Relatively prime values of polynomials

Avi Zeff

This is an exposition of some research that I did the summer after my freshman year of college [2], through the UROP+ program at MIT under the guidance of Professor Bjorn Poonen and the direct supervision of Soohyun Park; the problem in question is due to them, together with many helpful discussions.

Since then, I have learned that in fact this problem has been well-studied, by Poonen among others, and that the essence of the result, as well as various generalizations, is well-known (though the details of various results differ); for example, (†) is the Ekedahl-Poonen formula. I haven't seen the explicit lower bound written down elsewhere, but it may well exist in the literature and certainly would not be difficult to conclude from the well-known versions. The advantage here is the relatively elementary approach avoiding powerful algebra-geometric tools.

As such, my intention is that this exposition be readable without essentially any mathematical background beyond familiarity with exponents and logarithms, and the occasional limit; I'll refer to calculus occasionally, but the reader can gloss over those mentions if desired.

More significantly, I found while doing this exposition that **my original paper actually has significant errors and the main result is not correct as written**, or at least that paper does not prove it. (Bolded to ensure notice.) I haven't done anything about that paper yet since it isn't published anywhere beyond the UROP+ website and no one is likely to read the corresponding section, but possibly I should. Hopefully the significant errors at least are corrected here, but very possibly more remain; please let me know if you find others.

§1. Introduction

The driving question is about *coprimality*. Given two (let's say positive) integers, we can ask whether they are *coprime* (also called *relatively prime*): does there exist an integer greater than 1 dividing both of them? For example, 12 and 21 are *not* relatively prime, since both are divisible by 3; on the other hand 12 *is* relatively prime to 25. A related concept is that of the *greatest common divisor*, written $\gcd(x, y)$, which is what it sounds like: the largest integer which divides both x and y. For our examples above, $\gcd(12, 21) = 3$, while $\gcd(12, 25) = 1$. In general two integers are relatively prime if and only if their greatest common divisor is 1.[1]

Next, we can think about polynomials, i.e. combinations of numbers and some number of variables through addition, subtraction, or multiplication (but *not*, in general, division). For example, $x^2 + 3y$, $2x(24 - 9y^2 + z)^3$, or x^7 are all polynomials, in different numbers of variables. In the case we'll be interested in, the allowed numbers are all integers; the set of integers is written as $\mathbb{Z}$, and so we write the set of polynomials in some number of variables, say $x_1, x_2, \ldots, x_n$, as $\mathbb{Z}[x_1, x_2, \ldots, x_n]$, to denote that this is the set of things you can make by

[1] If we want to include negative numbers, which we often will, we can generalize this definition easily by saying that x and y are relatively prime if $|x|$ and $|y|$ are, so that for example -12 and 25 are also relatively prime. Similarly we can set $\gcd(x, y) = \gcd(|x|, |y|)$.

combining elements of $\mathbb{Z}$ with these variables (and multiplying, adding, and subtracting). A useful notion is that of the *degree* of a polynomial, which is the highest number of variables multiplied together in any one term, counting multiplicity: for example x^3 has degree 3, since there are three copies of x; $x + 2y^2 + yz^3$ has degree 4, since the terms have degrees 1, 2, and 4 respectively (since there is one copy of y and three copies of z in the final term).

With these notions in hand, we can ask our question of interest: given two polynomials $f(x_1, x_2, \ldots, x_n)$ and $g(x_1, x_2, \ldots, x_n)$, if we pick a "random"[2] point $x = (x_1, x_2, \ldots, x_n)$ where all the x_i are integers, how likely is it that $f(x)$ and $g(x)$ will be relatively prime?[3]

Of course, the answer to this can only be "It depends on the polynomials." After all, choosing $f(x_1, \ldots, x_n) = x_1$ and $g(x_1, \ldots, x_n) = x_1 + 1$, the values will always be relatively prime: if d is a positive integer dividing both x_1 and $x_1 + 1$, then we can write $x_1 = ad$ and $x_1 + 1 = bd$ for some integers a and b, and so $ad + 1 = x_1 + 1 = bd$, so $bd - ad = (b - a)d = 1$ and so we see that d divides 1, i.e. $\frac{1}{d}$ is an integer. Since d is a positive integer, it follows that $d = 1$, and so $\gcd(f(x), g(x)) = 1$ for every $x = (x_1, \ldots, x_n)$.

On the other hand, consider the case where f and g are the same polynomial, say $f(x_1, \ldots, x_n) = g(x_1, \ldots, x_n) = x_1$. Then their values are essentially never relatively prime: $\gcd(f(x), g(x)) = \gcd(x_1, x_1) = x_1$, and so the values are relatively prime only when $x_1 = \pm 1$. Since x_1 can take on infinitely many values, each equally likely, the "probability" of these values being coprime is therefore 0.

Okay, so we've learned that this probability can be all the way up to 1 or all the way down to 0. But it seems like in the latter case that was kind of a degenerate example: can we rule it out by just requiring that f and g be different?

Well, no, not quite. Consider the case $f(x_1, \ldots, x_n) = x_1$ and $g(x_1, \ldots, x_n) = x_1^2$. Then x_1 always divides both of them, and so $\gcd(f(x), g(x)) = x_1$ again and so we have probability 0 again.

At this point we're starting to pin down what the problem is, though: in each case there's some nontrivial polynomial[4] which divides both f and g. Indeed, if this kind of thing ever holds we can work out in the same way that the probability will also be 0. Here's a better question, then: suppose that we require that our polynomials f and g are *relatively prime*, which we define roughly in the same way as for integers: no (nontrivial) polynomial divides both of them. Is this enough to ensure that the probability that their values are relatively prime is nonzero?

Alas, the answer is still: not quite. Consider the example $f(x_1, \ldots, x_n) = x_1^2 + x_1$, $g(x_1, \ldots, x_n) = 2$. These are relatively prime *as polynomials*: $g = 2$ is not divisible by anything other than 1 and itself (since 2 is prime), and so the only possibility for coprimality to fail is that f is divisible by 2; but $\frac{1}{2}f = \frac{1}{2}x_1^2 + \frac{1}{2}x_1$, which does not have integer coefficients, so they are genuinely relatively prime.

However, let's look at the values of $f(x)$; since they depend only on x_1, we'll forget the others. At $x_1 = 0$, we have $f(x) = 0^2 + 0 = 0$; at $x_1 = 1$, $f(x) = 1^2 + 1 = 2$; at $x_2 = 2$,

[2]The reason for the scare quotes is that picking a "random" point out of infinitely many options doesn't really make sense, at least in this context; we'll give a better definition of what we really mean by this in a bit.

[3]Here x denotes a tuple of integers, and so we write $f(x)$ and $g(x)$ where we really mean $f(x_1, \ldots, x_n)$ and $g(x_1, \ldots, x_n)$ to save time.

[4]i.e. different from the constant polynomials ± 1, which can always be said to divide every polynomial.

$f(x) = 2^2 + 2 = 6$; and so forth. Every value of $f(x)$ is even! You can see this abstractly by observing that x_1^2 and x_1 have the same parity, so $f(x)$ is always the sum of two things which are either both even or both odd and therefore the sum is even. Since $f(x)$ is always divisible by 2, $\gcd(f(x), g(x)) = 2$ for every x and so the probability of getting coprime values is again 0.

This, however, is a kind of special situation. It can occur for other numbers besides 2 as well—for example, $x_1^3 - x_1$ is always divisible by 3—but in particular it can't occur for any nonconstant polynomials. This leaves us with the following (somewhat vague) conjecture:

Conjecture. *Let $f, g \in \mathbb{Z}[x_1, \ldots, x_n]$ be relatively prime polynomials which satisfy the following property: there exists no integer $m > 1$ such that* every *value of both f and g is divisible by m. Then the probability that $f(x)$ and $g(x)$ are coprime, for x chosen "randomly" among tuples of integers, is nonzero.*

We don't yet really have a good reason to believe this conjecture, other than that we haven't thought of a way to make counterexamples yet. Feel free to spend some time trying and convince yourself that it seems likely to be true. If you believe this conjecture, the next question that might occur to you is: how low can this probability be? Can it be arbitrarily close to 0? Given a pair of polynomials, is there a way to immediately give a bound on how low the probability of coprime values can be? If so, what properties of the polynomials does it depend on?

What we will prove is that not only does our conjecture hold, but there is an explicit lower bound on the probability, which depends only on the degrees of f and g. In order to formally state the result we'll need a substantial amount of notation, so I'll defer the presentation of the result until §4.

In §2, we'll transform this question about relatively prime values into a more analytic question about a certain infinite series involving number-theoretic and algebraic quantities. In §3 and §4 we'll do some number theory to reduce this to a bound in algebraic geometry modulo p; and finally in §5 we'll apply the theory of resultants to prove this result.

The case of one-variable polynomials is actually somewhat different, so in general we'll assume that $n \geq 2$, especially in §4 and §5; we'll address the case $n = 1$ separately in §6.

§2. Analytic reformulation

The first thing to do is replace the idea of choosing a point $x = (x_1, \ldots, x_n)$ "randomly," which does not actually make sense. We'll do this as follows: the problem is that there are infinitely many possibilities, so we'll just restrict the selection so that there aren't. Pick a large integer N, and think about the n-dimensional box consisting of tuples of integers $(x_1, \ldots, x_n)$ with $|x_i| \leq N$ for every x_i. We'll call this box $B(N)$. It's $2N + 1$ points long along each axis (N in each direction, positive and negative, plus the point at 0) and so the total number of points in $B(N)$ is $(2N + 1)^n$. Therefore the probability of a point $x \in B(N)$[5] satisfying $\gcd(f(x), g(x)) = 1$ is the number of such points, written $\#\{x \in B(N) : \gcd(f(x), g(x)) = 1\}$, divided by the total number of points, which we've just computed. We

[5]The symbol $\in$ is read "in," so this means "x in $B(N)$."

write this probability as

$$\delta(f, g; N) = \frac{\#\{x \in B(N) : \gcd(f(x), g(x)) = 1\}}{(2N+1)^n}.$$

(The symbol is a delta for "density," reflecting the fact that really we're replacing this notion of a probability, which doesn't behave well on infinite sets, with the notion of a density of such points, which does.)

Now, ultimately we don't want our density to depend on a choice of N, and we want to be able to choose any x, not just one in our particular box $B(N)$. Therefore we take the limit and define

$$\delta(f, g) = \lim_{N \to \infty} \delta(f, g; N),$$

i.e. the point that the density approaches as $N \to \infty$.[6] In practice most of the time we'll fix some N and work in the finite setting, and just take the limit at the end.

Now we know what it is that we want to bound; we still don't know how to bound it. One thing that we can do is to rewrite it using summation notation, in the hopes that this'll make it easier to work with. A useful piece of notation here is Iverson bracket notation: we have some statement, which can be true or not (for us this'll be $\gcd(f(x), g(x)) = 1$) and we write it in brackets to denote a quantity which is 1 if this is true and 0 if it's false. Thus

$$\#\{x \in B(N) : \gcd(f(x), g(x)) = 1\} = \sum_{x \in B(N)} [\gcd(f(x), g(x)) = 1],$$

where $\sum_{x \in B(N)}$ means that we sum over all x in $B(N)$.

How is this supposed to help? Well, it turns out that there is a key identity which is useful for checking whether a given positive integer is equal to 1. To state it we need to introduce the Möbius function $\mu(x)$.

For each positive integer x, we can factor it as a product of primes $x = p_1 \cdot p_2 \cdots p_r$ for some integer r. For some numbers x some of these factors may be repeated, e.g. $12 = 2 \cdot 2 \cdot 3$; for all of these numbers we define $\mu(x) = 0$. The remaining numbers with no repeated prime factors, e.g. $15 = 3 \cdot 5$, are called squarefree; for these numbers we define $\mu(x) = (-1)^r$, i.e. $\mu(x)$ is 1 if x has an even number of prime factors and -1 if it has an odd number. For example, 1 has zero prime factors, so $\mu(1) = 1$; any prime number p has exactly one prime factor, so $\mu(p) = -1$. Observe that if x and y are relatively prime and squarefree, then they have no prime factors in common and so the total number of prime factors of xy is the sum of the number of prime factors of each of x and y, so that $\mu(xy) = \mu(x)\mu(y)$. (If either of x and y is not squarefree, then neither is their product and so both sides are 0, so the equation holds in any case so long as x and y are relatively prime.) This property is called being multiplicative.

The key identity here is that the quantity

$$\sum_{d|x} \mu(d)$$

[6] The reader might wonder how we know that this limit converges at all; the answer is that at least off the top of my head I don't think we do, but since we're only going to worry about lower bounding it it doesn't matter; the reader can replace the limit above by a limit inferior if it bothers them.

is always 0 unless $x = 1$, in which case it is 1, where the notation $\sum_{d|x}$ means that we sum over all divisors of x (including x itself). For example, we can quickly verify the case where $x = 1$, since the only divisor of 1 is 1 and $\mu(1) = 1$; another example is $x = 10$, where the divisors are 1, 2, 5, and 10, and we have $\mu(1) + \mu(2) + \mu(5) + \mu(10) = 1 - 1 - 1 + 1 = 0$ as predicted.

Now, why is this identity true? Let's call this key quantity $F(x)$, i.e.

$$F(x) = \sum_{d|x} \mu(d).$$

Suppose that x and y are relatively prime. Then

$$F(xy) = \sum_{d|xy} \mu(d) = \sum_{d_1|x} \sum_{d_2|y} \mu(d_1 d_2)$$

since every divisor d of xy can be factored into the portion d_1 dividing x and the portion d_2 dividing y. Since x and y are coprime, so are d_1 and d_2, so $\mu(d_1 d_2) = \mu(d_1)\mu(d_2)$ by the above, and so

$$F(xy) = \sum_{d_1|x} \sum_{d_2|y} \mu(d_1)\mu(d_2) = \sum_{d_1|x} \mu(d_1) \sum_{d_2|y} \mu(d_2) = F(x)F(y)$$

by rearranging the order of summation, so F is also multiplicative.

Since we can factor every integer as a product of prime powers and powers of different primes are always coprime, it would be enough to prove that $F(p^k) = 0$ for every prime p and positive integer k; then we would have

$$F(x) = F(p_1^{k_1} \cdot p_2^{k_2} \cdots p_r^{k_r}) = F(p_1^{k_1}) \cdot F(p_2^{k_2}) \cdots F(p_r^{k_r}) = 0 \cdot 0 \cdots 0 = 0$$

for every $x > 1$.

But the divisors of p^k are pretty simple: they're 1, p, p^2, ..., p^k; for example the divisors of 16 are 1, 2, 4, 8, and 16. Therefore

$$F(p^k) = \mu(1) + \mu(p) + \mu(p^2) + \cdots + \mu(p^k).$$

But for $j > 1$, p^j has repeated prime factors and so $\mu(p^j) = 0$, so

$$F(p^k) = \mu(1) + \mu(p) = 1 - 1 = 0,$$

which is what we wanted to show.

So, why is this helpful? Well, we have this term in our summation $[\gcd(f(x), g(x)) = 1]$, which is 0 if $\gcd(f(x), g(x)) = 1$ and 0 otherwise. Using our key identity, we can rewrite this as

$$[\gcd(f(x), g(x)) = 1] = \sum_{d|\gcd(f(x),g(x))} \mu(d).$$

Why is *this* helpful? Because the divisors of the greatest common divisor of $f(x)$ and $g(x)$ can be phrased in another way: they're exactly the numbers that divide both $f(x)$ and $g(x)$,

and so we can rewrite this as a sum over all positive integers d which divide both $f(x)$ and $g(x)$.

Thus we have all in all

$$\delta(f,g;N) = \frac{1}{(2N+1)^n} \sum_{x \in B(N)} \sum_{\substack{d|f(x) \\ d|g(x)}} \mu(d).$$

We can think of this equivalently as summing over pairs (x,d) where $x \in B(N)$ and d is a positive integer such that d divides both $f(x)$ and $g(x)$; rearranging, we can write this as

$$\delta(f,g;N) = \frac{1}{(2N+1)^n} \sum_{d=1}^{\infty} \mu(d) \sum_{x \in B(N)} [d|f(x), d|g(x)],$$

using Iverson bracket notation again.

To proceed further, we need to know something about modular arithmetic. This is the idea of looking at numbers modulo some fixed number, say m; that is, we only care about numbers up to multiples of m. Specifically, we say that two numbers x and y are *congruent modulo* m if $x - y$ is a multiple of m; this relation is written $x \equiv y \pmod{m}$. For example, $15 \equiv 7 \pmod{4}$, since $15 - 7 = 8$ is divisible by 4. Note that $x \equiv 0 \pmod{m}$ is equivalent to the statement $m|x$. For each m, we can partition the integers into equivalence classes modulo m; there are m of these, and we can think of them as the set of integers congruent to 0 modulo m, those congruent to 1 modulo m, etc., all the way up to those congruent to $m-1$ modulo m.

Using this notation, we can rewrite the innermost summation above as

$$\sum_{x \in B(N)} [f(x) \equiv g(x) \equiv 0 \pmod{d}],$$

i.e. the number of $x \in B(N)$ for which $f(x)$ and $g(x)$ are 0 modulo d.

The key property of modular arithmetic, for us, is that it behaves well with respect to addition and multiplication: in other words, if $x \equiv y \pmod{m}$ and $a \equiv b \pmod{m}$, then $x + a \equiv y + b \pmod{m}$ and $ax \equiv by \pmod{m}$.[7]

Since polynomials are, by definition, things built out of multiplication and addition (and subtraction, which is just addition with negative numbers), this means that they also behave well with respect to modular arithmetic. In particular, the value of $f(x) = f(x_1, \ldots, x_n)$ modulo m depends only on the values of the x_i modulo m: if $x_i \equiv y_i \pmod{m}$ for every i, then $f(x_1, \ldots, x_n) \equiv f(y_1, \ldots, y_n) \pmod{m}$. Therefore if we imagine that N is much bigger than d, rather than checking whether $f(x) \equiv g(x) \equiv 0 \pmod{d}$ for every $x \in B(N)$ it's much easier to just check for some representatives of the equivalence classes, of which there are d^n (since there are d in each variable) and then use this property of polynomials and modular arithmetic.

Explicitly, this works like this. Each axis of $B(N)$ has length $2N+1$, while each axis of the box modulo d—which we can think of as the set of tuples of integers $(x_1, \ldots, x_n)$ such

[7] These properties are not hard to check from the definition, and doing so is a good way to get more familiar with modular arithmetic.

that $0 \leq x_i \leq d-1$ for each x_i, and for which we'll write B_d—has length d. The ratio is $\frac{2N+1}{d}$, and since each box is n-dimensional it follows that roughly

$$\left(\frac{2N+1}{d}\right)^n = \frac{(2N+1)^n}{d^n}$$

boxes modulo d fit into $B(N)$. Therefore if we can count how many $x \in B_d$ satisfy $f(x) \equiv g(x) \equiv 0 \pmod{d}$ and call this quantity $\nu(d)$, then the total number of $x \in B(N)$ satisfying the same property should be approximately

$$\frac{(2N+1)^n}{d^n}\nu(d).$$

The word "approximately" above is a concerning one, but actually we can be precise, because we don't need to compute $\delta(f,g;N)$ for every N, just for enough N going to infinity. There are infinitely many N such that $2N+1$ is divisible by d if d is odd, in which case this approximation is exact; if d is even, we can choose $2N$ to be divisible by d, so that the estimate of $\frac{2N+1}{d}$ for the number of times d goes into $2N+1$ is only off by $\frac{1}{d}$. Therefore the number of times B_d goes into $B(N)$ differs from the estimate $\left(\frac{2N+1}{d}\right)^n$ by at most

$$\left(\frac{2N+1}{d}\right)^n - \left(\frac{2N}{d}\right)^n.$$

Generally speaking for any large number x we have

$$x^n - (x-1)^n = nx^{n-1} + \text{some lower degree terms},$$

and we'll see that the lower degree terms will be negligible. (In fact, the main term here will be negligible too.) We write $O(x^{n-1})$ to denote that this is something of order at most x^n; for us, taking $x = 2N+1$ and dividing everything by d^n this means that our estimate is off by at most $O\left(\frac{(2N+1)^{n-1}}{d^n}\right)$. Combining this with the above, we conclude that the total number of $x \in B(N)$ satisfying $f(x) \equiv g(x) \equiv 0 \pmod{d}$ is

$$\left(\frac{(2N+1)^n}{d^n} + O\left(\frac{(2N+1)^{n-1}}{d^n}\right)\right)\nu(d) = \frac{(2N+1)^n}{d^n}\nu(d) + O\left(\frac{(2N+1)^{n-1}}{d^n}\nu(d)\right).$$

We'll see in §5 that $\nu(d)$ is at most order d^{n-2} and so we can cancel terms to get

$$\frac{(2N+1)^n}{d^n}\nu(d) + O\left(\frac{(2N+1)^{n-1}}{d^2}\right),$$

and since we divide everything by $(2N+1)^n$ this will turn into

$$\frac{\nu(d)}{d^n} + O\left(\frac{1}{Nd^2}\right).$$

Therefore all in all we have

$$\delta(f,g;N) = \sum_{d=1}^{\infty} \mu(d)\left(\frac{\nu(d)}{d^n} + O\left(\frac{1}{Nd^2}\right)\right).$$

By the theory of sums and series from calculus, $\sum_{d=1}^{\infty} \frac{1}{d^2}$ converges (e.g. by the integral test), and by comparison the same is true with the extra factor of $\mu(d)$, so the bound becomes

$$\delta(f, g; N) = \sum_{d=1}^{\infty} \mu(d) \frac{\nu(d)}{d^n} + O\left(\frac{1}{N}\right).$$

Taking the limit as $N \to \infty$, the error term vanishes and we're left with

$$\delta(f, g) = \sum_{d=1}^{\infty} \mu(d) \frac{\nu(d)}{d^n}. \tag{$*$}$$

The hardest part of this series is the Möbius function μ. In the next section we'll use the theory of Euler products to get this into a different form which doesn't involve μ, and has the added benefit of only needing to evaluate ν at prime numbers.

§3. Euler products

The classic Euler product is in a simpler situation than our case, namely the Riemann zeta function: for a complex number s with real part greater than 1,[8] we define

$$\zeta(s) = \sum_{k=1}^{\infty} \frac{1}{k^s}.$$

The Euler product formula gives another expression for $\zeta(s)$:

$$\zeta(s) = \prod_{p} \frac{1}{1 - p^{-s}},$$

where the notation $\prod_p$ denotes the product over all primes p, i.e.

$$\zeta(s) = \frac{1}{1 - 2^{-s}} \cdot \frac{1}{1 - 3^{-s}} \cdot \frac{1}{1 - 5^{-s}} \cdots.$$

Why does this hold? First, recall the geometric series formula: for any number x with $|x| < 1$, we have

$$\sum_{j=0}^{\infty} x^j = \frac{1}{1 - x}.$$

For $x = p^{-s}$, which is less than 1 since the real part of s is at least 1, we therefore have

$$\frac{1}{1 - p^{-s}} = \sum_{j=0}^{\infty} p^{-sj} = 1 + p^{-s} + p^{-2s} + \cdots.$$

[8] For the reader unfamiliar with complex numbers, you can just think of s as a real number greater than 1, or indeed for our purposes an integer.

Writing out

$$(1+2^{-s}+2^{-2s}+\cdots)(1+3^{-s}+3^{-2s}+\cdots)(1+5^{-s}+5^{-2s}+\cdots)\cdots,$$

we can start multiplying this out like a usual product: first we have the term which takes the $1 = p^{0s}$ term from each factor, so this is $1 \cdot 1 \cdots = 1$. Next, for each prime p we have the term which takes the p^{-s} term from that factor and 1 from every other factor; this gives us a term of p^{-s} for every prime p. We can do the same thing for prime powers p^j, taking the $p^{-js} = (p^j)^{-s}$ term and all other factors 1. Then we can start looking at the terms which for any two primes p and q take the terms p^{-s} and q^{-s}, with all other factors 1; this gives $p^{-s}q^{-s} = (pq)^{-s}$. Continuing in this fashion, we see that we get a term of k^{-s} where k is given by a product of prime powers. By the fundamental theorem of arithmetic, which states that every integer can be uniquely written as the product of prime powers, we see that we get exactly one term k^{-s} for each integer k, since k corresponds to a unique combination of prime powers, and so expanding the infinite product in this way gives exactly

$$\prod_p \frac{1}{1-p^{-s}} = \sum_{k=1}^{\infty} \frac{1}{k^s} = \zeta(s).$$

Now, the Riemann zeta function is not the same as the series $(*)$ we arrived at in §2. However, we can generalize the approach above to get a more general formula, which we will be able to apply to our case.

Proposition (Euler product formula)**.** *Let $F(x)$ be a multiplicative function, i.e. for relatively prime positive integers x and y we have $F(xy) = F(x)F(y)$. Then*

$$\sum_{k=1}^{\infty} \frac{F(k)}{k^s} = \prod_p \sum_{j=0}^{\infty} F(p^j)p^{-js}.$$

By taking $F(x) = 1$, which is multiplicative since $F(xy) = 1 = F(x)F(y)$, we recover the formula above for the Riemann zeta function (after applying the geometric series formula again).

Proof. We can mimic the strategy above, except that we replace factors of p^{-js} by $F(p^j)p^{-js}$. Since F is multiplicative, if $k = p_1^{a_1} \cdots p_r^{a_r}$ then $F(p_1^{a_1}) \cdots F(p_r^{a_r}) = F(k)$ and so we can combine these terms as well as the p^{-js} to get $\frac{F(k)}{k^s}$ as desired. □

In order to apply this proposition, we need to show that $\mu(x)\nu(x)$ is multiplicative. We already know that μ is multiplicative, so it remains to show that ν is multiplicative, i.e. the number of points x in B_{ab} such that for our chosen polynomials f and g we have $f(x) \equiv g(x) \equiv 0 \pmod{ab}$ is equal to the product of the number of such points in B_a modulo a and in B_b modulo b for any relatively prime integers a and b.

This follows from the Chinese remainder theorem, which states that for relatively prime integers a, b and any x, y, there is a unique integer z up to equivalence modulo ab such that $z \equiv x \pmod{a}$ and $z \equiv y \pmod{b}$. For example, take $a = 3$ and $b = 7$, and let's look for a number which is congruent to 2 modulo 3 and congruent to 5 modulo 7. There are

three numbers between 0 and 20 which are congruent to 5 modulo 7, namely 5, 12, and 19; of these, only 5 is congruent to 2 modulo 3, so the unique equivalence class satisfying both properties is the set of numbers congruent to 5 modulo 21. We won't prove this theorem, but it's a good exercise and not hard to find proofs for online.

Given this theorem, our desired claim is not hard: each solution x for $f(x) \equiv g(x) \equiv 0 \pmod{a}$ and y for $f(y) \equiv g(y) \pmod{b}$ yields a unique solution modulo ab, so the total number of solutions modulo ab is the number of pairs of solutions modulo a and modulo b, which is the product of the numbers of solutions modulo a and b. In other words, $\nu(ab) = \nu(a)\nu(b)$.

Having checked this, we can apply our proposition, with $s = n$:

$$\delta(f,g) = \sum_{d=1}^{\infty} \frac{\mu(d)\nu(d)}{d^n} = \prod_p \sum_{j=0}^{\infty} \mu(p^j)\nu(p^j)p^{-jn}.$$

Since $\mu(p^j) = 0$ for $j > 1$, noting that $\nu(1) = 1$ (since $B_1 = \{(0,0,\ldots,0)\}$ and $f(0,\ldots,0) \equiv g(0,\ldots,0) \equiv 0 \pmod{1}$ since there is only one equivalence class modulo 1) and $\mu(p) = -1$ we can rewrite this as

$$\delta(f,g) = \prod_p \left(1 - \frac{\nu(p)}{p^n}\right). \tag{$\dagger$}$$

Now, we want to show that this quantity is nonzero, and ultimately bound it from below. Observe that the smaller $\nu(p)$ is, the larger $\delta(f,g)$ is, so we might expect that proving a lower bound for $\delta(f,g)$ boils down somehow to proving an *upper* bound for $\nu(p)$. This is indeed the case. In the following section we'll state an upper bound on $\nu(p)$, and derive a lower bound on $\delta(f,g)$ from it; after that in the final section we'll prove the lower bound.

§4. Proof of the theorem

Recall the Riemann zeta function

$$\zeta(s) = \sum_{k=1}^{\infty} \frac{1}{k^s} = \prod_p \frac{1}{1-p^{-s}}.$$

As stated above, this converges to a finite value whenever s has real part greater than 1, and in particular converges for integer values greater than 1, i.e. $s = 2, 3, 4, \ldots$. Therefore the inverse of $\zeta(s)$ is nonzero at these values:

$$\frac{1}{\zeta(s)} = \prod_p (1 - p^{-s})$$

is nonzero for $s = 2, 3, 4, \ldots$.

Comparing to $\delta(f,g)$, we expect to therefore be able to say that $\delta(f,g)$ converges to a nonzero value if each factor $1 - \nu(p)p^{-n}$ is lower bounded by something of the form $1 - p^{-s}$ for $s > 1$.[9] This is the same thing as *upper* bounding $\nu(p)p^{-n}$ by p^{-s} for some $s > 1$, and the

[9]We don't need to worry about the upper bound for convergence because each factor $1 - \nu(p)p^{-s}$ is automatically bounded above by 1, since $\nu(p)$ and p^{-s} are positive (for real $s > 1$).

statement $\nu(p)p^{-n} \leq p^{-s}$ is equivalent to $\nu(p) \leq p^{n-s}$. We could just hope that we have a bound of this form for some $s > 1$, e.g. $s = 1.01$, but it seems most natural that s should be an integer, or at least it would be nice if that were the case; the easiest guess then is $s = 2$, so we hope to bound $\nu(p)$ by something like p^{n-2}.

Why do we say "something like"? Because like before, we don't necessarily have a literal inequality $\nu(p) \leq p^{n-2}$; we really care about the order of magnitude. We'll get around this by assuming a bound of the following form.

Lemma. *There exist integers A and B, depending on n, f, and g (but* not *on p), such that $\nu(p) \leq Ap^{n-2}$ for every prime number $p > B$.*

The proof of this lemma will be deferred to §5, where we'll also pin down exactly what A and B are.

Assuming this lemma, we can now try to explicitly lower bound $\delta(f, g)$. There are two different cases depending how big p is compared to B. If p is large, then our bound applies and $\nu(p) \leq Ap^{n-2}$ implies that

$$\frac{\nu(p)}{p^n} \leq \frac{A}{p^2}$$

is very small and therefore

$$1 - \frac{\nu(p)}{p^n} \geq 1 - \frac{A}{p^2}$$

is very close to 1, which is good for the purpose of bounding the product from below. (We can assume that $B \geq \sqrt{A}$, so that $\frac{A}{p^2} < \frac{A}{B^2} \leq 1$, since if not we can replace B by a higher bound.) If $p \leq B$, though, we know nothing.

In this case, we're saved by our assumption that there is no number m dividing every value of both f and g. What this means in our terms is that there is no m such that $f(x) \equiv g(x) \equiv 0 \pmod{m}$ for every m, and so it is impossible to have $\nu(m) = p^n$: this would imply that every x in B_m satisfies $f(x) \equiv g(x) \equiv 0 \pmod{m}$, and by the modular invariance for polynomials this would imply this was true for *every* x, a contradiction. Therefore we always have $\nu(p) \leq p^n - 1$, and so

$$1 - \frac{\nu(p)}{p^n} \geq 1 - \frac{p^n - 1}{p^n} = \frac{1}{p^n}.$$

Therefore we can split the primes into two cases: the good case where $p > B$, where we can use the bound $1 - \frac{\nu(p)}{p^n} \geq 1 - \frac{A}{p^2}$; and the bad case where $p \leq B$, where we only have the bound $1 - \frac{\nu(p)}{p^n} \geq \frac{1}{p^n}$. Splitting the product into these two factors, we get

$$\begin{aligned}\delta(f, g) &= \left(\prod_{p \leq B} \left(1 - \frac{\nu(p)}{p^n}\right)\right)\left(\prod_{p > B} \left(1 - \frac{\nu(p)}{p^n}\right)\right) \\ &\geq \left(\prod_{p \leq B} \frac{1}{p^n}\right)\left(\prod_{p > B} \left(1 - \frac{A}{p^2}\right)\right)\end{aligned}$$

applying the respective bound on each term.

We can now analyze these factors separately. First, let's look at the bad term. Products are annoying, so let's convert it to a sum using the properties of exponents and logarithms:

$$\prod_{p \leq B} \frac{1}{p^n} = \exp\left(\sum_{p \leq B} \log\left(\frac{1}{p^n}\right)\right) = \exp\left(-n \sum_{p \leq B} \log p\right).$$

Now, the sum

$$\sum_{p \leq x} \log p$$

for any number x has a name: it's called Chebyshev's first function, and written $\vartheta(x)$. Therefore the first factor can be rewritten as

$$\exp(-n\vartheta(B)).$$

Okay, that wasn't so bad. What about this second factor for large p?

If A and B were equal to 1, then this would be

$$\prod_{p} \left(1 - \frac{1}{p^2}\right)$$

since all primes are already greater than 1. This is exactly the inverse of the product formula for the Riemann zeta function, and so would just give $\frac{1}{\zeta(2)}$, which we know is a finite nonzero number (in fact, it's known to be equal to $\frac{6}{\pi^2} \approx 0.6079$). Therefore we would be able to conclude that this second factor is an explicit nonzero constant, which together with the first part would give us our result.

Unfortunately, A and B aren't (necessarily) equal to 1, so our lives are more complicated; but we can try to use trickery to change the situation into one which looks more like the Riemann zeta case. One way of doing this is by noticing that

$$1 - \frac{A}{p^2} = \left(1 - \frac{1}{p^2 - (A-1)}\right)\left(1 - \frac{A-1}{p^2}\right),$$

which can be verified by explicitly multiplying everything out; I am honestly not sure how I came up with this in the first place though, it seems pretty unmotivated to me (possibly my supervisor suggested it). In any case we can then repeat the trick for $A-1$ to get

$$1 - \frac{A}{p^2} = \left(1 - \frac{1}{p^2 - (A-1)}\right)\left(1 - \frac{1}{p^2 - (A-2)}\right)\left(1 - \frac{A-2}{p^2}\right)$$

and then for $A-2$ and so on until we get down to 1:

$$1 - \frac{A}{p^2} = \left(1 - \frac{1}{p^2 - (A-1)}\right)\left(1 - \frac{1}{p^2 - (A-2)}\right) \cdots \left(1 - \frac{1}{p^2 - 1}\right)\left(1 - \frac{1}{p^2}\right).$$

Therefore for each k between 0 and $A-1$, taking all primes greater than B gives a factor of

$$\prod_{p > B} \left(1 - \frac{1}{p^2 - k}\right).$$

For the $k = 0$ factor, this is closely related to the Riemann zeta function, minus a certain number of factors from the small primes; for the other k we want to relate the corresponding factors to something more familiar.

For $k > 0$, the k-term $\prod_{p>B}\left(1 - \frac{1}{p^2-k}\right)$ is smaller than the $k = 0$ factor $\prod_{p>B}\left(1 - \frac{1}{p^2}\right)$, so directly comparing them isn't helpful. However, one thing we could try is looking for a number c_k for each k such that

$$\prod_{p>B}\left(1 - \frac{1}{p^2-k}\right) \geq \prod_{p>B}\left(1 - \frac{1}{p^2}\right)^{c_k}.$$

To find such a c_k, it would be enough to find a c_k such that

$$1 - \frac{1}{p^2-k} \geq \left(1 - \frac{1}{p^2}\right)^{c_k}$$

for every prime $p > B$, since then the products would have the same relation.

To understand this requirement a little better, let's take logarithms, since that always simplified things with exponents in them. This gives

$$\log\left(1 - \frac{1}{p^2-k}\right) \geq c_k \log\left(1 - \frac{1}{p^2}\right),$$

i.e.

$$c_k \geq \frac{\log\left(1 - \frac{1}{p^2-k}\right)}{\log\left(1 - \frac{1}{p^2}\right)}$$

(since $\log(1 - p^{-2}) < 0$ since $1 - p^{-2} < 1$, so dividing by it reverses the inequality). For $p^2 > k$, the right-hand side is a decreasing function of p, and since we assume $p^2 > B^2$ and both are integers we have $p^2 \geq B^2 + 1$ and so the right-hand side is always at most its value given by replacing p^2 with $B^2 + 1$. Therefore it's enough to assume

$$c_k \geq \frac{\log\left(1 - \frac{1}{B^2+1-k}\right)}{\log\left(1 - \frac{1}{B^2+1}\right)}.$$

For each k, the corresponding factor is

$$\prod_{p>B}\left(1 - \frac{1}{p^2-k}\right) \geq \prod_{p>B}\left(1 - \frac{1}{p^2}\right)^{c_k} = \left(\frac{1}{\zeta(2)}\prod_{p\leq B}\frac{1}{1-p^{-2}}\right)^{c_k}.$$

The $\zeta(2)$ part is relatively easy to deal with; let's focus on the product part. Again, we'll make our lives easier by taking logarithms:

$$\left(\prod_{p\leq B}\frac{1}{1-p^{-2}}\right)^{c_k} = \exp\left(-c_k \sum_{p\leq B}\log(1-p^{-2})\right).$$

A fact from calculus is that for any positive number x, we have

$$\log(1-x) \leq -x,$$

and so

$$\exp\left(-c_k \sum_{p \leq B} \log(1-p^{-2})\right) \geq \exp\left(c_k \sum_{p \leq B} p^{-2}\right).$$

Now,

$$\sum_p \frac{1}{p^s}$$

is a known function converging for complex numbers s with real part at least 1, sometimes called the prime zeta function $P(s)$ due to its similarity to the Riemann zeta function except summing only over the primes. Therefore

$$\sum_{p \leq B} p^{-2} = P(2) - \sum_{p > B} p^{-2}.$$

On the other hand the latter sum is upper bounded by the same sum but over all integers greater than B, which in turn is bounded by corresponding integral

$$\sum_{j > B} j^{-2} \leq \int_B^\infty x^{-2}\, dx = \frac{1}{B}$$

by methods from calculus. Therefore

$$\exp\left(c_k \sum_{p \leq B} p^{-2}\right) \geq \exp\left(c_k \left(P(2) - \frac{1}{B}\right)\right).$$

Putting everything together, we have

$$\begin{aligned}\delta(f,g) &\geq \exp(-n\vartheta(B)) \prod_{k=0}^{A-1} \exp\left(c_k \left(P(2) - \frac{1}{B}\right)\right) \zeta(2)^{-c_k} \\ &= \exp\left(-n\vartheta(B) + \sum_{k=0}^{A-1} c_k \left(P(2) - \log\zeta(2) - \frac{1}{B}\right)\right).\end{aligned}$$

The last thing left to do is to make the c_k terms explicit. Notice that the terms of the sum in the parentheses do not depend on k, so the only thing we need to bound is

$$\sum_{k=0}^{A-1} c_k.$$

When we introduced the c_k, we found the bounds

$$c_k \geq \frac{\log\left(1 - \frac{1}{B^2+1-k}\right)}{\log\left(1 - \frac{1}{B^2+1}\right)},$$

and summing gives

$$\sum_{k=0}^{A-1} c_k \geq \sum_{k=0}^{A-1} \frac{\log\left(1-\frac{1}{B^2+1-k}\right)}{\log\left(1-\frac{1}{B^2+1}\right)}.$$

Since the denominator does not depend on k, we can pull it out and apply the rule $\log(1-x) \leq -x$ to get

$$\sum_{k=0}^{A-1} \frac{\log\left(1-\frac{1}{B^2+1-k}\right)}{\log\left(1-\frac{1}{B^2+1}\right)} \geq -(B^2+1)\sum_{k=0}^{A-1} \log\left(1-\frac{1}{B^2+1-k}\right).$$

Rewriting $1-\frac{1}{B^2+1-k} = \frac{B^2+1-1-k}{B^2+1-k}$, by the properties of logarithms this is

$$\begin{aligned}
&-(B^2+1)\sum_{k=0}^{A-1}\big(\log(B^2+1-1-k)-\log(B^2+1-k)\big)\\
=&-(B^2+1)\big((\log(B^2+1-1)-\log(B^2+1))\\
&\quad+(\log(B^2+1-2)-\log(B^2+1-1))\\
&\quad\quad+(\log(B^2+1-3)-\log(B^2+1-2))\\
&\quad\quad\quad+\cdots+(\log(B^2+1-A)-\log(B^2+1-A+1))\big).
\end{aligned}$$

Canceling terms leaves only the first and last terms

$$(B^2+1)(\log(B^2+1)-\log(B^2+1-A)) = (B^2+1)\log\left(1+\frac{A}{B^2+1-A}\right).$$

Therefore in all we've proven the following:

Theorem. *Suppose that f and g are polynomials in n variables over the integers satisfying the hypotheses of the conjecture above, and that the above lemma holds for an integer A, depending on n, f, and g. Then*

$$\delta(f,g) \geq \exp\left(-n\vartheta(B)+\left(P(2)-\log\zeta(2)-\frac{1}{B}\right)(B^2+1)\log\left(1+\frac{A}{B^2+1-A}\right)\right).$$

Since the exponential function $\exp(x) = e^x$ always has positive values, this shows that $\delta(f,g)$ is never zero. Notice that since $P(2) \approx 0.4522$ is less than $\log\zeta(2) \approx 0.4977$, the argument of the exponential function is always negative and so there is no risk of predicting a density greater than 1.

Thus up to the lemma we have proven our conjecture, plus a pleasantly explicit bound. To make it even more explicit, we could remove the dependence on Chebyshev's first function ϑ: the prime number theorem, which states that there are roughly $\frac{x}{\log x}$ prime numbers less than or equal to a given positive number x, can be phrased in terms of this function, for which it says that $\vartheta(x)$ is roughly x. To get a concrete bound, Theorem 2.4 of [1][10] shows that

$$\vartheta(x) \leq x + \frac{0.15}{(\log x)^3},$$

[10]The title of this paper has changed slightly since I first worked on this, but this is the same as the first citation in [2].

and so using this together with our theorem we could get the more explicit (but less precise) bound

$$\delta(f,g) \geq \exp\left(-nB - \frac{0.15n}{(\log B)^3} + \left(P(2) - \log\zeta(2) - \frac{1}{B}\right)(B^2+1)\log\left(1 + \frac{A}{B^2+1-A}\right)\right).$$

Of course, we still have no idea what A and B are in terms of f and g; we haven't even proven that they exist yet (and satisfy the properties required of them in the lemma). Thus it's hard to say that our theorem is really explicit. This is the last thing to do, then: prove our lemma for some A and B. (Unfortunately our value for B will not be very explicit, but at least we can define it properly and show that it exists.)

§5. Proof of the bound

For this section we'll need to change our notation slightly: previously we had fixed our polynomials f and g while varying the prime p, and so we used the notation $\nu(p)$, which depends on p explicitly and on f and g only implicitly. Now we're going to fix a prime p and use slightly varying polynomials f and g, and so we write $\nu(f,g)$ to denote the same thing, i.e. the number of common zeros of f and g modulo p.

The idea of the proof is by induction on n. In other words, first we'll prove the lemma for a base case (for us, $n = 2$), and then we show that assuming the lemma holds for n, it also holds for $n+1$. Therefore by applying this second part to the base case, we show that the lemma also holds for $n = 3$; applying again, it holds for $n = 4$; and so on, so that it holds for all n.

In order for the induction step to work, we need a way to go between polynomials in different number of variables. For us, this will work like this: given an integer a (which we really only care about modulo p) and a polynomial f in $n+1$ variables, we define a polynomial f_a in n variables by $f_a(x_1,\ldots,x_n) = f(x_1,\ldots,x_n,a)$. For example if f is a polynomial in x,y,z defined by $f(x,y,z) = 2x - yz^2$ and $a = 5$, then $f_a(x,y) = 2x - 25y$. For notational convenience, we'll use $\mathbb{F}_p$ to denote a fixed set of representatives for the equivalence classes modulo p; for example, we could take $\mathbb{F}_p = \{0,1,2,\ldots,p-2,p-1\}$.

One general lemma which will be useful for us is this:

Lemma. *Let f be a nonzero polynomial in n variables. The number of solutions to*

$$f(x_1,\ldots,x_n) \equiv 0 \pmod{p}$$

for any prime p is at most $\deg(f)p^{n-1}$.

Proof. This follows from the fundamental theorem of algebra, which says that any polynomial g in one variable has at most $\deg g$ zeros (over the real or complex numbers, not just modulo p). If we fix $x_1,\ldots,x_{n-1}$ and set $g(x) = f(x_1,\ldots,x_{n-1},x)$ as a polynomial in one variable, this has at most $\deg g \leq \deg f$ solutions for x; now taking all possible combinations of $x_1,\ldots,x_{n-1}$ modulo p, of which there are p^{n-1}, gives the result. □

This is sometimes called the Schwartz-Zippel lemma.

Now, we start off with our two polynomials f and g, which satisfy the properties of the conjecture; in particular they're relatively prime, which implies for one thing that they're both nonzero (since 0 is divisible by everything). There's no guarantee that any of this will hold after passing to f_a and g_a. In particular there are three possibilities: either f_a and g_a will still be relatively prime, they'll no longer be relatively prime but are still both nonzero, exactly one is zero, or both are zero. Since everything is now modulo p, we actually want to modify our notion of coprimality a little: we say that two polynomials f and g are relatively prime modulo p if there is no nontrivial polynomial h such that h divides both f and g *modulo* p, i.e. there do not exist polynomials j_1, j_2, k_1, k_2 such that $f = hj_1 + pk_1$ and $g = hj_2 + pk_2$. We no longer care about constants much, so "nontrivial" here just means that h is not constant.

Let's split $\mathbb{F}_p$ into four sets, S_1, S_2, S_3, and S_4, such that the first case (f_a and g_a are relatively prime modulo p) happens for $a \in S_1$, the second (they are not coprime modulo p but are both nonzero modulo p, i.e. neither is divisible by p) for $a \in S_2$, the third (exactly one of f_a and g_0 is zero modulo p) for $a \in S_3$, and the fourth ($f_a \equiv g_a \equiv 0 \pmod{p}$) for $a \in S_4$. We have

$$\nu(f, g) = \sum_{a \in \mathbb{F}_p} \nu(f_a, g_a) = \sum_{a \in S_1} \nu(f_a, g_a) + \sum_{a \in S_2} \nu(f_a, g_a) + \sum_{a \in S_3} \nu(f_a, g_a) + \sum_{a \in S_4} \nu(f_a, g_a).$$

First, we'll look at the base case, $n = 2$. Therefore each of f_a and g_a is a one-variable polynomial. Now, one-variable polynomials modulo p satisfy a useful property: if $f_a(b) \equiv 0 \pmod{p}$, then f_a is divisible by $x - b$ as a polynomial modulo p; in other words, there exists some other polynomial h such that $f_a(x) \equiv (x - b)h(x) \pmod{p}$ as polynomials. (This can be seen by long division of polynomials, for example.)

Therefore for any $b \in \mathbb{F}_p$, if $f_a(b) \equiv g_a(b) \equiv 0 \pmod{p}$ then both f_a and g_a are divisible by $x - b$ modulo p, and therefore are not coprime modulo p. Therefore for $a \in S_1$, there exist no such b; in other words, $\nu(f_a, g_a) = 0$.

This takes care of S_1; let's skip S_2 for now and move on. For $a \in S_3$, if (say) $f_a \equiv 0 \pmod{p}$ then by definition this means that $f(x_1, a) \equiv 0 \pmod{p}$ for every $x_1 \in \mathbb{F}_p$. By a similar argument, this implies that $f(x_1, x_2)$ must be divisible (modulo p) by $x_2 - a$ as a polynomial. Doing the same thing for g_a, we see that for every $a \in S_3$ we must have a factor of $x_2 - a$ dividing either f or g or both (modulo p). Each factor of $x_2 - a$ contributes 1 to the degree of either f or g. To count $\nu(f_a, g_a)$ for $a \in S_3$, note that if say $f_a \equiv 0 \pmod{p}$ then every $b \in \mathbb{F}_p$ satisfies $f_a(b) \equiv 0 \pmod{p}$, so $\nu(f_a, g_a)$ is just counting the number of solutions to $g_a \equiv 0 \pmod{p}$, which is at most the degree of g_a by our lemma. Therefore $a \in S_3$ such that $x_2 - a$ divides f contribute a term of at most $\deg g$ and there are at most $\deg f$ of them, since each contributes a factor of 1 to the degree of f; the same holds for g, and so the total contribution from S_3 is at most $2 \deg f \deg g$.

For $a \in S_4$, computing $\nu(f_a, g_a)$ is easy: both are zero modulo p and so every $b \in \mathbb{F}_p$ satisfies $f_a(b) \equiv g_a(b) \equiv 0 \pmod{p}$, so $\nu(f_a, g_a) = p$. This is too big!

This is a real problem, but fortunately it's one we can salvage: for any f and g, there will be only finitely many primes p such that S_4 is nonempty. We'll come back to this once we understand the theory of resultants.

Now, let's think about the case $a \in S_2$, i.e. f_a and g_a are both nonzero but are not coprime modulo p. This sort of thing is detected by the *resultant*: the resultant of two

(one-variable) polynomials f and g (over the integers) is a number $\mathrm{Res}(f, g)$ which is zero if and only if f and g have a common zero, i.e. there exists some x such that $f(x) = g(x) = 0$. Equivalently, $\mathrm{Res}(f, g) = 0$ if and only if f and g are *not* coprime.

The resultant is itself a polynomial in the coefficients of f and g, of degree $\deg f \deg g$, and therefore behaves well with respect to reduction modulo p. In particular, f and g have a common zero modulo p, i.e. there exists $x \in \mathbb{F}_p$ such that $f(x) \equiv g(x) \equiv 0 \pmod{p}$, or equivalently f and g are *not* coprime modulo p, if and only if $\mathrm{Res}(f, g) \equiv 0 \pmod{p}$, i.e. if and only if $\mathrm{Res}(f, g)$ is divisible by p. One immediate consequence is that for one-variable f and g relatively prime (so that the resultant is nonzero), since $\mathrm{Res}(f, g)$ is divisible only by finitely many primes since it is a nonzero finite number there are only finitely many primes p such that f and g have a common zero modulo p!

How does this help us? Well, by assumption f_a and g_a are *not* coprime modulo p for $a \in S_2$, so $\mathrm{Res}(f_a, g_a) \equiv 0 \pmod{p}$. Since $\mathrm{Res}(f_a, g_a)$ is a polynomial of degree $\deg f_a \deg g_a \leq \deg f \deg g$ in the coefficients of f_a and g_a, which include powers of a (at most $a^{\max(\deg f, \deg g)}$), it is a polynomial in a of degree at most $\deg f \deg g \max(\deg f, \deg g)$ and so by our lemma it has at most $\deg f \deg g \max(\deg f, \deg g)$ zeros, i.e. there are at most $\deg f \deg g \max(\deg f, \deg g)$ values of $a \in \mathbb{F}_p$ such that f_a and g_a are not coprime (and nonzero), so that $a \in S_2$. Therefore S_2 has at most $\deg f \deg g \max(\deg f, \deg g)$ elements. For each such a, the common zeros of f_a and g_a are also zeros of each of f_a and g_a, and so there are at most $\min(\deg f_a, \deg g_a)$ of them; therefore

$$\sum_{a \in S_2} \nu(f_a, g_a) \leq \deg f \deg g \max(\deg f, \deg g) \min(\deg f, \deg g) = (\deg f \deg g)^2.$$

This sort of thing can also help us with the S_4 case. Ideally, we would find some number N depending on f and g, like the resultant, such that there exists a such that $f_a \equiv g_a \equiv 0 \pmod{p}$ if and only if p divides N; then we can rule out the existence of such problematic $a \in S_4$ for $p > N$. How can we do this?

An equivalent phrasing of coprimality for one-variable polynomials f and g over the *rational* numbers is that for any one-variable polynomial F we can find polynomials h and j such that $F = hf + jg$. In particular we can do this for the constant polynomial $F = 1$. Unfortunately, this doesn't work for higher numbers of variables for algebraic reasons.

However, if we extend to *rational* functions in x_1, i.e. fractions of polynomials, then it does work: as polynomials in x_2 with coefficients which are rational functions in x_1, f and g are coprime if and only if we can find similar such functions h and j such that $1 = hf + jg$. Clearing denominators (minimally, i.e. choosing h and j such that they are not both divisible by the same prime, just like f and g), we end up with something of the form $F = hf + jg$ where we can now assume that h and j are polynomials in x_1 and x_2 and F is a polynomial in x_1, and does not depend on x_2.

We want to know when both f and g are divisible by $x_2 - a$ modulo p for some a. Reducing both sides modulo p, since F does not depend on x_2 it can only be divisible by $x_2 - a$ if it is 0, since 0 is divisible by everything. Therefore the only p for which S_4 can ever be nonempty are those dividing F, i.e. those dividing the greatest common divisor of the coefficients of F. We call this greatest common divisor $G(f, g)$. In particular for $p > G(f, g)$, we conclude that S_4 is empty.

This completes the case $n = 2$: putting all the cases together and assuming $p > G(f, g)$, we get

$$\nu(f, g) \leq (\deg f \deg g)^2 + 2 \deg f \deg g.$$

This suggests choosing $A = (\deg f \deg g)^2 + 2 \deg f \deg g$ for our lemma, i.e. attempting to prove $\nu(f, g) \leq Ap^{n-2}$ with this value of A. This turns out to not be quite right, for reasons that we'll see; instead we'll call this quantity $T(f, g)$ and choose $A = (n - 1)T(f, g)$, which for the case $n = 2$ gives the same thing.

We can move on to the induction part. Suppose that the result holds for n; we want to prove it for $n + 1$. For $a \in S_1$, by the inductive hypothesis (i.e. the result we assumed for n) each $\nu(f_a, g_a)$ is at most $(n - 1)T(f_a, g_a)p^{n-2} \leq (n - 1)T(f, g)p^{n-2}$; and there are at most p possible a that could be in S_1, so the contribution from S_1 is at most

$$(n - 1)T(f, g)p^{n-1}.$$

Next, for S_2 we work similarly to above. Now f_a and g_a are multivariable polynomials, so we write $\mathrm{Res}_{x_i}(f_a, g_a)$ to denote the resultant when f_a and g_a are considered as single-variable polynomials in x_i, with the rest of the variables fixed. For $a \in S_2$, by assumption f_a and g_a are not coprime, so we can find a variable x_i such that fixing all the other variables $\gcd(f_a, g_a)$ is a nonconstant polynomial in x_i; therefore $\mathrm{Res}_{x_i}(f_a, g_a) = 0$ for every value of the other variables. Viewing $\mathrm{Res}_{x_i}(f_y, g_y)$ as a polynomial in y, it follows that it has a zero at a and therefore is divisible by $y - a$, so by a similar argument to above there are at most $\deg f \deg g \max(\deg f, \deg g)$ such a. By our lemma above, each such a gives rise to at most $\min(\deg f, \deg g)p^{n-1}$ simultaneous zeros of f_a and g_a modulo p, so the total contribution from S_2 is at most

$$(\deg f \deg g)^2 p^{n-1}.$$

Next, we look at S_3: exactly the same argument as for $n = 2$ applies, now with the number of zeros of the nonzero polynomial (say g_a) bounded by $\deg(g)p^{n-1}$ by our lemma above rather than just $\deg g$, and so we get a contribution of at most

$$2 \deg(f) \deg(g)p^{n-1}.$$

Combined with the S_2-contribution, they together give a contribution bounded by $T(f, g)p^{n-1}$.

Finally, the S_4 contribution works essentially the same as in the $n = 2$ case. The argument we used for $n = 2$ to show that there exists some polynomial $F(x_1, \ldots, x_{n-1})$ such that $x_n - a$ divides both f and g modulo p if and only if p divides all the coefficients of F works for any n, and in particular for $n + 1$ here; we similarly define $G(f, g)$ to be the greatest common divisor of the coefficients of F, and conclude that S_4 is empty for $p > G(f, g)$. (In fact, it's not too hard to see that actually $F = \mathrm{Res}_{x_n}(f, g)$.) Therefore the total is bounded by

$$\nu(f, g) \leq (n - 1)T(f, g)p^{n-1} + T(f, g)p^{n-1} = nT(f, g)p^{n-1},$$

which is exactly the desired bound for $n + 1$. Thus by induction our lemma—and therefore our theorem in the previous section—holds for

$$A = (\deg f \deg g)^2 + 2 \deg f \deg g$$

and

$$B = G(f, g),$$

or to ensure that the theorem holds as stated with the assumption that $B \geq \sqrt{A}$ we can take

$$B = \max(G(f, g), \sqrt{A}).$$

(I suspect that in most "natural" cases $G(f, g)$ will actually be 1 and so we can just take $B = \sqrt{A}$ (we assumed B was an integer, but actually the only thing we needed was for B^2 to be an integer) and get a result more similar to that of [2] which is simply a slight improvement.)

This concludes our proof! Let's try to think through an example, to get a sense of how good we expect this bound to be. (It really shouldn't be very tight: we just want to be sure it always holds.)

Consider $n = 3$ and $f(x, y, z) = 2x + y^3 - yz$, $g(x, y, z) = 11xyz - 5$. Experimentally, the average probability that $f(x, y, z)$ and $g(x, y, z)$ are relatively prime seems to be around 0.69. What would our bound give us?

First, we need to compute A and B. The easy one is A: $\deg f = \deg g = 3$, so

$$A = (3 \cdot 3)^2 + 2 \cdot 3 \cdot 3 = 99.$$

Computing B is a little trickier, but we can figure out that actually

$$11xy^3 + 22x^2 - 5 = 11xf(x, y, z) + g(x, y, z),$$

so since the left-hand side is independent of z and we have $\gcd(11, 22, 5) = 1$, we actually have $G(f, g) = 1$ and so we want to take B to be the smallest integer greater than $\sqrt{A}$, namely $B = 10$. To plug this into our theorem, observe first that

$$\vartheta(B) = \vartheta(10) = \log 2 + \log 3 + \log 5 + \log 7 = \log 210 \approx 5.3471,$$

and so $\exp(-n\vartheta(B)) = 210^{-3}$. Therefore in all our bound is

$$\delta(f, g) \geq \frac{1}{210^3} \exp\left(\left(P(2) - \log \zeta(2) - \frac{1}{10}\right) \cdot 101 \log\left(1 + \frac{99}{2}\right)\right) \approx 1.025 \times 10^{-32}.$$

In other words, this is a very very very bad bound. However, it *is* nonzero, and we were able to actually compute it and be certain that in the unlikely event that actually for large numbers the density does decrease greatly, it at least never gets below 1.025×10^{-32}.

Incidentally, we could also use (†) to estimate the true density, or at least to upper bound it: since each factor $1 - \frac{\nu(p)}{p^n}$ is less than 1, using only finitely many of them gives a result greater than the true one. We can compute directly (e.g. by enumerating all the possible points (x, y, z) modulo p and checking each one) that $\nu(2) = 1$, $\nu(3) = 2$, $\nu(5) = 13$, and $\nu(7) = 12$, so

$$\delta(f, g) \leq \left(1 - \frac{1}{2^3}\right)\left(1 - \frac{2}{3^3}\right)\left(1 - \frac{13}{5^3}\right)\left(1 - \frac{12}{7^3}\right) = \frac{662}{945} \approx 0.7005,$$

not too far from our original estimate. Notice that since this takes care of all the primes up to $\sqrt{A} \approx 9.95$, this allows us to replace the factor of $\frac{1}{210^3}$ with $\frac{662}{945}$, which improves the lower bound to $\approx 6.65 \times 10^{-26}$.

Using all the primes up to 100, we can get a more accurate estimate of

$$\delta(f, g) \leq 0.692.$$

Finally, although as we've seen computing B in practice isn't too hard we'd like to be able to bound it in order to make our final bound more explicit. Since $B = \max(A, G(f, g))$ and we know how to bound A, this amounts to bounding $G(f, g)$.

We know that $\mathrm{Res}_{x_n}(f, g)$ is a polynomial of degree $\deg f + \deg g$ in the a_i and b_i; thinking of it as the determinant, which as a polynomial has all coefficients ± 1 (or 0), it follows that each coefficient of $\mathrm{Res}_{x_n}(f, g)$ is given by the sum of at most $\deg f + \deg g$ terms coming from expansions of products of the a_i and b_j. Since the a_i and b_j have degrees at most $\deg f$ and $\deg g$ respectively, there are at most $\max(\deg f, \deg f)^{\deg f + \deg g}$ terms coming from any one product of a_i and b_j terms. If each of these entries has absolute value at most H, i.e. f and g have *height* at most H, then it follows that each entry of $\mathrm{Res}_{x_n}(f, g)$ is at most $(\deg f + \deg g)(H \max(\deg f, \deg g))^{\deg f + \deg g}$, and so $G(f, g) \leq (\deg f + \deg g)(H \max(\deg f, \deg g))^{\deg f + \deg g}$.

In practice this will usually be a very bad bound: in our example above, we have $H = 11$ and so this bound gives

$$G(f, g) \leq 6 \cdot (11 \cdot 3)^6 = 7748807814,$$

while the true value is just 1!

§6. The case $n = 1$

Finally, let's think about the special case $n = 1$. In this case, f and g have a common zero modulo p if and only if p divides $\mathrm{Res}(f, g)$, so in particular for only finitely many p; thus $\nu(p) = 0$ for all p sufficiently large. It follows from (†) that

$$\delta(f, g) = \prod_{p \mid \mathrm{Res}(f,g)} \left(1 - \frac{\nu(p)}{p}\right)$$

is nonzero, since we assume that $\nu(p) < p$ in the $n = 1$ case. To get a precise bound, we can observe that $\nu(p) < \min(\deg f, \deg g)$ by the Schwartz-Zippel lemma, and so we can apply similar reasoning to above, splitting the primes into the case $p < \min(\deg f, \deg g)$ and $p \geq \min(\deg f, \deg g)$; writing down the resulting bound is left as an exercise for the interested reader.

References

[1] Christian Axler. New estimates for some functions defined over primes. *arXiv preprint arXiv:1703.08032*, 2017.

[2] Avi Zeff. Counting points parametrized by infinitely many polynomial conditions. 2017.