Does Fermat's Last Theorem hold for polynomials?

Sachin Kumar
University of Waterloo, Faculty of Mathematics

Before going into the context of the polynomial case, I would really wish you all to read my previous essays on the Fermat's last theorem. Fermat's last theorem states that there are no non-trivial integer solutions to,

$$a^n + b^n = c^n$$

for $n \geq 3$ and $abc \neq 0$. We will be considering a similar statement, where the solutions are polynomials.

Theorem 1. Let $f, g, h \in \mathbb{C}[x]$ be non-constant polynomials such that no irreducible divides all of f, g, h. Then if

$$f^n + g^n = h^n$$

implies that $n \leq 2$.

Note that the statement that no irreducible divides all of f, g, h is equivalent to seemingly stronger statement that the three being pairwise coprime beacuse of the linear relation between, f^n, g^n, h^n. Also that this statement implies the more general result that if f, g, h are non-constant and satisfy $f^n + g^n = h^n$ then $n \leq 2$ because if there is some factor dividing all three, then we can divide out by the GCD of the three polynomials and reduce to the case in Theorem 1.

Note that we cannot improve on this bound of $n \leq 2$. Indeed, if $a \in \mathbb{C}[x]$ then it is always true that,

$$(1 - a^2)^2 + (2a)^2 = (1 + a^2)^2$$

Finally, note that Theorem 1 implies thee result for polynomials in any finite number of variables. To see this, let $f, g, h \in \mathbb{C}[x_1, \ldots, x_n]$ be polynomials that satisfy for $n \geq 3$,

$$f^n + g^n = h^n$$

Reordering variables if necessary, we can always choose $\alpha_1, \ldots, \alpha_n \in \mathbb{C}$ such that,

$$f(x_1, \alpha_2, \ldots, \alpha_n), g(x_1, \alpha_2, \ldots, \alpha_n), h(x_1, \alpha_2, \ldots, \alpha_n) \in \mathbb{C}[x_1]$$

are non-constant, violating Theorem 1. I will try to present three proofs, by using traditional techniques of number theory (Algebraic and computation), and geometric linear algebraic proofs.

We first note that it suffices to prove the result for $n = p$ a prime because all $n \geq 3$ are divisible by some prime p and if we have a solution for n, we replace

$$(f, g, h) \to \left(f^{\frac{n}{p}}, g^{\frac{n}{p}}, h^{\frac{n}{p}}\right)$$

to get a solution for p. Because we are working over $\mathbb{C}$ we have all roots of unity. Thus we can factor,

$$h^p = f^p + g^p = (f+g)(f+\zeta_p g)(f+\zeta_p^2 g)\cdots(f+\zeta_p^{p-1}g) = \prod_{i=0}^{p-1}(f+\zeta_p^i g)$$

Note that for $i \neq j$, $f+\zeta_p^i g$ and $f+\zeta_p^j g$ are coprime. Indeed,

$$\begin{aligned} f+\zeta_p^i g - (f+\zeta_p^j g) &= (\zeta_p^i - \zeta_p^j)g \\ f+\zeta_p^i g - (\zeta_p^i - \zeta_p^j)\frac{\zeta_p^i}{\zeta_p^i - \zeta_p^j} g &= f \\ \gcd(f+\zeta_p^i g, f+\zeta_p^j g) &= \gcd(f,g) = 1 \end{aligned}$$

Now, we have laid a surface for the first proof.

Proof. We will use the idea of induction and roots of unity. Let $\deg f + \deg g = d$ and $p \geq 3$. By above,

$$h^p = \prod_{i=0}^{p-1}(f+\zeta_p^i g)$$

These are pairwise coprime polynomials and h^p factors uniquely into irreducible because $\mathbb{C}[x]$ is a UFD, so they must be p^{th} powers. We induct on d. When $d = 2$, f, g are linear and this is clearly impossible by degree considerations. Now suppose, Theorem 1 holds for all degree less than d where $d > 2$. Now, $p \geq 3$, so $p - 1 \geq 2$ and so we must have some $a, b, c \in \mathbb{C}[x]$ such that,

$$\begin{aligned} f+g &= x^p \\ f+\zeta_p g &= y^p \\ f+\zeta_p^2 g &= z^p \end{aligned}$$

But then,

$$\begin{aligned} g &= \frac{1}{\zeta_p - 1}(y^p - x^p) \\ f &= \frac{1}{\zeta_p = 1}(y^p - \zeta_p x^p) \end{aligned}$$

Combining the above two equations with $f + \zeta_p^2 g = z^p$ yields,

$$\begin{aligned} \left(\frac{1}{\zeta_p - 1}(y^p - \zeta_p x^p)\right) + \zeta_p^2\left(\frac{1}{\zeta_p - 1}(y^p - x^p)\right) &= z^p \\ (-\zeta_p)x^p + (1+\zeta_p)y^p &= z^p \end{aligned}$$

Because we are working over $\mathbb{C}$, there exists $u, v \in \mathbb{C}$ such that $u^p = -\zeta_p$ and $v^p = 1 + \zeta_p$. Let $x' = ux$, $y' = vy$ and so substituting back in we get,

$$x'^p + y'^p = z^p$$

Note that x, y are non-constant of smaller degree than f, g respectively, so $\deg x + \deg y < \deg f + \deg g = d$ and thus this violates the inductive hypothesis and hence we are done. $\square$

In this proof, we use Mason's theorem as a lemma and prove Theorem 1 using this. A fun and serious fact about the Mason's theorem, if it holds for $\mathbb{Z}$, then Fermat's Last theorem's proof wouldn't be extremely complicated as it was. First, we define the derivative operator. Note that we can use the analytic definition as a guide, but that over general rings, the lacking analytic structure requires an algebraic definition for use to use the concept of a derivative. Let us define a linear function, $D : \mathbb{C}[x] \to \mathbb{C}[x]$ with $Dx = 1$, $D\alpha = 0$ for all $\alpha \in \mathbb{C}$ and if $f, g \in \mathbb{C}[x]$, let $D(fg) = fDg + gDf$. Then by induction we get $Dx^n = nx^{n-1}$ and in general Df is just the derivative as we think about it.

In $\mathbb{C}[x]$, we have unique factorization. Thus if $f \in \mathbb{C}[x]$, there exists irreducible $p_1, \ldots, p_n \in \mathbb{C}[x]$, $e_1, \ldots, e_n \in \mathbb{N}$, and $u \in \mathbb{C}$ such that

$$f = u \prod p_i^{e_i}$$

We define $\mathrm{rad} f = u \prod p_i$. Before, we begin the proof of Theorem 1, we will prove the Mason's theorem.

Proposition 4 (Mason's Theorem). Let $f, g, h \in \mathbb{C}[x]$ be non-constant and coprime such that $f + g = h$. Then,

$$\max\{\deg f, \deg g, \deg h\} \leq \deg \mathrm{rad}(fgh) - 1$$

Lemma 5. If $f \in \mathbb{C}[x]$, then we have the inequality

$$\deg \gcd(f, Df) \geq \deg f - \deg \mathrm{rad} f$$

Proof. We will prove Lemma 5. By unique factorization, there is are irreducibles $p_i \in \mathbb{C}[x]$, $u \in \mathbb{C}$, and natural numbers e_i such that $f = u \prod p_i^{e_i}$, making $\mathrm{rad} f = u \prod p_i$. For any i, let $f = p_i^{e_i} q_i$, so we have,

$$Df = D(p_i^{e_i} q_i) = p_i^{e_i} Dq_i + e_i p_i^{e_i - 1} q_i Dp_i = p_i^{e_i}(p_i Dq_i + e_i q_i Dp_i)$$

Thus for each i, $p_i^{e_i - 1} \mid Df$ and because the p_i are pairwise coprime, we have $\prod p_i^{e_i - 1} \mid Df$ and so $\prod p_i^{e_i - 1} \mid \gcd(f, Df)$. Let $g = \prod p_i^{e_i - 1}$. Then we have $\deg g \leq \deg \gcd(f, Df)$. But we have $g \mathrm{rad} f = f$ so $\deg g + \deg \mathrm{rad} f = \deg f$. Hence proved. □

Now, we are prepared to prove Mason's Theorem.

Proof. Remember, we discuseed that f, g, h are pairwise coprime. Now, notice that we have

$$f + g = h \tag{0.1}$$

Applying D and noting that D is linear gives,

$$Df + Dg = Dh \tag{0.2}$$

Multiplying (0.1) by Dg and (0.2) by g and subtracting yields,

$$fDg + gDg = hDg$$
$$gDf + gDg = gDh$$

We get,

$$fDg - gDf = hDg - gDh \tag{0.3}$$

To see that $fDg - gDf$ is non-zero, note that if $fDg = gDf$, by the fact that f, g are coprime, we must have $f \mid Df$ but Df is of lower degree so we must have $Df = 0$ so f is constant, contradicting the assumption that f is non-constant in Mason's theorem. Now, let

$$d_f = \gcd(f, Df)$$
$$d_g = \gcd(g, Dg)$$
$$d_h = \gcd(h, Dh)$$

Note that d_f, $d_g \mid fDg - gDf$ and that $d_h \mid hDg - gDh = fDg - gDf$ by (0.3) and that because f, g, h are pairwise coprime, so must d_f, d_g, d_h be. Thus, we have $d_f d_g d_h \mid fDg - gDf$. Clearly, $\deg(fDg - gDf) \leq \deg f + \deg g - 1$. By Lemma 5,

$$\deg d_f \geq \deg f - \deg \mathrm{rad} f$$
$$\deg d_g \geq \deg g - \deg \mathrm{rad} g$$
$$\deg d_h \geq \deg h - \deg \mathrm{rad} h$$

Thus, we have

$$\deg(d_f d_g d_h) = \deg d_f + \deg d_g + \deg d_h \geq \deg f + \deg g + \deg h - \deg \mathrm{rad} f - \deg \mathrm{rad} g - \deg \mathrm{rad} h$$

But $\deg(d_f d_g d_h) \leq \deg(fDg - gDf)$, so we have

$$\deg f + \deg g - 1 \geq \deg d_f + \deg d_h + \deg d_g \geq \deg f + \deg g + \deg h - \deg \mathrm{rad} f - \deg \mathrm{rad} g - \deg \mathrm{rad} h$$

Rearranging and cancelling, we get

$$\deg h \leq \deg \mathrm{rad} f + \deg \mathrm{rad} g + \deg \mathrm{rad} h - 1 = \deg \mathrm{rad}(fgh) - 1$$

where we get $(\mathrm{rad} f)(\mathrm{rad} g)(\mathrm{rad} h) = \mathrm{rad}(fgh)$ because f, g, h are pairwise coprime. We can now apply the same argument to the equations,

$$h + (-f) = g$$
$$h + (-g) = f$$

to bound $\deg f, \deg g$ with the same bpund. Thus, we are done. □

Now that we have proven this useful, albeit somewhat technical, lemma, we are prepared for the second proof of Theorem 1.

Proof. Suppose there exists $f, g, h \in \mathbb{C}[x]$ non-constant, coprime such that

$$f^n + g^n = h^n$$

By Mason's theorem, we have

$$\max\{\deg f^n, \deg g^n, \deg h^n\} \leq \deg \mathrm{rad}(fgh) - 1 \leq \deg f + \deg g + \deg h - 1$$

Because clearly $\mathrm{rad}(q^n) = \mathrm{rad} q$. The maximum of a finite set is at least the mean, so we have

$$\frac{\deg f^n + \deg g^n + \deg h^n}{3} = \frac{n}{3}(\deg f + \deg g + \deg h) \leq \max\{\deg f^n, \deg g^n, \deg h^n\}$$

Combining the above inequalities and letting $\deg f + \deg g + \deg h = d$, we get

$$\frac{nd}{3} \leq d - 1$$

Rearranging, we get

$$3 < d(3 - n)$$

By the fact that f, g, h are non-constant, we have $d > 0$ so $n < 3$. □

The final proof, we provide new modern approach, which is a culmination of recent theories in advanced linear algebra and algebraic geometry. Before, we begin the proof, we need to develop some prerequisite concepts. For any field k, we define $\mathbb{P}^n_k$ to be the set of lines in an $n+1$ dimensional k-vector space. This can be realized as the set,

$$\mathbb{P}^n_k = \{[x_0 : x_1 : \cdots : x_n] \in k^{n+1} \setminus \{0, 0, \ldots, 0\}\}/\sim$$

where $[x_0 : x_1 : \cdots : x_n] \sim [x'_0 : \cdots : x'_n]$ if and only if there is some $\lambda \in k^\times$ such that $x'_i = \lambda x_i$ for $1 \leq i \leq n$. We will be concerning ourselves with the projective line, $\mathbb{P}^1$ and the projective plane, $\mathbb{P}^2$. We can consider the sets,

$$U_i = \{[x_0 : x_1 : \cdots : x_n] \in \mathbb{P}^n \mid x_i \neq 0\}$$

and note that all points in U_i are equivalent to a unique.

$$\left[\frac{x_0}{x_i} : \frac{x_1}{x_i} : \cdots : \frac{x_{i-1}}{x_i} : 1 : \frac{x_{i+1}}{x_i} : \cdots : \frac{x_n}{x_i}\right]$$

Thus we can embed any n-dimensional vector space V in U_i for any i. Moreover, the union of all of the U_i is $\mathbb{P}^1$. There are many cool facts about projective space and projective geometry in a rich field of study, but will restrict ourselves to facts relevant to the subsequent proof. We will assume that all fields $k = \mathbb{C}$ so we will abbreviate $\mathbb{P}^n_{\mathbb{C}}$ as $\mathbb{P}^n$. We can consider polynomials as functions on points in k^n be evaluation in the usual way. For example, if $f \in \mathbb{C}[c, y]$, $f = x^2 + y^2$, then $f(1, 1) = 2$. We might wish to extend this to functions on projective coordinates. We can attempt to naively do the same thing, but we quickly run into a problem of the function being well defined. For instance, let $f = x^2 + y$. We have that $[1 : 2] \sim [2 : 4]$ but $f(1, 2) = 3 \neq 8 = f(2, 4)$. This leads us to homogeneous polynomials. The degree of a monomial $x_0^{\alpha_0} x_1^{\alpha_1} \cdots x_n^{\alpha_n}$ is defined to be $\alpha_0 + \alpha_1 + \cdots + \alpha_n$. We call a polynomial homogeneous if all monomials have the same degree. It is easy to check that for any $\lambda \in k$, if f is a homogeneous polynomial of degree d, then $f(\lambda x_0, \lambda x_1, \ldots, \lambda x_n) = \lambda^d f(x_0, x_1, \ldots, x_n)$. Thus it makes sense to talk about when a polynomial evaluates to 0 on projective space, for if $v = [x_0 : x_1 : \cdots : x_n] \sim v' = [x'_0 : x'_1 : \cdots : x'_n]$ then there is some $\lambda \neq 0 \in k$ such that $x'_i = \lambda x_i$. Given a polynomial in n variables, we can homogenize the polynomial by adding an extra variable t, letting the top degree monomial be of degree d and multiply each monomial of degree α by $t^{d-\alpha}$. For example, if we wish to homogenize the polynomial,

$$f = x^3 + y^3 + xy + x^2 + y$$

we add a variable z and notice that the top degree is 3 and get the polynomial,

$$f' = x^3 + y^3 + xyz + x^2 z + yz^2$$

Now we need to talk about some topological invariants. Heuristically the genus of a surface is the number of holes in the surface. We will take it on faith that the genus satisfies a number of nice properties, including that we can define this number on curves. We can define the Euler Characteristic, χ as $2 - 2g$ as a starting point. For the purposes of this lecture, we will gloss over much of deep theory, but the interested (and advanced) reader is directed to William Fulton's Intersection theory. One important property of χ is that if P is a point then $\chi(P) = 1$. Also, if U, V are disjoint, then $\chi(U \cup V) = \chi(U) + \chi(V)$. Thus, if S is a finite set of points of size n, then $\chi(S) = n$ by an easy induction argument. We can think of the degree of a map as the size of the inverse image at a suitably general point, assuming a variety of conditions that lie outside the scope of this essay. The interested reader is directed to learn algebraic geometry if he wished more rigor in this proof, with a suggested reference of Elements of Algebraic Geometry by Alexander Grothendieck. The relevant fact is that if if we have a degree n surjective map, $f : X \to Y$, then $\chi(X) = n\chi(Y)$. This can be thought of intuitively as coming directly from additivity. This idea is that in a sifficiently small neighbourhood U of a point P, we have $f^{-1}(U)$ consists of n disjoint copies of U. Thus by additivity, $\chi(f^{-1}(U)) = n\chi(U)$. Finally, we state without the proof that $g(\mathbb{P}^1) = 0$. This means that $\chi(\mathbb{P}^1) = 2 - 2 \cdot 0 = 2$. Finally, we are now prepared to prove Theorem 1 geometrically.

Proof. Let $Y \subset \mathbb{P}^2$ be,

$$Y = \{[x : y : z] \mid x^n + y^n = z^n\}$$

Let $\phi : Y \to \mathbb{P}^1$ sending $[x : y : z] \mapsto [x : y]$. Let,

$$Y_{[s:t]} = Y \cap \phi^{-1}([s : t]) = \{z \mid z^n = s^n + t^n\}$$

Because there are n roots of unity in $\mathbb{C}$, if $s^n + t^n \neq 0$, we have $|Y_{[s:t]}| = n$. Let $Z = \{[s : t] \mid s^n + t^n = 0\}$. Note that because $[0 : 0] \notin \mathbb{P}^1$, we must have that $s, t \neq 0$ if $[s : t] \in Z$. Thus $[s : t] = [1 : t']$, where $t' = \frac{t}{s}$. Then Z is just the set of n points $[1 : \rho^i]$ where ρ is a primitive n-th root of -1, so $\chi(Z) = n$. We now consider $\phi : Y \setminus \phi^{-1}(Z) \to \mathbb{P}^1 \setminus Z$. By the multiplicativity of the Euler characteristic, we have

$$\chi(Y \setminus \phi^{-1}(Z)) = n \cdot \chi(\mathbb{P}^1 \setminus Z) = n(\chi(\mathbb{P}^1) - \chi(Z)) = n(2 - n)$$

Note that $\phi : \phi^{-1}(Z) \to Z$ is a bijection so $\chi(\phi^{-1}(Z)) = \chi(Z) = n$. By the additivity of the Euler Characteristic, we have

$$\chi(Y) = \chi(Y \setminus \phi^{-1}(Z)) + \chi(\phi^{-1}(Z)) = n(2 - n) + n = n(3 - n)$$

By the computation $2 - 2g = \chi$, we get that

$$g(Y) = \frac{(n-2)(n-1)}{2}$$

Now, suppose that there are non-constant $f, g, h \in \mathbb{C}[x]$ such that

$$f^n + g^n = h^n$$

Let $f', g', h' \in \mathbb{C}[s, t]$ be the homogenized f, g, h. On U_1 Fermat's equation is clearly satisfied because all points are equivalent to $[s' : 1]$ and evaluating f', g', h' at $t = 1$ gives f, g, h. On U_0, every point is equivalent to $[1 : t']$ and evaluating f', g', h' at $s = 1$ gives polynomials $f, g, h \in \mathbb{C}\left[\frac{1}{t}\right]$ so the identity still holds. But $\mathbb{P}^1 = U_0 \cup U_1$ so the identity holds on all of $\mathbb{P}^1$. Let $\psi : \mathbb{P}^1 \to Y \subset \mathbb{P}^2$ send $[s : t] \mapsto [f'(s, t) : g'(s, t) : h'(s, t)]$. This is non-constant so we have

$$0 = g(\mathbb{P}^1) \geq g(Y) = \frac{(n-1)(n-2)}{2}$$

Thus, we must have $n = 1$ or $n = 2$. □