Vinogradov's Mean Value Theorem

Ciprian Demeter (Indiana University)
joint with Jean Bourgain (IAS) and Larry Guth (MIT)

Madison, May 2016

We write $A \lesssim B$ if there is an implicit constant C which depends on fixed parameters such as n (dimension of $\mathbb{R}^n$) and p (the index of L^p) such that

$$A \leq CB.$$

C will never depend on the scales δ, N.

We often write $A \lesssim_\epsilon N^\epsilon$ to denote the fact that the implicit constant depends on ϵ. For example $\log N \lesssim_\epsilon N^\epsilon$, for each $\epsilon > 0$

We will use the notation $e(z) = e^{2\pi i z}$, $z \in \mathbb{R}$.

For each integers $s \geq 1$ and $n, N \geq 2$ denote by $J_{s,n}(N)$ the number of integral solutions for the following system

$$X_1^i + \ldots + X_s^i = Y_1^i + \ldots + Y_s^i, \quad 1 \leq i \leq n,$$

with $1 \leq X_1, \ldots, X_s, Y_1, \ldots, Y_s \leq N$.

Example: n=2

$$\begin{cases} X_1 + \ldots + X_s = Y_1 + \ldots + Y_s \\ X_1^2 + \ldots + X_s^2 = Y_1^2 + \ldots + Y_s^2 \end{cases}.$$

Theorem (Vinogradov's Mean Value "Theorem")

For each $s \geq 1$, $\epsilon > 0$ and $n, N \geq 2$ we have the upper bound

$$J_{s,n}(N) \lesssim_\epsilon N^{s+\epsilon} + N^{2s - \frac{n(n+1)}{2} + \epsilon}.$$

The number $J_{s,n}(N)$ has the following analytic representation

$$J_{s,n}(N) = \int_{[0,1]^n} |\sum_{j=1}^{N} e(x_1 j + x_2 j^2 + \ldots + x_n j^n)|^{2s} dx_1 \ldots dx_n.$$

Theorem (Vinogradov's Mean Value "Theorem" (VMVT))

For each $p \geq 2$, $\epsilon > 0$ and $n, N \geq 2$ we have the upper bound

$$\Big(\int_{[0,1]^n} |\sum_{j=1}^{N} e(x_1 j + x_2 j^2 + \ldots + x_n j^n)|^p dx_1 \ldots dx_n\Big)^{1/p} \lesssim_\epsilon$$

$$\begin{cases} N^{\frac{1}{2}+\epsilon}, & \text{if } 2 \leq p \leq n(n+1) \\ N^{1-\frac{n(n+1)}{2p}+\epsilon}, & \text{if } p \geq n(n+1) \end{cases}.$$

When $p = 2, \infty$ we have sharp estimates

$$\|\sum_{j=1}^{N} e(x_1 j + x_2 j^2 + \ldots + x_n j^n)\|_{L^p(\mathbb{T}^n)} = \begin{cases} N^{\frac{1}{2}}, & p = 2 \\ N, & p = \infty \end{cases}$$

Given n, the full range of estimates in VMVT will follow if we prove the case $p = n(n+1)$ (**critical exponent**)

• **n=2** is easy and has been known (folklore?). It has critical exponent $p = 2(2+1) = 6$. One needs to check that

$$\begin{cases} X_1 + X_2 + X_3 = Y_1 + Y_2 + Y_3 \\ X_1^2 + X_2^2 + X_3^2 = Y_1^2 + Y_2^2 + Y_3^2 \end{cases}.$$

has $O(N^{3+\epsilon})$ integral solutions in the interval $[1, N]$. Note that $(X_1, X_1, X_3, X_1, X_2, X_3)$ is always a (trivial) solution, so we have at least N^3 solutions. The required estimate says that fixing X_1, X_2, X_3 will determine Y_1, Y_2, Y_3 within $O(N^\epsilon)$ choices. Using easy algebraic manipulations this boils down to the fact that a circle of radius N contains at most $O(N^\epsilon)$ lattice points.

- $n \geq 3$: Only partial results have been known until ~ 2012

Theorem (Vinogradov (1935), Karatsuba, Stechkin)

VMVT holds for $p \geq n^2(4\log n + 2\log\log n + 10)$, *and in fact one has a sharp asymptotic formula*

$$\|\sum_{j=1}^{N} e(x_1 j + x_2 j^2 + \ldots + x_n j^n)\|_{L^p(\mathbb{T}^n)} \sim C(p,n) N^{1-\frac{n(n+1)}{2p}}$$

Wooley developed the efficient congruencing method which led to the following progress

Theorem (Wooley, 2012 and later)

VMVT holds for

- $n = 3$ *and all values of* p
- $p \leq n(n+1) - \frac{2n}{3} + O(n^{2/3})$,
- $p \geq 2n(n-1)$, *all* $n \geq 3$

Theorem (Bourgain, D, Guth 2015)

VMVT holds for all $n \geq 2$ and all p.

Moreover, when combining this with known sharp estimates on major arcs, there will be no losses in the supercritical regime $p > n(n+1)$

$$\|\sum_{j=1}^{N} e(x_1 j + x_2 j^2 + \ldots + x_n j^n)\|_{L^p(\mathbb{T}^n)} \leq C(p,n) N^{1-\frac{n(n+1)}{2p}}.$$

Our method does not seem to say anything meaningful about the implicit constant $C(p,n)$, so we can't say anything new about the zero-free regions of the Riemann zeta. But the are at least two other classical applications.

Weyl sums

$$\mathbf{x} = (x_1, \dots, x_n)$$

$$f_n(\mathbf{x}, N) = \sum_{j=1}^{N} e(x_1 j + x_2 j^2 + \dots + x_n j^n)$$

Theorem (H. Weyl)

Assume $|x_n - \frac{a}{q}| \le \frac{1}{q^2}$, $(a, q) = 1$. *Then*

$$|f_n(\mathbf{x}, N)| \lesssim N^{1+\epsilon}(q^{-1} + N^{-1} + qN^{-n})^{2^{1-n}}$$

As a consequence of VMVT we can now replace 2^{1-n} with $\sigma(n) = \frac{1}{n(n-1)}$ (best known bounds for large n).

The asymptotic formula in Waring's problem

$R_{s,k}(n)$ = number of representations of the integer n as a sum of s kth powers. Based on circle method heuristics, the following asymptotic formula is conjectured

$$R_{s,k}(n) = \frac{\Gamma(1+\frac{1}{k})^s}{\Gamma(\frac{s}{k})}\mathfrak{G}_{s,k}(n)n^{\frac{s}{k}-1} + o(n^{\frac{s}{k}-1}), \quad n \to \infty$$

for $s \geq k+1$, $k \geq 3$. Let $\tilde{G}(k)$ (**Waring number**) be the smallest s for which the formula holds.

Wooley showed that VMVT would imply for all $k \geq 3$

$$\tilde{G}(k) \leq k^2 + 1 - \max_{\substack{1 \leq j \leq k-1 \\ 2^j \leq k^2}} \left[\frac{kj - 2^j}{k + 1 - j} \right].$$

In particular, we get

$$\tilde{G}(k) \leq k^2 + 1 - \left[\frac{\log k}{\log 2} \right]$$

This improves all previous bounds on $\tilde{G}(k)$, except for Vaughan's $\tilde{G}(3) \leq 8$ (1986).

Further improvements are possible. Our VMVT leads (rather immediately) to progress on Hua's lemma, which leads (Bourgain 2016) to a further improvement

$$\tilde{G}(k) \leq k^2 - k + O(\sqrt{k}).$$

$$f(x) = \sum_{j \sim N} e(j^n x)$$

Conjecture: $\int_0^1 |f(x)|^p dx \lesssim N^{p-n+\epsilon}$, for $p \geq 2n$

Lemma (Hua)

For $l \leq n$

$$\int_0^1 |f(x)|^{2^l} dx \lesssim N^{2^l - l + \epsilon}, \text{ sharp when } l = n$$

Theorem (Bourgain, 2016)

For $s \leq n$

$$\int_0^1 |f(x)|^{s(s+1)} dx \lesssim N^{s^2 + \epsilon}, \text{ sharp when } s = n$$

Motivated in part by investigations by T. Wolff from late 1990s, Bourgain and I have developed a decoupling theory for L^p spaces. In a nutshell, our theorems go as follows:

Theorem (Abstract decoupling theorem)

Let $f : \mathcal{M} \to \mathbb{C}$ be a function on some compact manifold $\mathcal{M}$ in $\mathbb{R}^n$, with natural measure σ. Partition the manifold into caps τ of size δ (with some variations forced by curvature) and let $f_\tau = f 1_\tau$ be the restriction of f to τ. Then there is a critical index $p_c > 2$ and some $q \geq 2$ (both depending on the manifold) so that we have

$$\|\widehat{fd\sigma}\|_{L^p(B_{\delta^{-q}})} \lesssim_\epsilon \delta^{-\epsilon}\Big(\sum_{\tau:\delta-cap} \|\widehat{f_\tau d\sigma}\|^2_{L^p(B_{\delta^{-q}})}\Big)^{1/2}$$

for each ball $B_{\delta^{-q}}$ in $\mathbb{R}^n$ with radius δ^{-q} and each $2 \leq p \leq p_c$.

For a "non-degenerate" d-dimensional smooth, compact graph manifold in $\mathbb{R}^n$

$$\mathcal{M} = \{(t_1, \dots, t_d, \phi_1(t_1, \dots, t_d), \dots, \phi_{n-d}(t_1, \dots, t_d))\}$$

it seems reasonable to expect (at least for l^p decouplings)

(1) $p_c = \frac{4n}{d} - 2$ and $q = 2$, if $d > \frac{n}{3}$. This should be achieved with purely quadratic ϕ_i. When $d = n - 1$, $p_c = \frac{2(n+1)}{n-1}$.

(2) $p_c = 3 \cdot 4$ and $q = 3$, if $\frac{n}{4} < d \leq \frac{n}{3}$. The cubic terms become relevant. Examples include

$$(t, t^2, t^3) \text{ in } \mathbb{R}^3, \quad (t_1, t_1^2, t_1^3, t_2, t_2^2, t_2^3, 0) \text{ in } \mathbb{R}^7$$

(3) $p_c = 4 \cdot 5$ and $q = 4$, if $\frac{n}{5} < d \leq \frac{n}{4}$. The quartic terms become relevant. One example is (t, t^2, t^3, t^4) in $\mathbb{R}^4$.

It is clear how to continue.

Bourgain's observation (2011): To get from...

Theorem (Abstract decoupling theorem)

$$\|\widehat{fd\sigma}\|_{L^p(B_{\delta^{-q}})} \lesssim_\epsilon \delta^{-\epsilon}\Big(\sum_{\tau:\delta-cap}\|\widehat{f_\tau d\sigma}\|^2_{L^p(B_{\delta^{-q}})}\Big)^{1/2}$$

for each ball $B_{\delta^{-q}}$ in $\mathbb{R}^n$ with radius δ^{-q} and each $2 \le p \le p_c$.

...to the exponential sum estimate

Theorem (Abstract exponential sum estimate)

For each cap τ let $\xi_\tau \in \tau$ and $a_\tau \in \mathbb{C}$. Then

$$|B_{\delta^{-q}}|^{-1/p}\|\sum_\tau a_\tau e(\xi_\tau \cdot \mathbf{x})\|_{L^p(B_{\delta^{-q}})} \lesssim_\epsilon \delta^{-\epsilon}\Big(\sum_\tau |a_\tau|^2\Big)^{1/2}$$

for each ball $B_{\delta^{-q}}$ in $\mathbb{R}^n$ with radius δ^{-q} and each $2 \le p \le p_c$,

simply use (a smooth approximation of) $f(\xi) = \sum_\tau a_\tau \delta_{\xi_\tau}$

We have so far established the optimal decoupling theory for the following manifolds $\mathcal{M}$, with the following applications

• Hypersurfaces in $\mathbb{R}^n$ with nonzero Gaussian curvature ($p_c = \frac{2(n+1)}{n-1}$). **Many applications:** Optimal Strichartz estimates for Shrödinger equation on both rational and irrational tori in all dimensions, improved L^p estimates for the eigenfunctions of the Laplacian on the torus, etc

• The cone (zero Gaussian curvature) in $\mathbb{R}^n$ ($p_c = \frac{2n}{n-2}$). **Many applications:** progress on Sogge's "local smoothing conjecture for the wave equation", etc

• (Bourgain) Two dimensional surfaces in $\mathbb{R}^4$ ($p_c = 6$). **Application:** Bourgain used this to improve the estimate in the Lindelöf hypothesis for the growth of Riemann zeta

• (with Larry, too) Curves with torsion in $\mathbb{R}^n$ ($p_c = n(n+1)$). **Application:** Vinogradov's Mean Value Theorem.

Here is some insight on why we need to work on "big" balls $B_{\delta^{-q}}$.

Typically, working with $q = 1$ does not produce interesting results, decoupling only works at this scale for $p = 2$. The very standard (L^2 almost orthogonality) estimate is that, for any δ- separated points ξ in $\mathbb{R}^n$.

$$\left(\frac{1}{|B_{\delta^{-1}}|}\int_{B_{\delta^{-1}}} |\sum_{\xi} a_\xi e(\xi \cdot \mathbf{x})|^2 d\mathbf{x}\right)^{1/2} \lesssim \|a_\xi\|_{l^2}.$$

One can **not** replace the L^2 average with an L^p ($p > 2$) average if no additional restrictions are imposed.

Even under the curvature assumption $\Lambda \subset S^{n-1}$, when $p = \frac{2(n+1)}{n-1}$ the expected estimate is (equivalent form of Stein-Tomas)

$$\left(\frac{1}{|B_{\delta^{-1}}|}\int_{B_{\delta^{-1}}} |\sum_{\xi\in\Lambda} a_\xi e(\xi\cdot\mathbf{x})|^p d\mathbf{x}\right)^{1/p} \lesssim \delta^{\frac{n}{p}-\frac{n-1}{2}} \|a_\xi\|_{l^2}.$$

Note that the exponent $\frac{n}{p} - \frac{n-1}{2}$ is negative.

However, by averaging the same exponential sum on the larger ball $B_{\delta^{-2}}$ (this allows more room for the oscillations to annihilate each other), we get a stronger estimate (reverse Hölder)

$$\left(\frac{1}{|B_{\delta^{-2}}|}\int_{B_{\delta^{-2}}} |\sum_{\xi\in\Lambda} a_\xi e(\xi\cdot\mathbf{x})|^p d\mathbf{x}\right)^{1/p} \lesssim \delta^{-\epsilon} \|a_\xi\|_{l^2}.$$

This perhaps explains why early attempts to prove optimal Strichartz estimates on $\mathbb{T}^n$ using the Stein-Tomas approach failed.

Recap: Decouplings need **separation, curvature** and **large enough spatial balls**.

Theorem (Bourgain, D, Guth, 2015)

Let $\bar{\xi} = (\xi, \dots, \xi^n)$ *be* $\delta-$*separated points on the curve*

$$\{(t, t^2, \dots, t^n) : 0 \le t \le 1\}.$$

Then for each $2 \le p \le n(n+1)$

$$(\frac{1}{|B_{\delta^{-n}}|}\int_{B_{\delta^{-n}}} |\sum_{\bar{\xi}} a_{\bar{\xi}} e(\xi x_1 + \xi^2 x_2 + \dots \xi^n x_n)|^p d\mathbf{x})^{1/p} \lesssim_\epsilon \delta^{-\epsilon} \|a_{\bar{\xi}}\|_{l^2}$$

Apply this with $\xi = \frac{j}{N}$, $1 \le j \le N$. Change variables $\frac{x_1}{N} = y_1, \dots, \frac{x_n}{N^n} = y_n$. Then we get ($\delta = \frac{1}{N}$)

$$(\frac{1}{|C|}\int_C |\sum_{j=1}^{N} a_j e(jy_1 + j^2 y_2 + \dots j^n y_n)|^p d\mathbf{y})^{1/p} \lesssim_\epsilon N^\epsilon \|a_j\|_{l^2}$$

$$C = [-N^{n-1}, N^{n-1}] \times [-N^{n-2}, N^{n-2}] \times \dots \times [-1, 1]$$

$$\left(\frac{1}{|C|}\int_C |\sum_{j=1}^N a_j e(jy_1 + j^2y_2 + \dots j^n y_n)|^p d\mathbf{y}\right)^{1/p} \lesssim_\epsilon N^\epsilon \|a_j\|_{l^2}$$

$$C = [-N^{n-1}, N^{n-1}] \times [-N^{n-2}, N^{n-2}] \times \dots \times [-1, 1]$$

Next cover C with translates of $[0,1]^n$ **and use periodicity** to get

$$\left(\int_{\mathbb{T}^n} |\sum_{j=1}^N a_j e(jy_1 + j^2y_2 + \dots j^n y_n)|^p d\mathbf{y}\right)^{1/p} \lesssim_\epsilon N^\epsilon \|a_j\|_{l^2}$$

Conclusions

1. Periodicity is the only fact that we exploit about integers j. We have no other number theory in our argument. In fact, **integers** can be replaced with well separated **real** numbers.

2. We recover a more general theorem, with coefficients a_j.

The proof of our decoupling theorem (n=3)...

$$\mathcal{M} = \{(t, t^2, t^3) : 0 \le t \le 1\}.$$

Theorem

Let $f : \mathcal{M} \to \mathbb{C}$. Partition $\mathcal{M}$ into caps τ of size δ. Then

$$\|\widehat{fd\sigma}\|_{L^{12}(B_{\delta^{-3}})} \lesssim_\epsilon \delta^{-\epsilon}\Big(\sum_\tau \|\widehat{f_\tau d\sigma}\|^2_{L^{12}(B_{\delta^{-3}})}\Big)^{1/2}$$

for each ball $B_{\delta^{-3}}$ in $\mathbb{R}^3$ with radius δ^{-3}.

...goes via gradually decreasing the size of the caps τ and at the same time increasing the radius of the balls. This is done using the following tools.

- L^2 **decoupling:** This is a form of L^2 orthogonality

$$\|\widehat{fd\sigma}\|_{L^2(B_{\delta^{-1}})} \lesssim (\sum_{\tau} \|\widehat{f_\tau d\sigma}\|^2_{L^2(B_{\delta^{-1}})})^{1/2}$$

It only works for L^2 but it decouples efficiently, into caps of very small size, equal to

$$\frac{1}{\text{radius of the ball}}$$

• **Lower dimensional decoupling:** We use induction on dimension. We assume and use the $n = 2$ decoupling result at L^6.

The **weakness** of this is that the critical exponent $p_c = 6$ for $n = 2$ is small compared to 12 ($n = 3$).

The **strength** is the fact that it decouples into small intervals, of length $\frac{1}{R^{1/2}}$ as opposed to $\frac{1}{R^{1/3}}$ (R is the radius of the spatial ball).

At the right spatial scale, arcs of the twisted cubic look planar. One can treat them with L^6 decoupling. For example, the $\sim \delta^{-3}$ neighborhood of

$$\{(t, t^2, t^3) : 0 \le t \le \delta\}$$

is essentially the same as the $\sim \delta^{-3}$ neighborhood of the arc of parabola

$$\{(t, t^2, 0) : 0 \le t \le \delta\}$$

so there is an L^6 decoupling of this into $\delta^{\frac{3}{2}}$ arcs on $B_{\delta^{-3}}$

• **Multilinear Kakeya type inequalities:** Do a wave packet decomposition of $\widehat{fd\sigma}$ using plates.

There is a hierarchy of incidence geometry inequalities about how these plates intersect, ranging from easy to hard. These inequalities have only been clarified in the last two years.

Theorem (Multilinear Kakeya in disguise)

Fix $1 \le k \le n-1$, $p \ge 2n$ and $n!$ separated intervals $I_i \subset [0,1]$. Let B be an arbitrary ball in $\mathbb{R}^n$ with radius $\delta^{-(k+1)}$, and let $\mathcal{B}$ be a finitely overlapping cover of B with balls Δ of radius δ^{-k}. Then (# denotes an average)

$$\frac{1}{|\mathcal{B}|}\sum_{\Delta\in\mathcal{B}}\left[\prod_{i=1}^{n!}\Big(\sum_{\substack{J_i\subset I_i\\ |J_i|=\delta}}\|\widehat{g_{J_i}d\sigma}\|^2_{L_\sharp^{\frac{pk}{n}}(\Delta)}\Big)^{1/2}\right]^{p/n!} \lesssim$$

$$\delta^{-\epsilon}\left[\prod_{i=1}^{n!}\Big(\sum_{\substack{J_i\subset I_i\\ |J_i|=\delta}}\|\widehat{g_{J_i}d\sigma}\|^2_{L_\sharp^{\frac{pk}{n}}(B)}\Big)^{1/2}\right]^{p/n!}.$$

Our first attempt (Jean and I) to prove VMVT only used the $k=1$ result and resulted in the poor range $2 \le p \le 4n-2$.

• **Parabolic rescaling:** Each arc on $(t, t^2, \ldots, t^n)$ can be mapped via an affine transformation to the full arc $(0 \le t \le 1)$.

• **Lots of induction on scales:** Let C_δ be the best constant in some decoupling inequality at scale δ. How does C_δ relate to $C_{\delta^{1/2}}$?

• **Lots of Hölder's inequality and ball inflations**