Introduction to p-adic Numbers

Benjamin Church and Matthew Lerner-Brecher

December 3, 2022

Contents

1 Introduction

Consider the infinite sum (series):

$$1 + 2 + 4 + 8 + 16 + \cdots = S$$

we want to find the value of this expression. Undaunted by the criticism "you can't count up infinitely many things" , we proceed by taking,

$$S - 2S = (1+2+4+\cdots) - 2(1+2+4+\cdots) = (1+2+4+8+\cdots) - (2+4+8+\cdots) = 1$$

Therefore, $-S = 1$ so $S = -1$ so we have the marvelous result that,

$$1 + 2 + 4 + 8 + 16 + \cdots = -1$$

This is an application of the more general formula that,

$$1 + x + x^2 + x^3 + \cdots = \frac{1}{1-x}$$

were we have plugged in $x = 2$. Now, your calculus teacher may object that this formula is only defined for $|x| < 1$ but if we listened to every naysayer screaming that our ideas don't quite make sense how far can we really get? In this class we are going to consider an alternative to the real numbers in which this formula makes sense. First, we need to think about what an infinite sum really means and how we can define its value.

2 Equivalence Relations

Definition: A relation $\sim$ on a set X is an equivalence relation if for every $a, b, c \in X$,

1. $a \sim a$
2. if $a \sim b$ then $b \sim a$
3. if $a \sim b$ and $b \sim c$ then $a \sim c$

Definition: Under and equivalence relation $\sim$ on the set X, the equivalence class of $x \in X$ is $[x] = \{y \in X \mid x \sim y\}$ the set of all equivalent elements.

Lemma 2.1. *if $x \sim y$ then $[x] = [y]$.*

Proof. If $a \in [x]$ then $a \sim x$ but $x \sim y$ so $a \sim y$ so $a \in [y]$. Likewise, if $a \in [y]$ then $a \sim y$ but $y \sim x$ so $a \sim x$ so $a \in [x]$. Therefore, $[x] = [y]$. □

Lemma 2.2. *If $[x] \cap [y] \neq \varnothing$ then $[x] = [y]$.*

Proof. Take $a \in [x] \cap [y]$ then $a \sim x$ and $a \sim y$ so $x \sim y$. Therefore, $[x] = [y]$. □

Lemma 2.3. *If $a \in [x]$ then $[a] = [x]$.*

Proof. Since $a \in [x]$ we have $a \sim x$ so $[a] = [x]$. □

Lemma 2.4. *Equivalence classes of $\sim$ over a set X partition X.*

Proof. Because for any $x \in X$, $x \sim x$ so $x \in [x]$ and therefore, the union of all equivalence classes is X. Also, if $[x] \cap [y] \neq \varnothing$ then $[x] = [y]$. Therefore, distinct equivalence classes are disjoint. □

3 Norms

We begin by defining a notion of distance from 0 in the rational numbers $\mathbb{Q}$.

Definition: A *norm* on $\mathbb{Q}$ is a function $||\cdot|| : \mathbb{Q} \to \mathbb{Q}$ satisfying:

1. $||x|| \geq 0$
2. $||x|| = 0$ if and only if $x = 0$
3. $||xy|| = ||x|| \cdot ||y||$
4. (Triangle Inequality) $||x+y|| \leq ||x|| + ||y||$

Remark 3.0.1. We have $||-x|| = ||x||$

Example 3.1. The standard absolute value:

$$||x|| = |x|$$

Example 3.2. The p-adic norm: Let $x = \frac{a}{b}$ where $a, b \in \mathbb{Z}$. Let $v_p(a), v_p(b)$ be the exponent of highest power of p dividing a, b respectively, then:

$$||x|| = |x|_p = p^{v_p(b) - v_p(a)}$$

For example,

- $54 = 2 \cdot 3^3$ so $v_3(54) = 3$ and therefore $|54|_3 = \frac{1}{3^3}$
- $\frac{24}{25} = \frac{2^3 \cdot 3}{5^2}$ so $v_3(\frac{24}{25}) = 3$ and therefore $|\frac{24}{25}|_3 = \frac{1}{3^3}$
- $\frac{24}{25} = \frac{2^3 \cdot 3}{5^2}$ so $v_5(\frac{24}{25}) = -2$ and therefore $|\frac{24}{25}|_5 = 5^2$

Proposition. The p-adic norm is non-archimedean, that is, $|a+b|_p \leq \max\{|a|_p, |b|_p\}$.

Proof. First, write the numbers in reduced form, $a = p^{v_p(a)}a'$ and $b = p^{v_p(b)}b'$ where a and b are rational numbers with both numerator and denominators not containing multiples of p. Now, because $v_p(a), v_p(b) \geq \min\{v_p(a), v_p(b)\}$ we can write $a + b = p^{\min\{v_p(a), v_p(b)\}}(a'p^x + b'p^y)$ where x and y are nonnegative. Thus, $v_p(a+b) \geq \min\{v_p(a), v_p(b)\}$ because $(a'p^x + b'p^y)$ can only contain positive powers of p. Therefore,

$$|a+b|_p \leq p^{-\min\{v_p(a), v_p(b)\}} = \max\{p^{-v_p(a)}, p^{-v_p(b)}\} = \max\{|a|_p, |b|_p\}$$

where I have used the fact that $-\min\{x, y\} = \max\{-x, -y\}$. □

Theorem 3.1 (Ostrowski)**.** *The only norms on $\mathbb{Q}$ are the p-adic norms for any prime p and the absolute value norm up to raising a given norm to a power greater than one.*

We will not provide a proof of this deep result here, however, Ostrowski's Theorem motivates the study of the p-adic numbers. Together with the real numbers, they exhaust the possible normed completions of the rational numbers.

This notion of distance from 0, gives us a notion of distance between any two points x, y through the expression $||x - y||$. We'll now use this notion of 0, to define more rigorously the notion of the terms of a sequence approaching some value. The intuition here is that the distance between points of the sequence and the value has to get closer and closer to zero.

4 Cauchy Sequences and Completion

Definition: A *metric space* is a set M and a function $d : M \times M \to \mathbb{R}$ which satisfies the following properties for any $x, y \in M$,

1. $d(x, y) \geq 0$
2. $d(x, y) = 0$ if and only if $x = y$
3. $d(x, y) = d(y, x)$
4. for any $z \in M$, $d(x, y) \leq d(x, z) + d(z, y)$

Example 4.1. Any norm on $\mathbb{Q}$ gives a metric defined by $d(x, y) = ||x - y||$. In particular, the standard distance on $\mathbb{Q}$ is given by taking the metric defined by the absolute value norm, $d(x, y) = |x - y|$.

Definition: A sequence a_n tends to a value a (written $a_n \to a$) if for all $\epsilon > 0$, there exists N such that for all $n \geq N$ we have

$$d(a_n, a) < \epsilon$$

Example 4.2. For sequences in $\mathbb{Q}$ with respect to the absolute value norm, the sequences $\frac{1}{n} \to 0$ and $\sum\limits_{k=0}^{n} \frac{1}{2^k} \to 2$

Remark 4.0.1. If the sequence $\sum\limits_{k=0}^{n} a_k \to L$ then we write $\sum\limits_{k=0}^{\infty} \frac{1}{2^k} = L$.

This gives us a notion of some sequence approaching a value. However, we want a further notion of *convergence* which loosely means that the sequence should approach a value, we just don't necessarily know what that value is. Such a sequence is known as a Cauchy sequence:

Definition: A sequence a_n is called a *Cauchy sequence* if for all ϵ there exists N such that for all $m, n \geq N$ we have:

$$d(a_n, a_m) < \epsilon$$

Theorem 4.1. *If $a_n \in M$ is a sequence such that there exists $a \in M$ for which $a_n \to a$, then a_n is a Cauchy sequence.*

Proof. Take the N such that for all $k \geq N$ we have $d(a_k - a) \leq \frac{\epsilon}{2}$. If $m, n \geq N$ we have by the triangle inequality:

$$d(a_m, a_n) \leq d(a_m, a) + d(a, a_n) < \frac{\epsilon}{2} + \frac{\epsilon}{2} = \epsilon$$

□

Now we feel like if we have a Cauchy sequence, there should be some value it approaches. Cauchy is the technical way of capturing the idea that a sequence *should* converge. We define a complete set to reflect this:

Definition: A set is *complete* if all Cauchy sequences converge to some value. A *completion* of $\mathbb{Q}$ with respect to some norm, is the set of all possible values a Cauchy sequence could converge to.

Remark 4.0.2. A more rigorous way to do this is to define the completion as the set of all possible Cauchy sequences and call two sequences equivalent if they tend to the same value.

Example 4.3. An important thing to note though is that all Cauchy sequences tend to some value in $\mathbb{R}$. For instance, partial expressions for π and the newton's method approximation for $\sqrt{2}$. This means that $\mathbb{R}$ is a complete metric space.

Definition: Sequences a_n and b_n are equivalent if $d(a_n, b_n) \to 0$ as a sequence in $\mathbb{Q}$ with the standard absolute value distance. We write $a_n \sim b_n$. Note that $d(a_n, b_n) \to 0$ is equivalent to the statement: for any $\epsilon > 0$ there exists N such that for all $n \geq N$ we have,

$$d(a_n, b_n) < \epsilon$$

Proposition. Sequence equivalence is an equivalence relation.

Proof. $d(a_n, a_n) = 0$ so clearly $d(a_n, a_n) \to 0$ therefore $a_n \sim a_n$. Let $a_n \sim b_n$ then $d(a_n, b_n) = d(b_n, a_n)$ so $d(b_n, a_n) \to 0$ and thus $b_n \sim a_n$.

Suppose that $a_n \sim b_n$ and $b_n \sim c_n$ then for any $\epsilon > 0$, by the definition of convergence, there exist N_1, N_2 corresponding to $\frac{\epsilon}{2}$ such that $n > N_1 \implies d(a_n, b_n) < \frac{\epsilon}{2}$ and $n > N_2 \implies d(b_n, c_n) < \frac{\epsilon}{2}$. Thus for any $n > N$ we have,

$$d(a_n, c_n) \leq d(a_n, b_n) + d(b_n, c_n) < \epsilon$$

so $d(a_n, c_n) \to 0$ and thus $a_n \sim c_n$. □

Definition: The metric completion of a metric space is the set of equivalence classes of Cauchy sequences under sequence equivalence.

Remark 4.0.3. The set $\mathbb{R}$ can be defined as the completion of $\mathbb{Q}$ with respect to the distance given by the standard absolute value. This is one of many equivalent constructions of the real numbers from $\mathbb{Q}$.

5 The p-adic Numbers

Definition: The set $\mathbb{Q}_p$ is the completion of $\mathbb{Q}$ with respect to the p-adic norm.

Remark 5.0.1. Two sequences are p-adic equivalent $a_n \sim b_n$ if $|a_n - b_n|_p \to 0$ as a sequence of distances in $\mathbb{Q}$ with the standard absolute value notion of convergence. This is equivalent to the statement that the sequence $(a_n - b_n) \to 0$ under the metric derived from the p-adic norm.

Definition: The set $\mathbb{Z}_p$ is the completion of $\mathbb{Z}$ with respect to the p-adic norm. Equivalently, it is the set of all $\alpha \in \mathbb{Q}_p$ such that α can be represented by a Cauchy sequence with only integer terms.

Remark 5.0.2. There is no equivalent notion of "real" integers in $\mathbb{R}$ viewed as the completion of $\mathbb{Q}$. The only Cauchy sequences with integer terms under the absolute value distance are eventually constant and therefore approach a standard integer in $\mathbb{Z}$. This is because no distinct integers get arbitrarily close under the usual notion of distance. However, under the p-adic notion of distance, powers of p do get arbitrarily close so we can have always non-constant Cauchy sequences with integer terms. Therefore, there can be p-adic integers which are not standard integers in $\mathbb{Z}$.

Example 5.1. As 2-adic numbers, $1 + 2 + 4 + 8 + \cdots = -1$. The technical statement is that the sequence $a_n = \sum_{k=0}^{n} 2^k$ and the sequence $b_n = -1$ are equivalent so the equivalence classes are equal: $\left[\sum_{k=0}^{n} 2^k\right] = [-1]$.
To prove this fact, note that,

$$a_n = \sum_{k=0}^{n} 2^k = 2^{n+1} - 1$$

Therefore, $|a_n - (-1)|_2 = |2^{n+1}|_2 = \frac{1}{2^{n+1}} \to 0$ so the sequences are equivalent.

Example 5.2. In general, in the p-adics, $1 + p + p^2 + p^3 + \cdots = \frac{1}{1-p}$.
To prove this fact, note that,

$$a_n = \sum_{k=0}^{n} p^k = \frac{p^{n+1} - 1}{p - 1}$$

Therefore, $|a_n - \frac{1}{1-p}|_p = |\frac{p^{n+1}}{p-1}|_p = \frac{1}{p^{n+1}} \to 0$ so the sequences a_n and $\frac{1}{1-p}$ are equivalent. Thus,

$$\sum_{k=0}^{n} p^k \to \frac{1}{1-p}$$

so we can say,

$$\sum_{k=0}^{\infty} p^k = 1 + p + p^2 + p^3 + \cdot = \frac{1}{1-p}$$

This fact shows that the validity of the formula $1 + x + x^2 + x^3 + \cdots = \frac{1}{1-x}$ extends somewhat beyond its interpretation as a Taylor series or a convergent sum in $\mathbb{R}$.

Lemma 5.1. *If $a_n \to a$ and $b_n \to b$ for $a, b \in \mathbb{Q}$ then $(a_n + b_n) \to a + b$.*

Proof. Let $\epsilon > 0$ then, by the definition of convergence, there exist N_1, N_2 corresponding to $\frac{\epsilon}{2}$ such that $n > N_1 \implies ||a_n - a|| < \frac{\epsilon}{2}$ and $n > N_2 \implies ||b_n - b|| < \frac{\epsilon}{2}$. Then for any $n > N$ we have,

$$|(a_n + b_n) - (a + b)|_p \leq |a_n - a|_p + |b_n - b|_p < \epsilon$$

so $(a_n + b_n) \to a + b$. □

A similar result holds for the product of two sequences. This suggests the following definition of the sums and products of p-adic numbers.

Definition: Let $\alpha, \beta \in \mathbb{Q}_p$ so there are Cauchy sequences a_n and b_n such that $\alpha = [a_n]$ and $\beta = [b_n]$ then $\alpha + \beta = [a_n + b_n]$ and similarly $\alpha\beta = [a_n b_n]$.

It follows immediately from this definition that if f is a polynomial and $\alpha \in \mathbb{Q}_p$ then $f(\alpha) = [f(a_n)] \in \mathbb{Q}_p$.

Theorem 5.2. *The series $\sum\limits_{n=0}^{\infty} a_n$ exists if and only if the sequence $a_n \to 0$.*

Proof. First suppose that the sum exists. Then, because the sum exists, its terms form a Cauchy sequence. For any $\delta > 0$ we can choose N so that $k-1, k > N$ implies that

$$\left| \sum_{n=0}^{k} a_n - \sum_{n=0}^{k-1} a_n \right|_p = |a_k|_p < \delta$$

which is exactly the definition of $a_k \to 0$. Conversely, let $a_n \to 0$ then for any $\delta > 0$ there exists N such that $n > N \implies |a_n|_p < \delta$. Now, we apply the ultrametric inequality. Suppose that $m > n > N$ then,

$$|a_{n+1} + a_{n+2} + \cdots + a_m|_p \leq \max\{|a_{n+1}|_p, |a_{n+2}|_p, \cdots, |a_m|_p\} < \delta$$

because each term is less than δ. Therefore,

$$\left| \sum_{n=0}^{m} a_n - \sum_{n=0}^{n} a_n \right|_p = |a_k|_p < \delta$$

which implies that $\left\{ \sum\limits_{n=0}^{k} a_n \right\}$ is a Cauchy sequence. However, the p-adics are complete meaning that every Cauchy sequence converges to some limit. Therefore, the series $\sum\limits_{n=0}^{\infty} a_n$ exists. □

Remark 5.0.3. For those who have studied the convergence of series in the real numbers, this theorem should come as quite a shock. In $\mathbb{R}$, the conditions for when a given series converges are extremely subtle and complex. However, in $\mathbb{Q}_p$ there is a simple to check necessary and sufficient condition. The analogous statement for the real numbers is emphatically false. For example, $\frac{1}{n} \to 0$ but $\sum\limits_{n=1}^{\infty} \frac{1}{n} \to \infty$. Due it its marvelous simplicity and power compared to the much less appealing situation for the more widely studied series over $\mathbb{R}$, this theorem is sometimes referred to as "The Freshman's Dream".

Example 5.3. In the p-adics, the sequence $p^n \to 0$ because $|p^n|_p = p^{-n} \to 0$ in the real numbers. Therefore, $\sum\limits_{n=0}^{\infty} p^n$ exists and in fact equals -1

Example 5.4. In the p-adics, the sequence $\frac{p^n}{n!} \to 0$ because $n!$ has at most $\frac{n}{p-1}$ powers of p so $v_p(\frac{p^n}{n!}) > n\frac{p-2}{p-1}$ and therefore the sequence goes to zero. Therefore, $e^p = \sum\limits_{n=0}^{\infty} \frac{p^n}{n!}$ exists in the p-adics

6 Ultrametric Geometry

We have seen that the p-adic norm satisfies a stronger version of the triangle inequality than the standard euclidean "length" does, namely, $|a+b|_p \leq \max\{|a|_p, |b|_p\}$. A space with a distance function which satisfies this inequality is called an *ultrametric space.* Formally,

Definition: An *ultrametric space* is a metric space satisfying the ultrametric inequality: $d(x,y) \leq \max\{d(x,z), d(z,y)\}$ for any x, y, z. Explicitly, an *ultrametric space* is a set M and a function $d : M \times M \to \mathbb{R}$ which satisfies the following properties for any $x, y \in M$,

1. $d(x,y) \geq 0$
2. $d(x,y) = 0$ if and only if $x = y$
3. $d(x,y) = d(y,x)$
4. for any $z \in M$, $d(x,y) \leq \max\{d(x,z), d(z,y)\}$

The p-adic numbers with the function $d(x,y) = |x-y|_p$ satisfies these conditions. We will deduce geometric properties of a general ultrametric space.

Definition: A *ball* centered at x with radius δ is $B_\delta(x) = \{y \in M \mid d(x,y) < \delta\}$

Proposition. In an ultrametric space, all triangles are isosceles.

Proof. Take three points $x, y, z \in M$ and let the side lengths of the triangle be $a = d(x,y), b = d(y,z), c = d(z,x)$. We can suppose without loss of generality that $a \leq b \leq c$. Using the Ultrametric inequality, $c \leq \max\{a,b\} = b$ so $c \leq b$. Therefore, $c \leq b$ and $b \leq c$ so $c = b$. Therefore, two sides of the triangle have equal lengths. Furthermore, $a \leq b = c$ so no isosceles triangles that are more obtuse than an equilateral triangle exist. □

Proposition. In an ultrametic space, every point inside a ball is the center.

Proof. Take the ball $B_\delta(c)$ and the point $a \in B_\delta(c)$. I claim that a is the center. This is true if $B_\delta(a) = B_\delta(c)$. We know that $d(a,c) < \delta$ because $a \in B_\delta(c)$. Take any $x \in B_\delta(a)$ then $d(a,x) < \delta$ but $d(c,x) < \max\{d(c,a), d(a,x)\} < \delta$ because both terms are less than δ. Thus, $d(c,x) < \delta$ so $x \in B_\delta(c)$. We have shown that every point in $B_\delta(a)$ is in $B_\delta(c)$. The reverse holds by the exact same argument. Take any $x \in B_\delta(c)$ then $d(c,x) < \delta$ but $d(a,x) < \max\{d(a,c), d(c,x)\} < \delta$ because both terms are less than δ. Thus, $d(a,x) < \delta$ so $x \in B_\delta(a)$. Therefore $B_\delta(a) = B_\delta(c)$. □

Definition: The boundary of a set A, denoted by ∂A, is the set of points x such that for any positive radius δ, the ball $B_\delta(x)$ contains points in A and points not in A.

Proposition. In an ultrametric space, no ball has boundary points.

Proof. Let x be a boundary point of $B_r(c)$ then for any $\delta > 0$ we must have $a, b \in B_\delta(x)$ with $a \in B_r(c)$ and $a \notin B_r(c)$. Therefore, $d(x,a) < \delta$ and $d(x,b) < \delta$ so $d(a,b) \leq \max\{d(a,x), d(x,b)\} < \delta$. We take $\delta < r$ so $d(c,b) \leq \max\{d(c,a), d(a,b)\} < r$ because $d(c,a) < r$ (since $a \in B_r(c)$) and $d(a,b) < \delta < r$. Therefore $d(c,b) < r$ so $b \in B_r(c)$ which contradicts the definition of a boundary point. Therefore, no ball can have any boundary points. □

The property that no balls have a boundary makes the topology of an ultrametric space *totally disconnected*. Basically this means that any subset that contains more than one point is disconnected in the sense that it can be broken up into disjoint parts that have no boundary points.

7 Polynomials with roots in $\mathbb{Q}_p$

Lemma 7.1 (Bezout). *There exists $x, y \in \mathbb{Z}$ such that $ax + by = \gcd(a,b)$.*

Proof. Consider the set,

$$T_{a,b} = \{ax + by \mid x, y \in \mathbb{Z} \text{ and } ax + by > 0\}$$

Because $T_{a,b} \subset \mathbb{Z}^+$ it has a least element, namely $g = ax_0 + by_0 \in T_{a,b}$. Consider any element $ax + by \in T_{a,b}$. Because $g > 0$, by the division algorithm, we can write,

$$ax + by = qg + r$$

where $0 \leq r < g$. Therefore,

$$r = ax + by - gq = (ax + by) - q(ax_0 + by_0) = a(x - qx_0) + b(y - qy_0)$$

Thus, if $r > 0$ then $r \in T_{a,b}$. However, $r < g$ which is the least element of $T_{a,b}$. This is a contradiction unless $r = 0$. Therefore the remainder is zero so, $g \mid ax + by$. In particular, $a, b \in T_{a,b}$ so $g \mid a$ and $g \mid b$. Furthermore, if any $c \mid a$ and $c \mid b$ then $c \mid ax_0 + by_0 = g$. Thus, any divisor of both a and b is a divisor of g so g is maximal. Thus, $g = \gcd(a, b)$. □

Lemma 7.2. *If a and n are coprime, which means $\gcd(a, b) = 1$, then there exists $s \in \mathbb{Z}$ such that $as \equiv 1 \pmod{n}$ so we can write, $s \equiv a^{-1} \pmod{n}$.*

Proof. If a and n are coprime then by Bezout's identity, there are integers $x, y \in \mathbb{Z}$ such that $ax + ny = 1$. Therefore, $n \mid ax - 1$ so $ax \equiv 1 \pmod{n}$. □

Theorem 7.3 (Hensel)**.** *If $f(x) \in \mathbb{Z}_p[x]$ and $a \in \mathbb{Z}_p$ satisfies*

$$f(a) \equiv 0 \pmod{p}$$
$$f'(a) \not\equiv 0 \pmod{p}$$

then there exists a p-adic integer α such that $f(\alpha) = 0$ and $\alpha \equiv a \pmod{p}$

Proof. We construct a series r_n such that $r_1 = a$ and for $n \geq 1$ we have:

$$r_{n+1} = r_n - f(r_n) \cdot s$$

where $s \equiv [f'(r_n)]^{-1} \pmod{p^{n+1}}$. We'll show by induction that

$$f(r_n) \equiv 0 \pmod{p^n}$$

and that such an integer s exists at every step. Our base case holds because $f(a) \equiv 0 \pmod{p}$ and $f'(a) \not\equiv 0 \pmod{p}$ implies that $f'(a)$ is coprime with p and therefore an inverse exists modulo p^2. We'll now do the inductive step. Assume that for all $k \leq n$ for some n the above equation holds. By hypothesis,

$$f(r_n) \equiv 0 \pmod{p^n}$$

Therefore, we can write,

$$f(r_n) \cdot s + mp^n = 0$$

and thus

$$r_{n+1} = r_n - f(r_n) \cdot s = r_n + mp^n$$

for some integer m. In particular,

$$r_{n+1} \equiv r_n \pmod{p^n}$$

Now consider the polynomial in t: $f(r_n + tp^n)$. Suppose f is of degree d and let a_i be the coefficients of f. Then we have:

$$f(r_n + tp^n) = \sum_{k=0}^{d} a_k (r_n + tp^n)^k$$

This can be written as

$$f(r_n + tp^n) = \sum_{k=0}^{d} c_k (tp^n)^k$$

for some terms c_k. We can get c_0 by just plugging in $t = 0$. This gives: $f(r_n) = c_0$ Now if we take the derivative with respect to t, the constant term goes to 0 so we get:

$$f'(r_n + tp^n) = \sum_{k=1}^{d} kc_k (tp^n)^{k-1}$$

If we plug in $t = 0$ this time we get: $c_1 = f'(r_n)$. First, we use this to show that $s \equiv [f'(r_{n+1})]^{-1} \pmod{p^{n+2}}$ exists. We know that $r_{n+1} = r_n + mp^n$ so,

$$f'(r_{n+1}) = f'(r_n + mp^n) = \sum_{k=1}^{d} kc_k (mp^n)^{k-1}$$

every term past $k = 1$ contains a factor of p so,

$$f'(r_{n+1}) \equiv c_1 = f'(r_n) \not\equiv 0 \pmod{p}$$

therefore, $f'(r_{n+1})$ is coprime with p and therefore also coprime with every power of p. In particular, $s \equiv [f'(r_{n+1})]^{-1} \pmod{p^{n+2}}$ exists.

Now, taking the next term in the sequence modulo p^{n+1}:

$$\begin{aligned} f(r_{n+1}) &= f(r_n + mp^n) \\ &= f(r_n) + f'(r_n)mp^n + \sum_{k=2}^{d} c_k m^k (p^n)^k \\ &\equiv f(r_n) + f'(r_n)mp^n \pmod{p^{n+1}} \end{aligned}$$

From its definition:

$$mp^n = -f(r_n)s$$

So we get:

$$\begin{aligned} f(r_{n+1}) &\equiv f(r_n) + f'(r_n)(-f(r_n)s) \pmod{p^{n+1}} \\ &\equiv f(r_n) - f(r_n)f'(r_n)[f'(r_n)]^{-1} \pmod{p^{n+1}} \\ &\equiv f(r_n) - f(r_n) \pmod{p^{n+1}} \\ &\equiv 0 \pmod{p^{n+1}} \end{aligned}$$

As desired. Now note that $r_{n+1} \equiv r_n \pmod{p^n}$. This means for all $m, n \geq N$ we have $r_m \equiv r_n \pmod{p^N}$. Thus,

$$|r_m - r_n|_p \leq \frac{1}{p^N}$$

Because the sequence is Cauchy, r_n converges to some p-adic number $\alpha = [r_n]$. The p-adic number $f(\alpha)$ is defined (by the definitions of multiplication and addition of p-adic numbers) as the limit of the sequence $f(r_n)$ (technically $f(\alpha) = [f(r_n)]$ the equivalence class of this sequence). However, we know that $f(r_n) \equiv 0 \pmod{p^n}$ so $|f(r_n)| \leq \frac{1}{p^n}$ implying $f(r_n) \to 0$. Thus, $f(\alpha) = 0$. $\square$

Example 7.1. The cool thing about Hensel's Lemma is it implies that $\sqrt{-1}$ is a p-adic number if $p \equiv 1 \pmod 4$. If we let $f(x) = x^2 + 1$. If $p \equiv 1 \pmod 4$ then there exists a such that

$$a^2 + 1 \equiv 0 \pmod p$$

Furthermore $f'(x) = 2x$, and we cannot have $2a \equiv 0 \pmod p$. By Hensel's lemma the function $f(x) = x^2 + 1$ has a root in the p-adics.

8 Appendix: Derivatives of Polynomials

Definition: A derivative $\frac{d}{dx}$ is a function that takes polynomials to other polynomials that satisfies

1. $\frac{d}{dx}x^n = nx^{n-1}$
2. $\frac{d}{dx}cP = c \cdot \frac{d}{dx}P$
3. $\frac{d}{dx}(P + Q) = \frac{d}{dx}P + \frac{d}{dx}Q$

Remark 8.0.1. For simplicity sake, we often denote $\frac{d}{dx}P$ as P'

Example 8.1. The derivative of $x^2 + 2x + 1$.