Waring's problem with restrictions on q-additive functions *

Stephan Wagner

Department of Mathematics

Graz University of Technology

A-8010 Graz, Steyrergasse 30

wagner@finanz.math.tugraz.at

Abstract. In a paper of Thuswaldner and Tichy, a version of Waring's problem with restrictions on the sum of digits was considered. This paper is devoted to a generalization of their result to arbitrary completely q-additive functions.

Key words: Waring's problem, q-additive functions, correlation

2000 Mathematics Subject Classification: 11P05, 11A63

*This work was supported by Austrian Science Fund project no. S-8307-MAT.

1 Notation and Introduction

Throughout this paper, the same notation as in [6] will be used. A set $A \subseteq \mathbb{N}$ is said to be a *basis* (*asymptotic basis*) of order s if every positive integer (sufficiently large positive integer) n can be represented as

$$n = x_1 + \ldots + x_s \text{ with } x_1, \ldots, x_s \in A.$$

The classical problem of Waring corresponds to the question whether the set A_k of k-th powers is a basis (resp. asymptotic basis). There is a vast amount of literature on this topic, we refer to [4, 7] for instance; [8] gives a comprehensive survey on Waring's problem. In [6], the authors discuss a generalization of Waring's problem with restrictions on the sum of digits. In particular, they show that the set

$$A_{k,h,m} := \{n^k | s_q(n) \equiv h \mod m\}$$

forms an asymptotic basis of order 2^k+1, where $s_q(n)$ denotes the q-adic sum of digits. The sum of digits is the classical example of a q-*additive* function, i.e.

$$s_q(aq^h + b) = s_q(aq^h) + s_q(b)$$

whenever $b < q^h$. In fact, it is even *completely* q-*additive*, which means that

$$s_q(aq^h + b) = s_q(a) + s_q(b)$$

whenever $b < q^h$. Thus it is natural to consider the analogous problem for general (completely) q-additive functions. The current note is devoted to this generalization. Let w be some integer-valued weight function on the set $\{0,\dots,q-1\}$ of q-adic digits, and define $v(n)$ by

$$v(n)=\sum_{j=0}^{l} w(d_j), \text{ where } n=\sum_{j=0}^{l} d_j q^j. \tag{1}$$

All integer-valued completely q-additive functions are of this form with $w(0)=0$. The sum of digits corresponds to $w(d)=d$, the q-adic length to $w(d)=1$.

2 Main Result

Theorem 2.1 *Let $s,k\in\mathbb{N}$, $s>k^2(\log k+\log\log k+O(1))$, $h_i,m_i,q_i\in\mathbb{N}$ $(1\le i\le s)$ with $m_i,q_i\ge 2$, and let $v_i(n)$ be defined by some weight function $w_i(d)$ on the q_i-adic digits for all i. Suppose that for all i the following holds true:*

> *There is no prime $P|m_i$ such that $w_i(0),\dots,w_i(q_i-1)$ is an arithmetic progression modulo P and $w_i(0)\equiv w_i(q_i-1) \mod P$.*

Then if $r(N)$ is the number of representations of N in the form

$$N=x_1^k+\dots+x_s^k \qquad (v_i(x_i)\equiv h_i \mod m_i),$$

there is a positive constant δ such that

$$r(N)=\frac{1}{m_1\dots m_s}\mathfrak{S}(N)\Gamma\Big(1+\frac{1}{k}\Big)^s\Gamma\Big(\frac{s}{k}\Big)^{-1}N^{s/k-1}+O(N^{s/k-1-\delta}). \tag{2}$$

The implied constant depends only on s, k and m_i. $\mathfrak{S}(N)$ is the singular series for the classical Waring problem – it is an arithmetic function for *which there exist positive constants* $0 < c_1 < c_2$ *depending only on* k *and* s *such that* $c_1 < \mathfrak{S}(N) < c_2$.

The proof will be essentially the same as in the paper of Thuswaldner and Tichy, apart from some minor changes. It is based on the following correlational result generalizing Theorem 3.3 of [6]:

Theorem 2.2 *Let* k, m, h, q *and* N *be positive integers with* $m \geq 2$, $q \geq 2$, *and let* $v(n)$ *be a function defined by (1) for some weight* w. *Suppose that there is a prime* P *in the factorization of the denominator of* $\frac{h}{m}$ *(in its lowest terms) such that either*

- $P \nmid (w(q-1) - w(0))$

or

- $w(0), \ldots, w(q-1)$ *is not an arithmetic progression modulo* P *(which is equivalent to the fact that* s *is a linear combination of the* q*-adic digit sum and the* q*-adic length modulo* P*).*

Now let $I_1, \ldots, I_k, J$ *be intervals of integers with* $\sqrt{N} \leq |I_j|, |J| \leq N$ $(1 \leq j \leq k)$. *Set*

$$Y(I_1, \ldots, I_k, J) := \sum_{h_1 \in I_1} \cdots \sum_{h_k \in I_k} \left| \sum_{n \in J} e\Big(\frac{h}{m} \Delta_{h_k, \ldots, h_1}(v)(n)\Big) \right|^2.$$

Then

$$Y(I_1, \ldots, I_k, J) \ll |I_1| \ldots |I_k| |J|^2 N^{-\eta} \tag{3}$$

holds with $\eta > 0$ *depending on* m, k *and* q.

The proof will be given in Section 3. Having proved this theorem, one can obtain the following result in litterally the same way as in [6] and finally prove Theorem 2.1 by means of the circle method. The original version of the proof given in [6] was modified by Pfeiffer and Thuswaldner in [5] – they used the results of Ford [2] to improve the bound for s from 2^k to $k^2(\log k + \log\log k + O(1))$. Their proof can easily be adapted to the current problem. Note also that the results of this paper can be generalized to systems of congruences in just the same way as in the paper of Pfeiffer and Thuswaldner.

Theorem 2.3 *Let k, m, h, q, N be positive integers with the same properties as in Theorem 2.2 and let $v(n)$ be defined as before. Then the estimate*

$$\Big|\sum_{n=1}^{N} e\Big(\theta n^k + \frac{h}{m} v(n)\Big)\Big| \ll N^{1-\gamma} \tag{4}$$

holds uniformly in $\theta \in [0,1)$ with $\gamma := \eta 2^{-(k+1)}$ (η as in Theorem 2.2).

3 Proof of Theorem 2.2

First, let's repeat the basic definitions and lemmas of [6]:

Definition 3.1 *Let $\mathcal{M} := \{1, 2, \ldots, k\}$ and $\mathcal{M}' := \{0, 1, \ldots, k+1\}$, and define the class of functions $\mathcal{F} := \{f : 2^{\mathcal{M}} \to \mathcal{M}'\}$. By F_0 and F_1, we denote the special functions*

$$F_0(S) := 0 \textit{ for all } S \subseteq \mathcal{M}$$

$$F_1(S) := \begin{cases} 1 & S = \mathcal{M} \\ 0 & \textit{otherwise.} \end{cases}$$

Furthermore, the operator Ξ *is defined by*

$$\Xi_{\mathbf{r},i}(f)(S) := \left\lfloor \frac{i + \sum_{j\in S} r_j + f(S)}{q} \right\rfloor$$

for each vector $\mathbf{r} = (r_1, \ldots, r_k) \in \{0, \ldots, q-1\}^k$ *and each* $0 \le i < q$.

The following result is easy to show:

Lemma 3.2 *For each pair* $\mathbf{r}, i$ *we have* $\Xi_{\mathbf{r},i}(\mathcal{F}) \subseteq \mathcal{F}$. *Furthermore, let*

$$\Xi_{\{\mathbf{r}_l, i_l\}_{1\le l\le L}} := \Xi_{\mathbf{r}_L, i_L} \circ \ldots \circ \Xi_{\mathbf{r}_1, i_1}$$

denote the iterates of Ξ. *Then for arbitrary* $f \in \mathcal{F}$,

$$\Xi_{\{\mathbf{0},0\}_{1\le l\le L'}}(f) = F_0$$

if $L' := \left\lfloor \frac{\log(k+1)}{\log q} \right\rfloor + 1$, *and with* $L'' := \left\lfloor \frac{k-1}{q-1} \right\rfloor + 1$,

$$\Xi_{\{\mathbf{r}_l^*, i_l^*\}_{1\le l\le L''}}(F_0) = F_1$$

for certain special values $\{\mathbf{r}_l^*, i_l^*\}$ *depending on* k *and* q.

Definition 3.3 *Let* $I_1, \ldots, I_k, J$ *be intervals of integers and* $f, f_1, f_2 \in \mathcal{F}$. *Define*

$$\Phi(h_1, \ldots, h_k; J; f) := \sum_{n\in J} e\left(\frac{h}{m} \sum_{S\subseteq\mathcal{M}} (-1)^{k-|S|} v\Big(n + \sum_{t\in S} h_t + f(S)\Big)\right),$$

$$\Psi(h_1,\ldots,h_{k-1};I_k,J;f_1,f_2) := \sum_{h_k\in I_k} \Phi(h_1,\ldots,h_k;J;f_1)\overline{\Phi(h_1,\ldots,h_k;J;f_2)},$$

$$X(I_1,\ldots,I_k,J;f_1,f_2) := \sum_{h_1\in I_1}\cdots\sum_{h_{k-1}\in I_{k-1}} \Psi(h_1,\ldots,h_{k-1};I_k,J;f_1,f_2).$$

Then $Y(I_1,\ldots,I_k,J) = X(I_1,\ldots,I_k,J,F_0,F_0)$.

Proposition 3.4 *Let* $f_1,f_2\in\mathcal{F}$ *and let* $I_1,\ldots,I_k,J$ *be intervals of integers. Then*

$$X(qI_1,\ldots,qI_k,qJ;f_1,f_2) = \sum_{r_1=0}^{q-1}\cdots\sum_{r_k=0}^{q-1}\sum_{i_1=0}^{q-1}\sum_{i_2=0}^{q-1}\alpha(f_1,f_2,\mathbf{r},i_1,i_2) \cdot X(I_1,\ldots,I_k,J;\Xi_{\mathbf{r},i_1}(f_1),\Xi_{\mathbf{r},i_2}(f_2)) + O\left(|I_1|\ldots|I_k||J|\right). \quad (5)$$

The implied constant depends only on q *and* k*. Here,*

$$\alpha(f_1,f_2,\mathbf{r},i_1,i_2) := e\left(\frac{h}{m}\sum_{S\subseteq\mathcal{M}}(-1)^{k-|S|}\Big(w(b(f_1,S,\mathbf{r},i_1)) - w(b(f_2,S,\mathbf{r},i_2))\Big)\right),$$

and $b(f,S,\mathbf{r},i)\in\{0,\ldots,q-1\}$ *is defined as the remainder of* $i+\sum_{t\in S}r_t + f(S)$ *modulo* q*.*

Proof: We exploit the fact that $v(qa+b)=v(a)+w(b)$ for $a>0$, $b<q$, to derive from

$$i+\sum_{t\in S} r_t + f(S) = q\Xi_{\mathbf{r},i}(f)(S) + b(f,S,\mathbf{r},i)$$

(which follows from the definition of Ξ) the identity

$$\begin{aligned} v\Big(qn+\sum_{t\in S}qh_t + i + \sum_{t\in S}r_t + f(S)\Big) &= v\Big(qn+\sum_{t\in S}qh_t + q\Xi_{\mathbf{r},i}(f)(S) + b(f,S,\mathbf{r},i)\Big) \\ &= v\Big(n+\sum_{t\in S}h_t + \Xi_{\mathbf{r},i}(f)(S)\Big) + w(b(f,S,\mathbf{r},i)) \end{aligned}$$

whenever $n > 0$. This yields

$$\Phi(q\mathbf{h}+\mathbf{r};qJ;f) = \sum_{i=0}^{q-1} e\Big(\frac{h}{m}\sum_{S\in\mathcal{M}}(-1)^{k-|S|}w(b(f,S,\mathbf{r},i))\Big)\Phi(\mathbf{h};J;\Xi_{\mathbf{r},i}(f))+O(q).$$

Applying this to Ψ and X in turn gives us the desired result. Next, we need some special values of α:

Lemma 3.5 *For $0 \le i < q-1$, we have*

$$\alpha(F_0, F_0, \mathbf{0}, 0, 0) = e(0) = 1, \tag{6}$$

$$\alpha(F_1, F_0, \mathbf{0}, i, 0) = e\Big(\frac{h}{m}(w(i+1) - w(i))\Big), \tag{7}$$

$$\alpha(F_1, F_0, \mathbf{0}, q-1, 0) = e\Big(\frac{h}{m}(w(0) - w(q-1))\Big). \tag{8}$$

The proof is the same as in [6, Lemma 5.1]. Now, iterating Proposition 3.4 gives us (using the notation $\mathcal{Q}_l := \{0, \dots, q-1\}^l$)

$$X(q^L I_1, \dots, q^L I_k, q^L J; f_1, f_2) =$$
$$\sum_{r_1,\dots,r_L\in\mathcal{Q}_k}\ \sum_{i_1,\dots,i_L\in\mathcal{Q}_2}\Big(\prod_{l=1}^{L}\alpha(\Xi_{(\mathbf{r}_j,i_{j1})_{1\le j\le l-1}}(f_1), \Xi_{(\mathbf{r}_j,i_{j2})_{1\le j\le l-1}}(f_2), \mathbf{r}_l, i_{l1}, i_{l2})\Big)$$
$$\cdot X(I_1, \dots, I_k, J; \Xi_{(\mathbf{r}_l,i_{l1})_{1\le l\le L}}(f_1), \Xi_{(\mathbf{r}_l,i_{l2})_{1\le l\le L}}(f_2)) + O\left(|I_1|\dots|I_k||J|\right),$$

where the implied constant depends on q, k and L. We select $L := L' + L'' + 3$ (L', L'' as in Lemma 3.2) and extract two summands from the above sum in analogy to [6]. Let P be a prime satisfying the conditions of Theorem 1. If $w(0), \dots, w(q-1)$ is not an arithmetic progression modulo P, then the sequence $w(1) - w(0), \dots, w(q-1) - w(q-2)$ is not constant, so we may choose $0 \le i_1, i_2 < q-1$ in such a way that $w(i_1+1)-w(i_1) \not\equiv w(i_2+1)-w(i_2)$

mod P. If on the other hand $w(0),\dots,w(q-1)$ is an arithmetic progression modulo P, then $w(0)\not\equiv w(q-1) \mod P$.

In the first case, we let the first summand V_1 correspond to the selection

$$\begin{aligned}
\mathbf{r}_l &= (0,\dots,0), & \mathbf{i}_l &= (0,0) & &(1\le l\le L'),\\
\mathbf{r}_l &= \mathbf{r}^*_{l-L'}, & \mathbf{i}_l &= \mathbf{i}^*_{l-L'} & &(L'+1\le l\le L-3),\\
\mathbf{r}_l &= (0,\dots,0), & \mathbf{i}_l &= (q-1,0) & &(l=L-2),\\
\mathbf{r}_l &= (0,\dots,0), & \mathbf{i}_l &= (i_1,0) & &(l=L-1),\\
\mathbf{r}_l &= (0,\dots,0), & \mathbf{i}_l &= (0,0) & &(l=L),
\end{aligned}$$

and let the second summand V_2 correspond to the same selection with $\mathbf{i}_{L-1} = (i_2,0)$. Then, using the same abbreviations as in [6], we arrive at

$$V_1 = A(f_1,f_2)\,\alpha(F_1,F_0,\mathbf{0},i_1,0)\,\alpha(F_0,F_0,\mathbf{0},0,0)\,X(I_1,\dots,I_k,J;F_0,F_0)$$

and

$$V_2 = A(f_1,f_2)\,\alpha(F_1,F_0,\mathbf{0},i_2,0)\,\alpha(F_0,F_0,\mathbf{0},0,0)\,X(I_1,\dots,I_k,J;F_0,F_0).$$

Now, by Lemma 3.5,

$$V_1 = A(f_1,f_2)e\Big(\frac{h}{m}(w(i_1+1)-w(i_1))\Big)X(I_1,\dots,I_k,J;F_0,F_0)$$

and

$$V_2 = A(f_1,f_2)e\Big(\frac{h}{m}(w(i_2+1)-w(i_2))\Big)X(I_1,\dots,I_k,J;F_0,F_0).$$

Therefore,

$$V_1 + V_2 = A(f_1, f_2) \left(e\Big(\frac{h}{m}(w(i_1+1) - w(i_1))\Big) + e\Big(\frac{h}{m}(w(i_2+1) - w(i_2))\Big) \right)$$
$$X(I_1, \dots, I_k, J; F_0, F_0).$$

Since $P \nmid \Big((w(i_1+1) - w(i_1)) - (w(i_2+1) - w(i_2))\Big)$, we are now able to apply the same argument as in [6] to prove a matrix inequality of the form

$$(|X(q^L I_1, \dots, q^L I_k, q^L J; f_1, f_2)|)_{(f_1,f_2)\in\mathcal{F}^2} \leq B \cdot (|X(I_1, \dots, I_k, J; g_1, g_2)|)_{(g_1,g_2)\in\mathcal{F}^2} + O\left(|I_1| \dots |I_k||J|\right), \tag{9}$$

where B is a matrix whose row sums are $\leq q^{L(k+2)}(1-\varepsilon)$ for a certain $\varepsilon > 0$ depending on q, k and m. In the second case (i.e., $P \nmid w(0) - w(q-1)$), we may even use the same parameters as in [6], and the argument stays the same. Iterating this matrix inequality and specializing $f_1 = f_2 = F_0$ then gives the estimate of Theorem 2.2. The O-term in (9) is of no harm since it can be included in the estimate (note also that it appears only if $w(0) \neq 0$).

Remark 1: The crucial tool in the proof of Thuswaldner and Tichy is the fact that the application of inequality (9) saves a factor of $(1-\varepsilon)$ from the trivial estimate. Now, let us consider arbitrary (not completely) q-additive functions, which can be written as

$$v(n) = \sum_{j=0}^{l} w^{(j)}(d_j), \text{ where } n = \sum_{j=0}^{l} d_j q^j, \tag{10}$$

i.e. the weight depends on the position of a digit, too. Then, it is necessary that a "positive percentage" of the weights satisfies the condidion of Theorem 2.2 so that the argument can still be applied. Formally, if $\omega(l)$ denotes the number of weights $w^{(i)}$ with $i \le l$ which satisfy the condition, the proof still works (with some technical and notational inconveniences) if

$$\liminf_{l\to\infty} \frac{\omega(l)}{l} > 0.$$

4 Final Remarks and Conclusion

Remark 2: First, we are going to explain why the condition posed on m_i is also necessary. Suppose there is a $P|m_i$ such that $w_i(0), \dots, w_i(q_i-1)$ is an arithmetic progression modulo P and $w_i(0) \equiv w_i(q_i-1) \mod P$. Then either $w_i(0), \dots, w_i(q-1)$ is constant modulo P, which means that the congruence condition for v_i is in fact a condition on the length of the q_i-adic expansion, or $w_i(d) \equiv A \cdot d + B \mod P$ for some $A \not\equiv 0 \mod P$.

In that case, the condition $w_i(0) \equiv w_i(q_i-1) \mod P$ turns into $P|(q_i-1)$, so $v_i(n)$ is a linear combination of the digit sum and length of n modulo P, and since also $s_{q_i}(n) \equiv n \mod P$ for all $P|(q_i-1)$, it is in fact a linear combination of n and its length, so the restriction is actually equivalent to congruence restrictions on intervals of the form $[q_i^k, q_i^{k+1})$.

In both cases, the asymptotics cannot hold any longer.

REMARK 3: Second, we discuss the size of the asymptotic order in the case $k = 1$ shortly. Theorem 2.1 (with the weaker estimate $s > 2^k$) tells us that it must be either 2 or 3 in the case that the conditions of the theorem are satisfied. It seems to be a nontrivial problem to determine whether it is 2 or 3 given some function v. Note, however, that it must be 3 for the q-adic sum of digits in view of the integers of the form

$$n_K := q^K - 1 = \sum_{i=0}^{K-1} (q-1)q^i.$$

If we write n_K as the sum of two integers n_1, n_2 with $s_q(n_1) \equiv s_q(n_2) \equiv h \mod m$, there cannot be any carry, so we would have $s_q(n_K) = K(q-1) \equiv 2h \mod m$, which is impossible for infinitely many values of K.

Note also that the set defined by $v(n) \equiv h \mod m$ can still be an asymptotic basis of $\mathbb{N}$ even if the conditions are violated. However, if we consider the q-adic length modulo m for instance, the order as an asymptotic basis might be as large as q^m.

REMARK 4: It is very difficult to give information about the order of the set $A = \{n^k | v(n) \equiv h \mod m\} \cup \{0, 1\}$ as a basis of $\mathbb{N}$ (0 and 1 have to be added to the set so that it is really a basis). In fact, the order depends highly on the parameters even in the very special case that $k = 1$ and $a = s_q$ is the q-adic sum of digits:

- If we take $q = 2$ and $h = 0$, $2^m - 1 = \sum_{i=0}^{m-1} 2^i$ is the smallest positive integer whose sum of digits is $\geq m$. Therefore, it is also the smallest element of the set $\{n^k | v(n) \equiv h \mod m\}$. All smaller integers can only be represented as the sum of 0's and 1's, thus at least $2^m - 2$ summands are needed.

- On the other hand, let $r \geq 2$ be arbitrary, $h = r$ and q sufficiently large, e.g. $q = 3m + r$. Then the distance between two subsequent elements of the set $\{n^k | v(n) \equiv h \mod m\}$ is at most $2m - 1$ (which is easy to verify). Such a gap can be filled with $\leq \lfloor \frac{m}{r} \rfloor + r - 1$ summands from the set $\{1, r, m + r\}$; so a total of $\lfloor \frac{m}{r} \rfloor + r$ summands is sufficient. Taking $r = \lfloor \sqrt{m} \rfloor$ thus gives an order $\leq 2\sqrt{m} + O(1)$.

These two examples show that the order of the studied set as a basis of $\mathbb{N}$ may grow exponentially in terms of the modulus as well as sublinearly. So there is probably not much hope that one can give any precise information on the order in general.

References

[1] H. Delange. Sur les fonctions q-additives ou q -multiplicatives. *Acta Arith.*, 21:285–298. (errata insert), 1972.

[2] K. B. Ford. New estimates for mean values of Weyl sums. *Internat. Math. Res. Notices*, (3):155–171 (electronic), 1995.

[3] D.-H. Kim. On the joint distribution of q-additive functions in residue classes. *J. Number Theory*, 74(2):307–336, 1999.

[4] M. B. Nathanson. *Additive number theory. The classical bases*, volume 164 of *Graduate Texts in Mathematics*. Springer-Verlag, New York, 1996.

[5] O. Pfeiffer and J. M. Thuswaldner. Waring's problem restricted by a system of sum of digits congruences. *Preprint*, 2005.

[6] J. M. Thuswaldner and R. F. Tichy. Waring's problem with digital restrictions. To appear in *Israel J. Math*, 2004.

[7] R. C. Vaughan. *The Hardy-Littlewood method*, volume 125 of *Cambridge Tracts in Mathematics*. Cambridge University Press, Cambridge, second edition, 1997.

[8] R. C. Vaughan and T. D. Wooley. Waring's problem: a survey. In *Number theory for the millennium, III (Urbana, IL, 2000)*, pages 301–340. A K Peters, Natick, MA, 2002.