A remark on the number field analogue of Waring’s constant $g(k)$

Paul Pollack*

Department of Mathematics, University of Georgia, Athens, Georgia 30602, United States of America

Received XXXX, revised XXXX, accepted XXXX
Published online XXXX

Key words Waring’s problem, Waring’s constant, Pythagoras number

MSC (2010) 11P05 (primary), 11R04, 11R47 (secondary)

Let $K$ be a number field, and let $k$ be an integer with $k \geq 2$. Let $\mathcal{O}^{\geq 0}$ be the collection of totally nonnegative integers in $K$ (i.e., the totally positive integers together with zero). We let $g(k, K)$ denote the smallest positive integer with the following property: Every element of $\mathcal{O}^{\geq 0}$ that is a sum of $k$th powers of elements of $\mathcal{O}^{\geq 0}$ is the sum of $g$ such $k$th powers. Work of Siegel in the 1940s shows that $g(k, K)$ is well-defined for all $k$ and $K$. In this note, we prove that $g(k, K)$ cannot be bounded by a function of $k$ alone: For each $k \geq 2$,

$$\sup_{k} g(k, K) = \infty.$$ 

Copyright line will be provided by the publisher

1 Introduction

Let $R$ be a semiring. For each pair of positive integers $k$ and $s$, we let $R_k[s]$ denote the collection of elements of $R$ that can be written in the form $\sum_{i=1}^{s} \alpha_i^k$ with all $\alpha_i \in R$, and we let $R_k[\infty] = \bigcup_{s=1}^{\infty} R_k[s]$. The $k$th Waring constant of $R$ is the number

$$g_R(k) := \inf\{s : R_k[s] = R_k[\infty]\}.$$ 

That is, $g_R(k)$ is the least positive integer $s$ with the property that every sum of $k$th powers in $R$ is a sum of $s$ $k$th powers, if any such $s$ exists; otherwise, $g_R(k) = \infty$. An alternative notation, introduced by Joly [1], is $w(k; R)$. This same quantity is also sometimes called the $k$th Pythagoras number of $R$.

Waring’s original conjecture, proposed in 1770 [2, p. 336] and settled by Hilbert in 1909 [3], is that $g_{\mathbb{Z}^{\geq 0}}(k) < \infty$ for every positive integer $k$. We now know several families of $R$ for which $g_R(k) < \infty$ for all $k$. For instance, this holds if $R = F[x]$ for $F$ a finite field (Paley [4]), for the matrix rings $R = M_n(\mathbb{Z})$ (Richman [5]), and for all $p$-adic rings $R$ (Ramanujam [6]). In fact, in each of these families, the quantity $g_R(k)$ can be bounded by a constant depending only on $k$ and not on the particular $R$ in question.

The subject of this note is Waring’s problem for rings of integers of number fields. The most important early results in this direction are due to Siegel, with the following theorem proved in [7, 8]. For a number field $K$, we let $\mathcal{O}$ denote its ring of integers. We use a superscript “$\geq 0$” to indicate a restriction to totally nonnegative elements, meaning totally positive or zero.

**Theorem A** Let $k \geq 2$. Let $K$ be a number field, and let $R$ denote the subring of $\mathcal{O}$ generated by the $k$th powers of elements of $\mathcal{O}$. There is a positive integer $G = G(k, K)$ for which the following holds: Every totally positive element of $R$ of sufficiently large norm is the sum of $G$ $k$th powers of elements of $\mathcal{O}^{\geq 0}$. In fact, if $K$ is not totally real, then the phrase “of sufficiently large norm” may be omitted.

On its face, Theorem A does not appear to fit into our notational framework for discussing Waring-type problems. But in fact, it easily implies a result of this kind, namely that

$$g_{\mathcal{O}^{\geq 0}}(k) < \infty \quad \text{for all pairs of } K, k.$$ 

* Corresponding author E-mail: pollack@uga.edu, Phone: +01 706 621 3275, Fax: +01 706 542 5907

1 By a $p$-adic ring, we mean the ring of integers belonging to the completion of a number field at a finite place.
(When $K = \mathbb{Q}$, this is the Waring–Hilbert theorem.) To see this, given $k$ and $K$, first choose $G$ and $C$ so that all totally positive $\alpha \in R$ with $N\alpha \geq C$ are sums of $G$ $k$th powers of elements of $\mathcal{O}^{\geq 0}$. Since $(\mathcal{O}^{\geq 0})_{K^{\infty}} \subset R$, certainly all $\alpha \in (\mathcal{O}^{\geq 0})_{K^{\infty}}$ with norm exceeding $C$ are sums of $G$ $k$th powers of elements of $\mathcal{O}^{\geq 0}$. Now consider those $\alpha \in (\mathcal{O}^{\geq 0})_{K^{\infty}}$ with $N\alpha \leq C$. Let $U$ denote the group of totally positive units in $\mathcal{O}$. Then the set of $\alpha$ being considered here is finite, up to multiplication by an element of $U$. Since $U$ is finitely generated, $U/U^k$ is finite. This allows us choose a finite set of $\beta$’s such that any $\alpha$ under consideration is a multiple of some $\beta$ by the $k$th power of a totally positive unit. Now letting $G_0$ be the maximum number of $k$th powers of elements of $\mathcal{O}^{\geq 0}$ needed to represent any $\beta$, we have $g_{\mathcal{O}^{\geq 0}}(k) \leq \max \{G_0, G\}$.

Siegel conjectured in [7] that the $G$ in Theorem A could be taken to depend only on $k$, and this was confirmed by later results of Birch [9] and Ramana (op. cit.; see also [10]); according to their work, we may take $G = \max \{8k^5, 2^k + 1\}$. Here we prove the contrasting result that the Waring constants $g_{\mathcal{O}^{\geq 0}}(k)$ cannot be bounded solely in terms of $k$.

**Theorem 1.1** For every $k \geq 2$, we have $\sup_K g_{\mathcal{O}^{\geq 0}}(k) = \infty$, where the supremum is taken over all number fields $K$.

**Remarks.**

(i) We have already mentioned that $g_R(k)$ is bounded in terms of $k$ uniformly across all $p$-adic rings $R$. It follows that the phenomenon described in Theorem 1.1 is a genuinely global one.

(ii) Let $K$ be an arbitrary number field. Siegel showed already in 1921 [11, Satz 2] that for any integer $k \geq 2$, every totally positive element of $K$ is a sum of $k$th powers of totally positive elements of $K$, with the number of summands bounded solely in terms of $k$. Consequently, $\sup_K g_{\mathcal{O}^{\geq 0}}(k) < \infty$.

The proof of Theorem 1.1 is inspired by an argument of Scharlau [12], who showed (in our notation) that $\sup_K g_{\mathcal{O}^{\geq 0}}(2) = \infty$. (This disproved a conjecture of Peters [13].) Scharlau’s theorem does not immediately imply the case $k = 2$ of Theorem 1.1, nor does it appear to be trivially implied by it.

## 2 Proof of Theorem 1.1

Throughout this section, $k \geq 2$ is fixed.

Let $q_1, q_2, q_3, \ldots$ be a sequence of distinct primes congruent to 1 modulo 4. Put

$$K_t = \mathbb{Q}(\sqrt{q_1}, \sqrt{q_2}, \ldots, \sqrt{q_t})$$

here, by $\sqrt{q_i}$, we mean the positive square root of $q_i$ living inside $\mathbb{R}$. Thus, each $K_t$ is a subfield of $\mathbb{R}$.

Since the discriminants of the fields $\mathbb{Q}(\sqrt{q_i})$ are pairwise relatively prime, $[K_t : \mathbb{Q}] = 2^t$, and $K_t$ has ring of integers $\mathbb{Z}[\frac{1+\sqrt{-q_i}}{2} : i = 1, 2, \ldots, t]$. With $Q_t$ the smallest odd integer exceeding $\sqrt{q_t}$, we let

$$\eta_t = \sum_{i=1}^{t} \left(\frac{Q_t + \sqrt{q_i}}{2}\right)^k.$$ 

Clearly, $\eta_t$ is a sum of $t$ $k$th powers of totally positive integers from $K_t$.

We will prove the following proposition.

**Proposition 2.1** We can select the sequence $q_1, q_2, q_3, \ldots$ in such a way that each $\eta_t$ is not a sum of fewer than $t$ $k$th powers of totally nonnegative integers in $K_t$.

So with $K = K_t$ and $\mathcal{O}$ the ring of integers of $K$, considering the element $\eta_t$ reveals that $g_{\mathcal{O}^{\geq 0}}(k) \geq t$. Since $t$ can be taken arbitrarily large, Theorem 1.1 follows immediately.

---

2. Siegel used Hilbert’s method to prove this. The Birch–Ramanujam result implies the analogous (and stronger) theorem with $\mathcal{O}[1/k^t]$ replacing $K$. Indeed, let $\alpha$ be a totally positive element of $\mathcal{O}[1/k^t]$. If $j$ is a large positive integer, then $k^{kj}\alpha$ is a totally positive element of $k^j\mathcal{O}$. Since $k^j\mathcal{O} \subset R$ (see top of p. 134 in [7]), the Birch–Ramanujam theorem implies that if $j$ is large enough, then $k^{kj}\alpha$ is a sum of $\max\{2^k + 1, 8k^5\}$ $k$th powers in $\mathcal{O}^{\geq 0}$. Now divide this representation of $k^{kj}\alpha$ through by $k^k$, absorbing this factor into the $k$th powers.
Proof of Proposition 2.1. We let \( q_1 = 5 \) and select \( q_2, q_3, q_4, \ldots \) inductively. Suppose that \( t \geq 2 \) and that we have already selected \( q_1, \ldots, q_{t-1} \) in such a way that \( \eta_{t-1} \) is not a sum of fewer than \( t-1 \) \( k \)-th powers of totally nonnegative integers in \( K_{t-1} \). (Obviously, this is satisfied when \( t = 2 \).) We will show how to select \( q_t \) so that \( \eta_t \) is not a sum of fewer than \( t \) \( k \)-th powers of totally nonnegative integers in \( K_t \). In fact, we will prove that it suffices to choose \( q_t \) sufficiently large in terms of \( q_1, \ldots, q_{t-1} \).

To see what constraints on \( q_t \) are necessary, let us suppose we have a representation

\[
\eta_t = \sum_i \alpha_i^k,
\]

where each \( \alpha_i \) is a totally nonnegative integer in \( K_t \) and the right-hand sum has fewer than \( t \) terms. We can (and will) assume that no \( \alpha_i = 0 \); thus, each \( \alpha_i \) is totally positive.

To begin with, observe that for each \( i = 1, 2, \ldots, t \),

\[
\left( \frac{Q_i + \sqrt{q_t}}{2} \right)^k = \frac{1}{2^k} \left( \sum_{\ell \text{ even}} \binom{k}{\ell} Q_i^{k-\ell} q_t^{\ell/2} + \sum_{\ell \text{ odd}} \binom{k}{\ell} Q_i^{k-\ell} q_t^{(\ell-1)/2} \sqrt{q_t} \right).
\]

Therefore, using \( \text{Tr}(\cdot) \) for the trace from \( K_t \) down to \( \mathbb{Q} \),

\[
\text{Tr}(\eta_t) = \frac{2^t}{2^k} \sum_{t=1}^t \sum_{\ell \text{ even}} \binom{k}{\ell} Q_i^{k-\ell} q_t^{\ell/2}
= C_{t-1} + \frac{2^t}{2^k} \sum_{\ell \text{ even}} \binom{k}{\ell} Q_i^{k-\ell} q_t^{\ell/2},
\]

say. Here \( C_{t-1} \) depends only on \( q_1, \ldots, q_{t-1} \). Noting that \( \sqrt{q_t} < Q_t < 2 + \sqrt{q_t} \), we see that

\[
\text{Tr}(\eta_t) \leq C_{t-1} + \frac{2^t}{2^k} \sum_{\ell \text{ even}} \binom{k}{\ell} Q_i^{k-\ell} q_t^{\ell/2} = C_{t-1} + 2^{t-1} Q_t^{k/2} \leq C_{t-1} + 2^{t-1} q_t^{k/2} (1 + 2/\sqrt{q_t})^k. \tag{2}
\]

For comparison, we estimate the trace of the right-hand side of (1). Write each

\[
\alpha_i = \frac{1}{2}(\mu_i + \nu_i \sqrt{q_t}),
\]

where \( \mu_i, \nu_i \in K_{t-1} \). From our description of the integers of \( K_t \), it follows that both \( \mu_i \) and \( \nu_i \) are integers of \( K_{t-1} \). Moreover,

\[
\alpha_i^k = \frac{1}{2^k} \left( \sum_{\ell \text{ even}} \binom{k}{\ell} \mu_i^{k-\ell} \nu_i^\ell q_t^{\ell/2} + \sum_{\ell \text{ odd}} \binom{k}{\ell} \mu_i^{k-\ell} \nu_i^\ell q_t^{(\ell-1)/2} \sqrt{q_t} \right),
\]

and so

\[
\text{Tr}(\alpha_i^k) = \frac{1}{2^k} \sum_{\ell \text{ even}} \binom{k}{\ell} q_t^{\ell/2} \text{Tr}(\mu_i^{k-\ell} \nu_i^\ell) = \text{Tr}(\mu_i^{k-\ell} \nu_i^\ell) = \text{Tr}(\mu_i^{k-\ell} \nu_i^\ell) + \text{Tr}(\mu_i^{k-\ell} \nu_i^\ell) = \text{Tr}(\mu_i^{k-\ell} \nu_i^\ell).
\]

To handle the trace terms on the right-hand side, we use that \( \alpha_i \) is totally positive. For any \( \sigma \in \text{Gal}(K_{t-1}/\mathbb{Q}) \), we can extend \( \sigma \) in two ways to an element of \( \text{Gal}(K_t/\mathbb{Q}) \); one of these fixes \( \sqrt{q_t} \) while the other sends it to \( -\sqrt{q_t} \). Since the image of \( \alpha_i \) is \( \frac{1}{2}(\mu_i + \nu_i \sqrt{q_t}) \) under both extensions is positive, it follows that

\[
\sigma(\alpha_i) \geq |\sigma(\nu_i)| \sqrt{q_t}. \tag{3}
\]

We proved this inequality for \( \sigma \in \text{Gal}(K_{t-1}/\mathbb{Q}) \), but of course it remains valid for all \( \sigma \in \text{Gal}(K_t/\mathbb{Q}) \), since both sides depend only on the restriction of \( \sigma \) to \( K_{t-1} \). For even \( \ell \leq k \), we raise both sides of (3) to the power \( k-\ell \) and multiply by the (nonnegative) number \( \sigma(\nu_i) \ell = |\sigma(\nu_i)| \ell \) to find that

\[
\sigma(\mu_i^{k-\ell} \nu_i^\ell) \geq |\sigma(\nu_i)| q_t^{(k-\ell)/2}.
\]
Summing on $\sigma$,
\[
\text{Tr}(\mu_i^{-t} \nu_i^t) \geq q_t^{(k-t)/2} \cdot S(\nu_i^k),
\]
where $S: K_t \to \mathbb{R}$ is defined by
\[
S(\theta) = \sum_{\sigma} |\sigma(\theta)|.
\]
(Unless otherwise specified, sums on $\sigma$ run over all elements of $\text{Gal}(K_t/\mathbb{Q})$.) Inserting this estimate above,
\[
\text{Tr}(\alpha_i^k) \geq \frac{1}{2} q_t^{k/2} \sum_i S(\nu_i^k).
\]
Hence,
\[
\text{Tr} \left( \sum_i \alpha_i^k \right) \geq \frac{1}{2} q_t^{k/2} \sum_i S(\nu_i^k).
\]
Recall that $\nu_i$ is an algebraic integer. So using $N(\cdot)$ for the norm from $K_t$ to $\mathbb{Q}$, we see that if $\nu_i \neq 0$, then
\[
S(\nu_i^k) = \sum_{\sigma} |\sigma(\nu_i^k)| \geq 2^t \cdot \left( \prod_{\sigma} |\sigma(\nu_i^k)| \right)^{1/2t} \geq 2^t \cdot |N(\nu_i^k)|^{1/2t} \geq 2^t.
\]
In the first inequality, we used the so-called “AM-GM inequality” connecting the arithmetic and geometric means.

Since $\eta_i \notin K_{t-1}$, there is at least one value of $i$ with $\nu_i \neq 0$. If there are at least two such values of $i$, then
\[
\text{Tr}(\eta_i) = \text{Tr} \left( \sum_i \alpha_i^k \right) \geq 2 \cdot \frac{1}{2} q_t^{k/2} \cdot 2^t = 2^t q_t^{k/2}.
\]
If $q_t$ is sufficiently large in terms of $q_1, \ldots, q_{t-1}$, then the final expression here exceeds the final term in (2). Thus, assuming that $q_t$ is chosen sufficiently large, there must be exactly one index $i$ with $\nu_i \neq 0$. If there are at least two such values of $i$, then
\[
\text{Tr}(\eta_i) = \text{Tr} \left( \sum_i \alpha_i^k \right) \geq 2 \cdot \frac{1}{2} q_t^{k/2} \cdot 2^t = 2^t q_t^{k/2}.
\]

Claim 2.2 If $\nu_j \neq \pm 1$, then $S(\nu_j^k) \geq (1+\delta)2^t$, where $\delta$ is a positive constant depending only on $q_1, \ldots, q_{t-1}$.

To see the claim, start by noting that for $\alpha \in K_{t-1}$, we have
\[
S(\alpha) = 2 \sum_{\sigma \in \text{Gal}(K_{t-1}/\mathbb{Q})} |\sigma(\alpha)|.
\]
In particular, the value of $S(\alpha)$, for $\alpha \in K_{t-1}$, does not depend on the choice of $q_t$. If $\alpha$ is an integer of $K_{t-1}$ and $S(\alpha)$ lies below a given bound, then $\prod_{\sigma \in \text{Gal}(K_{t-1}/\mathbb{Q})} (X - \sigma(\alpha))$ is a polynomial with bounded integer coefficients having $\alpha$ as a root; hence, there are only finitely many possibilities for $\alpha$. We quickly deduce that, as $\alpha$ ranges over integers of $K_{t-1}$, there is a smallest value of $S(\alpha)$ exceeding $2^t$, and the claim follows. So if $\nu_j \neq \pm 1$, then
\[
\text{Tr}(\eta_i) \geq \frac{1}{2} q_t^{k/2} \cdot S(\nu_j^k) \geq \frac{1}{2} q_t^{k/2} \cdot (1+\delta)2^t = (1+\delta)2^{t-1} \cdot q_t^{k/2}.
\]
But if $q_t$ is sufficiently large in terms of $q_1, \ldots, q_{t-1}$, this contradicts (2).
Hence, for large enough \( q_t \), the equality (1) forces \( \nu_i = 0 \) for all \( i \neq j \) and \( \nu_j = \pm 1 \). Therefore \( \alpha_i \in K_{t-1} \) for \( i \neq j \), and (1) assumes the form
\[
 \left( \frac{Q_1 + \sqrt{q_1}}{2} \right)^k + \cdots + \left( \frac{Q_t + \sqrt{q_t}}{2} \right)^k = \sum_{i \neq j} \alpha_i^k + \left( \frac{1}{2} (\nu_j \pm \sqrt{q_t}) \right)^k,
\]
where the \( \pm \) is the sign of \( \nu_j \). Comparing the \( K_{t-1} \)-coefficients of \( \sqrt{q_t} \) on both sides,
\[
 \sum_{\ell \text{ odd}} \left( \frac{k}{\ell} \right) Q_{t-1}^{k-\ell} q_t^{(\ell-1)/2} = \sum_{\ell \text{ odd}} \left( \frac{k}{\ell} \right) \mu_j^{k-\ell} q_t^{(\ell-1)/2}.
\]
Recall that \( \mu_j \) is nonnegative (in fact, from (3), totally nonnegative). If \( \nu_j = -1 \), then the immediately preceding right-hand side is nonpositive, whereas the left-hand side is clearly positive. So \( \nu_j = +1 \), and
\[
 \sum_{\ell \text{ odd}} \left( \frac{k}{\ell} \right) Q_{t-1}^{k-\ell} q_t^{(\ell-1)/2} = \sum_{\ell \text{ odd}} \left( \frac{k}{\ell} \right) \mu_j^{k-\ell} q_t^{(\ell-1)/2}.
\]
Viewed as a function of the real variable \( \mu_j \), the right-hand side is strictly increasing for \( \mu_j \geq 0 \). So the displayed equality forces \( \mu_j = Q_t \). Putting these deductions together, we see that the final summand on the left-hand side of (4) is the same as the final summand on its right-hand side. Subtracting this common value reveals that
\[
 \eta_{t-1} = \sum_{i \neq j} \alpha_i^k.
\]
So we have expressed \( \eta_{t-1} \) as a sum of fewer than \( t-1 \) \( k \)th powers of totally positive integers from \( K_{t-1} \). This contradicts our induction hypothesis.

3 Concluding remarks

It seems reasonable to also consider the Waring numbers (or higher Pythagoras numbers) of \( O \) and not only \( O^{\geq 0} \). As remarked in the introduction, Scharlau showed [12] that \( \sup_{K} g_{O}(2) = \infty \), and his construction was the inspiration for our proof of Theorem 1.1. It is possible to modify his construction to prove that \( \sup_{K} g_{O}(k) = \infty \) for every positive even integer \( k \). In fact, if \( q_1, q_2, q_3, \ldots \) is a suitably-chosen sequence of primes congruent to 1 mod 4, then
\[
 \sum_{i=1}^{t} \left( \frac{1 + \sqrt{q_i}}{2} \right)^k
\]
cannot be expressed as a sum of fewer than \( t \) \( k \)th powers of integers of \( K_t \). The proof is similar to, but simpler than, our proof of Theorem 1.1. By contrast, for odd values of \( k \), it is known that
\[
 \sup_{K} g_{O}(k) \leq 2^{k-1} + 8k^5.
\]
This follows from elementary arguments of Stemmler [14], with the local analysis there replaced by the results of Ramanujam alluded to previously (see also [15]).

Acknowledgements The author is supported by a grant from the National Science Foundation, award DMS-1402268. He thanks Enrique Treviño and the anonymous referee for several useful comments.

References