 d_2 \ldots d_s \leq C_{h+s}^{h} = \frac{(h+1)(h+2)\ldots(h+s)}{s!}$$

By *lemma 4*, $p_1 p_2 \ldots p_s \leq d_1 d_2 \ldots d_s$. By *lemma 3* and *lemma 4*, we get $s \leq h$. Otherwise, $s - 1 \geq d_i$, for $i = 1, \ldots, s$. So we have s pairwise distinct integers taking values between 2 and $s - 1$. This implies that there are i and j such that $d_i = d_j$, contradiction with the *lemma 4*. Finally $s! s^s (\log^s s) \alpha^{-s} \leq (2h)^s$. By *lemma 5* we get the result.

Using the above lemma, we can now prove the main theorem.

Assume for contradiction that $s > C\sqrt{\frac{h}{\log h}}$.
Then $s^2 > C^2 \frac{h}{\log h}$ and $\log s > \log C + \frac{1}{2}\log h - \frac{1}{2}\log\log h$. So

$$s^2 \log s > \frac{C^2}{2}h - \frac{C^2}{2}h\frac{\log\log h}{\log h} + C^2 \log C \frac{h}{\log h}$$

Applying lemma 8 and simplifying:

$$(\frac{4e\alpha}{C^2} - 1)\log h + \log\log h - 2\log C > 0$$

Considering the function $f :]1, \infty[\to \mathbb{R}$ defined by

$$f(x) = (\frac{4e\alpha}{C^2} - 1)\log x + \log\log x - 2\log C$$

and calculating his derivative one can easily obtain that f has an absolute maximum at the point $x_0 = \exp(\frac{C^2}{C^2 - 4e\alpha})$. This fact implies the wanted contradiction, since $f(x_0) = 0$. This completes the proof of theorem 1.

Remark. Before, the study of $\#(A \backslash A^*)$ was undertaken in a series of works and in particular in [Gre2], where it was shown that $\#(A \backslash A^*)$ is less than $(h-1)$.

3. Proof of theorem 2

In order to prove the main theorem, we need the following proposition.

Consider the set:

$$A_n = p_1 p_2 \dots p_n \mathbb{N} \bigcup \{q_i = \prod_{j \neq i, j=1}^{n} p_j; \quad i = 1, 2, \dots, n\}$$

where p_n is the $n-$th prime number.

Proposition. *A_n is a basis set of order $1 - n + \sum_{k=1}^{n} p_k$ and*

$$A_n \backslash A_n^* = \{q_1, q_2, \dots, q_n\}$$

With this result, we can show now **the second theorem**.
As a consequence of the prime number theorem, one gets an asymptotic

expression for the $n-$th prime number : $p_n \cong n \log n$. This implies that there exists a constant $w > 1$ such that: $\forall n \geq 2$, $p_n \leq wn \log n$.
Thus, denoting h_n the order of A_n, we have: $h_n \leq \sum_{k=1}^{n} p_k \leq wn^2 \log n$.
Using the proposition: $s_n = \#(A_n \backslash A_n^*) = n$ and replacing n by s_n, we get:

$$h_n \leq w s_n^2 \log s_n.$$

Now, we choose the constant w such that $w > \frac{2}{\log 2}$ and we put $K = \sqrt{\frac{2}{w}}$.
If we assume by contradiction that for a positive integer n, $s_n < K\sqrt{\frac{h_n}{\log h_n}}$.
Applying this inequality to s_n^2, to $\log s_n$, and then to $s_n^2 \log s_n$, and using the fact $\frac{1}{w} h_n \leq s_n^2 \log s_n$, we obtain: $h_n < e^{K^2} < e^{\log 2} = 2$, hence $h_n < 2$. This is not possible, since $h_n \geq 2$.

Now, it only remains to show **the proposition**. First of all, we notice that $q = q_1 + q_2 + \ldots + q_n$ generates $\mathbb{Z}_{p_1 p_2 \ldots p_n}$. Therefore A_n is a basis set of order less than or equal to $np_1 p_2 \ldots p_n + 1$. Let $i_0 \in \{1, 2, \ldots, n\}$. For any $i \neq i_0$, p_{i_0} is a divisor of q_i, so $\# h\Phi_{p_{i_0}}(A_n \backslash \{q_{i_0}\}) = 1 \neq \mathbb{Z}_{p_{i_0}}$ and hence $A_n \backslash A_n^* = \{q_1, q_2, \ldots, q_n\}$.
Finally, we want to find the order of A_n. For that, we introduce, for any h, the set

$$\Omega_h = \Phi_{p_1 p_2 \ldots p_n} \left\{ \sum_{k=1}^{n} \alpha_k q_k; \quad (\alpha_1, \alpha_2, \ldots, \alpha_n) \in \mathbb{N}^n \quad and \quad \sum_{k=1}^{n} \alpha_k \leq h \right\}.$$

If h_n indicates the order of A_n, it is then clear that $h_n = h + 1$, where h is the smallest integer such that $\Omega_h = \mathbb{Z}_{p_1 p_2 \ldots p_n}$, i.e. $\#\Omega_h = p_1 p_2 \ldots p_n$.
We consider now the set: $E = \{(\alpha_1, \alpha_2, \ldots, \alpha_n) \in \mathbb{N}^n; \quad \sum_{k=1}^{n} \alpha_k \leq h\}$ and let us define the equivalence relation "$\sim$":

$$(\alpha_1, \alpha_2, \ldots, \alpha_n) \sim (\beta_1, \beta_2, \ldots, \beta_n) \Leftrightarrow \sum_{k=1}^{n} \alpha_k q_k \equiv \sum_{k=1}^{n} \beta_k q_k \, [p_1 p_2 \ldots p_n] \, .$$

Then $\#\Omega_h = \#E/_\sim$. In fact, this class is characterized in a more precise way:

Lemma. *If* $(\alpha_1, \alpha_2, \ldots, \alpha_n)$ *and* $(\beta_1, \beta_2, \ldots, \beta_n)$ *are two elements of* E, *then the following propositions are equivalent* :
i) $(\alpha_1, \alpha_2, \ldots, \alpha_n) \sim (\beta_1, \beta_2, \ldots, \beta_n)$; ii) $\forall i = 1, 2, \ldots, n$, $\alpha_i \equiv \beta_i \, [p_i]$.

Proof. "*i*) $\Rightarrow$ *ii*)" If $\sum_{k=1}^{n} (\alpha_k - \beta_k) q_k \equiv 0 \, [p_1 p_2 \ldots p_n]$, then

$$\sum_{k=1}^{n} (\alpha_k - \beta_k) q_k \equiv 0 \, [p_i] \, , \quad \text{for} \quad \text{all} \quad i = 1, 2, \ldots, n.$$

Since p_i/q_k for any $k \neq i$, we get $(\alpha_i - \beta_i)q_i \equiv 0\,[p_i]$, hence $(\alpha_i - \beta_i) \equiv 0\,[p_i]$. The second implication "$ii) \Rightarrow i)$" is analogous to the first one.

Consequently, a class of representants of $E/_\sim$ is defined by the $n-$uples $(\alpha_1, \alpha_2, \ldots, \alpha_n)$ such that: $0 \leq \alpha_i \leq p_i - 1$ and $\sum_{k=1}^{n} \alpha_k \leq h$. Denoting ϵ_h the set of all these $n-$uples, it is then clear that $\#\epsilon_h = \#\Omega_h$.
If we suppose $h < \sum_{k=1}^{n} p_k - n$, then at least one α_i is less than $p_i - 1$ and then α_i takes less than p_i values. Thus $\#\epsilon_h = \#\Omega_h < p_1 p_2 \ldots p_n$ and $\Omega_h \neq \mathbb{Z}_{p_1 p_2 \ldots p_n}$.
If $h \geq \sum_{k=1}^{n} p_k - n$, $\sum_{k=1}^{n} \alpha_k \leq \sum_{k=1}^{n} p_k - n \leq h$ and then

$$\#\epsilon_h = \#\left\{(\alpha_1, \alpha_2 \ldots, \alpha_n); \quad 0 \leq \alpha_i \leq p_i - 1; \forall i = 1, 2, \ldots, n\right\} = p_1 p_2 \ldots p_n.$$

Moreover $\#\Omega_h = p_1 p_2 \ldots p_n$ as desired. We are looking for the least h with this property, hence $h = \sum_{k=1}^{n} p_k - n$.
Finally, the order of A_n is $h_n = h + 1 = 1 + \sum_{k=1}^{n} p_k - n$.

Conclusion

The study of the size of the exceptional set $A \backslash A^*$ is one fundamental problem in the study of an asymptotic basis. Another interesting problem is the study of the function $\mathbf{X}$ defined by:

$$\mathbf{X}(h) = \max_{hA \sim \mathbb{N}} x(A) \quad \text{where} \quad x(A) = \max_{a \in A^*} \mathrm{order}(\mathrm{A} \backslash \{\mathrm{a}\})$$

and a long standing open question is to prove that the limit of $\frac{\mathbf{X}(h)}{h^2}$, for h tending to infinity, exists; and, if this is the case, to determine its value.

Answers to some questions concerning asymptotic bases are given in a series of papers written by: P. Erdős, R. Graham, G. Grekos, M. Nathanson, X-D. Jia, J. Nash, A. Plagne, J. Cassaigne, B. Deschamps.

References:

[**DG**] G. Grekos and B. Deschamps, *Estimations du nombre d'exceptions à ce qu'un ensemble de base privé d'un point reste un ensemble de base*, J. reine angew. Math. 539 (2001), 45–53.

[**EG**] P. Erdős and R.L. Graham, *On bases with an exact order*, Acta Arithmetica 37 (1980), 201–207.

[**Gre1**] G. Grekos, *Sur l'ordre d'une base additive*, Séminaire de théorie des nombres de Bordeaux (1987-1988), exposé 31.

[**Gre2**] G. Grekos, *Quelques aspects de la théorie additive des nombres*, Thèse de doctorat, Paris 6, 1976.

[**Nas**] J.C.M. Nash, *Some applications of a theorem of M. Kneser*, Journal of Number Theory 44 (1993), 1–8.

[**Rob**] G. Robin, *Estimation de la fonction de Tchebychev*, Acta Arith. 42 (1983), 367–389.