

The Vinogradov Method and Weyl Sums

Andrei Jorza

August 9, 2004

Иван Матеевич Виноградов came up with a very efficient way of estimating Weyl sums. If $f(x) = a_k x^k + \dots + a_1 x + a_0$ is a polynomial then the Weyl sum associated to it is

$$S(q) = \sum_{n=a+1}^{a+q} e^{2\pi i f(n)} \quad (1)$$

for $a \in \mathbb{N}$.

Виноградов had the idea to view this Weyl sum as a function on the unit hypercube $\mathcal{C} = [0, 1]^k$, where the k -dimensional variable is the vector of coefficients of f , i.e., $\alpha = (a_1, \dots, a_k)$ (note that the value of the sum only depends on the values of the coefficients mod 1). We shall approximate the Weyl sum by the average on a small neighborhood of the coefficient vector. Then we will patch the unit hypercube with such neighborhoods and we will get a bound on $S(q)$ using the average on \mathcal{C} .

1 Average on the Unit Hypercube

Consider

$$\begin{aligned} J(q, l) &= \int_{\mathcal{C}} |S(q)|^{2l} d\alpha = \int_{\mathcal{C}} S(q)^l \overline{S(q)}^l d\alpha = \int_{\mathcal{C}} \sum_{n_i, m_i} e^{2\pi i \sum_i f(n_i) - 2\pi i \sum_i f(m_i)} d\alpha \\ &= \int_{\mathcal{C}} \sum_{n_i, m_i, j} e^{2\pi i a_j (\sum_i n_i^j - \sum_i m_i^j)} d\alpha = \sum_{n_i, m_i} \prod_j \int_0^1 e^{2\pi i a_j (\sum_i n_i^j - \sum_i m_i^j)} da_j \end{aligned}$$

The integral is 1 if the exponent is 0 and 0 otherwise. Therefore we get that $J(q, l)$ is the number of solutions to the simultaneous (because of \prod) system

$$\begin{aligned} n_1 + \dots + n_l &= m_1 + \dots + m_l \\ &\vdots \\ n_1^k + \dots + n_l^k &= m_1^k + \dots + m_l^k \end{aligned}$$

(with $a + 1 \leq n_i, m_i \leq a + q$).

Witty Observation 1 *The number of solutions to the system in the hypercube $[a+1, a+q]^k$ is the same as in the hypercube $[1, q]^k$*

Proof: Write $n_i = a + n'_i, m_i = a + m'_i$ and expand using the binomial formula to get the same system for n'_i, m'_i . ■

Lemma 2 *Let $1 < G < q$ and $1 < u_1 < u_2 < \dots < u_k \leq G$ so that no two u_i are consecutive. Prove that the number of k -tuples (m_1, \dots, m_k) so that $m_i \in A_i = (\frac{q(u_i-1)}{G}, \frac{qu_i}{G}]$ with $s_l = m_1^l + \dots + m_k^l$ lying in a fixed interval of length q^{l-1} is at most $M_k = (4kG)^{\binom{k}{2}}$.*

Proof: Consider two such k -tuples $(m_1, \dots, m_k), (n_1, \dots, n_k)$. Let s_l, s'_l be the two sums associated to them. Then $|s_l - s'_l| < q^{l-1}$. Also, let σ_l, σ'_l be the l -th symmetric polynomials associated to the two k -tuple. From Newton's formulae ($s_j - \sigma_1 s_{j-1} + \dots + (-1)^j j \sigma_j = 0$) one gets by induction that

$$|\sigma_l - \sigma'_l| < 3(2kG)^{l-1}/4 \quad (2)$$

(use the obvious bound $s_l \leq kq^l$).

For $x < q$ get

$$|(x - m_1) \dots (x - m_l) - (x - n_1) \dots (x - n_l)| \leq \sum_{i \geq 1} |\sigma_i - \sigma'_i| x^{l-i} \leq q^{l-1} + (3/4) \sum_{i \geq 2} ((2kG)^{i-1} q^{l-i}) \quad (3)$$

and since $G < q$ this is $\leq q^{l-1} (1 + \frac{3}{4} \frac{(2k)^{l-2k}}{2k-1}) \leq (2kq)^{l-1}$.

So $|(n_l - m_1) \dots (n_l - m_l)| \leq (2kq)^{l-1}$ and by construction $n_l - m_i > q/G$ (if $i \neq l$; here is where you use the condition on the u_i) so we get $m_l - n_l \leq (2kG)^{l-1}$. So there are at most $1 + (2kG)^{l-1} < (4kG)^{l-1}$ values for m_l so in all at most M_k k -tuples clearly. ■

Witty Observation 3 *If in this lemma, the s_l 's lie in given intervals of length $cq^{l-1/k}$ rather than q^{l-1} , then the number of k -tuples is at most $N_k = (2c)^k q^{(k-1)/2} M_k$.*

To see this, divide each interval of length $cq^{l-1/k}$ into at most $2cq^{1-1/k}$ intervals of length at most q^{l-1} . Then apply the lemma for each k -tuple of such subintervals and add all the numbers. ■

Lemma 4 *We are in the context of the previous lemma with $G = 2^m$ and $l > k$. If $S_{m,i} = \sum_{A_i} e^{2\pi i f(n)}$ then*

$$I = \int_{\mathcal{C}} |S_{m,1} \dots S_{m,k}|^2 |S(q^{1-1/k})|^{2(l-k)} d\alpha \leq C_{m,l,k} J(q^{1-1/k}, l-k). \quad (4)$$

Proof: Open the parentheses in the definition of $|S_{m,i}|^2$ and $|S(q^{1-1/k})|^{2(l-k)}$; using the same trick as in the first evaluation of $J(q, l)$ we get that I is the number of solutions to

$$\sum_1^k m_i^j - \sum_1^k n_i^j = \sum_1^{l-k} m_i^j - \sum_1^{l-k} n_i^j \quad (5)$$

where $m_i, n_i \in A_i$ and $m'_i, n'_i \in (a, a + q^{1-1/k}]$. The RHS takes $2(l-k)q^{j(1-1/k)}$ values (using witty observation 1) and using witty observation 3 there are at most N_k (with $c = 2(l-k), G = 2^m$) k -tuples (m_i) for each (n_i) . Trivially there are at most $(q/2^m)^k$ (n_i) 's so there are at most $N_k(q/2^m)^k$ pairs of k -tuples with that property.

Go back to the expression for I and open up the parentheses of $|S_{m,i}|^2$. Bound each exponential by 1 so $I \leq J(q^{1-1/k}, l-k)$ times the number of k -tuples. Therefore we choose

$$\mathcal{C}_{m,l,k} = N_k(q/2^m)^k = 2^{2k+(m+2)\binom{k}{2}-mk}(l-k)^k q^{3k/2-1/2} k^{\binom{k}{2}} \quad (6)$$

Witty Observation 5 *Using trivial bounds on $|S_{m,i}| \leq 1 + |A_i| \leq 2^{1-M}q$, if $m = M = \lfloor (\log_2 q)/k \rfloor$ then the previous lemma holds whether or not the u_i 's satisfy the conditions of lemma 2.*

Exercise 1 *The set of integers $\{u_1, \dots, u_l\} \subset \{1, \dots, G\}$ is called good if there are k of them which, when written in increasing order of indices, satisfy the conditions in lemma 2. The number of not good sets is $\leq B = l!3^l G^{k-1}$.*

Hint: Write in increasing order (which accounts for $l!$). Look at pairwise differences, at most $h < k-1$ of which are > 1 . Then h differences can be chosen in $\binom{l-1}{h}$ ways and can take $\leq G$ values, while all the other differences can be 0 or 1. Sum up (varying the first number too) and get q.e.d.

2 Inductive Estimate of the Average

Lemma 6 *For $l \geq k(k+5)/4$ and M defined in witty observation 5 we have*

$$J(q, l) \leq \mathcal{C}'_{q,l,k} J(q^{1-1/k}, l-k) \quad (7)$$

Proof: If $M \geq 2$ take $m < M$ (i.e., $2^{m+1} \leq q^{1/k}$). Let $S(q) = \sum_{i=1}^{2^m} S_{m,i}$. Then $S^l(q) = \sum S_{m,i_1} \cdots S_{m,i_l}$. Among these terms there are $G_m \leq 2^{ml}$ whose second indices form a good set $\{i_1, i_2, \dots, i_k\}$. Denote any of them by \mathcal{G}_m .

Using $(\frac{(u-1)q}{2^m}, \frac{uq}{2^m}] = (\frac{(2u-2)q}{2^{m+1}}, \frac{(2u-1)q}{2^{m+1}}] \cup (\frac{(2u-1)q}{2^{m+1}}, \frac{(2u)q}{2^{m+1}}]$ divide each of the terms whose indices are not good into terms of type $S_{m+1,j_1} \cdots S_{m+1,j_l}$. Here, $G_{m+1} \leq 2^{(m+1)l}$ terms have good indices, and denote any of them by \mathcal{G}_{m+1} . Divide the terms with not good indices as above and continue this process until reached M .

Get $S^l(q) = \sum_{n=m}^M \sum \mathcal{G}_n$. So

$$|S(q)|^{2l} \leq M \sum G_n \sum |\mathcal{G}_n|^2. \quad (8)$$

G_n is at most the number of bad sets among the $G_{n-1} \leq 2^{(n-1)l}$. So by exercise 1 (with $G = 2^{n-1}$), $G_n \leq l!6^l 2^{(n-1)(k-1)}$. The extra factor 2^l comes from the fact that each $S_{n-1,i}$ is divided into two parts.

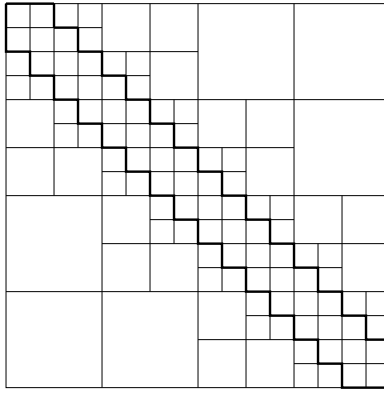


Figure 1: Inductive Estimates

Now, if i_1, \dots, i_k are the indices that make the set of indices good then by Hölder we have

$$|S_{n,i_{k+1}} \cdots S_{n,i_l}|^2 \leq \frac{1}{l-k} \sum |S_{n,i_j}|^{2(l-k)}. \quad (9)$$

Since $q/2^n > q^{1-1/k}$, divide each $S_{n,i}$ into at most $h = 2^{1-n}q^{1/k}$ terms of type $S(q^{1-1/k})$ (each $S_{m,i}$ has $q/2^n$ terms; here is where we use $M \geq 2$).

Again by Hölder,

$$|S_{n,i}|^{2(l-k)} \leq h^{2(l-k)-1} \sum |S(q^{1-1/k})|^{2(l-k)} \quad (10)$$

so

$$\sum |\mathcal{G}_n|^2 \leq \frac{h^{2(l-k)-1}}{l-k} \sum |S_{n,i_1} \cdots S_{n,i_k}|^2 \sum_{k+1}^l \sum_1^h |S(q^{1-1/k})|^{2(l-k)}. \quad (11)$$

From lemma 4,

$$\int_{\mathcal{C}} \sum |\mathcal{G}_n|^2 d\alpha \leq \frac{h^{2(l-k)-1}}{l-k} G_n h(l-k) \mathcal{C}_{n,l,k} J(q^{1-1/k}, l-k), \quad (12)$$

because there are $|G_n|^2$ terms in the sum of \mathcal{G}_n and $l-k$ terms in the sum of h terms of form $|S(q^{1-1/k})|^{2(l-k)}$.

Therefore

$$J(q, l) \leq M \sum_{n=m}^M h^{2(l-k)} G_n^2 \mathcal{C}_{m,l,k} J(q^{1-1/k}, l-k) \quad (13)$$

Using the bounds on $G_n \leq l! 6^l 2^{(n-1)(k-1)}$ and $l \geq k(k+5)/4$ we can bound the part that has the variable n in it: $\sum_{n=m}^M 2^{n(k(k+1)/2-2l)} G_n^2 \leq 2(l!)^2 6^{2l}$.

If $M < 2$ (i.e. $q < 2^{2k}$) divide $S(q)$ into four parts of equal length, use Hölder's inequality and the fact that $q/4 \leq q^{1-1/k}$ to get the same answer, less the factor of M . So may take $(K = 48^{2l}(l!)^2 l^k k^{\binom{k}{2}})$

$$\mathcal{C}'_{q,l,k} = \max(1, M) K q^{2(l-k)/k+3k/2-1/2} \quad (14)$$

Now we can use this inductively to get the following result:

Lemma 7 1. If $r \geq 0$ is an integer and $l \geq k(k + 4r + 1)/4$ then

$$J(q, l) \leq K^r q^{2l - k(k+1)/2 - \delta_r} \log^r q \quad (15)$$

where $\delta_r = k(k + 1)(1 - 1/k)^r/2$.

2. If we take $l = \lfloor k^2 \log(k(k + 1)) + k(k + 5)/4 \rfloor + 1$ then this becomes

$$J(q, l) \leq e^{64k \log^2 k} q^{2l - k(k+1)/2 + 1/2} \log^{2l} q \quad (16)$$

Proof: Inductively use the previous lemma to decrease r by 1 until $r = 0$ when the lemma is obvious.

For the second part note that $\log K \leq 16l \log k$ (Stirling and the definition of l). Also note that $\delta_r \leq 1/2$ because $l < k^3$.

3 Estimates on Weyl Sums

Let $\langle x \rangle$ be the distance to the nearest integer to $x \in \mathbb{R}$.

Exercise 2 Let ϕ be a function so that $\delta \leq \phi(n + 1) - \phi(n) \leq c\delta$ ($0 < \delta, 1 \leq c, c\delta \leq 1/2$). If $T \geq 1$ then the number of $A \leq n \leq A + B$ ($A, B \in \mathbb{N}$) so that $\langle \phi(n) \rangle \leq T\delta$ is at most $(Bc\delta + 1)(2T + 1)$ (for proof refer to the van der Corput handout, or to the Appendix).

Now we get to the main theorem of this paper, the estimate on the Weyl sums:

Theorem 8 Let $k \geq 7, Q \geq 2$ and $f : \mathbb{R} \rightarrow \mathbb{R}$ so that $1 < \lambda \leq \frac{f^{(k+1)}(x)}{(k+1)!} \leq 2\lambda$ for $P + 1 \leq x \leq P + Q$, $\lambda^{-1/4} < Q < \lambda^{-1}$. Prove that

$$\left| \sum_{n=P+1}^{P+Q} e^{2\pi i f(n)} \right| \ll e^{32k \log^2 k} Q^{1-\rho} \log Q \quad (17)$$

where $\rho = 1/(56k^2 \log k)$.

Proof: Write $S = \sum_{n=P+1}^{P+Q} e^{2\pi i f(n)}$, $T(n) = \sum_{m=P+1}^{P+Q} e^{2\pi i (f(m+n) - f(n))}$. Let $e = 1/(6k^2 \log k)$ and $q = \lfloor \lambda^{-\frac{1-e}{k+1}} \rfloor + 1$.

Then

$$q|S| = \left| \sum_1^m \sum e^{2\pi i f(n)} \right| \leq \left| \sum_1^m \sum_{P+1+m}^{P+Q-q+m} e^{2\pi i f(n)} \right| + \sum_1^m q \leq \sum_{P+1}^{P+Q-q} |T(n)| + q^2. \quad (18)$$

By Hölder's inequality this is $\leq q^2 + Q^{1-1/(2l)} \left(\sum_{P+1}^{P+Q-q} |T(n)|^{2l} \right)^{1/(2l)}$.

Taylor expand $f(m+n) - f(n) = \sum_1^k b_i m^i + 2\lambda t q^{k+1}$ with $t \in (0, 1)$. Then for $|\alpha_i - b_i| \leq \lambda q^{k+1}/(2q^i)$ we have

$$|e^{2\pi i(f(m+n)-f(n))} - e^{2\pi i(\sum_j \alpha_j m^j)}| \leq 2\pi k \lambda q^{k+1} = s/2 \quad (19)$$

(use $|e^{ix} - e^{iy}| \leq |x - y|$)

So by Cauchy-Schwarz $\implies |T(n)|^{2l} \leq 2^{2l} |S(q)|^{2l} + s^{2l}$. Since the choice of α_i 's is ours, we can take this inequality for the average on the domain \mathcal{D}_n of the α_i . The volume of the domain is $\mathcal{V}_n = \lambda^k q^{k(k+1)/2}$. So

$$|T(n)|^{2l} \leq \frac{2^{2l}}{\mathcal{V}_n} \int_{\mathcal{D}_n} |S(q)|^{2l} d\alpha + s^{2l} \quad (20)$$

The integral depends only on $\mathcal{D}_n \bmod \mathbb{Z}^k$ so let a, b two integers in $(P+1, P+Q-q)$ so that $\mathcal{D}_a \bmod \mathbb{Z}^k$ and $\mathcal{D}_b \bmod \mathbb{Z}^k$ intersect. Then $\langle b_k(a) - b_k(b) \rangle \leq \lambda q$. Put $\phi(a) = b_k(a) - b_k(b)$.

Then $\phi(n+1) - \phi(n) = f^{(k+1)}(\epsilon)/k!$ so may apply exercise 2 with $c = 2, \delta = \lambda(k+1)$. ($\Delta\phi(n) = \Delta f^{(k)}(n)/k!$). If $T = \frac{q}{k+1}$ then there are at most $3kq$ (from exercise 2) numbers n so that \mathcal{D}_n can intersect \mathcal{D}_b . So each integral over \mathcal{D}_n is covered at most $3kq$ times so

$$\sum_{P+1}^{P+Q} \int_{\mathcal{D}_n} |S(q)|^{2l} d\alpha \leq 3kq \int_{\mathcal{C}} |S(q)|^{2l} d\alpha = 3kq J(q, l) \quad (21)$$

Choose l as in lemma 7, second part. Use the bound on $|S|$ previously found to get that $|S| \ll e^{32k \log^2 k} Q^{1-\rho} \log Q$ (see Appendix). \blacksquare

Observation 9 *Note that if f in the theorem satisfies the inequalities only on an interval $(P+1, P+T)$ with $T < Q$, but $\lambda^{-1/3} < Q < \lambda^{-1}$ then the conclusion still holds.*

This needs explanation if $N < \lambda^{-1/4}$. If that happens, use a trivial bound on the Weyl sum to get the same answer.

4 Note on the Vinogradov Method

So why is this called the *Виноградов* method, rather than the *Виноградов* theorem?

Within the larger context of *Виноградов*'s work, the underlying principle of his method of estimating exponential sums is the possibility to efficiently estimate sums of the form

$$\sum_{u,v} e^{2\pi i h(u,v)} \quad (22)$$

for sufficiently *nice* h, u, v .

In the presented estimate of Weyl sums, the *method* is applied in equation (18), where it relates the value of the Weyl sum to the average on \mathcal{C} . There the variables were $u = m, v = n$ and $h(u, v) = f(m+n) - f(n)$.

Виноградов's method has many applications of this sort. I list a few of them:

1. Estimates of *special* Weyl sums, e.g., sums of the form

$$\sum_{p \leq N} e^{2\pi i f(p)}$$

where p runs over the primes.

2. Estimating the smallest number $r = r(n)$ so that any large enough N can be written as $N = x_1^n + x_2^n + \cdots + x_r^n$.

These problems are covered in [Vinogradov 1958].

5 Applications

I end by giving a few applications:

Exercise 3 *Using the same kinds of partial summation methods we used when we used van der Corput estimates of the Weyl sum, prove that*

$$\zeta(1 + it) = \mathcal{O}((\log t \log \log t)^{\frac{3}{4}}) \quad (23)$$

Exercise 4 *If $0 < \sigma < 1$ then prove that*

$$\sum_a^b \frac{1}{n^{\sigma+it}} = \mathcal{O}(a^{1-\sigma} e^{-c \log^{\frac{1}{4}} t (\log \log t)^{\frac{5}{4}}} \log t) \quad (24)$$

Exercise 5 *Use this to get a zero-free region for the ζ function*

$$\sigma \geq 1 - \frac{c}{(\log t \log \log t)^{\frac{3}{4}}}. \quad (25)$$

[Titchmarsh 1951]

Exercise 6 *Use this to prove that*

$$\pi(x) = li(x) + \mathcal{O}(xe^{-c(\log x)^{4/7}}) \quad (26)$$

Observation: In class we showed this for $\frac{1}{2}$ instead of $\frac{4}{7}$. The best proven constant is $\frac{3}{5}$.
(Hint: $\frac{1}{2} < \frac{4}{7} < \frac{3}{5}$.)

References

[Titchmarsh 1951] Titchmarsh, E.C., *The Theory of the Riemann Zeta-Function*, Oxford: Clarendon, 1951.

[Vinogradov 1958] Vinogradov, I.M., *The Method of Trigonometrical Sums in the Theory of Numbers*, London, New York: Interscience, 1958.

6 Appendix

1. Hölder's Inequality:

If $a_i, b_i \geq 0, p, q > 0, \frac{1}{p} + \frac{1}{q} = 1$ then

$$\left(\sum_1^n a_i^p \right)^{\frac{1}{p}} \left(\sum_1^n a_i^q \right)^{\frac{1}{q}} \geq \sum_1^n a_i b_i \quad (27)$$

2. Hint to exercise 2. For a real x look at the number G of n so that $x+h < \phi(n) \leq x+h+\delta$. Since any h corresponds to at most one n (from the hypothesis). So $G \leq h_2 - h_1$ where h_1, h_2 are the extremal values of h . Also $\phi(A) \leq x + h_1 + \delta, x + h_2 < \phi(A + B)$ so $G \leq Bc\delta + 1$. Divide the domain of $\{\phi(n)\}$ into $2T + 1$ parts of length $< \delta$.

3. Calculations for theorem 8. From the inequality $q|S| \leq \sum_{P+1}^{P+Q-q} |T(n)| + q^2$ we get that

$$|S| \leq q + (2/q)Q^{1-1/(2l)} \left(3kq\lambda^{-k} q^{-k(k+1)/2} J(q, l) + (s/2)^{2l} Q \right)^{1/(2l)} \quad (28)$$

and by Cauchy-Schwarz this is equivalent to

$$|S| \leq q + 4Q^{1-1/(2l)} \left(3kq\lambda^{-k} q^{3/2-k(k+1)/2} e^{64lk \log^2 k} \right)^{1/(2l)} + 2sQ \quad (29)$$

Now use $q \leq 2Q^{4/(k+1)}, Q^{-4e} \leq \lambda q^{k+1} \leq 2^{k+1} Q^{-e}$. Then get

$$|S| \ll e^{32k \log^2 k} Q^{1-1/(2l)+3/((k+1)l)+2ek/l} \log Q + 8\pi k Q^{1-4e} + 2Q^{4/(k+1)} \quad (30)$$

Since $1/2 - 3/(k+1) - 2ek \geq 1/14$ and $l < 4k^2 \log k$ we get the bound sought.