STEPANOV'S METHOD FOR HYPERELLIPTIC CURVES

BRANDON HANSON

1. Introduction

In this this note we give a description of how to apply Stepanov's method to get a good estimate for the number of points on the hyperelliptic curve

$$y^2 = f(x)$$

over a finite field with p elements, where $f(x) \in \mathbb{F}_p[x]$ is a polynomial of degree d which is not a square in $\overline{\mathbb{F}_p}[x]$.

Let N denote the number of points on the curve, i.e. the number of solutions $(x, y) \in \mathbb{F}_p^2$ with $y^2 = f(x)$. If (x, y) is a solution with $y = 0$ then x is a root of $f(x)$, in which case there are at most d choices for x. If d is small, we might think of this as an error term. For any other choice of x, $f(x) \neq 0$, and $f(x) = y^2$ is only possible if $f(x)$ is a quadratic residue. The is no obvious reason for $f(x)$ to be a quadratic residue (after all, $f(x)$ is not the square of some other polynomial) so we think that it has about a 50/50 chance of being a quadratic residue. But if there is a solution, there are in fact two solutions, namely (x, y) and $(x, -y)$. So we expect the number of solutions (x, y) with $y \neq 0$ to be about $2\frac{p-1}{2} = p - 1$. All in all, we expect there to be about p solutions to the equation. What we will show is that this is indeed the case:

Theorem 1. *Let $f(x) \in \mathbb{F}_p[x]$ be a polynomial of degree $d \geq 3$ which is not a square in $\overline{\mathbb{F}_p}[x]$. Then, if $p > 4d^2$, we have*

$$|N - p| \leq 8d\sqrt{p}.$$

We are going to actually deduce this from an upper bound on the size of the set

$$X_a = \{x \in \mathbb{F}_p : f(x) = 0 \text{ or } f(x)^{\frac{p-1}{2}} = a\}.$$

The reason for looking at this set is that any non-zero element $z \in \mathbb{F}_p$ satisfies $z^{p-1} = 1$ and so if $f(x) = y^2$ then $f(x)^{\frac{p-1}{2}} = y^{p-1} = 1$. In this case $x \in X_1$. Meanwhile, if x is such that $f(x)$ is a quadratic non-residue then $f(x)^{\frac{p-1}{2}} = -1$ and $x \in X_{-1}$. Since X_1 and X_{-1} satisfy

$$|X_1| + |X_{-1}| = p + |\{x : f(x) = 0\}|$$

we will be able to turn use upper bounds to prove lower bounds.

Stepanov's method uses the following simple idea, pioneered by Thue, in a beautiful way: if $r(x)$ is a non-zero polynomial of degree D and $r(x)$ has a zero of order l at distinct values $x_1, \ldots, x_n$ then $n \leq D/l$. This fact is basically just the prime factorization of polynomials. By using the relation

$$f(x)^{\frac{p-1}{2}} = a$$

we will build a polynomial (using linear algebra) to create a low-degree polynomial $r(x)$ which has zeros of high order at each element of X_a. This will help us bound $|X_a|$ from above.

2. Hasse Derivatives

There is a bit of a snag however. Usually, Taylor expansion tells us that a polynomial $r(x)$ has a zero of order l at x_0 if all of the $l-1$'th derivatives of r vanish at x_0. We run into trouble with this fact over $\mathbb{F}_p$ because $\frac{d}{dx}(x^p) = px^{p-1} = 0$. This affects the Taylor expansion of a polynomial since one usually needs to divide by $n!$ which is no longer non-zero. So to get zeros of high order and not have to deal with the characteristic of the field, we have to work with a slightly more complicated differential operator: the Hasse derivatives.

Definition (Hasse Derivative)**.** *We define the Hasse derivative of order k, E^k, by setting $E^k(x^n) = \binom{n}{k}x^{n-k}$ and extending linearly to all polynomials.*

One big downside of these operators is that now we don't have the usual convention that a k'th order derivative is just a first derivative applied k times. Said differently, the operator E^k is not k applications of E^1. However there are other formulae that come out nicer in the language of Hasse derivatives. For instance, by the binomial theorem,

$$x^n = ((x-a+a)^n) = \sum_{k=0}^{n} \binom{n}{k} a^{n-k}(x-a)^k$$

so that the coefficient of x^k in it's expansion about a is $E^k(x^n)$ evaluated at k. Also, and this is crucial, $E^k(x^p) = \binom{p}{k} x^{p-k}$ which vanishes for $k = 0, \ldots, p-1$ but does not vanish at $k = p$.

Lemma 1. *For any two polynomials f and g we have*

$$E^k(fg) = \sum_{s=0}^{k} E^s(f)E^{k-s}(g).$$

In general,

$$E^k(f_1 \cdots f_r) = \sum_{j_1+\cdots+j_r=k}^{k} E^{j_1}(f_1)\cdots E^{j_r}(f_r).$$

Proof. If $f(x) = \sum a_i x^i$ and $g(x) = \sum b_j x^j$ then

$$E^k(fg) = \sum_{i,j} a_i b_j E^k(x^{i+j}) = \sum_{i,j} a_i b_j \binom{i+j}{k} x^{i+j-k}$$

meanwhile the right hand side is

$$\sum_{s=0}^{k} E^s(f)E^{k-s}(g) = \sum_{s=0}^{k}\sum_{i}\sum_{j} a_i \binom{i}{s} x^{i-s} b_j \binom{j}{k-s} x^{j-k+s}.$$

The first identity follows from the fact that

$$\binom{i+j}{k} = \sum_{s=0}^{k} \binom{i}{s}\binom{j}{k-s}.$$

The second claim follows by induction on r. □

We can use this lemma to derive some more natural properties of Hasse derivatives.

Lemma 2. *Let $a \in \mathbb{F}_p$. Then*

$$E^k\left((x-a)^r\right) = \binom{r}{k}(x-a)^{r-k}.$$

If $0 \le k \le r$ then for any polynomials f and g we have

$$E^k(fg^r) = hg^{r-k}$$

for some polynomial $h(x)$ with $\deg(h) \le \deg(f) + k\deg(g) - k$.

This last consequence is an analog of the familiar rule: if you take k derivatives of the r'th power of g then you still have something which is divisible by g^{r-k}.

Proof. For the first claim, apply part 2 of Lemma 1 with $f_i(x) = (x-a)$ for each i. Then the only way a derivative $E^j(x-a)$ is non-zero is if $j = 0$ or $j = 1$. In this way we have $\binom{r}{k}$ choices to place the derivatives with $k = 1$ and each such derivative is 1. The remaining choices have $j = 0$ so we are applying the identity operator and are left with a factor $(x-a)$.

For the second claim, again apply part 2 of Lemma 1 with $f_1 = f$ and $f_i = g$ for $i = 2, \ldots, r+1$. Since $k \le r$, there are at least $k - r$ values of j_i which must be zero in each summand. Hence we are left with a factor of g^{k-r} in each summand, and so the entire expression is divisible by g^{k-r}. The degree restriction on h follows from the fact that the Hasse derivative decreases the degree of the polynomial by at least k. Hence $\deg(h) \le (\deg(f) + r\deg(g)) - k - (r-k)\deg(g)$. $\square$

Finally, we can derive the fact that we really need Hasse derivatives to obey, which is that many vanishing derivatives of a polynomial means a high order zero at that point. Specifically,

Lemma 3. *Suppose f is a polynomial and $a \in \mathbb{F}_p$ is such that $(E^k(f))(a) = 0$ for $0 \le k \le l-1$. Then $(x-a)^l$ divides f.*

Proof. Write $f(x)$ in terms of the basis of polynomials $(x-a)^j$:

$$f(x) = \sum_j c_j(x-a)^j.$$

Then by Lemma 2,

$$E^k(f(x)) = \sum_j c_j \binom{j}{k} (x-a)^{j-k}.$$

Plugging in $x = a$, the only term which survives is $k = j$ and the constant and so we are left with a constant c_k which must therefore vanish (by our hypothesis). So $c_k = 0$ for $k < l$ and the lemma follows. □

The final lemma we will prove is a bit more technical. In the process of constructing our polynomial we will use polynomials in two variables. This, from a linear algebra perspective, gives us more free variables. Then we will collapse down to one variable by setting $y = x^p$ and using the relation $x^p = x$, which will reduce the number of linear equations we need to solve. So we will have need to take Hasse derivatives of polynomials of the form $h(x, x^p)$, where $h(x, y) \in \mathbb{F}_p[x, y]$. To help with that, we have the following.

Lemma 4. *Suppose $h(x, y)$ is a polynomial and $r(x) = h(x, x^p)$. Let $E_x^k(h)$ denote the k'th order Hasse derivative of h with respect to x (i.e. applied to $h(x, y)$ with the variable y treated as a constant). Then, for $k < p$,*

$$E^k(r(x)) = E_x^k(h)(x, x^p).$$

The lemma is essentially saying that a certain "diagram" commutes. We can substitute $y = x^p$ and then apply Hasse derivatives or else we can apply a Hasse derivative in the x variable only, and then make the substitution $y = x^p$.

Proof. First, if $h(x, y) = x^m y^n$, then $r(x) = x^m (x^p)^n$ and by Lemma 1

$$E^k(x^m x^{pn}) = \sum_{s=0}^{k} E^{s-k}(x^m) E^s(x^{pn})$$

while $E_x^k(h(x, y)) = y^n \binom{m}{k} x^{m-k}$ so that

$$E_x^k(h)(x, x^p) = \binom{m}{k} x^{m+np-k}.$$

For $s>0$, $E^s(x^{pn})=\binom{pn}{s}x^{pn-s}$ and since $s\le k<p$ the binomial coefficient is divisible by p. Thus

$$E^k(r(x))=\binom{m}{k}x^{m+np-k}=E^k_x(h)(x,x^p).$$

Now, if $h(x,y)=\sum_{m,n}c_{m,n}x^my^n$ then the result holds by linearity and the above case. $\square$

3. Constructing the auxiliary polynomial

Now assume as in the introduction that we have a polynomial f of degree $d\ge 3$ which is not a square in $\overline{\mathbb{F}_p}[x]$, and $a\in\mathbb{F}_p$. We want to find a polynomial which vanishes to high order on

$$X_a=\{x\in\mathbb{F}_p : f(x)=0 \text{ or } f(x)^{\frac{p-1}{2}}=a\}.$$

The next proposition nearly does this.

Proposition 1 (Existence of an auxiliary polynomial)**.** *Assume $p>8d$ and let l be an integer in the range $d<l\le p/8$. There is a non-zero polynomial $r\in\mathbb{F}_p[x]$ of degree*

$$\deg(r)<\frac{p-1}{2}l+2dl(l-1)+dp$$

which has a zero of order l at each $x_0\in X_a$.

The first step is to hone in on the right sort of polynomial we ought to look for. In this case, set $g(x)=f(x)^{\frac{p-1}{2}}$ and we try a polynomial of the form

$$r(x)=f^l\sum_{0\le j<J}(r_j(x)+g(x)s_j(x))x^{jp}$$

where $r_j,s_j\in\mathbb{F}_p[x]$ are to be determined.

Remark. *Why is this a good type of polynomial to try? First, the factor f^l is there mostly to counteract differentiation: if we take l derivatives of $f^{\frac{p-1}{2}}$ we get something divisible by $f^{\frac{p-1}{2}-l}$ and the extra f^l gives us a factor $f^{\frac{p-1}{2}}$ which collapses down to just a or 0 on substituting in $x\in S_a$. The rest of the expression for this polynomial is not too much* *worse: we are basically separating the terms according to the degrees in ranges $[jp,(j+1)p)$. Indeed, any polynomial can be written*

$$R(x) = \sum_{0\le j<J} R_j(x)$$

where all terms in $R_j(x)$ have degree in $[jp,(j+1)p)$. Factoring out x^{jp} from $R_j(x)$ we get $R_j(x) = x^{jp}S_j(x)$ where $\deg(S_j) < p$. Assume the r_j and s_j terms have low degree. Then, since $g(x)$ is a $\frac{p-1}{2}$ 'th power of f, if say $f(0) = 0$, each term in $s_j(x)g(x)$ has degree at least $\frac{p-1}{2}$ and each term in $r_j(x)$ is of degree at most $\frac{p-1}{2}$, we can sort of see this as breaking down the $S_j(x)$ into the high degree parts and low degree parts.

Assume now that each r_j and s_j has degree bounded by $\frac{p-1}{2} - d$. Then the degree of r satisfies

$$\deg r \le ld + Jp + \frac{p-1}{2} - d + \frac{p-1}{2}d \le (J+d)p.$$

Next, all the work will have been for not if the polynomial we construct is identically zero. This is where the hypothesis that f is not a square will come in.

Lemma 5 (The auxiliary polynomial is non-zero)**.** *Suppose*

$$r(x) = f^l \sum_{0\le j<J} (r_j(x) + g(x)s_j(x))x^{jp}$$

where each r_j and s_j has degree bounded by $\frac{p-1}{2} - d$. If f is not a square in $\overline{\mathbb{F}_p}[x]$ then $r = 0$ only if $s_j = r_j = 0$ for each j.

Proof. Assume, by making the change of variables $x \mapsto x + a$ that $f(0) \neq 0$. Suppose, by way of contradiction, that $r = 0$ but some s_j or r_j is non-zero and let k be the least index of such a j.. We can divide r by $f^l x^{kp}$ to get

$$\sum_{k\le j<J} (r_j(x) + s_j(x)g(x))x^{p(j-k)} = 0.$$

Group the terms with $g = f^{\frac{p-1}{2}}$ and rewrite this as

$$h_1 = -h_2 g$$

where

$$h_1(x) = \sum_{k\leq j<J} r_j(x)x^{p(j-k)},\ h_2(x) = \sum_{k\leq j<J} s_j(x)x^{p(j-k)}$$

so that upon squaring and multiplying by f, we get

$$h_1^2 f = h_2^2 f^p.$$

Reduce this equation modulo the polynomial x^p. Then

$$\begin{aligned} r_k(x)^2 f(x) &= h_1(x)^2 f(x) \bmod x^p \\ &= h_2(x)^2 f(x)^p \bmod x^p \\ &= h_2(x)^2 f(x^p) \bmod x^p \\ &= s_k(x)^2 f(0) \bmod x^p. \end{aligned}$$

We have used $f(x)^p = f(x^p)$, in light of the fact we are in characteristic p. Now, the degree constraints on s_k and r_k, plus the fact that one of them is non-zero, means that $r_k(x)^2 f(x) - s_k(x)^2 f(0)$ cannot be divisible by x^p unless it is zero. Thus we must in fact have

$$r_k(x)^2 f(x) = s_k(x)^2 f(0)$$

which is impossible since it would imply (by factoring $f(0) = t^2$ in some extension) that $f(x)$ is in fact a square in $\overline{\mathbb{F}_p}[x]$. □

Next we take derivatives of our polynomial.

Lemma 6. *Suppose*

$$r(x) = f^l \sum_{0\leq j<J} (r_j(x) + g(x)s_j(x))x^{jp}$$

where each r_j and s_j has degree bounded by $\frac{p-1}{2} - d$. For each k with $0 \leq k < l$ we have

$$E^k(r(x)) = f^{l-k} \sum_{0\leq j<J} (r_j^{(k)}(x) + g(x)s_j^{(k)}(x))x^{jp}$$

where $r_j^{(k)}(x)$ and $s_j^{(k)}$ are polynomials of degree at most

$$\frac{p-1}{2} - d + k(d-1).$$

Proof. To make things simple, we write $r(x) = h(x, x^p)$ where

$$\begin{aligned} h(x,y) &= f(x)^l \sum_{0\le j<J} (r_j(x) + g(x)s_j(x))y^j \\ &= \sum_{0\le j<J} (f(x)^l r_j(x) + f(x)^{\frac{p-1}{2}+l} s_j(x))y^j. \end{aligned}$$

By Lemma 4 and linearity

$$\begin{aligned} E^k(r(x)) &= E^k_x(h(x,y))(x,x^p) \\ &= f(x)^l \sum_{0\le j<J} (E^k(r_j(x)f(x)^l) + E^k(f(x)^{\frac{p-1}{2}+l} s_j(x)))x^{pj}. \end{aligned}$$

By Lemma 2 applied to $E^k(r_j(x)f(x)^l)$ and $E^k(f(x)^{\frac{p-1}{2}+l}s_j(x))$, there are polynomials $r_j^{(k)}$ and $s_j^{(k)}$, of degrees

$$\deg(r_j^{(k)}) \le \deg(r_j) + k\deg(f) - k \le \frac{p-1}{2} - d + k(d-1)$$

and

$$\deg(s_j^{(k)}) \le \deg(s_j) + k\deg(f) - k \le \frac{p-1}{2} - d + k(d-1),$$

and such that

$$E^k(r_j f^l) = r_j^{(k)} f^{l-k},\ E^k(s_j f^{\frac{p-1}{2}+l}) = s_j^{(k)} f^{\frac{p-1}{2}+l-k},$$

which is just what we wanted to prove. □

Now we can prove Proposition 1.

Proof of Proposition 1. Let $x_0 \in X_a$. We want to ensure that the polynomial $r(x)$ has a zero of order at least l at x_0. To that end, we consider (using Lemma 3) the value of $E^k(r(x))$ at x_0. By Lemma 6,

$$E^k(r(x)) = f^{l-k} \sum_{0\le j<J} (r_j^{(k)}(x) + g(x)s_j^{(k)}(x))x^{jp}$$

and if we substitute in x_0 we get

$$E^k(r(x_0)) = f^{l-k}(x_0) \sum_{0\le j<J} (r_j^{(k)}(x_0) + bs_j^{(k)}(x_0))x_0^j$$

where $b = 0$ or $b = a$. But $b = 0$ means that $f(x_0) = 0$ which means that the term $f(x_0)$ out front vanishes already. So $b = a$, and we can rewrite this as

$$E^k(r(x_0)) = f^{l-k}(x_0)\sigma_k(x_0)$$

where
$$\sigma_k(x) = \sum_{0\le j<J} (r_j^{(k)}(x) + as_j^{(k)}(x))x^j$$
which is a polynomial of *much smaller degree* than $r(x)$. So to satisfy $\sigma_k(x_0) = 0$ for each k and x_0, it is certainly sufficient that σ_k is the 0-polynomial for each k. This imposes $\deg(\sigma_k) + 1$ linear constraints (one for each coefficient of σ_k) for each k. Thus the total number of linear equations we wish to have vanish is
$$\sum_{k\le l-1} \deg(\sigma_k) \le l\left(J + \frac{p-1}{2} - d + \frac{1}{2}(l-1)(d-1)\right).$$
On the other hand we have $2\left(\frac{p-1}{2} - d\right)$ coefficients to choose from for each r_j and s_j, which gives us $2J\left(\frac{p-1}{2} - d\right)$ variables. Take
$$J = \left[\frac{l}{p}\left(\frac{p-1}{2} + 2d(l-1)\right)\right]$$
then
$$J \ge \frac{l}{p}\left(\frac{p-1}{2} + 2d(l-1)\right) - 1$$
which will be enough to have more variables than constraints. In this case,
$$\deg(r) \le ld + \frac{p-1}{2}d + \frac{p-1}{2} - d + \frac{l}{p}\left(\frac{p-1}{2} + 2d(l-1)\right)$$
which is good enough to prove the proposition. $\square$

We can now prove our theorem.

Proof. By Proposition 1, there is a non-zero polynomial r of degree at most
$$\frac{p-1}{2}l + 2dl(l-1) + dp$$
which has a zero of order l at each point of X_a. Thus $(x-x_0)^l$ divides $r(x)$ for each $x_0 \in X_a$ which means
$$|X_a| \le \frac{p-1}{2} + 2d(l-1) + \frac{dp}{l}.$$
We take $l = 1 + \left[\frac{\sqrt{p}}{2}\right]$. Then
$$|X_a| \le \frac{p-1}{2} + 4d\sqrt{p}.$$

Applying this upper bound to X_1 tells us that the curve has at most $p + 8d\sqrt{p}$ points on it. Using the fact that

$$|X_1| + |X_{-1}| = p + |\{x : f(x) = 0\}|$$

we see

$$N \geq 2|X_1| \geq 2(p - |X_{-1}|) \geq p - 8d\sqrt{p}$$

which gives the corresponding lower bound. □

4. Application: The Burgess Bound

We now use the above estimate to estimate very short character sums with the Legendre symbol. To keep things simple, we will just estimate the sum

$$S = \sum_{1 \leq n \leq N} \left(\frac{n}{p}\right),$$

but sums over other arithmetic progressions can be estimated the same way. Also, we can assume $N \leq p^{1/2+\varepsilon}$ since otherwise Polya-Vinogradov applies. Observe that since $\left(\frac{n}{p}\right)$ is bounded by 1, then

$$\left| S - \sum_{1 \leq n \leq N} \left(\frac{n+h}{p}\right) \right| \leq 2h.$$

Taking $h = ab$ and summing over all a in the range $1 \leq a \leq A$ and b in the range $1 \leq b \leq B$ we get

$$S = \frac{1}{AB} \sum_{1 \leq n \leq N} \sum_{1 \leq a \leq A} \sum_{1 \leq b \leq B} \left(\frac{n+ab}{p}\right) + O(AB).$$

Let's now focus on the new sum

$$T = \sum_{1 \leq n \leq N} \sum_{1 \leq a \leq A} \sum_{1 \leq b \leq B} \left(\frac{n+ab}{p}\right).$$

By the triangle inequality and the fact that the Legendre symbol is multiplicative,

$$|T| \leq \sum_{1 \leq n \leq N} \sum_{1 \leq a \leq A} \left| \sum_{1 \leq b \leq B} \left(\frac{na^{-1}+b}{p}\right) \right|.$$

Set $na^{-1} = x$. Then the inner most sum is

$$\sum_{1 \leq b \leq B} \left(\frac{x+b}{p}\right),$$

and we count this sum every time we have can represent x in this way. So let $r(x)$ denote the number of solutions to.

$$x = na^{-1},\ 1 \leq n \leq N,\ 1 \leq a \leq A$$

In short,

$$|T| \leq \sum_{x \in \mathbb{F}_p} r(x) \left| \sum_{1 \leq b \leq B} \left(\frac{x+b}{p} \right) \right|.$$

Now we apply Hölder's inequality in the form

$$\sum_x a_x b_x c_x \leq \left(\sum_x |a_x|^{q_1} \right)^{1/q_1} \left(\sum_x |b_x|^{q_2} \right)^{1/q_2} \left(\sum_x |c_x|^{q_3} \right)^{1/q_3}$$

which holds as long as $q_1^{-1} + q_2^{-1} + q_3^{-1} = 1$. In this case we take

$$a_x = r(x)^{(k-1)/k}, b_x = r(x)^{1/k}, c_x = \sum_{1 \leq b \leq B} \left(\frac{x+b}{p} \right)$$

and

$$q_1 = \frac{k}{k-1}, q_2 = 2k, q_3 = 2k$$

which gives

$$T \leq P_1^{1-1/k} P_2^{1/2k} P_3^{1/2k}$$

where

$$P_1 = \sum_x r(x),\ P_2 = \sum_x r(x)^2,\ P_3 = \sum_x \left(\sum_b \left(\frac{x+b}{p} \right) \right)^{2k}.$$

By double counting, $P_1 = AN$ since each a and each n contribute 1 to exactly one $r(x)$, namely $x = na^{-1}$. Now, each summand in P_2 counts a pair (a_1, n_1) and (a_2, n_2) with $n_1 a_1^{-1} = n_2 a_2^{-1} = x$. Summing over x eliminates the variable x and we get

$$\begin{aligned} P_2 &= |\{(a_1, a_2, n_1, n_2) : n_1 a_1^{-1} = n_2 a_2^{-1} \bmod p\}| \\ &= |\{(a_1, a_2, n_1, n_2) : a_2 n_1 = a_1 n_2 \bmod p\}| \end{aligned}$$

where $1 \leq a_i \leq A$ and $1 \leq n_i \leq N$. But by the same reasoning,

$$P_2 = \sum_x s(x)^2$$

where $s(x)$ is the number of representations of x as an modulo p. If $an = x_1 \equiv x \bmod p$, then there are only $AN/p + 1$ choices for x_1

(namely those congruent to x mod p and bounded by AN). This means that the congruence condition is at most $AN/p+1$ times the maximum number of solutions to $x_1 = an$, which is bounded by the divisor function $d(x_1)$. Thus $s(x) \leq (ANp^{-1}+1)d(x_1) \leq (ANp^{-1}+1)p^\delta$ for any $\delta > 0$ we like. This shows that

$$P_2 \leq p^\delta \sum_x s(x) = p^\delta (A^2N^2p^{-1} + AN),$$

again by double counting. Finally,

$$P_3 = \sum_{b_1,\ldots,b_{2k}} \sum_x \left(\frac{(x+b_1)\cdots(x+b_{2k})}{p}\right) = \sum_{b_1,\ldots,b_{2k}} \sum_x \left(\frac{f_{b_1,\ldots,b_{2k}}(x)}{p}\right)$$

where $f_{b_1,\ldots,b_{2k}}$ is a polynomial in x of degree $2k$ and is only a square if the numbers b_i can be arranged into pairs of equal values. If this does happen, then $f_{b_1,\ldots,b_{2k}}$ is a perfect square and the inner sum is about p. But this only happens in at most $k!\binom{2k}{k}B^k$ ways, which is at most $(2kB)^k$. So for these terms we get a bound of at most $p(2kB)^k$. If $f_{b_1,\ldots,b_{2k}}$ is not a square then then by our new-found knowledge of hyperelliptic curves, $f_{b_1,\ldots,b_{2k}}(x)$ is a quadratic residue about half of the time, and a quadratic non-residue about half of the time. Specifically

$$\left| \sum_{b_1,\ldots,b_{2k}} \sum_x \left(\frac{f_{b_1,\ldots,b_{2k}}(x)}{p}\right) \right| \leq 16k\sqrt{p}$$

which means that we get at most $16kB^{2k}\sqrt{p}$ for the other terms and

$$P_3 \leq (2kB)^k p + 16kB^{2k}\sqrt{p}.$$

Now choose $B \approx kp^{1/2k}$ and $A \approx N/(kp^{1/2k})$ and we have shown that

$$|S| \ll_\delta N^{1-1/k} p^{\frac{k+1}{4k^2}+\delta}.$$

This is smaller than N if $N > p^\theta$ with $\theta > 1/4$.

References

[IK] H. Iwaniec and E. Kowalski, *Analytic Number Theory.* AMS Colloquium Publications, 2004.

Pennsylvania State University, University Park, PA
E-mail address: bwh5339@psu.edu