Séminaire BOURBAKI Juin 2017
69^{e} année, 2016–2017, n^{o} 1134, p. 479 à 564
doi:10.24033/ast.1072

THE VINOGRADOV MEAN VALUE THEOREM
[after Wooley, and Bourgain, Demeter and Guth]

by **Lillian B. PIERCE**

INTRODUCTION

In 1770, Waring wrote:

> *Omnis integer numerus est quadratus; vel e duobus, tribus vel quatuor quadratis compositus. Omnis integer numerus vel est cubus; vel e duobus, tribus, 4, 5, 6, 7, 8, vel novem cubis compositus: est etiam quadrato-quadratus; vel e duobus, tribus, &c. usque ad novemdecim compositus, & sic deinceps: consimilia etiam affirmari possunt (exceptis excipiendis) de eodem numero quantitatum earundem dimensionum.*
>
> [**128**, Thm. XLVII p. 349]

From this we extrapolate Waring's problem, the assertion that for each $k \geq 2$, there exists an $s = s(k)$ such that every positive integer N may be expressed as

$$N = x_1^k + \cdots + x_s^k \tag{1}$$

with $x_1, \ldots, x_s$ non-negative integers. Hilbert proved this assertion in 1909. In the modern interpretation, Waring's problem also refers to the study of the number $r_{s,k}(N)$ of representations of N in the form (1) with $x_i \geq 1$, with the goal of finding the least $s = s(k)$ for which an asymptotic of the form

$$r_{s,k}(N) = \frac{\Gamma(1+1/k)^s}{\Gamma(s/k)} \mathfrak{S}_{s,k}(N) N^{s/k-1} + O_{s,k}(N^{s/k-1-\delta}) \tag{2}$$

holds, for some $\delta = \delta(s,k) > 0$, for all sufficiently large N; here $\mathfrak{S}_{s,k}(N)$ is an arithmetic quantity to which we will return to later. In the 1920's, Hardy and Littlewood were the first to prove such an asymptotic valid for all $k \geq 2$, with s at least exponentially large relative to k. Their general approach, via the circle method, relies critically on estimates for exponential sums. In 1935 Vinogradov introduced a new Mean Value Method for investigating such sums, which not only greatly reduced the number of variables required to obtain the asymptotic (2), but led to a new record for the zero-free region of the Riemann zeta function, which is still (in terms of its over-all shape) the best-known today.

Despite significant attention paid to sharpening the Vinogradov Mean Value Method since 1935, the cornerstone of the method, to which we will refer as the Main Conjecture, was not resolved in full until 2015. In this manuscript, we explore two approaches to the Main Conjecture: first, the work of Wooley using analytic number theory, which between 2010–2015 set significant new records very close to resolving the Main Conjecture in all cases, and resolved it in full in the first nontrivial case; second, the 2015 breakthrough of Bourgain, Demeter, and Guth using harmonic analysis, which resolved the Main Conjecture in full. Through this new work, connections have been revealed to areas far beyond arithmetic questions such as Waring's problem and the Riemann zeta function, stretching to core questions in harmonic analysis, restriction theory, geometric measure theory, incidence geometry, Strichartz inequalities, Schrödinger operators, and beyond.

1. THE MAIN CONJECTURE IN THE VINOGRADOV MEAN VALUE METHOD

Given integers $s, k \geq 1$, let $J_{s,k}(X)$ denote the number of integral solutions to the system of k equations

$$x_1^j + \cdots + x_s^j = x_{s+1}^j + \cdots + x_{2s}^j, \qquad 1 \leq j \leq k, \tag{3}$$

with $1 \leq x_i \leq X$ for $i = 1, \ldots, 2s$. This may be interpreted as a mean value for the exponential sum

$$f_k(\boldsymbol{\alpha}; X) = \sum_{1 \leq x \leq X} e(\alpha_1 x + \cdots + \alpha_k x^k)$$

upon observing that we may equivalently write

$$J_{s,k}(X) = \int_{(0,1]^k} |f_k(\boldsymbol{\alpha}; X)|^{2s} d\boldsymbol{\alpha}. \tag{4}$$

(Here and throughout, we use the notation $e(t) = e^{2\pi i t}$.) The foundational conjecture in the area of the Vinogradov Mean Value Method is as follows: [(1)]

CONJECTURE 1.1 (The Main Conjecture). — *For all integers $s, k \geq 1$,*

$$J_{s,k}(X) \ll_{s,k,\varepsilon} X^{\varepsilon}(X^s + X^{2s - \frac{1}{2}k(k+1)}), \tag{5}$$

for all $X \geq 1$, and every $\varepsilon > 0$.

[(1)] Here and throughout, we use the Vinogradov notation $A \ll_\varepsilon B$ to denote that there exists a constant C_ε such that $|A| \leq C_\varepsilon B$. Unless otherwise specified, any statement involving ε may be taken to hold for all arbitrarily small $\varepsilon > 0$, with associated implied constants. In addition, the implied constant is allowed to depend on other parameters, such as s, k in this section.

This conjecture may be refined to omit the factor X^ε if $k > 2$ (see [**124**, Eqn. 7.5], and § 3.3–§ 3.4 in this manuscript). Vinogradov's motivation for bounding the mean value $J_{s,k}(X)$ was to extract bounds for individual sums $f_k(\boldsymbol{\alpha}; X)$, which (as we will summarize later) would impact many number-theoretic problems, the most famous relating to Waring's problem and the Riemann zeta function. Vinogradov's mean value perspective [**127**], evidently inspired by an idea of Mordell [**99**], has been influential ever since.

Historically, any nontrivial result toward the Main Conjecture has been called the Vinogradov Mean Value Theorem. Instead, we will refer to partial results as the Vinogradov Mean Value Method, and reserve the terminology "Vinogradov Mean Value Theorem" for the newly proved theorems that verify the Main Conjecture in full.

The so-called critical case occurs when $s = s_k = \frac{1}{2}k(k+1)$: this is the index at which the two terms in (5) are of equivalent size. Importantly, if for a certain k the bound (5) has been proved for s_k, then it follows immediately for all $s \geq 1$. In particular, it is elementary to verify the cases of $k = 1, 2$ at the critical index. For a review of these facts, and heuristics leading to the Main Conjecture, see § 3.

Due to the role of the critical index, investigations naturally divide into the case of small s and large s. For small $s \leq k$, the diagonal solutions (those with $\{x_1, \ldots, x_s\} = \{x_{s+1}, \ldots, x_{2s}\}$ as sets) dominate, and the relation

$$J_{s,k}(X) = s!X^s + O(X^{s-1})$$

is an immediate consequence of the Newton-Girard identities. Hua [**71**] extended these considerations to verify the upper bound in the Main Conjecture when $s = k+1$, subsequently refined to an asymptotic in [**125**]. See e.g., [**139**, § 3] for further refinements and historic partial results for small s, before 2010.

We turn to the setting of large s. Vinogradov's original work [**127**] was taken up by Linnik [**92**] (who moved it to a p-adic setting) and polished by Karatsuba [**81**] and Stechkin [**117**]. In total, this approach showed that for $s \geq k$,

$$J_{s,k}(X) \leq D(s,k) X^{2s - \frac{1}{2}k(k+1) + \eta_{s,k}} \tag{6}$$

for an explicit constant $D(s,k)$, and $\eta_{s,k} = \frac{1}{2}k^2(1 - 1/k)^{[s/k]} \leq k^2 e^{-s/k^2}$ for $k \geq 2$. As a consequence of the decay of $\eta_{s,k}$, one can verify for $s \geq 3k^2(\log k + O(\log\log k))$ the bound in the Main Conjecture, and indeed obtain an asymptotic

$$J_{s,k}(X) \sim C(s,k) X^{2s - \frac{1}{2}k(k+1)} \tag{7}$$

for an explicit positive real constant $C(s,k)$. (Indeed, along the same lines of argument, the leading 3 can be improved to a 2; see [**1**, Thm. 3.9].) See [**1**, Ch. 3] for a treatment of these various historic methods, or [**124**, Ch. 7] for a modern overview.

In the 1990's, Wooley's thesis [**134, 135**] developed an *efficient differencing* method which allowed him to extract faster decay from $\eta_{s,k}$ for $s > k^2 \log k$, and as a result he obtained the next historic leap, showing the Main Conjecture held for $s \geq k^2(\log k + 2\log\log k + O(1))$.

1.1. The work of Wooley: Efficient Congruencing

This record remained untouched until the 2010's, when Wooley developed an *efficient congruencing* method. In his initial work on this method [**137**], Wooley set a startling new record, proving that the Main Conjecture held for $s \geq k(k+1)$, for every $k \geq 3$ (and in the asymptotic form (7) for $s \geq k(k+1)+1$). This was a landmark result, since for the first time the additional logarithmic factor was removed, and thus the limitation on s was only a constant multiple away from the expected truth. With this new method (and its "multigrade" version), Wooley held the Main Conjecture under siege, making continual progress on this and related consequences in a remarkable series of papers, including [**136, 138, 139, 140, 141, 142, 143, 144, 145, 146**], and in joint work with Ford [**51**].

By the end of 2015, Wooley had succeeded in proving the conjectured bound for $J_{s,k}(X)$ for s pushing very close to the critical index $s_k = \frac{1}{2}k(k+1)$. For s approaching the critical index from above, Wooley proved the Main Conjecture for $k \geq 3$ and $s \geq k(k-1)$ [**139**]. For s approaching the critical index from below, Wooley [**145**] proved the Main Conjecture for $1 \leq s \leq D(k)$, where $D(4) = 8$, $D(5) = 10, \dots$ and

$$D(k) \leq \frac{1}{2}k(k+1) - \frac{1}{3}k - 8k^{2/3}, \tag{8}$$

for large k. This landmark result was the first ever to prove the Main Conjecture for s differing from the critical index s_k by only a lower order term.

Moreover, in [**142**], Wooley proved the $k = 3$ case in full, establishing the Main Conjecture for the first nontrivial degree:

THEOREM 1.2 (Wooley: Vinogradov Mean Value Theorem, $k = 3$)
For $k = 3$, for every integer $s \geq 1$, the Main Conjecture holds,

$$J_{s,k}(X) \ll_{s,\varepsilon} X^{\varepsilon}(X^s + X^{2s-\frac{1}{2}k(k+1)}),$$

for all $X \geq 1$, and every $\varepsilon > 0$.

1.2. The work of Bourgain, Demeter and Guth: ℓ^2 decoupling

In December 2015, Bourgain, Demeter and Guth [**29**] announced the resolution of the final cases required for the Main Conjecture for $k \geq 4$:

Theorem 1.3 (Bourgain, Demeter, Guth: Vinogradov Mean Value Theorem, $k \geq 4$) *For every integer $k \geq 4$ and for every integer $s \geq 1$,*

$$J_{s,k}(X) \ll_{s,k,\varepsilon} X^{\varepsilon}(X^{s} + X^{2s-\frac{1}{2}k(k+1)}),$$

for all $X \geq 1$, and every $\varepsilon > 0$.

By standard methods, once Theorems 1.2 and 1.3 are known, the X^{ε} may be omitted and the asymptotic (7) obtained for all integers $k \geq 3$ and $s > \frac{1}{2}k(k+1)$; see § 3.4.

The resolution of the Main Conjecture, eighty years after its initiation by Vinogradov, is a spectacular achievement with many consequences. An additional striking feature is that the Bourgain-Demeter-Guth approach is rooted in harmonic analysis. Of course, even the expression (4) immediately indicates that the Main Conjecture is inextricably bound to ideas of Fourier analysis, as is the Hardy-Littlewood circle method, one of the core techniques of analytic number theory. But as we will see, the Bourgain-Demeter-Guth method (and work leading up to it) takes this quite a bit further, revealing fascinating connections between the Vinogradov Mean Value Method and deep open problems that have been motivating work in harmonic analysis over the past fifty years.

The Bourgain-Demeter-Guth work [**29**] is part of the new area of *decoupling* (and in particular ℓ^2 decoupling). This area was initiated in work of Wolff [**132**], who introduced the study of an ℓ^p decoupling inequality for the cone, motivated by the local smoothing conjecture for the wave equation, see e.g., the survey [**111**]. See also early work on decoupling, then called "Wolff's inequality," by Łaba, Pramanik, and Seeger [**91, 89, 108**].

In the present context, decoupling was deeply developed by Bourgain [**15**] and then Bourgain and Demeter [**19, 24, 26, 20, 27, 21, 22, 23, 25**], and is now a growing area of research, including for example [**31, 17, 18, 16, 28, 38, 40, 45, 39, 56**]. As a whole, the decoupling method has ramifications far broader than resolving the Main Conjecture in the Vinogradov Mean Value Method, several of which we will mention in § 2 and § 6.5.

While decoupling has deep ties to many aspects of harmonic analysis, it is worth noting (as per Wooley [**145**]) that the decoupling method shares similarities with Vinogradov's initial framework for mean value investigations in the 1930's, which used small real intervals. Indeed, it seems reasonable to speculate, as Wooley does, that efficient congruencing (in its most recent "nested" formulation) and ℓ^2 decoupling will ultimately be understood as p-adic and Archimedean perspectives of one unified method (see § 8.5 for a few clear parallels).