Fourier analytic aspects of Bhargava's proof of van der Waerden's conjecture

ANDREW O'DESKY

Our aim in this talk is to explain Bhargava's proof of van der Waerden's conjecture, with particular attention given to the Fourier analytic aspects of the proof.

Notation: Let $G \subset S_n$ be a permutation group. Let $E(H) = E(G, H)$ denote the number of monic integer polynomials of degree n with height $\leq H$ and Galois group G. Let V denote the affine space of monic degree n polynomials.

0.1. **van der Waerden's conjecture.** Van der Waerden conjectured that $E(H) = O(H^{n-1})$ if $G \neq S_n$. Thanks to earlier work of van der Waerden and Widmer, the remaining case to prove is when G is primitive. If G is primitive and $\neq S_n$, then its index is ≥ 2 (defined as $\min_{g \neq 1}(n - \#\mathrm{orb}(g))$).

0.2. **Polynomials with index at least k.** Certain subschemes of V which refine the discriminant locus of V play an important role in the proof. A *splitting type* σ is an unordered tuple of pairs of integers (f_i, e_i) written $(f_1^{e_1} \cdots f_r^{e_r})$. The *degree* of σ is $\sum_i f_i e_i$ and its *index* is $\sum_i f_i(e_i - 1)$. The index of a polynomial is the index of the splitting type defined by its irreducible factorization. The index of a polynomial is stable under separable field extensions. For any $k \geq 1$ let $V_k \subset V$ denote the closed subscheme parametrizing polynomials with index at least k. For example, V_1 is the discriminant locus.

Write $f((p))$ for the image of f in $V(\mathbb{F}_p)$.

Proposition 0.1. *If $f \in V(\mathbb{Z})$ has Galois group G, then $f((p)) \in V_{\mathrm{ind}(G)}$ for each prime p which is ramified in the root field K_f of f.*

Proof: (p odd, index 2). The action of inertia I_p on the set of complex embeddings of K_f can be used to show that $v_p(D_f) \geq \mathrm{ind}(G)$, so $v_p(\mathrm{disc}(f)) \geq v_p(D_f) \geq \mathrm{ind}(G) \geq 2$. If f has index one modulo p then $f = gq$ over $\mathbb{Q}_p$ where the reductions of g and q modulo p are coprime, the reduction of g modulo p is squarefree, and q is quadratic and Eisenstein. But this would imply that p ramifies with degree two in K_f. Assuming p is odd this is tame so $v_p(D_f) = e_p - 1 = 1$, a contradiction. As p is ramified f cannot have index zero modulo p, so the index is at least two. $\square$

Date: October 11, 2022. v1. Lecture notes for a talk in the arithmetic statistics seminar at Princeton University.

This proposition forms the basis for a sieve-type argument. Even though there are only finitely many local conditions (one for each ramified prime) and the local conditions depend on the point, these are strong local conditions since they are restriction to a codimension-k subset of the mod p fiber where $k = \mathrm{ind}(G) \geq 2$. (We are counting a "mod C_f Type I thin set".)

0.3. **Up to an epsilon.** Without Fourier analysis we can prove that $E(H) = O_\varepsilon(H^{n-1+\varepsilon})$ as follows.

Let $E^{sm}(H)$ be the number of f in $E(H)$ such that the product C of ramified primes in the root field K_f of f (without multiplicity) is $\leq H$, and let $E^{big}(H) = E(H) - E^{sm}(H)$. Let $C = p_1 \cdots p_r$ and let $D = p_1^{k_1} \cdots p_r^{k_r}$ for some positive integers $k_1, \ldots, k_r$. The fraction of polynomials in $V(\mathbb{F}_p)$ with index k is $O(p^{-k})$ (see the proof of Prop. 0.2 for justification), so by the Chinese remainder theorem the fraction of polynomials in $V(\mathbb{Z}/C\mathbb{Z})$ whose reduction modulo p_i has index k_i for each p_i is $O(c^{\omega(C)}/D)$ for some constant $c > 0$. If $C \leq H$, then the number of polynomials in $V(\mathbb{Z})$ with height $\leq H$ whose reduction modulo p_i has index k_i for each p_i is $O(H^n c^{\omega(C)}/D)$.

The form of this upper bound suggests forming a tail estimate for large D. By the proposition, the discriminant D_f of K_f is k-power-full, which gives

$$\begin{aligned}\#\{f \in E^{sm}(H) : D_f > H^2\} &= \sum_{D > H^2 \ k\text{-power-full}} O(H^n c^{\omega(C)}/D) \\ &= O(H^n c^{\omega(C)})(H^2)^{-(k-1)/k} = O_\varepsilon(H^{n-2(k-1)/k+\varepsilon})\end{aligned}$$

by partial summation (integration by parts). Namely, $\#\{D \leq X \ k\text{-power-full}\} \sim X^{1/k}$ so by partial summation

$$\begin{aligned}\sum_{Y < D \leq X \ k\text{-power-full}} D^{-s} &= O(X^{1/k-s}) + O(Y^{1/k-s}) + \int_Y^X O(Z^{1/k-s-1})\, dZ \\ &\xrightarrow{X \to \infty, s=1} O(Y^{-(k-1)/k}).\end{aligned}$$

For those $f \in E^{sm}(H)$ with $D_f < H^2$ we appeal to two known bounds: the number of K_f with discriminant $\leq X$ is at most $O(X^{(n+2)/4})$ by a result of Schmidt, and the number of f of height $\leq H$ for a given $K = K_f$ is at most $O_\varepsilon(H^{1+\varepsilon})$ by a result of Lemke-Oliver–Thorne. So the number of $f \in E^{sm}(H)$ with $D_f < H^2$ is at most $O((H^2)^{(n+2)/4}) O_\varepsilon(H^{1+\varepsilon})$. This has exponent $n/2 + 2 + \varepsilon$ which is $n - 1 + \varepsilon$ for $n = 6$ and $< n - 1$ for $n \geq 7$. For $n \leq 5$ the number of K_f with discriminant $\leq X$ is known to be $\sim cX$ and for such n one gets $O(H^2) O_\varepsilon(H^{1+\varepsilon})$. This works for $n = 4, 5$, Chow–Dietmann have shown it for $n = 3$, while $n = 2$ is trivial. This shows $E^{sm}(H) = O_\varepsilon(H^{n-1+\varepsilon})$ for all n.

Now we turn to $E^{big}(H)$ (recall this counts f with $C > H$). The key is the following observation, which can be proven by a clever elementary argument. If we specify the first r coefficients $a_1, \ldots, a_r$ of f in any field with characteristic

$> n$, then there are at most $r!$ many choices for $a_{r+1}, \ldots, a_n$ such that f has index $k = n - r$. In other words, the restriction of the projection map $V \to \mathbb{A}^r$ to V_k is quasi-finite away from primes dividing n, with uniformly bounded fiber cardinalities.

Say V_k is contained in the vanishing set of polynomials $g_1, \ldots, g_k$. Using successive resultants we can assume that g_1 only involves the variables $a_1, \ldots, a_{n-k+1}$. Fix coefficients $a_1, \ldots, a_{n-k+1} \in [-H, H]$. We claim that for almost all such choices, the rest of the coefficients of f are determined up to $O_\varepsilon(H^\varepsilon)$ many choices. First observe there are at most $O(H^{n-k})$ choices for these coefficients such that $g_1(f) = g_1(a_1, \ldots, a_{n-k+1}) = 0$. So assume $g_1(f) \neq 0$ in which case it still vanishes modulo C. There are $O_\varepsilon(H^\varepsilon)$ many divisors of $g_1(f)$ so C is determined up to $O_\varepsilon(H^\varepsilon)$ many possibilities. Once C is determined, then for any prime p dividing C and greater than n, there are only $O(1)$ many choices for $a_{n-k+2}, \ldots, a_n$ modulo p which obtain a polynomial $f \in V_k(\mathbb{F}_p)$, so f is determined modulo C up to $O_\varepsilon(H^\varepsilon)$ by the Chinese remainder theorem. Since $C > H$ this actually determines f. The number of choices modulo primes less than n is bounded also, so in total there are $O_\varepsilon(H^{n-k+1+\varepsilon})$ many choices for f.

Remark 1. The polynomial $g_1(a_1, \ldots, a_{n-1})$ for $k = 2$ was precisely the "double discriminant" $DD(f) = DD(a_1, \ldots, a_{n-1})$.

Remark 2. More generally, the proof shows that a subset $W \subset \mathbb{Z}^n$ has at most $O_\varepsilon(H^{n-k+1+\varepsilon})$ many elements in $[-H, H]$ if there is a hypersurface $H \subset \mathbb{A}^{n-k+1}$ with the property that for every $f \in W$ there is a positive integer $C > H$ such that $f((p)) \in H \times \mathbb{A}^{k-1}$ for all p dividing C.

So we've shown that $E^{sm}(H) = O_\varepsilon(H^{n-1+\varepsilon})$ and $E^{big}(H) = O_\varepsilon(H^{n-k+1+\varepsilon})$. Remember that this division was made on the basis of whether $C \leq H$. The Fourier analysis shows that in fact $E^{sm}(H) = O(H^{n-1-\mu})$ for some $\mu > 0$; equivalently, we can redefine the division between E^{sm} and E^{big} to $C \leq H^{1+\delta}$ and show that we have the same bound $E^{sm}(H) = O_\varepsilon(H^{n-1+\varepsilon})$.

0.4. **The proof.** Our goal is to prove that the bound

$$O(c^{\omega(C)} H^n / D)$$

still holds even if C is as large as $H^{1+\delta}$ for some positive δ.

Fix a prime p and a splitting type σ of degree n and index k.

Proposition 0.2. *Let 1_σ denote the characteristic function on the subset of polynomials in $V(\mathbb{F}_p)$ with type σ. Then for some positive constant c_σ,*

$$\widehat{1_\sigma}(g) = \begin{cases} c_\sigma p^{-k} + O(p^{-(k+1)}) & \text{if } g = 0, \\ O(p^{-(k+1/2)}) & \text{if } g \neq 0. \end{cases}$$

Proof. First we evaluate the number of polynomials in $V(\mathbb{F}_p)$ of type σ. Observe there is a surjective function (write $\sigma = (f_1^{e_1} \cdots f_r^{e_r})$)

$$\prod_{i=1}^{r} \{g_i \in \mathbb{F}_p[x] \text{ monic irred. of degree } f_i\} \to \{f \in \mathbb{F}_p[x] : \sigma(f) = \sigma\}$$
$$(g_1, \ldots, g_r) \mapsto g_1^{e_1} \cdots g_r^{e_r}$$

whose fibers have cardinalities that are bounded independently of p (e.g. $\leq r!$). Since $\#\{g \text{ irred. of degree } f\} = \frac{1}{f} p^f + O(p^{f-1})$, this gives

$$\#\{f \in \mathbb{F}_p[x] : \sigma(f) = \sigma\} = \prod_{i=1}^{r} (\tfrac{1}{f_i} p^{f_i} + O(p^{f_i - 1})) = c_\sigma p^{n-k} + O(p^{n-(k+1)}).$$

This proves the formula for $\widehat{1_\sigma}(0)$, which is p^{-n} times this quantity.

Now suppose $g \neq 0$. We have that

$$\widehat{1_\sigma}(g) = \frac{1}{p^n} \sum_{f : \sigma(f) = \sigma} \psi(f, g).$$

The key observation is that any translation $f(x + c)$ with $c \in \mathbb{F}_p$ has the same splitting type as $f(x)$, and that grouping summands according to these additive orbits leads to exponential sums of "Weil-type". It is easy to show that if m is the largest index such that $g(x^{n-m}) \neq 0$ then $g(f(x + c))$ is equal to

$$c^m g(x^{n-m} y^m) \binom{n}{n-m} + O(c^{m-1})$$

and is therefore a nonzero degree m polynomial in c if $g \neq 0$. A special case of an inequality of Weil (following from the Riemann hypothesis for curves over finite fields) says that for any non-constant polynomial $Q \in \mathbb{F}_p[c]$ we have

$$\left| \sum_{c \in \mathbb{F}_p} \psi(Q(c)) \right| \leq (\deg Q - 1)\sqrt{p}.$$

Thus

$$\begin{aligned} \widehat{1_\sigma}(g) = \frac{1}{p^n} \sum_{[f]} \sum_{c \in \mathbb{F}_p} \psi(f(x + c), g) = \frac{1}{p^n} \sum_{[f]} (m - 1)\sqrt{p} &= O(p^{-n} p^{n-k-1} p^{1/2}) \\ &= O(p^{-(k+1/2)}). \end{aligned}$$

□

Remark 3. Bhargava also considers a characteristic function w_σ with positive weights for σ of degree *less than* n. He shows that $\widehat{w_\sigma}(g) = O(p^{-(k+1)})$ for $g \neq 0$, but this more general context is not needed for the proof of van der Waerden's conjecture for monic polynomials.

Corollary 0.3. *Let $0 < \delta < 1/(2n-1)$. Let $D = p_1^{k_1} \cdots p_m^{k_m}$ be an integer such that $C = p_1 \cdots p_m < H^{1+\delta}$. Then the number of $f \in V(\mathbb{Z})$ of height $\leq H$ that, modulo p_i, have index at least k_i for every i is at most $O(c^{\omega(C)} H^n / D)$.*

Proof. Recall the twisted Poisson formula:

$$\sum_{f \in V(\mathbb{Z})} \Psi(f \pmod C)\phi(f/H) = H^n \sum_{g \in V(\mathbb{Z})^\vee} \widehat{\Psi}(g \pmod C)\widehat{\phi}(gH/C)$$

where ϕ is a Schwartz function on $V(\mathbb{R})$, $\widehat{\phi}$ is the Fourier transform of ϕ, and $\Psi \colon V(\mathbb{Z}/C\mathbb{Z}) \to \mathbb{C}$ is any set-theoretic function.

We will apply this with

$$\Psi(f \pmod C) = \prod_{i=1}^m 1_{V_{k_i}}(f \pmod{p_i}) = \prod_{i=1}^m \sum_{\sigma : \mathrm{ind}(\sigma) \geq k_i} 1_\sigma(f \pmod{p_i})$$

and ϕ with compact support and identically one on $[-1,1]^n \subset V(\mathbb{R})$. First observe that

$$\widehat{\Psi}(g \pmod C) = \prod_{i=1}^r \widehat{1_{V_{k_i}}}(g \pmod{p_i})$$

(for Fourier transforms modulo p_i with respect to suitably chosen additive characters). Then

$$\sum_{f \in V(\mathbb{Z})} \Psi(f \pmod C)\phi(f/H)$$
$$= H^n \left(\prod_{i=1}^m O(p^{-k_i}) \right) \widehat{\phi}(0) + H^n \left(\prod_{i=1}^m O(p^{-(k_i+1/2)}) \right) \sum_{0 \neq g \in V(\mathbb{Z})^\vee} |\widehat{\phi}(gH/C)|. \quad (1)$$

The first term gives the dominant term $O(H^n c^{\omega(C)}/D)$.

To bound the second term we collect the summands for which gH/C lies in a box $B(\varepsilon)$ of sidelength C^ε (for any $\varepsilon > 0$). For the summands outside the box, and any positive integer N, we have that

$$\sum_{gH/C \notin B(\varepsilon)} |\widehat{\phi}(gH/C)| \leq \sum_{gH/C \notin B(\varepsilon)} (gH/C)^{-N}$$

since $\widehat{\phi}$ is Schwartz. Since $\|gH/C\| > C^\varepsilon > 1$, by choosing $N \gg_{\varepsilon,n} 1$ we can arrange that this term is absorbed into the dominant term. Inside the box, there are at most $(C^\varepsilon)^n (C/H)^n$ many $0 \neq g \in V(\mathbb{Z})^\vee$ such that $gH/C \in B(\varepsilon)$. Now it may be the case that g can vanish modulo primes dividing C, in which case the worse bound on Fourier coefficients must be used, but for simplicity let's say $C = p$. Then this doesn't occur (since $0 \neq \|g\| \leq C^{1+\varepsilon}/H < C$) so the second term of (1) is at most

$$H^n \cdot O(c^{\omega(C)}/(D\sqrt{C})) \cdot O_\varepsilon(C^\varepsilon (C/H)^n) = O_\varepsilon(C^{n-1/2+\varepsilon}/D).$$

(If C is composite there is a bit more work to take care of the primes dividing C where g reduces to zero but one gets the same bound in the end.)

Altogether, (1) shows that the number of $f \in V(\mathbb{Z})$ of height $\leq H$ any that, modulo p_i, have index at least k_i for every i is at most $O(c^{\omega(C)}H^n/D) + O_\varepsilon(C^{n-1/2+\varepsilon}/D)$. The hypothesis that $C < H^{1+\delta}$ is optimally chosen so that the second term is smaller than the first. $\square$