Square Dependence in Random Integers

by

Michael C. Beck

(Under the direction of Andrew Granville)

Abstract

We begin by motivating and explaining the notion of square dependence. Then, given a sequence $S = \{s_1, \ldots, s_j\}$ composed of integers chosen independently and with uniform distribution from $\{1, \ldots, n\}$, we want to know how likely S is to be square dependent. We then ask how many subsets $I \subseteq \{1, \ldots, j\}$ we should expect for which $\prod_{i \in I} s_i$ is a square. To answer this, we find bounds on the function $a(n, k)$, which counts the number of such subsets I of size k. We do this for a few small specific choices of k, and then in a more general setting prove both an upper bound and, for a smaller range of k, a lower bound. We then apply this work to get the asymptotic for the original expected value.

Finally, we describe an algorithm for finding integer solutions to $x^3 + y^3 + z^3 = k$ for specific values of k, and present our computational results.

Index words: Analytic Number Theory, Factoring Algorithm, Square Dependence, Quadratic Sieve

Square Dependence in Random Integers

by

Michael C. Beck

B.S., Kennesaw State University, 1997

A Dissertation Submitted to the Graduate Faculty
of The University of Georgia in Partial Fulfillment
of the
Requirements for the Degree

Doctor of Philosophy

Athens, Georgia

2004

Square Dependence in Random Integers

by

Michael C. Beck

Approved:

Major Professor: Andrew Granville

Committee: Ed Azoff
Rodney Canfield
Dino Lorenzini
Robert Rumely

Electronic Version Approved:

Maureen Grasso
Dean of the Graduate School
The University of Georgia
May 2004

Acknowledgements

I would like to thank my advisor, Andrew Granville, for his help, support, and insight throughout my graduate career, and especially during the development of my dissertation.

I would like to thank Carl Pomerance for inspiring me to pursue number theory in the first place, and for always being so clearly passionate about his work.

I would like to thank my committee for their support and help in getting this dissertation into its final form, and for the many hurdles they have helped me jump along the way.

I would also like to thank Eric Pine for listening to many presentations of this material without giving away any hint of the fact that he was bored out of his mind, and for proofreading the most tedious chapter in the paper.

Finally, I would like to thank my wife, Wendy, without whose love, support, and inspiration I would never have made it this far. I owe you everything.

Table of Contents

List of Tables

Chapter 1

Introduction to Square Dependence

In many factoring algorithms, while attempting to factor the integer n, an intermediate goal is to find a congruence of the form $x^2 \equiv y^2 \pmod{n}$, where $x \not\equiv \pm y \pmod{n}$. If we can, then $\gcd(n, x-y)$ is a nontrivial factor of n. In the quadratic sieve, for example, we attempt to construct such a congruence from a given sequence $\{x_1, \ldots, x_j\}$ of integers by noting that $\prod_{i=1}^{j} {x_i}^2 \equiv \prod_{i=1}^{j} ({x_i}^2 - n) \pmod{n}$. Since the left-hand side of the congruence is already a perfect square, the problem is reduced to that of finding a particular sequence of integers (namely the sequence $\{{x_i}^2 - n\}$) whose product is a square.

We'll abstract this problem to one which is a bit more accessible: given a sequence $S = (s_1, \ldots, s_j)$ of integers chosen independently and with uniform distribution from $\{1, \ldots, n\}$, we hope to find some subset $I \subseteq \{1, \ldots, j\}$ such that $\prod_{i \in I} s_i$ is a perfect square. If such a subset exists, we call S *square dependent*. Note that each s_i can be any integer from $\{1, \ldots, n\}$, they are chosen independently, and each integer is equally likely to be chosen. How large, then, should j be in order to give us confidence that we can find such a subset I?

On page 2 of [19], Pomerance has provided bounds for the answer in the following way: Let $S = (s_1, \ldots, s_n)$ be a sequence of integers chosen independently from the set $\{1, \ldots, n\}$, with each integer being equally likely to be chosen as each s_i. Let $j = j(S)$ be the smallest integer such that there exists a nonempty subset $I \subseteq \{1, \ldots, j\}$ for which $\prod_{i \in I} s_i$ is a square. Then the probability that j is in the interval

$$\left(\exp\left((\sqrt{2} - \epsilon)\sqrt{\log n \log\log n}\right), \exp\left((\sqrt{2} + \epsilon)\sqrt{\log n \log\log n}\right)\right)$$

tends to 1 as $n \to \infty$.

We ask a somewhat different but related question. As we run through the set of sequences $S = (s_1, \ldots, s_j)$, so that $|S| = j$, where each s_i can be any integer from $\{1, \ldots, n\}$, and the integers s_i are chosen independently and with uniform distribution, we define $E(n, j)$ to be the expected value of:

$$\#\{I \subseteq \{1, \ldots, j\} : \prod_{i \in I} s_i \text{ is a square}\}$$

Then we can write:

$$\begin{aligned} E(n,j) &= \frac{1}{n^j} \sum_{\substack{S=(s_1,\ldots,s_j) \\ s_i \in \{1,\ldots,n\}}} \sum_{\substack{\emptyset \neq I \subseteq \{1,\ldots,j\} \\ \prod_{i \in I} s_i \text{ is a square}}} 1 \\ &= \frac{1}{n^j} \sum_{\substack{1 \leq k \leq j \\ T=(t_1,\ldots,t_k) \\ t_i \in \{1,\ldots,n\} \\ \prod_{i=1}^{k} t_i \text{ is a square}}} \binom{j}{k} n^{j-k} \\ &= \sum_{k=1}^{j} \binom{j}{k} \frac{a(n,k)}{n^k} \end{aligned}$$

where $a(n, k) = \#\{(t_1, \ldots, t_k) : t_i \in \{1, \ldots, n\}, \text{ and } \prod_{i=1}^{k} t_i \text{ is a square}\}$. Here each t_i can be any integer from $\{1, \ldots, n\}$, and they will be considered to be chosen independently and with uniform distribution. We will first find the asymptotic size of $a(n, k)$ for some small values of k, and then we'll bound the size of $a(n, k)$ in a more general setting.

It should be noted that, due to the result of Pomerance, we know we want to be looking at sequences of very small size relative to n — in particular, of size up to $\exp\left(\sqrt{2 \log n \log\log n}\right)$. Thus the likelihood of choosing the same element twice is small. So the difference between calculating $E(n, j)$ using $a(n, k)$, which looks at sequences chosen independently, in which each element can be any integer from $\{1, \ldots, n\}$, and calculating the related expectation using a function, say $b(n, k)$, which counts only sequences of distinct elements chosen from $\{1, \ldots, n\}$ whose product is a square, should also be small. To confirm this, we define:

$$\alpha_{n,k} := \frac{a(n,k)}{n^k}$$

so that $\alpha_{n,k}$ is the probability that the elements of a sequence of k elements chosen independently and with uniform distribution from $\{1,\ldots,n\}$ have product a square. We also define:

$$b(n,k) := \#\left\{(m_1,\ldots,m_k) : m_i \in \mathbf{Z}, 1 \leq m_i \leq n, \prod_{i=1}^{k} m_i \text{ is a square, } m_i \text{ distinct}\right\}$$

and:

$$\beta_{n,k} := \frac{b(n,k)}{\prod_{i=0}^{k-1}(n-i)}$$

so that $\beta_{n,k}$ is the probability that the elements of a sequence of size k composed of distinct integers chosen from $\{1,\ldots,n\}$ will multiply together to form a perfect square. We'd like to bound $|\alpha_{n,k} - \beta_{n,k}|$, and in section 4.3 we find:

$$|\alpha_{n,k} - \beta_{n,k}| \ll \frac{k^2}{n}$$

and with k bounded appropriately, we get:

$$|\alpha_{n,k} - \beta_{n,k}| \ll \frac{1}{n^{1-\epsilon}}$$

The significance here is that this is bounded well below the asymptotic function for $E(n,j)$, which we will see in section 4.2 is $\frac{j}{\sqrt{n}}$, for $1 \leq j \leq \frac{\log n}{2\log\log n}$.

Chapter 2

Bounding $a(n,k)$ for $k = 1, 2$

2.1 The Bound

Trivially, $a(n,1) = \lfloor \sqrt{n} \rfloor$. For $a(n,2)$ we show the following

Proposition 2.1

$$a(n,2) = \frac{6}{\pi^2} n \log n + O(n)$$

Proof:

To begin, note that two integers t_1 and t_2 for which $t_1 t_2$ is a square can be written in the form $t_1 = wr^2$, $t_2 = ws^2$, where w is squarefree. Then:

$$\begin{aligned}
a(n,2) &= \sum_{\substack{w \le n \\ w \text{ squarefree}}} \sum_{1 \le r,s \le \sqrt{\frac{n}{w}}} 1 \\
&= \sum_{\substack{w \le n \\ w \text{ squarefree}}} \left(\sqrt{\frac{n}{w}} + O(1) \right)^2 \\
&= \sum_{\substack{w \le n \\ w \text{ squarefree}}} \left(\frac{n}{w} + O\left(\sqrt{\frac{n}{w}} \right) \right) \\
&= n \sum_{\substack{w \le n \\ w \text{ squarefree}}} \frac{1}{w} + \sqrt{n} \sum_{\substack{w \le n \\ w \text{ squarefree}}} O\left(\frac{1}{\sqrt{w}} \right)
\end{aligned}$$

The error term is easily $O(n)$. Turning to the main term, we first state Perron's formula, from page 104 of [5]:

$$\frac{1}{2i\pi}\int_{c-i\infty}^{c+i\infty} y^s \frac{ds}{s} = \begin{cases} 0 & \text{if } y < 1 \\ \frac{1}{2} & \text{if } y = 1 \\ 1 & \text{if } y > 1 \end{cases} \tag{2.1}$$

where $c > 0$.

We will use make use of this now, as well as several times in the next chapter. Here we will choose $\frac{n}{w}$ to take the place of y in the formula. Thus multiplying the sum in our main term by this integral allows us to extend the sum to include all squarefree w greater than 1:

$$\begin{aligned} n \sum_{\substack{w \leq n \\ w \text{ squarefree}}} \frac{1}{w} &= n\frac{1}{2i\pi} \sum_{\substack{w \geq 1 \\ w \text{ squarefree}}} \frac{1}{w} \int_{c-i\infty}^{c+i\infty} \left(\frac{n}{w}\right)^s \frac{ds}{s} \\ &= n\frac{1}{2i\pi} \int_{c-i\infty}^{c+i\infty} \frac{n^s}{s} \sum_{\substack{w \geq 1 \\ w \text{ squarefree}}} \frac{1}{w^{s+1}} ds \\ &= n\frac{1}{2i\pi} \int_{c-i\infty}^{c+i\infty} \frac{n^s}{s} \prod_p \left(1 + \frac{1}{p^{s+1}}\right) ds \end{aligned} \tag{2.2}$$

We choose c to be $\frac{\log\log n}{\log n}$.

Now define:

$$\begin{aligned} A(s) &= \prod_p \left(1 + \frac{1}{p^{s+1}}\right)\left(1 - \frac{1}{p^{s+1}}\right) \\ &= \prod_p \left(1 - \frac{1}{p^{2s+2}}\right) \end{aligned}$$

so $A(s) = \zeta(2s+2)^{-1}$, which converges absolutely in the region $Re(s) > -1/2$, and our integral becomes:

$$\frac{1}{2i\pi}\int_{c-i\infty}^{c+i\infty}\frac{n^s}{s}\frac{\zeta(s+1)}{\zeta(2s+2)}ds$$

Instead, we'll evaluate

$$\frac{1}{2i\pi}\int_{c-iT}^{c+iT}\frac{n^s}{s}\frac{\zeta(s+1)}{\zeta(2s+2)}ds$$

where we choose $T=\log n$. This change is justified by:

Lemma 2.1

$$\left|\frac{1}{2i\pi}\int_{c-i\infty}^{c+i\infty}\frac{n^s}{s}A(s)\zeta(s+1)ds-\frac{1}{2i\pi}\int_{c-iT}^{c+iT}\frac{n^s}{s}A(s)\zeta(s+1)ds\right|=O(1)$$

when $A(s)$ is absolutely convergent in $Re(s)>-1/2$, $T=\log n$, and c is chosen to be $\frac{\log\log n}{\log n}$.

Proof:

$$\begin{aligned}
&\left|\frac{1}{2i\pi}\int_{c-i\infty}^{c+i\infty}\frac{n^s}{s}A(s)\zeta(s+1)ds-\frac{1}{2i\pi}\int_{c-iT}^{c+iT}\frac{n^s}{s}A(s)\zeta(s+1)ds\right|\\
&=\left|\frac{1}{2i\pi}\sum_{\substack{w\geq 1\\ w\text{ squarefree}}}\frac{1}{w}\int_{c-i\infty}^{c+i\infty}\left(\frac{n}{w}\right)^s\frac{ds}{s}-\frac{1}{2i\pi}\sum_{\substack{w\geq 1\\ w\text{ squarefree}}}\frac{1}{w}\int_{c-iT}^{c+iT}\left(\frac{n}{w}\right)^s\frac{ds}{s}\right|\\
&\leq\sum_{\substack{w\geq 1\\ w\text{ squarefree}}}\frac{1}{w}\left|\frac{1}{2i\pi}\int_{c-i\infty}^{c+i\infty}\left(\frac{n}{w}\right)^s\frac{ds}{s}-\frac{1}{2i\pi}\int_{c-iT}^{c+iT}\left(\frac{n}{w}\right)^s\frac{ds}{s}\right|\\
&\leq\sum_{\substack{w\geq 1\\ w\text{ squarefree}\\ w\neq n}}\left[\frac{1}{w}\left(\frac{n}{w}\right)^c\min\left(1,T^{-1}\left|\log\left(\frac{n}{w}\right)\right|^{-1}\right)\right]+\frac{c}{nT},
\end{aligned}$$

where the last inequality is a result from page 105 of [5].

It is clear from our choice of c and T that the last term, $\frac{c}{nT}$, is $o(1)$. The rest of the sum breaks naturally into three pieces, and we need each of these to be $o(\log n)$. Setting $\alpha = \exp\left(\frac{1}{\log n}\right)$, the pieces are:

1. $\frac{n}{\alpha} < w < \alpha n$, $w \neq n$, where $1 < T^{-1}\left|\log \frac{n}{w}\right|^{-1}$,
2. $1 \leq w \leq \frac{n}{\alpha}$, and
3. $w \geq \alpha n$

In each of the last two cases, $T^{-1}\left|\log \frac{n}{w}\right|^{-1} < 1$.

For the first piece:

$$\begin{aligned} n^c \sum_{\substack{\frac{n}{\alpha} \leq w \leq \alpha n \\ w \text{ squarefree}}} \frac{1}{w^{1+c}} &= O\left(\log n \left[-w^{-c}\right]_{\frac{n}{\alpha}}^{\alpha n}\right) \\ &= O\left(\log n \left(\frac{1}{\log n}\left(\frac{\alpha^{2c}-1}{\alpha^c}\right)\right)\right) \\ &= o(1) \end{aligned}$$

For the second piece:

$$\begin{aligned} \frac{n^c}{T} \sum_{\substack{1 \leq w \leq \frac{n}{\alpha} \\ w \text{ squarefree}}} \frac{1}{w^{1+c}\log \frac{n}{w}} &= O\left(\sum_{1 \leq w \leq \frac{n}{\alpha}} \frac{1}{w^{1+c}}\right) \\ &= O\left(\left[-w^{-c}\right]_1^{\frac{n}{\alpha}}\right) \\ &= O\left(1 - \frac{\alpha^c}{\log n}\right) \\ &= O(1) \end{aligned}$$

And for the third piece:

$$\begin{aligned}\frac{n^c}{T}\sum_{\substack{w\geq\alpha n\\ w \text{ squarefree}}}\frac{1}{w^{1+c}\log\frac{w}{n}} &= O\left(\left[-w^{-c}\right]_{\alpha n}^{\infty}\right)\\ &= O\left((\alpha n)^{-c}\right)\\ &= O\left(\frac{1}{\log n}\right)\end{aligned}$$

This completes the proof of Lemma 2.1.

We'll evaluate our new integral

$$\frac{1}{2i\pi}\int_{c-iT}^{c+iT}\frac{n^s}{s}\frac{\zeta(s+1)}{\zeta(2s+2)}ds$$

as part of a contour integral around a rectangle with corners at $c-iT$, $c+iT$, $-a+iT$, and $-a-iT$, where $a=1/4$. So we need to bound the three sides of the integral we're not interested in. For this, we make use of the following bound for $\zeta(s+1)$ from pages 95 and 96 of [23]:

$$|\zeta(\sigma+iT)| = O\left(|T|^{\frac{1}{2}-\frac{1}{2}\sigma}\right) \tag{2.3}$$

for $0<\sigma<1$.

Choosing $T=\log n$, the vertical side of the integral inside the critical strip is:

$$\begin{aligned}\left|\int_{a-iT}^{-a+iT}\frac{\zeta(s+1)}{\zeta(2s+2)}\frac{n^s}{s}ds\right| &\ll \left|\int_{a-iT}^{-a+iT}T^{\frac{1}{2}+\frac{1}{2}a}n^{-a}ds\right|\\ &\ll \frac{T^{13/8}}{n^a}\\ &= o(1)\end{aligned} \tag{2.4}$$

For the two horizontal sides of the integral, we have:

$$\begin{aligned}
\left| \int_{a+iT}^{c+iT} \frac{\zeta(s+1)}{\zeta(2s+2)} \frac{n^s}{s} ds \right| &\ll \frac{1}{T} \int_{-a+iT}^{c+iT} |\zeta(s+1)| \, n^{Re(s)} ds \\
&\ll \frac{1}{T} \int_{-a}^{c} T^{\frac{1}{2}-\frac{1}{2}\sigma} n^{\sigma} d\sigma \\
&\ll \frac{1}{\log\left(\frac{n}{\sqrt{T}}\right)} \frac{n^{\sigma}}{T^{\frac{1}{2}+\frac{1}{2}\sigma}} \Bigg]_{-a}^{c} \\
&\ll \frac{\log n}{(\log n)^{\frac{3}{2}+\epsilon}} \\
&= o(1) \qquad (2.5)
\end{aligned}$$

Then combining (2.4) and (2.5) with the residue theorem, we have:

$$\frac{1}{2i\pi} \int_{c-iT}^{c+iT} \frac{\zeta(s+1)}{\zeta(2s+2)} \frac{n^s}{s} ds = (\text{Residue at } 0) + o(1)$$

Expanding the factors of the integrand, the residue is:

$$\frac{1}{\zeta(2)} \log n + O(1) = \frac{6}{\pi^2} \log n + O(1)$$

and combining this with (2.2) gives us our result.

This completes the proof of Proposition 2.1.

Chapter 3

Bounding $a(n,k)$ for $k=3$

3.1 Setup

In this chapter, we prove the following:

Proposition 3.1

$$a(n,3) = \frac{1}{4}\left(\prod_p \left(1+\frac{3}{p}\right)\left(1-\frac{1}{p}\right)^3\right) n^{3/2}\log^3 n + O\left(n^{3/2}\log^2 n\right) \tag{3.1}$$

In Section 3.2, we'll handle the main term, and in Section 3.3, we'll bound the error.

First we consider three integers whose product is a square They may be written as $t_1 = u^2xy$, $t_2 = v^2xz$, and $t_3 = w^2yz$, where $\frac{t_1}{u^2}$, $\frac{t_2}{v^2}$, and $\frac{t_3}{w^2}$ are all squarefree. Thus x, y, and z will be squarefree and pairwise coprime.

We will first consider the few cases in which at least two of x, y, and z are equal. Since xy, xz, and yz are squarefree, this can only happen when:

1. $x = y = 1$,
2. $x = z = 1$,
3. $y = z = 1$

These of course overlap when $x = y = z = 1$, but as we are only concerned with an upper bound, this is not an issue. Each of these cases will have the same form, so we begin by using case one as a model, and we have:

$$\begin{aligned}
&\sum_{\substack{z\le n\\ z \text{ squarefree}}} \sum_{1\le u\le\sqrt{n}} \sum_{1\le v\le\sqrt{\frac{n}{z}}} \sum_{1\le w\le\sqrt{\frac{n}{z}}} 1 \\
&= \sum_{\substack{z\le n\\ z \text{ squarefree}}} \left(\sqrt{n}+O(1)\right)\left(\sqrt{\frac{n}{z}}+O(1)\right)\left(\sqrt{\frac{n}{z}}+O(1)\right) \\
&= \sum_{\substack{z\le n\\ z \text{ squarefree}}} \left(\frac{n^{3/2}}{z}+O\left(\frac{n}{\sqrt{z}}\right)\right) \\
&= \left(n^{3/2} \sum_{\substack{z\le n\\ z \text{ squarefree}}} \frac{1}{z}\right)+O\left(n^{3/2}\right)
\end{aligned}$$

This sum is the same one we handled via (2.1) in the last chapter. Using those results, the count in each of these three cases is $O\left(n^{3/2}\log n\right)$, which is within our stated error term.

When x, y, and z are distinct, we order them as $x<y<z$, and then multiply by the number of rearrangements, $3!=6$. Let $d=yz$, and write:

$$\begin{aligned}
a(n,3) &= 6 \sum_{\substack{d\le n\\ d \text{ squarefree}}} \sum_{\substack{yz=d\\ y<z\\ \gcd(y,z)=1}} \sum_{\substack{1\le x<y\\ \gcd(x,yz)=1\\ x \text{ squarefree}}} \sum_{1\le u\le\sqrt{\frac{n}{xy}}} \sum_{1\le v\le\sqrt{\frac{n}{xz}}} \sum_{1\le w\le\sqrt{\frac{n}{yz}}} 1 \\
&= 6 \sum_{\substack{d\le n\\ d \text{ squarefree}}} \sum_{\substack{yz=d\\ y<z\\ \gcd(y,z)=1}} \sum_{\substack{1\le x<y\\ \gcd(x,yz)=1\\ x \text{ squarefree}}} \left(\sqrt{\frac{n}{xy}}+O(1)\right)\left(\sqrt{\frac{n}{xz}}+O(1)\right)\left(\sqrt{\frac{n}{yz}}+O(1)\right) \\
&= 6 \sum_{\substack{d\le n\\ d \text{ squarefree}}} \sum_{\substack{yz=d\\ y<z\\ \gcd(y,z)=1}} \sum_{\substack{1\le x<y\\ \gcd(x,yz)=1\\ x \text{ squarefree}}} \left(\frac{n^{3/2}}{xd}+O\left(\frac{n}{x\sqrt{d}}\right)\right) \qquad (3.2)
\end{aligned}$$

Our next step is to evaluate the inner sum. To do so, we use:

Lemma 3.1

$$\sum_{\substack{1\le x\le y\\ \gcd(x,d)=1\\ x \text{ squarefree}}} \frac{1}{x} = \frac{6}{\pi^2}\log y \prod_{p|d}\left(\frac{p}{p+1}\right) + O(1) \tag{3.3}$$

Proof:

We again invoke (2.1) to get:

$$\begin{aligned}\sum_{\substack{1\le x\le y\\ \gcd(x,d)=1\\ x \text{ squarefree}}} \frac{1}{x} &= \sum_{\substack{x\ge 1\\ \gcd(x,d)=1\\ x \text{ squarefree}}} \frac{1}{x}\frac{1}{2i\pi}\int_{c-i\infty}^{c+i\infty}\left(\frac{y}{x}\right)^s\frac{ds}{s}\\ &= \frac{1}{2i\pi}\int_{c-i\infty}^{c+i\infty}\frac{y^s}{s}\sum_{\substack{x\ge 1\\ \gcd(x,d)=1\\ x \text{ squarefree}}}\frac{1}{x^{s+1}}ds,\end{aligned}$$

where we choose $c = \frac{\log\log y}{\log y}$. The inner sum can be rewritten as $\prod_{p\nmid d}\left(1+\frac{1}{p^{s+1}}\right)$. So we'll define

$$\begin{aligned}A(s) &= \prod_{p\nmid d}\left(1+\frac{1}{p^{s+1}}\right)\left(1-\frac{1}{p^{s+1}}\right)\prod_{p|d}\left(1-\frac{1}{p^{s+1}}\right)\\ &= \prod_{p|d}\left(1-\frac{1}{p^{s+1}}\right)\prod_{p\nmid d}\left(1-\frac{1}{p^{2s+2}}\right),\end{aligned}$$

which converges absolutely in $Re(s) > -1/2$. This allows us to write the inner sum as $A(s)\zeta(s+1)$, so that we want to evaluate the integral:

$$\int_{c-i\infty}^{c+i\infty} A(s)\zeta(s+1)\frac{y^s}{s}ds$$

We've already done this work in the case $k = 2$, with a function $A(s)$ that, although different, was also absolutely convergent in the same region. This allows us to use those results, and so:

$$\begin{aligned}\sum_{\substack{1\le x\le y\\ \gcd(x,d)=1\\ x \text{ squarefree}}} \frac{1}{x} &= A(0)\log y + O(1)\\ &= \prod_{p\mid d}\left(1-\frac{1}{p}\right)\prod_{p\nmid d}\left(1-\frac{1}{p^2}\right)\log y + O(1)\\ &= \frac{1}{\zeta(2)}\prod_{p\mid d}\left(\frac{p}{p+1}\right)\log y + O(1)\end{aligned}$$

This completes the proof of Lemma 3.1.

3.2 The Main Term

Using (3.3) for the inner sum in (3.2), along with the results from section 3.3, we get

$$\begin{aligned}a(n,3) &= \frac{36}{\pi^2}n^{3/2}\sum_{\substack{d\le n\\ d \text{ squarefree}}}\frac{1}{d}\sum_{\substack{yz=d\\ y<z\\ \gcd(y,z)=1}}\left(\log y\prod_{p\mid d}\frac{p}{p+1}\right)+O\left(n^{3/2}\log^2 n\right)\\ &= \frac{36}{\pi^2}n^{3/2}\sum_{\substack{d\le n\\ d \text{ squarefree}}}\prod_{p\mid d}\frac{1}{p+1}\sum_{\substack{y\mid d\\ y<z\\ \gcd(y,z)=1}}\log y+O\left(n^{3/2}\log^2 n\right)\\ &= \frac{36}{\pi^2}n^{3/2}\sum_{\substack{y\le\sqrt{n}\\ y \text{ squarefree}}}\frac{\log y}{\prod_{p\mid y}(p+1)}\sum_{\substack{y<z\le n/y\\ \gcd(y,z)=1\\ z \text{ squarefree}}}\frac{1}{\prod_{p\mid z}(p+1)}+O\left(n^{3/2}\log^2 n\right)\end{aligned}\tag{3.4}$$

We will evaluate the inner sum for z up to an arbitrary bound, M, and then take the difference of the results when $M=\frac{n}{y}$ and when $M=y$. To do this, we will again make use of (2.1):

$$\sum_{\substack{z\leq M\\ \gcd(y,z)=1\\ z \text{ squarefree}}} \frac{1}{\prod_{p|z}(p+1)} = \sum_{\substack{z\geq 1\\ \gcd(y,z)=1\\ z \text{ squarefree}}} \frac{1}{\prod_{p|z}(p+1)} \frac{1}{2i\pi} \int_{c-i\infty}^{c+i\infty} \frac{M^s}{z^s}\frac{ds}{s}$$

$$= \frac{1}{2i\pi}\int_{c-i\infty}^{c+i\infty} \frac{M^s}{s} \sum_{\substack{z\geq 1\\ \gcd(y,z)=1\\ z \text{ squarefree}}} \frac{1}{z^s\prod_{p|z}(p+1)} ds$$

where c is chosen to be $\frac{\log\log M}{\log M}$.

We can rewrite the inner sum as $\prod_{p\nmid y}\left(1+\frac{1}{p^s(p+1)}\right)$, and define:

$$\begin{aligned} A(s) &= \prod_{p\nmid y}\left(1+\frac{1}{p^s(p+1)}\right)\left(1-\frac{1}{p^{s+1}}\right)\prod_{p|y}\left(1-\frac{1}{p^{s+1}}\right) \\ &= \prod_{p|y}\left(1-\frac{1}{p^{s+1}}\right)\prod_{p\nmid y}\left(1-\frac{1}{p^{s+1}}+\frac{1}{p^s(p+1)}-\frac{1}{p^{2s+1}(p+1)}\right) \\ &= \prod_{p|y}\left(1-\frac{1}{p^{s+1}}\right)\prod_{p\nmid y}\left(1-\frac{1}{p^{s+1}(p+1)}-\frac{1}{p^{2s+1}(p+1)}\right) \end{aligned}$$

which converges absolutely in the region $Re(s) > -1/2$. Then we have:

$$\prod_{p\nmid y}\left(1+\frac{1}{p^s(p+1)}\right) = A(s)\zeta(s+1)$$

So we now want to evaluate the integral:

$$\frac{1}{2i\pi}\int_{c-i\infty}^{c+i\infty} A(s)\zeta(s+1)\frac{M^s}{s}ds$$

As in the case $k=2$, we'll instead look at:

$$\frac{1}{2i\pi}\int_{c-iT}^{c+iT} A(s)\zeta(s+1)\frac{M^s}{s}ds$$

where T is chosen to be $\log M$. This change is justified by Lemma 2.1, which tells us that the difference between these two integrals is $O(1)$.

As before, we'll evaluate our new integral by integrating around the rectangle with vertices at $c - iT$, $c + iT$, $-a + iT$, and $-a - iT$, where T is chosen to be $\log M$. Again we choose $a = 1/4$.

So we need to bound the integral along the three sides we've added to the contour, but again the work has been done in the previous section. The results (2.4) and (2.5) still hold, as the change from the $A(s)$ used there to the current choice of $A(s)$ does not affect the region in which it is absolutely convergent. So we have that the integral along each of the three extra sides is $o(1)$.

Expanding each function in the integrand, the significant term in the residue is $A(0)(\log M + O(1))$, and:

$$\begin{aligned} A(0) &= \prod_{p|y}\left(1-\frac{1}{p}\right)\prod_{p\nmid y}\left(1+\frac{1}{p+1}\right)\left(1-\frac{1}{p}\right) \\ &= \prod_{p|y}\left(1+\frac{1}{p+1}\right)^{-1}\prod_{p}\left(1-\frac{1}{p}\right)\left(1+\frac{1}{p+1}\right) \\ &= \prod_{p|y}\frac{p+1}{p+2}\prod_{p}\frac{(p+2)(p-1)}{p(p+1)} \end{aligned}$$

So we have that

$$\begin{aligned} \sum_{\substack{y<z\le n/y \\ \gcd(y,z)=1 \\ z \text{ squarefree}}} \frac{1}{\prod_{p|m}(p+1)} &= \prod_{p|y}\frac{p+1}{p+2}\prod_{p}\frac{(p+2)(p-1)}{p(p+1)}\left(\log\left(\frac{n}{y}\right)-\log y+O(1)\right) \\ &= \log\left(\frac{n}{y^2}\right)\prod_{p|y}\frac{p+1}{p+2}\prod_{p}\frac{(p+2)(p-1)}{p(p+1)}+O(1) \end{aligned}$$

We'll now substitute this into (3.4). First we see that the $O(1)$ gives us:

$$\begin{aligned} & O\left(n^{3/2} \sum_{\substack{y \leq \sqrt{n} \\ y \text{ squarefree}}} \frac{\log y}{\prod_{p|y}(p+1)}\right) \\ = \; & O\left(n^{3/2} \sum_{y \leq \sqrt{n}} \frac{\log y}{y}\right) \\ = \; & O\left(n^{3/2} \log^2 n\right) \end{aligned}$$

which is the size of the stated error term.

For the main term, we'll set $C = \frac{36}{\pi^2} \prod_p \frac{(p+2)(p-1)}{p(p+1)}$, and we have:

$$\begin{aligned} & a(n,3) \\ = \; & Cn^{3/2} \sum_{\substack{y \leq \sqrt{n} \\ y \text{ squarefree}}} \frac{\log y}{\prod_{p|y}(p+1)} \log\left(\frac{n}{y^2}\right) \prod_{p|y} \frac{p+1}{p+2} + O\left(n^{3/2} \log^2 n\right) \\ = \; & Cn^{3/2} \sum_{\substack{y \leq \sqrt{n} \\ y \text{ squarefree}}} \frac{\log n \log y - 2\log^2 y}{\prod_{p|y}(p+2)} + O\left(n^{3/2} \log^2 n\right) \\ = \; & Cn^{3/2} \sum_{\substack{y \geq 1 \\ y \text{ squarefree}}} \frac{\log n \log y - 2\log^2 y}{\prod_{p|y}(p+2)} \frac{1}{2i\pi} \int_{c-i\infty}^{c+i\infty} \left(\frac{\sqrt{n}}{y}\right)^s \frac{ds}{s} + O\left(n^{3/2} \log^2 n\right) \\ = \; & Cn^{3/2} \frac{1}{2i\pi} \int_{c-i\infty}^{c+i\infty} \sum_{\substack{y \geq 1 \\ y \text{ squarefree}}} \left(\frac{\log n \log y - 2\log^2 y}{\prod_{p|y} p^s (p+2)}\right) \frac{n^{s/2}}{s} ds + O\left(n^{3/2} \log^2 n\right) \quad (3.5) \end{aligned}$$

where we have once again made use of (2.1), choosing $c = \frac{\log\log n}{\log n}$. Now we'll define the function:

$$A_1(s) = \sum_{\substack{y \geq 1 \\ y \text{ squarefree}}} \frac{1}{\prod_{p|y} p^s(p+2)}$$

Its first and second derivatives are:

$$A_1'(s) = \sum_{\substack{y \geq 1 \\ y \text{ squarefree}}} \frac{-\log y}{\prod_{p|y} p^s(p+2)}$$

$$A_1''(s) = \sum_{\substack{y \geq 1 \\ y \text{ squarefree}}} \frac{\log^2 y}{\prod_{p|y} p^s(p+2)}$$

so the integral can be written as:

$$\int_{c-i\infty}^{c+i\infty} \left(-\log n A_1'(s) - 2A_1''(s)\right) \frac{n^{s/2}}{s} ds$$

Now write

$$A_1(s) = \prod_p \left(1 + \frac{1}{p^s(p+2)}\right) = \zeta(s+1)B_1(s)$$

so that:

$$\begin{aligned} B_1(s) &= \prod_p \left(1 + \frac{1}{p^s(p+2)}\right)\left(1 - \frac{1}{p^{s+1}}\right) \\ &= \prod_p \left(1 - \frac{2}{p^{s+1}(p+2)} - \frac{1}{p^{2s+1}(p+2)}\right) \end{aligned}$$

$B_1(s)$ then converges absolutely in the region $Re(s) > -1/2$, and serves as the analog to $A(s)$ in the previous integrals.

Looking at the expansion of $A_1(s)$ about 0, we get:

$$A_1(s) = \frac{B_1(0)}{s} + C + Ds + Es^2 + Fs^3 + \ldots$$

so

$$A_1'(s) = -\frac{B_1(0)}{s^2} + D + 2Es + 3Fs^2 + \dots$$

and:

$$A_1''(s) = \frac{2B_1(0)}{s^3} + 2E + 6Fs + \dots$$

Then the integral becomes

$$\int_{c-i\infty}^{c+i\infty} \left(-\frac{4B_1(0)}{s^3} + \frac{B_1(0)\log n}{s^2} + (-D\log n - 4E) + (-2E\log n - 12F)s + \dots \right) \frac{n^{s/2}}{s} ds$$

which we replace with:

$$\int_{c-iT}^{c+iT} \left(-\frac{4B_1(0)}{s^3} + \frac{B_1(0)\log n}{s^2} + (-D\log n - 4E) + (-2E\log n - 12F)s + \dots \right) \frac{n^{s/2}}{s} ds$$

where we choose T such that $|T - \log n| < 1$. The reason for allowing this room in the choice of T will be explained in section 3.5.

The justification for changing the limits on the integral this time is a bit more involved than for the previous integrals, and will be carried out in Section 3.4. The results of that section show that the difference between the values of the two integrals is $O\left(\log^{3/2} n\right)$.

We next consider our integral as one side of the integral around the same contour as before, and so we need to bound the three other sides. This is done in section 3.5. We'll note here that we cannot simply choose $T = \log n$ as we have in previous cases, as we will now have to be careful to avoid zeros of $\zeta(s+1)$ as we integrate through the critical strip. This small change in the size of T will, of course, have no direct effect on the size of the integrals we must bound, but the functions of the integrand will require us now to use some results on the zeros of $\zeta(s)$, and these will be noted as they arise. The result of that section is that those three integrals are $O\left((\log n)^{9/8}(\log\log n)^2\right)$.

So we just need the residue of the integrand at $s = 0$. Since $n^{s/2} = 1 + \frac{s}{2}\log n + \frac{s^2}{8}\log^2 n + \frac{s^3}{48}\log^3 n + \ldots$, this residue is:

$$-\frac{4B_1(0)}{48}\log^3 n + \frac{B_1(0)}{8}\log^3 n + O(\log n) = \frac{B_1(0)}{24}\log^3 n + O(\log n)$$

Combining this with

$$B_1(0) = \prod_p \left(1 + \frac{1}{p+2}\right)\left(1 - \frac{1}{p}\right) = \prod_p \frac{(p+3)(p-1)}{p(p+2)}$$

we get that the main term of $a(n, 3)$ is:

$$\begin{aligned}
&\left(\frac{36}{\pi^2}n^{3/2}\prod_p \frac{(p+2)(p-1)}{p(p+1)}\right)\left(\frac{1}{24}\prod_p \frac{(p+3)(p-1)}{p(p+2)}\log^3 n\right) \\
&= \frac{1}{4\zeta(2)}\left(\prod_p \frac{(p+3)(p-1)^2}{p^2(p+1)}\right) n^{3/2}\log^3 n \\
&= \frac{1}{4}\left(\prod_p \left(1 - \frac{1}{p^2}\right)\frac{(p+3)(p-1)^2}{p^2(p+1)}\right) n^{3/2}\log^3 n \\
&= \frac{1}{4}\left(\prod_p \frac{(p+3)(p-1)^3}{p^4}\right) n^{3/2}\log^3 n \\
&= \frac{1}{4}\left(\prod_p \left(1 + \frac{3}{p}\right)\left(1 - \frac{1}{p}\right)^3\right) n^{3/2}\log^3 n
\end{aligned}$$

3.3 The Error Terms

We want both error terms dealt with here to be $O\left(n^{3/2}\log^2 n\right)$, the error term stated in Proposition 3.1. First we'll bound the error term in (3.2). We're looking at:

$$
\begin{aligned}
& O\left(n \sum_{\substack{d\le n \\ d \text{ squarefree}}} \sum_{\substack{yz=d \\ y<z \\ \gcd(y,z)=1}} \frac{\log y}{\sqrt{d}} \prod_{p\mid d} \frac{p}{p+1} \right) \\
= \; & O\left(n \sum_{\substack{y\le\sqrt{n} \\ y \text{ squarefree}}} \frac{\log y}{\prod_{p\mid y}(p+1)} \sum_{\substack{y<z\le n/y \\ \gcd(z,y)=1 \\ z \text{ squarefree}}} \frac{\sqrt{d}}{\prod_{p\mid z}(p+1)} \right) \\
= \; & O\left(n \sum_{\substack{y\le\sqrt{n} \\ y \text{ squarefree}}} \frac{\sqrt{y}\log y}{\prod_{p\mid y}(p+1)} \sum_{y<z\le n/y} \frac{1}{\sqrt{z}} \right) \\
= \; & O\left(n \sum_{\substack{y\le\sqrt{n} \\ y \text{ squarefree}}} \frac{(\sqrt{n}-y)\log y}{\prod_{p\mid y}(p+1)} \right) \\
= \; & O\left(n^{3/2} \sum_{y\le\sqrt{n}} \frac{\log y}{y} \right) \\
= \; & O\left(n^{3/2}\log^2 n\right)
\end{aligned}
$$

Next we'll bound the error contributed by using the results of Lemma 3.1 to evaluate the main term in (3.2):

$$O\left(n^{3/2}\sum_{\substack{d\le n\\ d\text{ squarefree}}}\sum_{\substack{yz=d\\ y<z\\ \gcd(y,z)=1}}\frac{1}{d}\right) = O\left(n^{3/2}\sum_{y\le n}\frac{1}{y}\sum_{z\le n}\frac{1}{z}\right)$$
$$= O\left(n^{3/2}\log^2 n\right)$$

So both error terms are bounded as needed.

3.4 Bounding the Difference Between the Integrals

We now justify changing the limits of the last integral in section 3.2. We've chosen $c = \frac{\log\log n}{\log n}$ and T such that $|T - \log n| < 1$, so T can avoid any zeros of $\zeta(s+1)$. Then:

$$\left|\frac{1}{2i\pi}\int_{c-i\infty}^{c+i\infty}\sum_{\substack{y\ge 1\\ y\text{ squarefree}}}\frac{\log n\log y - 2\log^2 y}{y^s\prod_{p|y}(p+2)}\frac{n^{s/2}}{s}ds\right.$$
$$\left. - \frac{1}{2i\pi}\int_{c-iT}^{c+iT}\sum_{\substack{y\ge 1\\ y\text{ squarefree}}}\frac{\log n\log y - 2\log^2 y}{y^s\prod_{p|y}(p+2)}\frac{n^{s/2}}{s}ds\right|$$
$$\le \sum_{\substack{y\ge 1\\ y\text{ squarefree}}}\frac{\log y\left|\log\left(\frac{n}{y^2}\right)\right|}{\prod_{p|y}(p+2)}\left|\frac{1}{2i\pi}\int_{c-i\infty}^{c+i\infty}\left(\frac{\sqrt{n}}{y}\right)^s\frac{ds}{s} - \frac{1}{2i\pi}\int_{c-iT}^{c+iT}\left(\frac{\sqrt{n}}{y}\right)^s\frac{ds}{s}\right|$$
$$\le \sum_{\substack{y\ge 1\\ y\text{ squarefree}\\ y\ne\sqrt{n}}}\frac{\log y\left|\log\left(\frac{n}{y^2}\right)\right|}{\prod_{p|y}(p+2)}\left[\left(\frac{\sqrt{n}}{y}\right)^c\min\left(1,\frac{1}{T\left|\log\left(\frac{\sqrt{n}}{y}\right)\right|}\right)\right]$$

where the last inequality is an application of a result on page 105 of [5].

Now we'll break this sum into three pieces, defining $\alpha = \exp\left(\frac{1}{\log n}\right)$:

1. $\frac{\sqrt{n}}{\alpha} < y < \alpha\sqrt{n}$,

2. $1 \leq y \leq \frac{\sqrt{n}}{\alpha}$, and

3. $y \geq \alpha\sqrt{n}$.

In the first case, $1 < T^{-1}\left|\log\left(\frac{\sqrt{n}}{y}\right)\right|^{-1}$, and in the second and third cases, the inequality is reversed. Then to bound the first piece:

$$\begin{aligned}
\sqrt{n}^c & \sum_{\substack{\frac{\sqrt{n}}{\alpha}<y<\alpha\sqrt{n} \\ y \text{ squarefree} \\ y\neq\sqrt{n}}} \frac{\log y \left|\log\left(\frac{n}{y^2}\right)\right|}{y^c \prod_{p|y}(p+2)} \\
& \leq \sqrt{n}^c \sum_{\substack{\frac{\sqrt{n}}{\alpha}<y<\alpha\sqrt{n} \\ y \text{ squarefree}}} \frac{\log y \left|\log\left(\frac{n}{y^2}\right)\right|}{y^{1+c}} \\
& \leq \sqrt{\log n} \sum_{\substack{\frac{\sqrt{n}}{\alpha}<y<\alpha\sqrt{n} \\ y \text{ squarefree}}} \frac{2\log y}{y} \\
& \leq \sqrt{\log n}\left(\log^2\left(\alpha\sqrt{n}\right) - \log^2\left(\frac{\sqrt{n}}{\alpha}\right)\right) \\
& = O\left(\log^{3/2} n\right)
\end{aligned}$$

For the second piece:

$$\begin{aligned}
&\frac{\sqrt{n}^c}{T} \sum_{\substack{1\le y\le \frac{\sqrt{n}}{\alpha} \\ y \text{ squarefree}}} \frac{\log y \log\left(\frac{n}{y^2}\right)}{y^c \prod_{p|y}(p+2)\log\left(\frac{\sqrt{n}}{y}\right)} \\
&\le \frac{\sqrt{\log n}}{T} \sum_{\substack{1\le y\le \frac{\sqrt{n}}{\alpha} \\ y \text{ squarefree}}} \frac{2\log y}{y^{1+c}} \\
&\le \frac{1}{\sqrt{\log n}} \sum_{\substack{1\le y\le \frac{\sqrt{n}}{\alpha} \\ y \text{ squarefree}}} \frac{2\log y}{y} \\
&= O\left(\frac{1}{\sqrt{\log n}} \log^2\left(\frac{\sqrt{n}}{\alpha}\right)\right) \\
&= O\left(\log^{3/2} n\right)
\end{aligned}$$

For the third piece:

$$\begin{aligned}
&\frac{\sqrt{n}^c}{T} \sum_{\substack{y\ge \alpha\sqrt{n} \\ y \text{ squarefree}}} \frac{\log y \log\left(\frac{y^2}{n}\right)}{y^c \prod_{p|y}(p+2)\log\left(\frac{y}{\sqrt{n}}\right)} \\
&\ll \frac{1}{\sqrt{\log n}} \sum_{\substack{y\ge \alpha\sqrt{n} \\ y \text{ squarefree}}} \frac{2\log y}{y^{1+c}} \\
&\ll \frac{1}{\sqrt{\log n}} \sum_{y\ge \alpha\sqrt{n}} \frac{1}{y^{1+\frac{c}{2}}} \\
&\ll \frac{1}{\sqrt{\log n}} \left(\left[-\frac{2}{c} y^{-c/2}\right]_{\alpha\sqrt{n}}^{\infty}\right) \\
&\ll \frac{1}{\sqrt{\log n}} \cdot \frac{1}{n^{c/4}} \\
&= o(1)
\end{aligned}$$

Our main term in this case was $\frac{B_1(0)}{24}\log^3 n$, so the difference is bounded as needed.

3.5 Bounding the Rest of the Contour Integral

Now we'll bound the three other sides of our contour integral. The constant a will be chosen with $\frac{1}{5} < a < \frac{1}{4}$ such that for all zeroes $s = \rho + i\beta$ of $\zeta(s)$, $|\rho - a| \gg \frac{1}{T \log T}$. This choice is to ensure we can bound the size of the integrand along the vertical side of the contour in the critical strip.

We start with the horizontal side:

$$\left| \int_{a+iT}^{c+iT} \left(-\log n A_1'(s) - 2A_1''(s)\right) \frac{n^{s/2}}{s} ds \right|$$

First, we'll rewrite $A_1'(s)$ as:

$$\begin{aligned} A_1'(s) &= \zeta'(s+1)B_1(s) + \zeta(s+1)B_1'(s) \\ &= \zeta(s+1)\left(\frac{\zeta'(s+1)}{\zeta(s+1)}B_1(s) + B_1'(s)\right) \end{aligned}$$

Along with (2.3), we'll need two results here. The first is from page 99 of [5]:

> The number of zeros $\rho = \beta + i\gamma$ of $\zeta(s)$ with $T - 1 < \gamma < T + 1$ is $O(\log T)$.

The second is from page 217 of [23]:

> If $\rho = \beta + i\gamma$ runs through zeros of $\zeta(s)$,
>
> $$\frac{\zeta'(s)}{\zeta(s)} = \sum_{|t-\gamma|\leq 1} \frac{1}{s-\rho} + O(\log t) \tag{3.6}$$
>
> uniformly for $-1 \leq \sigma \leq 2$.

We use the first result to ensure that we can choose our T such that, for any zero $\rho = \beta + i\gamma$ of $\zeta(s)$, $|\gamma - T| \gg (\log T)^{-1}$. Then we have:

$$\begin{aligned} |A_1'(s)| &\ll T^{\frac{1}{2}-\frac{1}{2}\sigma}\left(\log^2 T\right) \\ &\ll T^{5/8}\log^2 T \end{aligned}$$

and using (2.3):

$$\begin{aligned}
|A_1''(s)| &= |\zeta''(s+1)B_1(s) + 2\zeta'(s+1)B_1'(s) + \zeta(s+1)B_1''(s)| \\
&= \left|\zeta(s+1)\left(\frac{\zeta''(s+1)}{\zeta(s+1)}B_1(s) + 2\frac{\zeta'(s+1)}{\zeta(s+1)}B_1'(s) + B_1''(s)\right)\right| \\
&\ll T^{5/8}\left(\log^2 T + \frac{\zeta''(s+1)}{\zeta(s+1)}\right)
\end{aligned}$$

Since $|T - \log n| \leq 1$, and $c = \frac{\log\log n}{\log n}$, we have:

$$\begin{aligned}
&\left|\int_{a+iT}^{c+iT} \left(-\log n A_1'(s) - 2A_1''(s)\right)\frac{n^{s/2}}{s}ds\right| \\
&\ll \left|\int_{a+iT}^{c+iT}\left(-\log n\left((\log n)^{5/8}(\log\log n)^2\right)\right.\right. \\
&\qquad \left.\left. - 2(\log n)^{5/8}\left((\log\log n)^2 + \frac{\zeta''(s+1)}{\zeta(s+1)}\right)\right)\frac{n^{s/2}}{s}ds\right| \\
&\ll \frac{1}{\sqrt{\log n}}\left|\int_{a+iT}^{c+iT}\left((\log n)^{13/8}(\log\log n)^2 + \frac{\zeta''(s+1)}{\zeta(s+1)}\right)ds\right|
\end{aligned}$$

We need this expression to be at most $O\left((\log n)^2\right)$, the size of our error term. We have:

$$\begin{aligned}
\frac{1}{\sqrt{\log n}}\left|\int_{a+iT}^{c+iT}(\log n)^{13/8}(\log\log n)ds\right| &= O\left((\log n)^{9/8}(\log\log n)^2\right) \\
&= o\left((\log n)^2\right)
\end{aligned}$$

and since $\dfrac{\zeta''(s+1)}{\zeta(s+1)} = \dfrac{d}{ds}\dfrac{\zeta'(s+1)}{\zeta(s+1)} + \left(\dfrac{\zeta'(s+1)}{\zeta(s+1)}\right)^2$,

$$\begin{aligned}
&\frac{1}{\sqrt{\log n}}\left|\int_{a+iT}^{c+iT}\frac{\zeta''(s+1)}{\zeta(s+1)}ds\right| \\
&= \frac{1}{\sqrt{\log n}}\left(\left|\int_{a+iT}^{c+iT}\frac{d}{ds}\frac{\zeta'(s+1)}{\zeta(s+1)}ds\right|+\left|\int_{a+iT}^{c+iT}\left(\frac{\zeta'(s+1)}{\zeta(s+1)}\right)^2 ds\right|\right) \\
&\ll \frac{(\log T)^4}{\sqrt{\log n}} \\
&= o(1)
\end{aligned}$$

in which we have used (3.6).

For the vertical side, we use the fact that the number of zeros ρ of $\zeta(s)$ in the critical strip with $0 \leq \mathrm{Im}(\rho) \leq T$ is $O(T\log T)$, along with our choice of a and (3.6), to get:

$$\begin{aligned}
&\left|\int_{a-iT}^{-a+iT}\left(-\log n A_1'(s) - 2A_1''(s)\right)\frac{n^{s/2}}{s}ds\right| \\
&= \left|\int_{a-iT}^{-a+iT}\left(-\log n\left((\log n)^{5/8}\left((\log n)^2(\log\log n)^2\right)\right)\right.\right. \\
&\qquad \left.\left. - 2(\log n)^{5/8}\left(\frac{\zeta''(s+1)}{\zeta(s+1)}B_1(s) + 2(\log n)^2(\log\log n)^2 B_1'(s) + B_1''(s)\right)\right)\frac{n^{s/2}}{s}ds\right| \\
&\ll \frac{T(\log n)^{29/8}(\log\log n)^2}{n^{a/2}} + \frac{1}{n^{a/2}}\left|\int_{a-iT}^{-a+iT}\frac{\zeta''(s+1)}{\zeta(s+1)}ds\right|
\end{aligned}$$

Now the first term is $o(1)$, and we will split the integral as we have previously to get:

$$\begin{aligned}
&\ll \frac{1}{n^{a/2}}\left|\int_{a-iT}^{-a+iT}\frac{d}{ds}\frac{\zeta'(s+1)}{\zeta(s+1)}ds\right| + \frac{1}{n^{a/2}}\left|\int_{a-iT}^{-a+iT}\left(\frac{\zeta'(s+1)}{\zeta(s+1)}\right)^2 ds\right| + o(1) \\
&\ll \frac{\left(T^2(\log T)^2\right)^2}{n^{a/2}} + \frac{T\left(T^2(\log T)^2\right)^2}{n^{a/2}} \\
&= o(1)
\end{aligned}$$

Thus all three sides are bounded to be within our error term.

Chapter 4

More general k

4.1 Upper and Lower Bounds for $a(n,k)$

In this chapter we prove

Proposition 4.1 *We have the following bounds for $a(n,k)$:*

$$a(n,k) < n^{k/2}(1+\log n)^{k(k-1)/2}$$

for k in the range $1 \le k \le \frac{\log n}{\log(1+\log n)}$

$$a(n,k) > \frac{n^{k/2}}{2^k}\left(\frac{c\log n}{k\log(k\log n\log\log n)}\right)^{k(k-1)/2}$$

for some constant c and for k in the range $1 \le k \le \frac{\log n}{2\log\log n}$

Proof:

Recall that $a(n,k)$ is defined as:

$$\#\{T=\{t_1,\ldots,t_k\} : t_i \in \{1,\ldots,n\}, \text{ and} \prod_{i=1}^{k} t_i \text{ is a square}\}$$

To find an upper bound on $a(n,k)$, note that we can write $t_i := c_i {x_i}^2$, with each c_i square-free. We know $\prod_{i=1}^{k} c_i$ is a square, so write $c_i := \prod_{j\neq i} c_{i,j}$, where $c_{i,j}$ is square-free and $c_{i,j} = c_{j,i}$. Then given c_i, the number of possibilities for x_i is at most $\sqrt{\frac{n}{c_i}}$, so the number of k-tuples $x_1,\ldots,x_k$ given $c_1,\ldots,c_k$ is at most $\left(\frac{n^k}{c_1\cdots c_k}\right)^{1/2}$. But we can write $c_1\cdots c_k = \left(\prod_{1\le i<j\le k} c_{i,j}\right)^2$. Thus the number of k-tuples $x_1,\ldots,x_k$ given $c_1,\ldots,c_k$ is at most $\frac{n^{k/2}}{\prod_{1\le i<j\le k} c_{i,j}}$.

Thus $a(n,k)$, the number of sets of t_i's, is at most:

$$n^{k/2}\prod_{1\le i<j\le k}\sum_{c_{i,j}\le n}\frac{1}{c_{i,j}} \le n^{k/2}(1+\log n)^{k(k-1)/2}$$

This bound is only meaningful as long as it remains below n^k. This is true as long as:

$$k - 1 < \frac{\log n}{\log (1 + \log n)}$$

thus giving us the range for k in Proposition 4.1.

For a lower bound, using the same language as for the upper bound, we need $\prod_{i=1}^{k} t_i \leq n^k$, and thus we must have $\prod_{i=1}^{k} c_i \leq n^k$. So we limit each $c_{i,j}$ to be at most $n^{1/(k-1)}$, guaranteeing that each $c_i \leq n$, and thus:

$$\prod_{i=1}^{k} c_i = \left(\prod_{1 \leq i < j \leq k} c_{i,j} \right)^2 \leq \left(\left(n^{1/(k-1)} \right)^{\frac{k(k-1)}{2}} \right)^2 = n^k$$

So:

$$a(n,k) \geq \sum_{\substack{c_{i,j},\, 1 \leq i < j \leq k \\ \text{pairwise coprime integers} \leq n^{\frac{1}{k-1}}}} \prod_{l=1}^{k} \left\lfloor \sqrt{\frac{n}{c_l}} \right\rfloor$$

Now place a dictionary ordering on $\{c_{i,j}\}$, and define $m_{i,j} = \prod_{(i',j') < (i,j)} c_{i',j'}$. For all (i,j), we have $m_{i,j} \leq \left(n^{1/(k-1)} \right)^{\binom{k}{2}} = n^{k/2}$. Using this, and the fact that for any number $t \geq 1$, $\lfloor t \rfloor \geq \frac{t}{2}$, we have:

$$\begin{aligned} a(n,k) &\geq \frac{1}{2^k} \sum_{\substack{c_{i,j},\, 1 \leq i < j \leq k \\ \text{pairwise coprime integers} \leq n^{\frac{1}{k-1}}}} \frac{n^{k/2}}{\prod_{1 \leq i < j \leq k} c_{i,j}} \\ &= \frac{n^{k/2}}{2^k} \sum_{c_{1,2} \leq n^{1/(k-1)}} \frac{1}{c_{1,2}} \sum_{\substack{c_{1,3} \leq n^{1/(k-1)} \\ \gcd(c_{1,3}, m_{1,3}) = 1}} \frac{1}{c_{1,3}} \cdots\cdots \sum_{\substack{c_{k-1,k} \leq n^{1/(k-1)} \\ \gcd\left(c_{k-1,k}, m_{k-1,k}\right) = 1}} \frac{1}{c_{k-1,k}} \end{aligned} \tag{4.1}$$

Considering each individual sum, define $M = \prod_{p \leq k \log n \log \log n} p$, so that the primes dividing M are the smallest primes, and for $k < \frac{\log n}{\log \log n}$, at least $\frac{k}{2} \log n$ primes divide M. Now for each pair (i,j), we have $m_{i,j} \leq n^{k/2}$, so there are fewer than $\frac{k}{2} \log n$ primes dividing $m_{i,j}$.

Thus considering integers coprime to M allows fewer terms in each sum in comparison to considering integers coprime to each $m_{i,j}$. So we have:

$$\sum_{\substack{c_{i,j}\leq n^{1/(k-1)}\\ \gcd(c_{i,j},m_{i,j})=1}} \frac{1}{c_{i,j}} \geq \sum_{\substack{r\leq n^{1/(k-1)}\\ \gcd(r,M)=1}} \frac{1}{r} \tag{4.2}$$

Now in the interval $[x, 2x]$, there are at least $c\frac{x}{\log y}$ integers with all prime factors greater than y, provided $x > y$, for some constant $c > 0$, by the fundamental lemma of the sieve. So we can get a lower bound on the above sum by:

$$\begin{aligned}
\sum_{\substack{r\leq n^{1/(k-1)}\\ \gcd(r,M)=1}} \frac{1}{r} &\geq \sum_{i=1}^{\left\lceil \log_2\left(\frac{n^{1/(k-1)}}{k\log n\log\log n}\right)\right\rceil} \sum_{\frac{n^{1/(k-1)}}{2^i}\leq r\leq \frac{n^{1/(k-1)}}{2^{i-1}}} \frac{1}{r} \\
&\geq \sum_{i=1}^{\left\lceil \log_2\left(\frac{n^{1/(k-1)}}{k\log n log\log n}\right)\right\rceil} c\frac{\left(\frac{n^{1/(k-1)}}{2^i}\right)}{\log\left(k\log n\log\log n\right)}\cdot\frac{2^{i-1}}{n^{1/(k-1)}} \\
&\geq \frac{c}{2\log\left(k\log n\log\log n\right)} \sum_{i=1}^{\left\lceil \log_2\left(\frac{n^{1/(k-1)}}{k\log n\log\log n}\right)\right\rceil} 1 \\
&\geq \frac{c\log n}{k\log\left(k\log n\log\log n\right)}
\end{aligned}$$

provided $n^{1/(k-1)} > 2k\log n\log\log n$.

So then, provided $k < \frac{\log n}{2\log\log n}$, our lower bound is:

$$a(n,k) \geq \frac{n^{k/2}}{2^k}\left(\frac{c\log n}{k\log\left(k\log n\log\log n\right)}\right)^{k(k-1)/2}$$

This completes the proof of Proposition 4.1.

We can unify both bounds as:

$$a(n,k) = n^{k/2}(\log n)^{(1+o(1))k(k-1)/2}$$

as long as $k < (\log n)^{o(1)}$.

4.2 Application to $E(n, j)$

Now recall that:

$$E(n,j) = \sum_{k=1}^{j} \binom{j}{k} \frac{a(n,k)}{n^k}$$

We'll start by plugging in the upper bound from the previous section:

$$E(n,j) \leq \sum_{k=1}^{j} \binom{j}{k} \frac{(1+\log n)^{k(k-1)/2}}{n^{k/2}}$$

and then consider whether these terms are increasing or decreasing. Naming the kth term T_k:

$$\frac{T_{k+1}}{T_k} = \frac{(j-k)(1+\log n)^k}{(k+1)n^{1/2}}$$

For small k, this ratio is certainly less than one. It remains less than one as long as $k < \dfrac{\log n}{2\log\log n}$. So for j (and hence k) $< \dfrac{\log n}{2\log\log n}$, the largest term in the series is $\dfrac{j \cdot a(n,1)}{n}$, when $k = 1$. So we can write:

$$\begin{aligned} \left| E(n,j) - \frac{j \cdot a(n,1)}{n} \right| &= \sum_{k=2}^{j} \binom{j}{k} \frac{a(n,k)}{n^k} \\ &\leq \max_{2 \leq k \leq j} \left\{ \frac{a(n,k)}{n^k} \right\} \sum_{k=2}^{j} \binom{j}{k} \end{aligned}$$

Now $\sum_{k=2}^{j} \binom{j}{k} \leq 2^j$, and combining this with the upper bound for $a(n,k)$, we have:

$$\leq 2^j \max_{2 \leq k \leq j} \left\{ \frac{(1+\log n)^{k(k-1)/2}}{n^{k/2}} \right\}$$

As long as we keep $1 \leq j \leq (1-\epsilon)\dfrac{\log n}{\log(1+\log n)}$, we'll ensure that $(1+\log n)^k \leq n^{1-\epsilon}$, so we can then bound the above expression as:

$$\begin{aligned} &\leq 2^j \max_{2\leq k\leq j}\left\{\frac{n^{(k-1)(1-\epsilon)/2}}{n^{k/2}}\right\} \\ &< \frac{1}{n^{\frac{1}{2}+\epsilon}} \end{aligned}$$

So we have that:

$$E(n,j) \sim \frac{j}{\sqrt{n}} \tag{4.3}$$

4.3 Consideration of repeated elements

We have worked thus far assuming that each element in our sequences is chosen independently, so that we may have elements which appear more than once. In practice, the bound from page 2 of [19] tells us that we are choosing so few elements relative to n, from among $\{1,\ldots,n\}$, that we don't expect to choose the same element twice. So the results above should remain substantially the same if we consider only sequences with no repeated elements. To prove this, we begin by recalling the definitions given in chapter 1:

$$\begin{aligned} \alpha_{n,k} &:= \frac{a(n,k)}{n^k} \\ b(n,k) &:= \#\left\{(m_1,\ldots,m_k)\leq n : \prod_{i=1}^{k} m_i \text{ is a square, } m_i \text{ distinct}\right\} \\ \beta_{n,k} &:= \frac{b(n,k)}{\prod_{i=0}^{k-1}(n-i)} \end{aligned}$$

We also define here $\beta'_{n,k} := \frac{b(n,k)}{n^k}$. The difference between $a(n,k)$ and $b(n,k)$ is the set of sequences which contain at least one element which appears at least twice. Now:

$$\begin{aligned}\frac{1}{\prod_{i=0}^{k-1}(n-i)} &= \frac{1}{n^k\prod_{i=0}^{k-1}\left(1-\frac{i}{n}\right)} \\ &= \frac{1}{n^k}\left(1+O\left(\frac{k^2}{n}\right)\right)\end{aligned}$$

so that $\beta_{n,k} = \beta'_{n,k}\left(1+O\left(\frac{k^2}{n}\right)\right)$, and so $\beta_{n,k}-\beta'_{n,k} = O\left(\frac{k^2}{n}\right)$, as $\beta'_{n,k} \leq 1$.

Now clearly $\alpha_{n,k} \geq \beta'_{n,k}$. Then we have

$$\alpha_{n,k}-\beta'_{n,k} = \frac{1}{n^k}\sum_{1\leq i<j\leq k}\sum_{m_i=m_j\leq n} \#\left\{m_l \leq n \text{ for } 1\leq l\leq k, l\neq i,j : \prod_{l\neq i,j} m_l \text{ is a square}\right\}$$

The set we are counting inside the sum, then, consists of sequences of size $k-2$, in which each element is chosen independently and with uniform distribution from $\{1,\ldots,n\}$. Thus this count is just $a(n,k-2)$, so we can pull it out of the sums to get:

$$\alpha_{n,k}-\beta'_{n,k} = \binom{k}{2}\frac{1}{n}\alpha_{n,k-2} \leq \frac{k^2}{2n}$$

as $\alpha_{n,k-2} \leq 1$. So we have:

$$\begin{aligned}|\alpha_{n,k}-\beta_{n,k}| &\leq (\alpha_{n,k}-\beta'_{n,k})+(\beta_{n,k}-\beta'_{n,k}) \\ &= O\left(\frac{k^2}{n}\right)\end{aligned}$$

Now we define $L(n) := \exp\left(\sqrt{\log n \log\log n}\right)$. As long as we bound k by a fixed power of $L(n)$, say $L(n)^{10}$ (which is much larger than the bounds for k we have used in this chapter, all of which are of the form $(\log n)^{1+o(1)}$), so that we easily include the interval from Pomerance's result on page 2 of [19], we have:

$$|\alpha_{n,k}-\beta_{n,k}| = O\left(\frac{L(n)^{20}}{n}\right) = O\left(\frac{1}{n^{1-\epsilon}}\right)$$

which bounds the difference well below $\frac{j}{\sqrt{n}}$, which we saw, in (4.3), to be the asymptotic function for $E(n,j)$.

Chapter 5

New Integer Representations as the Sum of Three Cubes[1]

[1]Beck, M., E. Pine, W. Tarrant, and K. Yarbrough. Submitted to *Mathematics of Computation,* 2/9/2002.

5.1 Introduction

The table below lists the new solutions we found using the method described in this paper.

Table 5.1: Solutions to the equation $x^3 + y^3 + z^3 = k$

k	(x, y, z)	*Date*
30	(-283059965, -2218888517, 2220422932)	7/10/99
52	(60702901317, 23961292454, -61922712865)	2/6/00
195	(-2238006277, -5087472163, 5227922915)	12/30/99
588	(-3650204951, -5097345554, 5657478787)	5/23/00

For a given k, we wish to find integers x, y, z satisfying

$$x^3 + y^3 + z^3 = k. \tag{5.1}$$

After searching for solutions where all of $|x|, |y|, |z|$ are relatively small with respect to the size of k, we then focus on solutions where at least one of $|x|, |y|, |z|$ is large. In this case, x, y, z can not all be the same sign, so let z be of different sign from x and y. By letting $T = |x + y|$ we notice that T divides $x^3 + y^3 = k - z^3$. Then given T, z must satisfy

$$z^3 \equiv k \pmod{T}. \tag{5.2}$$

These ideas have been used in earlier searches, for example see [2] and [4]. We were able to impose another condition on z, namely that either $-T < z < \frac{-T}{2}$ or $\frac{T}{2} < z < T$. Therefore each solution for z to (5.2) modulo T yields at most one possible *integer* z. Moreover, fixing such an integer z also determines the integer values for x and y if they exist.

5.2 Previous Results

The question as to which integers are expressible as a sum of three integral cubes is over 150 years old. The first known reference to this problem is found in S. Ryley's article in the Ladies' Diary in 1825 [20], in which he gives a parametrization of rational solutions x, y, z

to $x^3 + y^3 + z^3 = k$, for $k \in \mathbf{Z}$, namely:

$$x = \frac{(9d^6 - 30k^2d^3 + k^4)(3d^3 + k^2) + 72k^4d^3}{6kd(3d^3 + k^2)^2}$$

$$y = \frac{30k^2d^3 - 9d^6 - k^4}{6kd(3d^3 + k^2)}$$

$$z = \frac{18kd^5 - 6k^3d^2}{(3d^3 + k^2)^2}.$$

In 1908, A.S. Werebrusov found the following parametrization of x, y, and z when $k = 2$ [17]:

$$(1 - 6t^3)^3 + (1 - 6t^3)^3 + (-6t^2)^3 = 2.$$

In 1936, Mahler [15] discovered a parametric solution for $k = 1$:

$$(9t^4)^3 + (3t - 9t^4)^3 + (1 - 9t^3)^3 = 1.$$

Mordell proved in 1942 [17] that for any k any other parametric solution with rational coefficients must have degree at least 5.

In 1954, Miller and Woollett discovered explicit representations for 69 values of k between 1 and 100. Their search exhausted the region $\{|x|, |y|, |z| \leq 3164\}$ [16].

In 1963 Gardiner, Lazarus, and Stein looked at the equation $x^3 + y^3 = z^3 - k$ in the range $0 \leq x \leq y \leq 2^{16}$, where $0 < z - x \leq 2^{16}$ and $0 < |k| \leq 999$. Their search left only 70 values of k between 1 and 1000 not congruent to 4 or 5 modulo 9 without a known representation including eight values less than 100 [7].

In 1992 another number less than 100 was finally removed from the list of excluded values. Heath-Brown, Lioen, and te Riele [10] determined that $39 = 134476^3 + 117367^3 + (-159380)^3$ with the rather deep algorithm of Heath-Brown [9], which used a new idea: in searching for solutions for a specific value of k they used the class number of $\mathbf{Q}(\sqrt[3]{k})$ to eliminate values of x, y, z which will not yield a solution.

In 1994 Koyama used modern computers to expand the search region to $\{|x|, |y|, |z| < 2^{21}\}$, and successfully erased 16 integers between 100 and 1000 from the excluded values list [11].

Also in 1994, Conn and Vaserstein used a simpler method than Heath-Brown et al. to try to eliminate more of the excluded values from the list. They also chose specific values of k to target, but then they used relations implied by each chosen value of k to limit the number of triples (x, y, z) searched. In so doing, they found first representations for 84 and 960. They also list a solution for each $k < 100$ for which a representation had been found [4].

Next, in 1995, Bremner [2] devised an algorithm which uses elliptic curve arguments to narrow the search space. He discovered a solution for 75 (and thus a solution for 600), leaving only five excluded values less than 100. Lukes then extended this search method to find the first representations for each of the values 110, 435, and 478 [14].

In 1997, Koyama, Tsuruoka, and Sekigawa [12] used a new algorithm to find first solutions for five more excluded values. Their method proceeded by taking x to be the smallest of the three variables in absolute value, and letting this be the parameter by introducing the variable $A = X^3 - k$, where $X = x$, if $x > 0$, or $A = X^3 + k$, where $X = -x$, if $x \leq 0$.

Also in this paper, the authors discuss the complexity of the above algorithms. To find solutions with $|x|, |y|, |z| \leq N$, the Heath-Brown algorithm has a running time of $c_k N(\log(N))^{O(1)}$, for a fixed value of k. The running time for each of the other algorithms is $O(N^2)$. Of these, the method of [12] fixes a value of k, while the others search for solutions over a range of values of k.

In an August 1999 email, Elkies informed us that Bernstein had implemented the method he had suggested in [6], and found solutions for 11 new values of k, including the same solution for $k = 30$ that we had found.

The values of $k < 1000$ for which no representation is yet known are listed in Table 5.2 at the end of this paper.

5.3 The Algorithm

The first two steps of the algorithm find any small solutions to (5.1). Steps three through five search for the larger solutions. In the next section we explain the various conditions on the small solutions sought in the first two steps, and show that the algorithm will find any solution to (5.1). To ease notation, we first make the following definition:

$$B_k = \frac{3 + \sqrt{12k - 3}}{6} \tag{5.3}$$

Step 1: Fix k and search for solutions to $x^3 + y^3 + z^3 = k$ with $\max\{|x|, |y|, |z|\} \leq B_k$.

Step 2: When $3|k$, search for solutions to $k = x^3 + y^3 + z^3 = 3xyz$ where $x + y + z = 0$. Since $k > 0$, we are concerned only with the case when $x, y < 0$ and $z > 0$. Thus for each distinct factorization of $k/3$ as xyz, simply test if both $x + y + z = 0$ and $x^3 + y^3 + z^3 = k$.

We then repeat steps three through five for increasing values of T, beginning with $T = 2$.

Step 3: Let $s = x + y$ and $T = |x + y|$ and re-write our equation as

$$s(s^2 - 3xy) = k - z^3 \tag{5.4}$$

and therefore $k \equiv z^3 \pmod{T}$. We find all solutions for z modulo T. For a description of an algorithm for finding cube roots modulo a positive integer see [1].

Step 4: For each of the cube roots modulo T found in the previous step, we must ensure they lie in the range $\frac{T}{2} < |z| < T$, (see Proposition 1, section 4). Fix one of the values of z found in step 3. If $\frac{T}{2} < z < T$, then we leave z unchanged and so $s = -T$, since $z > 0$ implies $x, y < 0$. If $0 < z < \frac{T}{2}$, we replace z with $z - T$ and thus $s = T$.

Step 5: After fixing a z and s pair from step 4, we re-write (5.4) as:

$$s^2 - 3xy = \frac{k - z^3}{s}$$

Substituting $y = s - x$ and solving for x, we see that:

$$x = \frac{3s \pm \sqrt{-3s^2 - 12(\frac{z^3-k}{s})}}{6}$$

Letting $D = -3s^2 - 12(\frac{z^3-k}{s})$, we see that in order for x to be an integer, we must have that D is a perfect square, say $D = d^2$, and further that $3|d$ and $3s \equiv \pm d \pmod 6$. If all of these conditions hold, we then have an integer triple (x, y, z) with $x^3+y^3+z^3 = k$. If not, we return to step 4 with the next possible z value given by step 3. When all cube roots of k modulo T have been checked, we return to step 3, after incrementing T by 1.

We now show the expected number of bit operations required for this algorithm to check all values of $T \leq N$ is $N(\log(N))^{O(1)}$, so long as $N > k$. We first split the algorithm into two parts. In order to make calculating more efficient, we first factor the numbers $1 \leq T \leq N$ by sieving with the primes $p \leq \sqrt{N}$. For each such prime there are $\frac{N}{p}$ steps, hence the total number of steps is $O(N \log\log(N))$. Each value of T has at most one prime factor $p > \sqrt{N}$, so accounting for these takes N steps. The total number of steps for factoring the integers T is then $O(N \log\log(N))$. Accounting for the size of N, the number of bit operations is therefore $N(\log N)^{O(1)}$.

The second part of the algorithm is finding modular cube roots, and testing if each leads to a solution. To calculate modular cube roots modulo T, we use the factorization given above and first calculate cube roots modulo the prime factors of T. We then use the Chinese Remainder Theorem to find cube roots modulo T. The expected running time of finding all cube roots modulo a prime can be bounded by $O(\log(N)^3)$ bit operations (see [1]). Extending this to the prime power dividing T is of the same order. Let $\omega(T)$ be the number of distinct prime powers dividing T. To calculate inverses modulo p^a of $\frac{T}{p^a}$ for each $p^a \parallel T$ using Euclid's

Algorithm takes $O(\omega(T)\log(N))$ steps. For each $p^a \parallel T$ there are at most three cube roots of k modulo p^a, so the number of steps required to check all of them for a particular T is $O(\omega(T)\log(N) + 3^{\omega(T)})$. Summing this for values of $T \leq N$ gives $O\left(N(\log N)^2\right)$ (see [22]). Thus the number of expected bit operations of this second part and hence the algorithm is $N(\log(N))^{O(1)}$. Note that to check for solutions with $x, y, z < N$ we need to check values of $T \leq 2N$, which yields the same running time.

5.4 Verifying The Algorithm

In this section we show that the algorithm presented will, for a fixed value of k, find any solution. The main result of this section is:

Proposition 5.1 *Fix a positive integer $k \not\equiv \pm 4 \pmod 9$. Suppose integers x, y, and z satisfy $x^3 + y^3 + z^3 = k$, where x and y are of different sign than z, and $|z| > B_k$. Then letting $s = x + y$ and $T = |s|$ we have either*

$$i)\ \ 3xyz = k \text{ with } x + y + z = 0, \text{ or}$$
$$ii)\ \ \tfrac{T}{2} < |z| < T$$

In the proof of Proposition 5.1 we will use the following lemma:

Lemma 5.1 *Fix a positive integer k and an arbitrary integer z. Suppose integers x and y each have different sign from z, with $|x| \leq |y|$, and let $s = x + y$. Then under the constraint $x^3 + y^3 + z^3 - k = 0$ the quotient $\frac{s}{z}$ achieves a minimum when $x = y$.*

Proof:

Let $f(x, y) = \frac{x+y}{z}$ and $g(x, y) = x^3 + y^3 + z^3 - k$. Using Lagrange Multipliers to find the critical points of $f(x, y)$ under the constraint of $g(x, y)$, we know that for some real λ,

$$\frac{1}{z} = 3\lambda x^2 \quad \text{and} \quad \frac{1}{z} = 3\lambda y^2$$

So, $x = \pm y$. Yet by construction x and y have the same sign, hence the only critical point is when $x = y$.

The minimum may also occur when the partials of g equal 0. The two possibilities for this are $x = 0$ or $y = 0$. Notice that these two conditions are equivalent by renaming variables. Therefore, to show that the critical point yields the minimum for $f(x, y)$ under our constraint, we simply compare values of $f(x, y)$ at the critical point and at $x = 0$.

If $x = y$ then $2y^3 = k - z^3$. So:

$$\frac{s}{z} = \frac{2y}{z} = \sqrt[3]{4}\sqrt[3]{\frac{k}{z^3} - 1} \tag{5.5}$$

If $x = 0$ we have $\frac{s}{z} = \frac{y}{z} = \sqrt[3]{\frac{k}{z^3} - 1}$. Since $k < |z|^3$, indeed the minimum occurs when $x = y$.

This completes the proof of Lemma 5.1.

We can now prove the main proposition.

Proof of Proposition 5.1:

First we will let z be fixed but arbitrary. By Lemma 5.1, and since $|z|^3 > k$,

$$\frac{s}{z} \geq \sqrt[3]{4}\sqrt[3]{\frac{k}{z^3} - 1} > \sqrt[3]{4}\sqrt[3]{-2} = -2$$

and so $|z| > \frac{T}{2}$.

For the other inequality we note that since z has the opposite sign from x and y, that $\frac{s}{z} < 0$. But

$$\left(\frac{x+y}{z}\right)^3 = \frac{x^3 + y^3}{z^3} + \frac{3xy(x+y)}{z^3} \leq \frac{x^3 + y^3}{z^3} = \frac{k}{z^3} - 1$$

so then:

$$\frac{s}{z} \leq \sqrt[3]{\frac{k}{z^3} - 1} \tag{5.6}$$

The proof then naturally splits into two cases.

Case 1: $x, y, s, k \geq 0$ and $z < 0$.

Since $\frac{k}{z^3} < 0$, using (5.6) we see that $|z| < T$.

Case 2: $z, k > 0$ and $x, y, s \leq 0$.

In this case, $\frac{k}{z^3} > 0$ so we do not get the bound immediately as above. From (5.6) we have

$T \geq |\sqrt[3]{k-z^3}|$. If $T = |z|$ then $k = x^3 + y^3 + z^3 = 3xyz$ with $x + y + z = 0$, satisfying condition (i) of the proposition.

Otherwise, if $T < |z|$ then since T and $|z|$ are integers, we have $|\sqrt[3]{k-z^3}| \leq |z| - 1$, that is, $\sqrt[3]{k-z^3} + z \geq 1$. This is equivalent to the inequality $3z^2 - 3z + (1-k) \leq 0$, which is false since $z > B_k$. Thus $|z| < T$, completing Case 2.

This completes the proof of Proposition 5.1.

In order to use Proposition 5.1, we must determine when $|z| > B_k$. The following lemma shows that unless $|x|, |y|, |z|$ are all small this is indeed true.

Lemma 5.2 *Let k be a fixed positive integer. Suppose that x, y, z are integers such that $x^3 + y^3 + z^3 = k$, with $max\{|x|, |y|, |z|\} > B_k$. Then one of x, y, z must be of different sign from the other two. Moreover, if we let z be the one with different sign, then $|z| > B_k$.*

Proof:

By expanding out the expression (5.3) for B_k, we see that $k \geq 1$ implies $B_k^3 \geq k$. Therefore not all of x, y, z can be non-negative. By relabeling if necessary, let z have different sign from x and y.

Without loss of generality, we will assume that $|x| \leq |y|$. If $|z| > B_k$ then the lemma holds trivially. So, suppose $|z| \leq B_k$ but $|y| > B_k$.

We split into two cases, depending on the sign of z. If $z > 0$ and $x, y \leq 0$, then $x^3 + y^3 + z^3 \leq y^3 + z^3 < 0$. But since $k > 0$, this is a contradiction.

Now suppose that $z < 0$ and $x, y \geq 0$. Certainly, if $0 < |z| \leq |x| \leq |y|$, then $x^3 + y^3 + z^3 \geq y^3 > k$. So we can further suppose that $0 \leq |x| < |z| \leq B_k < |y|$. In this case,

$$\begin{aligned} x^3 + y^3 + z^3 - k &\geq 0^3 + (\lfloor B_k \rfloor + 1)^3 + (-\lfloor B_k \rfloor)^3 - k \\ &= 3(\lfloor B_k \rfloor)^2 + 3(\lfloor B_k \rfloor) + 1 - k \\ &> 3(B_k - 1)^2 + 3(B_k - 1) + (1 - k) = 0. \end{aligned}$$

Hence, $x^3 + y^3 + z^3 > k$, which again is a contradiction. This completes the proof of Lemma 5.2.

The first step of our algorithm, then, is to perform an exhaustive search on a small region, namely $x, y, z \leq B_k$. After which, by Lemma 2, any remaining solutions have z with different sign from x and y, and $|z| > B_k$. After quickly searching for solutions satisfying condition (i) in Proposition 1, we can then sequentially check values of T for any remaining solutions. For each value of T, we solve $z^3 \equiv k \pmod{T}$. For each solution to this equation modulo T, Proposition 1 ensures there is at most one possible integer value for z. Fixing z then determines the values for x and y. If they are both integers, a solution to (5.1) has been found. Since for any solution, $|x + y|$ must be an integer, as soon as T reaches this integer, the solution will be found.

5.5 Practical Considerations

While this algorithm works as described, below we list a few implementation considerations.

1. Suppose for a particular T value, there is a prime $p|T$ for which k is not a cube modulo p, then k is also not a cube modulo T. This allows us to skip any value of T having a prime factor p for which k is not a cube modulo p.

2. A similar argument can be made for primes dividing k. Suppose that a prime $p \parallel k$, and $p^2|T$, then $p \parallel (k - s(s^2 - 3xy)) = z^3$. Which is impossible. So, for primes $p \parallel k$, we can exempt T for which $p^2|T$ from consideration. This can be extended to the case where $p^r \parallel k$ and $p^{r+1}|T$ if $r \not\equiv 0 \pmod{3}$.

3. We pre-computed a cube root modulo p for all primes $p \equiv 1 \pmod{3}$ up to some bound B. This ensures that for $T < B^2$ we will need to calculate a cube root modulo p at most once for each T. (That is, only for the prime divisor $p \equiv 1 \pmod{3}$ of T which is greater than B).

5.6 Results

Our original implementation for the algorithm was written using Magma. It was with this version of the code that the solution for $k = 30$ was found. In order to run on several machines and speed up the program, we wrote a version in C, using the gmp arbitrary precision arithmetic library. The rest of the results were found using this second version of the program. The bulk of the calculations were carried out on a 400 MHz Sun Ultra Enterprise 3000. Checking 1,000,000 values for T of size 10^{10} took 1 minute on this machine.

We searched for a solution for each of the integers less than 1000 which are not congruent to 4 or 5 modulo 9 and for which no solution is yet known. The search found representations for four new values of k, and they are listed in Table 5.1 at the beginning of the paper. Current search bounds for the other such values of $k < 1000$ are given in the table below. No integer solution to $x^3 + y^3 + z^3 = k$ for these values of k were found with an associated T-value smaller than the bound indicated.

Table 5.2: Search bounds on T for $k < 1000$

k	T
33	10^{12}
42	6.5×10^{11}
74	1.5×10^{11}
156, 165, 318, 366, 390, 420, 534, 564, 579, 609, 627, 633, 732, 758, 786, 789, 795, 834, 894, 903, 906, 921, 948, 975	10^{10}

Bibliography

[1] Eric Bach and Jeffrey Shallit, Algorithmic number theory (1996), 160–161.

[2] Andrew Bremner, *On sums of three cubes,* Canadian Mathematical Society Conference Proceedings **15** (1995), 87–91.

[3] J.W.S. Cassels, *A Note on the Diophantine Equation* $x^3 + y^3 + z^3 = 3$*,* Mathematics of Computation **44** (1985), 265.

[4] W. Conn and L.N. Vaserstein, *On Sums of Three Integral Cubes,* Contemporary Mathematics **166** (1994), 285–294.

[5] H. Davenport, Multiplicative Number Theory, 2000.

[6] Noam Elkies, "xˆ3 + yˆ3 + zˆ3 = d," 9 July 1996, <nmbrthry@listserv.nodak.edu> via <http://listserv.nodak.edu/archives/nmbrthry.html> (23 January 2004).

[7] V.L. Gardiner, R.B. Lazarus, and P.R. Stein, *Solutions of the Diophantine Equation* $x^3 + y^3 = z^3 - d$*,* Mathematics of Computation **18** (1964), 408–413.

[8] Carl Friedrich Gauss, Disquisitiones Arithmeticae, Art. 216–217.

[9] D.R. Heath-Brown, *Searching for Solutions of* $x^3 + y^3 + z^3 = k$*,* Seminaire de Theorie des Nombres, (1989–1990), 71–76.

[10] D.R. Heath-Brown, W.M. Lioen, and H.J.J. te Riele, *On Solving the Diophantine Equation* $x^3 + y^3 + z^3 = k$ *on a Vector Computer,* Mathematics of Computation **61** (1993), 235–244.

[11] Kenji Koyama, *Tables of solutions of the Diophantine equation* $x^3 + y^3 + z^3 = n$, Mathematics of Computation **62** (1994), 941–942.

[12] Kenji Koyama, Yukio Tsuruoka, and Hiroshi Sekigawa, *On Searching for Solutions of the Diophantine Equation* $x^3 + y^3 + z^3 = n$, Mathematics of Computation **66** (1997), 841–851.

[13] D.H. Lehmer, *On the Diophantine Equation* $x^3 + y^3 + z^3 = 1$, Journal of the London Mathematical Society **31** (1956), 275–280.

[14] Richard F. Lukes, *A Very Fast Electronic Number Sieve,* University of Manitoba doctoral thesis, 1995.

[15] Kurt Mahler, *Note On Hypothesis K of Hardy and Littlewood,* Journal of the London Mathematical Society **11** (1936), 136–138.

[16] J.C.P. Miller and M.F.C. Woollett, *Solution of the Diophantine Equation* $x^3+y^3+z^3 = k$, Journal of the London Mathematical Society **30** (1955), 101–110.

[17] L.J. Mordell, *On Sums of Three Cubes,* Journal of the London Mathematical Society **17** (1942), 139–144.

[18] L.J. Mordell, *On an Infinity of Integer Solutions of* $ax^3 + ay^3 + bz^3 = bc^3$, Journal of the London Mathematical Society **30** (1955), 111–113.

[19] C. Pomerance, *Multiplicative independence for random integers*, Analytic Number Theory, Proceedings of a Conference in Honor of Heini Halberstam, Vol 2, Progr. Math. 139, 703-711, Birkhäuser, 1996.

[20] S. Ryley, The Ladies' Diary **122** (1825), 35.

[21] Manny Scarowsky and Abraham Boyarsky, *A Note on the Diophantine Equation* $x^n + y^n + z^n = 3$, Mathematics of Computation **42** (1984), 235–237.

[22] Gerald Tenenbaum, Introduction to analytic and probablistic number theory (1995), 200–202.

[23] E.C. Titchmarsh, The Theory of the Riemann Zeta-Function, 1986.

[24] R.C. Vaughan, *A new iterative method in Waring's problem,* Acta Mathematica **162** (1989), 1–71.