Newton Polygons of Polynomials Composed with Eisenstein Polynomials

By

Uri Tomer

* * * * * * * * *

Submitted in partial fulfillment of the requirements for
Honors in the Department of Mathematics

Union College

March 18th 2024

ABSTRACT

URI TOMER Newton polygons of polynomials composed with Eisenstein polynomials.

Department of Mathematics, 3/18/24.

ADVISOR: Gajek-Leonard, Rylan

In this paper we construct the p-adic number field $\mathbb{Q}_p$ and discuss some of its properties. We then introduce the concept of a Newton polygon of a polynomial with coefficients in $\mathbb{Z}_p$. We prove a theorem about the Newton polygon of a one segment Newton polygon f composed with an Eisenstein Newton polygon g. The theorem states that the Newton polygon of $f \circ g$ will be the Newton polygon of f stretched by a factor of the degree of g. Finally we discuss a conjecture about the Newton polygon of $f \circ g$ if f has more than one segment. The conjecture states that the Newton polygon of $f \circ g$ will again be stretched by a factor of the degree of g. We provide some evidence to support this conclusion.

ACKNOWLEDGEMENT

I would like to thank professor Rylan Gajek Leonard without whom this paper would not have been imbued with the same enthusiasm and care which it contains and deserves. Furthermore I would like to thank the entire Union College math department who have been incredibly supportive and nurturing of my curiosity over the last four years. Finally I would like to thank my parents, without whom I would never be in this situation to begin with.

NOTATION

We shall use the following notation throughout this paper. We write $\mathbb{N}$ for the set of natural numbers, $\mathbb{R}$ for the set of real numbers, $\mathbb{C}$ for the set of complex numbers and $\mathbb{Q}$ for the set of rational numbers.

Contents

1. Introduction

Consider the sum of all powers of 2.

$$S = \sum_{i=0}^{\infty} 2^i = 1 + 2 + 4 + 8 + \cdots + 2^i + \cdots .$$

A student studying infinite series for the first time might tell you that this sum is divergent; it does not equal anything. This is a natural and in most ways correct instinct however as one matures mathematically, and starts to expand their toolbox, they may find that there is some situations in which we can make sense of this sum and even assign it a value. Consider the following argument:

We can think of the sum as the power series whose coefficients are the powers of two 2 evaluated at 1. That is $S = f(1)$ when,

$$f(x) = 1 + 2x + 4x^2 + 8x^3 + \cdots = \sum_{i=0}^{\infty} (2x)^i$$

We know through the use of geometric series that,

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

Evaluating f at 1 then clearly gives us $S = -1$. Quite the strange result! Unfortunately things are not quite so simple. The careful student will (rightfully) point out that the radius of convergence of f is $\frac{1}{2}$ centered around the origin. Thus our evaluation is divergent and certainly not equal to -1.

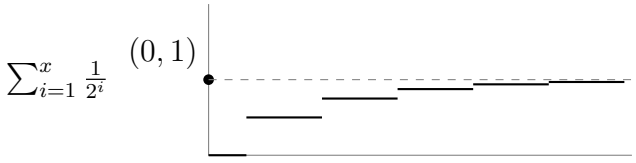
Satisfaction may come in at this point. Of course the notion that a strictly increasing sum would equal a negative number seems absolutely preposterous. Indeed how can such a sum end up being smaller than any of its partial sums

or even any of the individual numbers being added? However to a certain student dismissing this idea so early might seem rather *unsatisfying*. Maybe by backing up and thinking about those sums which we classically think of as convergent we might gain some insight into the relation between the sum S and -1.

The first convergent sum that most students encounter is that of the reciprocals of the powers of 2,

$$S_2 = \sum_{i=1}^{\infty} \frac{1}{2^i} = \frac{1}{2} + \frac{1}{4} + \frac{1}{8} + \cdots$$

Geometrically, it is clear that the partial sums of S_2 approach 1.



After years of math education we may take this fact for granted but a student encountering these objects for the first time might ask if it even makes rigorous sense for an infinite sum to equal some finite number and this question should not be taken lightly. We want to say that this sum *approaches* 1 but what is meant by approach is not so clear. We cannot say the distance between S_2 and 1 simply gets smaller with every partial sum as for that matter the distance between S_2 and 2 does the same, in fact it does so with any real number greater than 1. What makes 1 equivalent to S_2 is that any partial sum can get *arbitrarily* close to 1. We might express this fact using epsilons and inequalities but for now we will just say that no matter how small a number you choose, there is a point in S_2 such that the distance between between 1 and any partial sum after that point is less than your chosen number.

In hindsight this feels like an obvious notion but it is more clever than it initially seems. Archimedes, living on the island of Sicily in the third century BC, used such a definition however with his exception, mathematicians would take until the scientific revolution of the 16th century to recover the defintion of convergence.

Being mathematically inclined as we are, we might seek to generalize at this point. Generalization is often hindered by arbitrary decisions so we look for where we made such an arbitrary decision along the way and remedy it. One such decision is the series S_2 that we chose to consider. Aiming to find a sum that converges to the value 1 we could have just as easily chosen $S_3 = \sum_{i=1}^{\infty} \frac{2}{3^i} = \frac{2}{3} + \frac{2}{9} + \frac{2}{27} + \cdots$ or $S_{10} = \sum_{i=1}^{\infty} \frac{9}{10^i} = \frac{9}{10} + \frac{9}{10^2} + \cdots$ or any such sum that splits up the interval $(0, 1)$ into pieces with a fixed ratio, i.e.,

$$S_n = \sum_{i=1}^{\infty} \frac{n-1}{n^i}.$$

In fact, the sum S_{10} tells us that

$$S_{10} = 0.9 + 0.09 + 0.009 + \cdots = 0.999... = 1.$$

This is a very pretty result which has the tendency of infuriating those that have not taken the time to carefully examine mathematical definitions. They claim that 0.999... does not really equal 1 but only approaches it. Of course we know that by definition, those two statements, about equality and approach, are equivalent.

We can perform a similar summation that approaches 1 by choosing a number $p \in (0, 1)$ and splitting up the interval $(0, 1)$ into 2 pieces of length p and $1 - p$. Taking another step we split up p into two pieces of size p^2 and

$p(1-p)$. Continuing like this, we split p^2 into p^3 and $p^2(1-p)$ and p^3 into p^4 and $p^3(1-p)$, and we continue on and on always splitting the piece of length p^i into pieces of p^{i+1} and $p^i(1-p)$. Each of these partial sums will always equal 1.

$$1 = (1-p) + p$$

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

$$= 1 - p + p(1-p) + (p^2(1-p) + p^3)$$

$$= 1 - p + p(1-p) + p^2(1-p) + (p^3(1-p) + p^4)$$

$$\dots$$

This gives us the sum,

$$\sum_{i=1}^{n} p^i(1-p) = 1$$

for any $n \in \mathbb{N}$. In our quest to generalize, we may push this and consider the infinite sum of this form

$$\sum_{i=1}^{\infty} p^i(1-p) = 1.$$

We can manipulate this a little bit to get the following very pretty formula:

$$\sum_{i=1}^{\infty} p^i = \frac{1}{1-p}$$

for any $p \in (0,1)$. The choice to use p as our variable here is as suggestive as it is arbitrary.

Again we seek to generalize, hoping to extend our domain from $(0,1)$ to $\mathbb{R} - \{1\}$ if we can. If we do this blindly and just plug in interesting values we

get strange results such as,

$$\sum_{i=1}^{\infty}(-1)^i = 1 - 1 + 1 - 1 + 1 - 1 + \cdots = \frac{1}{1-(-1)} = \frac{1}{2}.$$

This seems incredibly strange and yet perfectly obvious at the same time. Plugging in 2 we get our old friend S,

$$\sum_{i=1}^{\infty} 2^i = \frac{1}{1-2} = -1$$

And indeed, we again find that it "equals" -1, pushing us to believe that there is some underlying meaning here and not just random noise. A student who is especially inclined to favor rigor might move to ignore these results for we cannot say that they approach any values; they are clearly divergent. Conversely, a more adventurous student might try to find a way to *make* such a tantalizing pattern make sense.

Following the path of the second student, we again look for arbitrary decisions that we made along the way. The arbitrary decision that we made here turns out to be incredibly subtle. So subtle, that it took until the late 20th century for anyone to notice. In 1897, the German mathematician Kurt Hensel introduced the p-adic numbers and in doing so pointed out a subtle assumption at the heart of number theory [3]. He noticed that the way that we measure distance between the rationals is an arbitrary decision. The usual way to measure the distance between two rational a and b is given by $|a - b|$. The tricky part has to do with the absolute value sign. We normally think of absolute value as the function which takes negative numbers to their positive counterparts and keeps the positive ones where they are. But if we aim to

capture the natural notion of distance, there are other ways that we can think about absolute value while retaining the essential properties.

After much experimentation, we might come to realize that there exist perfectly reasonable notions of absolute value that measure numbers not by their physical distance to the origin but by the reciprocal of how many times a certain prime p divides them. In order to generate some intuition through examples we can take $p = 2$ and produce the following absolute values for some arbitrary but easy numbers,

$$|2|_2 = 1, |16|_2 = 1/4, |-24|_2 = 1/2, |8/9|_2 = 1/3 \text{ etc.}$$

We call this notion of distance the 2-adic metric or more generally for any prime p, the p-adic metric. If we measure numbers in this way we can finally triumph over the tricky sum that began this exploration. Recall that we say that a sum approaches some number L if its partial sums get arbitrarily close to L. Measuring closeness with the 2-adic metric we find that if we add 1 to our sum S, then the nth partial sum of S is equal to 2^{n+1}. That is,

$$1 + \sum_{i=0}^{n} 2^i = 1 + (1 + 2 + 4 + 8 + \cdots + 2^n) = 2^{n+1}.$$

Taking the 2-adic absolute value of each of these partial sums we get

$$|1 + \sum_{i=0}^{n} 2^i|_2 = |2^{n+1}|_2 = \frac{1}{n+1}.$$

As n gets bigger, $\frac{1}{n+1}$ will approach 0. Therefore $1 + S = 0$ and thus,

$$S = \sum_{i=0}^{\infty} 2^i = 1 + 2 + 4 + 8 + \cdots = -1$$

(with respect to the 2-adic metric) just as we had hoped.

The subject of this thesis will be the exploration of these alternate absolute values and their corresponding distance functions. As alluded to before, they are called the p-adic metrics and give us the p-adic numbers. These numbers have been incredibly important in the development of math in the last 100 or so years. They have contributed to many proofs. Most famously they were integral to Andrew Wiles' famous proof of Fermat's last theorem where he used properties of both the 3-adics and the 5-adics. Additionally, they have far reaching applications within pure math and outside of it stretching as far as physics and computer science. Above all, they allow for some clever and even pretty techniques which give nice and sometimes beautifully counter intuitive results. This is the strength of the p-adics.

2. Setting the stage

In order to construct the p-adics we need only consider the field $\mathbb{Q}$. However, in order to gain an intuition it will be helpful to consider absolute values on an arbitrary field $\mathbb{K}$.

Definition 1. An **absolute value** on a field $\mathbb{K}$ is a function

$$|\ |: \mathbb{K} \to \mathbb{R}_{\geq 0}$$

that satisfies the following conditions:

$$i) |x| = 0 \text{ if and only if } x = 0$$

$$ii) |xy| = |x||y| \text{ for all } x, y \in \mathbb{K}$$

$$iii) |x + y| \leq |x| + |y| \text{ for all } x, y \in \mathbb{K}$$

Furthermore, an absolute value is **non-Archimedean** if it satisfies:

$$iv) |x + y| \leq \max\{|x|, |y|\} \text{ for all } x, y \in \mathbb{K}$$

Otherwise we will say that the absolute value is **Archimedean**.

At this point it will be useful to confirm that our usual notion of absolute value on $\mathbb{R}$ satisfies this definition.

Proposition 1. *Let* $|\ |: \mathbb{R} \to \mathbb{R}_{\geq 0}$ *be defined by*

$$|x| = \begin{cases} x & \text{if } x \geq 0 \\ -x & \text{if } x < 0 \end{cases}$$

Then $|\ |$ *is an absolute value.*

Proof. Condition (i) clearly holds. For conditions (ii) and (iii) there are 4 possible cases:

$$a)\ x, y \geq 0$$

$$b)\ x \geq 0, y < 0$$

$$c)\ y \geq 0, x < 0$$

$$d)\ x, y < 0$$

If (a) then,

$$|xy| = xy = |x||y|$$

So (ii) holds. Also,

$$|x+y| = x+y = |x|+|y| \leq |x|+|y|$$

So (iii) holds. Cases (b) and (c) are clearly symmetric so we will only consider (b). Then we have,

$$|xy| = -xy = x(-y) = |x||y|$$

So (ii) holds. Also,

$$|x|+|y| \geq x+y \geq |x+y|$$

So (iii) holds. Finally if we have (d) then,

$$|xy| = xy = (-x)(-y) = |x||y|$$

So (ii) holds. Also,

$$|x+y| = -(x+y) = (-x)+(-y) = |x|+|y| \leq |x|+|y|$$

Therefore our usual absolute value is indeed an absolute value. □

Notice that this absolute value is Archimedean. This is easy to see when considering $x = y = 1$. What we are capturing with the Archimedean property is the idea that we can have arbitrarily large numbers. Using our usual absolute value this is intuitively true but our definitions give us a way to formalize this idea. This insight, that numbers as we are used to them can be arbitrarily large, dates back aver 2200 years to Archimedes. Repeatedly referring to this function as "the usual notion of absolute value" will become tiring so for reasons that will become clear later we shall refer to this function as the **infinite absolute value**.

Another example worth considering is the **trivial absolute value**. It is defined as $|0| = 0$ and $|x| = 1$ if $x \neq 0$. Clearly this satisfies the definition of an absolute value and indeed works for any field $\mathbb{K}$. However its simplicity can actually be quite bothersome and it is excluded from many well known theorems on absolute values. This does not mean it is useless however. Consider the following proposition:

Proposition 2. *Let $\mathbb{K}$ be a finite field. Then the only absolute value on $\mathbb{K}$ is the trivial absolute value.*

Proof. Let $\mathbb{K}$ be a finite field and let $|\ |$ be an absolute value on it. We will rely on the following well known fact from field theory: $\mathbb{K}$ is finite implies that its multiplicative group $\mathbb{K}^\times = \mathbb{K} - \{0\}$ is generated by some $g \in \mathbb{K}^\times$. Thus there exists $n \in \mathbb{N}$ such that $g^n = 1$. Accordingly, $1 = |1| = |g^n|$. By property (ii) of absolute values, $|g^n| = |g|^n = 1$ and thus $|g|^n - 1 = 0$. Notice here that $|g| \in \mathbb{R}_{\geq 0}$ so $|g|^n = 1$ implies that $|g| = 1$. Finally, $|0| = 0$ by definition. So we have

$$|x| = \begin{cases} 0 & \text{if } x = 0 \\ 1 & \text{else} \end{cases}$$

This of course is the trivial absolute value. □

Thus ends our exploration of the trivial absolute value. Let us return to the non-Archimedian absolute values. A useful property of the non-Archimedean condition or strong triangle inequality as it is sometimes referred to is that given $|x + y| \leq \max\{|x|, |y|\}$, we can show equality if $|x| \neq |y|$.

Proposition 3. *Let $|\ |$ be a non-Archimedean absolute value on a field $\mathbb{K}$. Then for $x, y \in \mathbb{K}$ such that $|x| \neq |y|$, $|x + y| = \max\{|x|, |y|\}$.*

Proof. Without loss of generality assume $|x| > |y|$. We will show that $|x| \leq |x+y|$ and $|x| \geq |x+y|$. From the strong triangle inequality we have that,

$$|x+y| \leq \max\{|x|, |y|\} = |x|.$$

Now consider $|x| = |x+y-y|$. Again from the strong triangle inequality we have,

$$|x| = |x+y-y| \leq \max\{|x+y|, |-y|\} = \max\{|x+y|, |y|\} = |x+y|$$

by the assumption that $|x| > |y|$. Therefore $|x| \leq |x+y|$ and $|x| \geq |x+y|$ and thus $|x| = \max\{|x|, |y|\} = |x+y|$. □

We will now work towards the main subject of our investigation: the p-adic absolute values. In order to do this we must first introduce the related concept of a p-adic valuation. Intuitively, we want to measure how many times a certain prime p divides a number n. This is what we aim to represent with $v_p(n)$.

Definition 2. Let p be a prime. Then the **p-adic valuation** is the function

$$v_p : \mathbb{Z} - \{0\} \to \mathbb{R}$$

defined as follows: for any $n \in \mathbb{Z} - \{0\}$, $v_p(n)$ is the unique positive integer satisfying

$$n = p^{v_p(n)} n'$$

such that $n' \in \mathbb{Q}$ and p does not divide n'.

We note that by the fundamental theorem of arithmetic a positive integer $v_p(n)$ exists and is unique in the above definition.

At this point it will be helpful to work through several examples.

Example 1. Let $p = 5$ and $n = 375$. Then,

$$375 = 5^{v_5(375)}n'$$

Taking $n' = 3$ we have

$$125 = 5^{v_5(375)} \Rightarrow v_5(375) = 3$$

Notice here that for any n such that $5 \nmid n$, $125n$ will have a 5-adic valuation of 3.

Intuitively, we are measuring size by counting how many times our prime p divides our chosen number. Thus if our number is not divisible by p it should have a valuation of 0. Let us consider an example where this is the case.

Example 2. Let $p = 2$ and $n = 15$. Then,

$$15 = 2^{v_2(15)}n'$$

taking $n' = 15$ we have

$$1 = 2^{v_2(15)} \Rightarrow v_2(15) = 0.$$

Our theory works. Now consider $v_p(0)$. Intuitively we can divide 0 by any number infinitely many times and still be left with 0. So we will define $v_p(0) = \infty$ for all primes p with the usual conventions used to handle this

symbol. This is a nice start but we can go further. Let us extend the function to include all of $\mathbb{Q} - \{0\}$ in the following way:

Definition 3. let $x = \frac{a}{b} \in \mathbb{Q} - \{0\}$. Then

$$v_p(x) = v_p(a) - v_p(b)$$

Again an example may be useful.

Example 3. Consider $v_3(123/48)$. By the defintion above,

$$v_3(123/48) = v_3(123) - v_3(48).$$

First let us compute $v_3(123)$. We have,

$$123 = 3^{v_3(123)}n'.$$

Taking $n' = 41$ we have

$$3 = 3^{v_3(123)} \Rightarrow v_3(123) = 1$$

Next let us compute $v_3(48)$. We have,

$$48 = 3^{v_3(48)}n'$$

Taking $n' = 16$ we have

$$3 = 3^{v_3(48)} \Rightarrow v_3(48) = 1$$

Therefore,

$$v_3(123/48) = 1 - 1 = 0$$

Having familiarized ourselves with the concept let us explore some of the properties of the p-adic valuation.

Proposition 4. *For all $x, y \in \mathbb{Q}$ we have,*

$$i)\ v_p(xy) = v_p(x) + v_p(y)$$

$$ii)\ v_p(x+y) \geq \min\{v_p(x), v_p(y)\}$$

Proof. First let us consider the case when x and y are integers. Then we have

$$x = p^{v_p(x)}x' \text{ and } y = p^{v_p(y)}y'$$

where $x', y' \nmid p$. Then,

$$xy = p^{(v_p(x))+(v_p(y))}x'y'$$

Thus we have (i). Assume without loss of generality that $v_p(x) \leq v_p(y)$. Then $\min\{v_p(x), v_p(y)\} = v_p(x)$. We can see that,

$$x + y = p^{v_p(x)}x' + p^{v_p(y)}y' = p^{v_p(x)}(x' + p^{(v_p(y))-(v_p(x))}y') \geq p^{v_p(x)}$$

which implies that $v_p(x+y) \geq v_p(x)$. Thus we have (ii). To extend to the rationals, let $x = a/b \in \mathbb{Q} - \{0\}$ and $y = c/d \in \mathbb{Q} - \{0\}$. Then

$$v_p(xy) = v_p(\frac{ac}{bd}) = v_p(ac) - v_p(bd)$$

Here we apply (i) to ac and bd as they are integers to get,

$$v_p(ac) - v_p(bd) = v_p(a) + v_p(c) - v_p(b) - v_p(d) = v_p(\frac{a}{b}) + v_p(\frac{c}{d}) = v_p(x) + v_p(y)$$

Thus we have (i) for rationals. Because we have assumed that $v_p(x) \leq v_p(y)$, we have,

$$\min\{v_p(x), v_p(y)\} = v_p(x) = v_p(\frac{a}{b}) = v_p(a) - v_p(b)$$

We can see that,

$$x+y = p^{v_p(a)-v_p(b)}x' + p^{v_p(y)}y' = p^{v_p(a)-v_p(b)}(x' + p^{(v_p(y)-v_p(a)-v_p(b))}y') \geq p^{v_p(a)-v_p(b)}$$

This implies that

$$v_p(x+y) \geq v_p(a) - v_p(b) = \min\{v_p(x), v_p(y)\}$$

Therefore we have (ii) for the rationals as well. □

Comparing this proposition with properties (ii) and (iv) of the definition of absolute values we find that it is eerily close to showing that p-adic valuations are absolute values. Except of course, they are not. The product in the first property has turned into a sum and the inequality in the second property is reversed. No matter, we can remedy this by putting the sum in an exponent changing the sign of the inequality. Thus we introduce the p-adic absolute value.

Definition 4. Let p be a prime. Then the **p-adic absolute value** is a function

$$|\ |_p : \mathbb{Q} \to \mathbb{R}_{\geq 0}$$

Defined by $|0|_p = 0$ and,

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

otherwise.

Of course the next reasonable step is to show that this is indeed an absolute value.

Proposition 5. *For any prime p, $|\ |_p$ is a non-archimedean absolute value.*

Proof. Condition (i) clearly holds. Let $x, y \in \mathbb{Q}$. Then $|xy|_p = p^{-v_p(xy)}$. By proposition 4,

$$p^{-v_p(xy)} = p^{-v_p(x)-v_p(y)} = p^{-v_p(x)}p^{-v_p(y)} = |x|_p|y|_p$$

Thus we have (ii). Continuing, we have $|x+y|_p = p^{-v_p(x+y)}$. Again making use of proposition 4 we have,

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

Thus we have (iv). Notice here that $(iv) \Rightarrow (iii)$ as $\max\{|x|_p, |y|_p\} \leq |x|_p + |y|_p$. Therefore $|\ |_p$ is a non-archimedean absolute value. □

At this point it would be wise to step back and reexamine what it is we are measuring. As discussed before, the p-adic valuation of a number n is very large when n is very divisible by p. However we have reversed the situation due to the negative sign in the exponent. Accordingly, the p-adic absolute value of n will be very small when n is very divisible by p.

One question that we may have asked ourselves repeatedly up to this point is why we limit ourselves to primes when considering p-adic numbers. This question can be answered by Ostrowski's theorem which is discussed below. However this is not necessary. Using the above definition, we can show that allowing p to be a composite number does not work.

Example 4. Suppose we want to construct the 6-adic absolute value and find $|12|_6$. Using a slight variation of the definition above where we allow p to be composite we should have that

$$|12|_6 = 6^{-v_6(12)}.$$

Computing $v_6(12)$ we should get a unique integer where

$$12 = 6^{v_6(12)}n'$$

such that 6 does not divide $n' \in \mathbb{Q}$. Unfortunately this does not produce a unique answer. If $n' = 2$, then $v_6(12) = 1$ however if $n' = \frac{1}{3}$, $v_6(12) = 2$. Therefore $|12|_6 = \{6^{-1}, 6^{-2}\}$ Of course we cannot have this as then v_p would not be well defined. Thus we must stipulate that p is prime when discussing the p-adics.

Having dispelled of that possibility we can move forward with our construction. We are almost there but a few things remain in order to properly construct the p-adics. Lovers of analysis may at this point wonder if p-adic absolute value can be used to define a metric analogously to how the infinite absolute value is used to define the Euclidean metric. Naturally we find that it can.

Proposition 6. *Let* $d_p : \mathbb{Q} \times \mathbb{Q} \to \mathbb{R}_{\geq 0}$ *be defined by*

$$d_p(x, y) = |x - y|_p$$

Then d_p *is a metric.*

Proof. To show that d_p is a metric we must prove the following conditions for $x, y, z \in \mathbb{Q}$:

$$i)\ d_p(x,y) \geq 0$$

$$ii)\ d_p(x,y) = 0 \text{ if and only if } x = y$$

$$iii)\ d_p(x,y) = d_p(y,x)$$

$$iv)\ d_p(x,z) \leq d_p(x,y) + d_p(y,z)$$

The codomain of d_p guarantees (i). For (ii) we have,

$$d_p(x,y) = 0 \iff |x-y|_p = 0 \iff x - y = 0 \iff x = y$$

For (iii) we have,

$$d_p(x,y) = |x-y|_p = |(-1)(y-x)|_p = |-1|_p|y-x|_p$$

$$= p^{-v_p(-1)}|y-x|_p = p^0|y-x|_p = |y-x|_p = d_p(y,x)$$

Finally, to show the triangle inequality (iv) we have,

$$d_p(x,y) = |x-y|_p = p^{-v_p(x-y)} \text{ and } d_p(y,z) = |y-z|_p = p^{-v_p(y-z)}$$

Let $m = \min\{v_p(x-y), v_p(y-z)\}$. Then,

$$p^m \mid (x-y) \text{ and } p^m \mid (y-z)$$

Therefore, $p^m \mid (x-z)$ as $x - z = (x-y) + (y-z)$. Thus,

$$v_p(x,z) \geq m = \min\{v_p(x-y), v_p(y-z)\}$$

Therefore,

$$d_p(x,z) = |x-z|_p = p^{-v_p(x-z)} \leq p^m = p^{\min\{v_p(x-y),v_p(y-z)\}}$$

$$= \max\{p^{-v_p(x-y)}, p^{-v_p(y-z)}\} = \max\{|x-y|_p, |y-z|_p\}$$

$$= \max\{d_p(x,y), d_p(y,z)\} \leq d_p(x,y) + d_p(y,z)$$

So we have $d_p(x,z) \leq d_p(x,y) + d_p(y,z)$ giving us (iv). Thus d_p satisfies all metric axioms and is a metric. □

At this point we can begin to analyze the properties of these different absolute values. However in order to do so we need to make sure that there is indeed some sense in which they are fundamentally different. Furthermore, we would like to have a notion of when two absolute values are similar enough to be deemed equivalent. To do this we will invoke the topology induced by a certain absolute value. But first let us introduce the concept of a p-adic open ball.

Definition 5. Let p be a prime, $r \in \mathbb{R}_{\geq 0}$ and $q \in \mathbb{Q}$. Then **the p-adic open ball** with radius r and center q is

$$B_p(q,r) = \{x | x \in \mathbb{Q}, d_p(x,q) < r\}$$

Using the p-adic open ball we can define a topology for a given p-adic absolute value and set up a notion of equivalence between absolute values.

Definition 6. Let p be a prime. Then the **p-adic topology** is the topology whose basis consists of open balls with respect to d_p. In other words, the basis

$\mathcal{B}$ of the p-adic topology is:

$$\mathcal{B} = \{B_p(q, r) | r \in \mathbb{R}_{\geq 0},\ q \in \mathbb{Q}\}.$$

Furthermore, we will say that two absolute values are **equivalent** if they induce the same topology.

A natural question to ask at this point is why did we choose this particular construction of absolute value to study. Presumably there are many others to choose from. In fact, this turns out to be false and leads us to Ostrowski's theorem.

Theorem 1 (Ostrowski 1916)**.** *Every non-trivial absolute value on $\mathbb{Q}$ is equivalent to either the infinite absolute value or a p-adic absolute value.*

Proof. For this proof we will only consider the non-Archimedean case. The proof that an Archimedean absolute value must be the infinite absolute value is not particularly difficult to understand but it is long and not very enlightening. For a complete proof see page 46 of Gouvea [1].

Let $|\ |$ be a non-trivial non-Archimedean absolute value on $\mathbb{Q}$. Then $|n| \leq 1$ for all integers n as we have shown. Furthermore, $|\ |$ is non-trivial so there exists a $n \in \mathbb{N}$ such that $|n| < 1$. Let $\{p_i\}_{i \in \lambda}$ be the set of prime factors of n. By property (ii) of the definition of absolute value,

$$|\prod_{i \in \lambda} p_i| = |n| < 1$$

This implies that there exists $p \in \{p_i\}_{i \in \lambda}$ such that $|p| < 1$. We will show that in fact only one prime number can have this property. Assume for a contradiction that p_1 and p_2 are distinct primes such that $|p_1| < 1$ and $|p_2| < 1$.

let $m \in \mathbb{N}$ be large enough so that $|p_1|^m < \frac{1}{2}$ and $|p_2|^m < \frac{1}{2}$. Now we rely on the famous Bezout identity from number theory. Because $\gcd(p_1, p_2) = 1$, there exist integers a_1 and a_2 such that $a_1 p_1^m + a_2 p_2^m = 1$. So we have,

$$1 = |1| = |a_1 p_1^m + a_2 p_2^m| \leq |a_1 p_1^m| + |a_2 p_2^m|$$

by property (iii) of absolute values. Furthermore,

$$|a_1 p_1^m| + |a_2 p_2^m| = |a_1||p_1|^m + |a_2||p_2|^m < \frac{1}{2}(|a_1| + |a_2|)$$

Recall that $a_1, a_2 \in \mathbb{N}$ so $\frac{1}{2}(|a_1| + |a_2|) \leq 1$. This is a contradiction as

$$1 < \frac{1}{2}(|a_1| + |a_2|) \leq 1$$

Therefore there exists a unique prime p such that $|p| < 1$. Accordingly, for all other primes q, $|q| = 1$ as we have established that $|n| \leq 1$ for all integers n. Let $r = -\log_p(|p|)$. It follows then,

$$|p|_p^r = (p^{-v_p(p)})^r = (p^{-1})^r = p^{-r} = |p|$$

For all other primes q we have,

$$|q|_p^r = (p^{-v_p(q)})^r = (p^0)^r = 1^r = 1 = |q|$$

So $|\ |$ and $|\ |_p$ have the same values for all primes. Finally we will show that an absolute value on $\mathbb{Q}$ is completely determined by its values on the primes. Let $s \in \mathbb{N}$. Let $\mathbb{P}$ be the set of all primes. Then s is equal to the product of

all $p \in \mathbb{P}$ with each individual prime p raised to the power of $v_p(s)$. Therefore,

$$|s| = |\prod_{p\in\mathbb{P}} p^{v_p(s)}| = \prod_{p\in\mathbb{P}} |p|^{v_p(s)}$$

Thus the absolute value of positive integers is determined by the absolute value of the primes. We can extend this to the negative integers by noticing that,

$$|s| = |1||s| = |-1||s| = |-s|$$

Our last step is to extend to the rationals. We first notice that for all $k \in \mathbb{Z}$, $|1| = 1 = kk^{-1} = |kk^{-1}| = |k||k^{-1}|$. Therefore,

$$|k| = \frac{1}{|k^{-1}|} = |k^{-1}|^{-1} \Rightarrow |k^{-1}| = |k|^{-1}$$

Now let $t \in \mathbb{Q}$. There must exist two coprime positive integers x, y such that $t = xy^{-1}$. Hence,

$$|t| = |x||y|^{-1}$$

Therefore the absolute value of a rational is entirely determined by the absolute value of its numerator and denominator. The numerator and denominator are integers and thus their absolute value is entirely determined by the absolute values of the primes. The non-Archimedean absolute value $|\ |$ gives the same value for all primes as $|\ |_p$ for p a prime. Thus every non-trivial non-Archimedean absolute value on $\mathbb{Q}$ is equivalent to a p-adic absolute value. $\square$

Despite our abundance of p-adic objects which we have so far defined, we have not actually constructed the p-adic numbers yet. Using our metric, we might think to complete the space analogously to how the real numbers complete the space $\mathbb{Q}$ with respect to the Euclidean metric. Such a construction

might look something like this:

$$\mathbb{Q}_p = \{(x_n) | (x_n) \text{ is a Cauchy sequence with respect to } |\ |_p\}.$$

This definition almost works but critically it lacks a certain crucial property: it allows for multiple zero divisors. That is, there are multiple elements of a set, call them a and b, such that for any element of the set x, $ax = 0$ and $bx = 0$. For an example, let $p = 5$ and consider the sequences $a = (5^n)$ and $b = (10^n)$. According to the 5-adic absolute value, the terms in both a and b get arbitrarily close to 0. Thus their limit is 0 and algebraically we can treat them the same way we would normally treat the number 0. The reason this is a problem is that it means that for any x, we have $ax = 0 = bx$ which implies that $a = b$ or $(5^n) = (10^n)$ which will cause some problems if we accept it as is. In fancier term this contradicts the definition of an *integral domain* which we want the p-adic numbers to be. There is luckily an easy fix to this. We can associate all sequences who have the same limit together and divide out by this association. In other words we acknowledge that there exist sequences whose elements are the same and treat all of them as a single element. This will allow us to save the cancellation property that $ax = bx$ implies that $a = b$. In fact, this definition will give us several nice properties that will finally allow us to begin exploring the p-adic numbers.

Definition 7. Let $\mathbb{Q}'$ be the completion of the rationals with respect to the p-adic metric $d_p(x, y) = |x - y|_p$. More formally,

$$\mathbb{Q}' = \{(x_n) | (x_n) \text{ is a Cauchy sequence with respect to } |\ |_p\}.$$

Furthermore, let $\sim$ be a relation such that for any sequences $(a_n), (b_n) \in \mathbb{Q}'$,

$$(a_n) \sim (b_n) \text{ if } \lim_{n\to\infty}(a_n) = \lim_{n\to\infty}(b_n).$$

Then the **field of p-adic numbers**, $\mathbb{Q}_p$, is $\mathbb{Q}'$ with elements associated according to $\sim$. That is,

$$\mathbb{Q}_p = \mathbb{Q}'_{/\sim}.$$

This definition has 2 peculiarities that are begging to be proved. First, that $\sim$ is an *equivalence* relation and second that $\mathbb{Q}_p$ is indeed a *field*. In the following propositions we will indeed prove these assertions.

Proposition 7. *The relation $\sim$ on $\mathbb{Q}'$ where*

$$\mathbb{Q}' = \{(x_n) | (x_n) \textit{ is a Cauchy sequence with respect to } |\ |_p\}$$

and $\sim$ is defined by,

$$(a_n) \sim (b_n) \textit{ if } \lim_{n\to\infty}(a_n) = \lim_{n\to\infty}(b_n)$$

for any

$$(a_n), (b_n) \in \mathbb{Q}',$$

is an equivalence relation.

Proof. To show that $\sim$ is an equivalence relation we must show that for $(a_n), (b_n), (c_n) \in \mathbb{Q}'$,

(1) $(a_n) \sim (a_n)$ (reflexivity),

(2) $(a_n) \sim (b_n)$ if and only if $(b_n) \sim (a_n)$ (symmetry), and

(3) $(a_n) \sim (b_n)$ and $(b_n) \sim (c_n)$ implies that $(a_n) \sim (c_n)$ (transitivity). All of these properties come directly from the fact that we used the equivalence relation $=$ to define $\sim$. In more detail, we clearly have that (1) is true as $\lim_{n\to\infty}(a_n) = \lim_{n\to\infty}(a_n)$. We can also easily see that (2) is true as $\lim_{n\to\infty}(a_n) = \lim_{n\to\infty}(b_n)$ implies that $\lim_{n\to\infty}(b_n) = \lim_{n\to\infty}(a_n)$ and vice versa. Finally for (3) we have,

$$\lim_{n\to\infty}(a_n) = \lim_{n\to\infty}(b_n) = \lim_{n\to\infty}(c_n)$$

implies that $\lim_{n\to\infty}(a_n) = \lim_{n\to\infty}(c_n)$.

□

Proposition 8. $\mathbb{Q}_p$ *is a field*

Proof. Because we used $\mathbb{Q}$ to define $\mathbb{Q}_p$ we can see that it inherits associativity and commutativity from $\mathbb{Q}$. Also it is clear to see that $1 \neq 0$ in $\mathbb{Q}_p$. So $\mathbb{Q}_p$ is a commutative ring with $1 \neq 0$. All that is left to show is that every non-zero element has a multiplicative inverse. Let $x \in \mathbb{Q}_p$ such that $x \neq 0$. Then we can select a sequence $(x_n) \subset \mathbb{Q}$ such that for some $\epsilon > 0$, $x_n \neq 0$ and $|x_n|_p > \epsilon$ for all n. Let $x^{-1} = (x_n^{-1})$. This sequence is indeed Cauchy as for $n, m > N = \frac{1}{\epsilon}$,

$$d_p(\frac{1}{x_n}, \frac{1}{x_m}) = |\frac{1}{x_n} - \frac{1}{x_m}|_p = |\frac{x_n - x_m}{x_n x_m}|_p < N|x_n - x_m|_p = \frac{|x_n - x_m|_p}{\epsilon} < \epsilon.$$

Finally we have

$$(x_n x_n^{-1}) \sim (1) \Rightarrow x x^{-1} = 1$$

which gives us inverses proving that $\mathbb{Q}_p$ is a field. □

Because $\mathbb{Q}_p$ is a *completion* of $\mathbb{Q}$ it also contains $\mathbb{Q}$. This implies that all rational numbers are p-adic numbers for any prime p. However this is not to say that every p-adic number is a rational. Let us consider an example.

Example 5. The sequence

$$(a_n) = ((\sum_{i=0}^{n} 3^i)_n)$$

clearly diverges with respect to the infinite absolute value and thus is not a rational number. In $\mathbb{Q}_3$ however we can use the fact that $|3^i|_3 = 3^{-i}$ to show that this series is indeed Cauchy.

Let $\epsilon > 0$ and let $N \in \mathbb{N}$ such that $\frac{2}{3^N} < \epsilon$. Then for $m > n \geq N$,

$$d_3(a_m, a_n) = |a_m - a_n|_3 = |\sum_{i=0}^{m} 3^i - \sum_{i=0}^{n} 3^i|_3 = |\sum_{i=n+1}^{m} 3^i|_3$$

Using property (iii) of absolute values we have,

$$|\sum_{i=n+1}^{m} 3^i|_3 \leq \sum_{i=n+1}^{m} (|3^i|_3) = \sum_{i=n+1}^{m} 3^{-i} = \frac{1}{3^{n+1}} \sum_{i=0}^{m} 3^{-i}$$

The sum of a finite geometric series is given by

$$\sum_{i=0}^{b} ar^i = \frac{a(1-r^b)}{1-r}$$

In this case, we have $a = 1, r = \frac{1}{3}$ and $b = m$. Therefore,

$$\frac{1}{3^{n+1}} \sum_{i=0}^{m} 3^{-i} = (\frac{1}{3^{n+1}})(\frac{1-(\frac{1}{3})^m}{1-\frac{1}{3}}) = \frac{2(1-3^{-m})}{3^n} = 2(\frac{1}{3^n} - \frac{1}{3^{n+m}})$$

$$< \frac{2}{3^n} \leq \frac{2}{3^N} < \epsilon \Rightarrow d_3(a_m, a_n) < \epsilon$$

Therefore this series is Cauchy and converges to some limit L in regards to the 3-adic metric. Accordingly,

$$L = \sum_{i=0}^{\infty} 3^i \in \mathbb{Q}_3$$

Counter-intuitive is an understatement here. If we choose to represent L in base 3 we may even write,

$$L = \sum_{i=0}^{\infty} 3^i = 1 + 10 + 100 + \cdots = \cdots 111$$

Strange indeed.

3. Exploring the p-adics

Now that we have constructed $\mathbb{Q}_p$, we can begin to explore some of its properties. Let us start with topology. To begin, one of the more counter intuitive properties of the topology of $\mathbb{Q}_p$ is that any point in an open ball is a center of the open ball.

Proposition 9. *Let $x, y \in \mathbb{Q}_p$ and let $r \in \mathbb{R}_{\geq 0}$. If $y \in B_p(x, r)$, then $B_p(x, r) = B_p(y, r)$.*

Proof. Let $x, y \in \mathbb{Q}_p$ and $r \in \mathbb{R}_{\geq 0}$ such that $y \in B_p(x, r)$. Let $z \in B_p(y, r)$. By definition then, $|y - z|_p < r$ and $|x - y|_p < r$. By the non-Archimedean property of the p-adic absolute value,

$$|x - z|_p \leq \max\{|y - z|_p, |x - y|_p\} < r$$

and therefore, $z \in B_p(x, r)$. Thus $B_p(y, r) \subseteq B_p(x, r)$. Similarly, for all $w \in B_p(x, r)$, we have $|x - w|_p < r$ by definition and $|y - x| < r$ by $y \in B_p(x, r)$.

Again making use of the non-Archimedean property of the p-adic absolute value we have,

$$|y - w| \leq \max\{|x - w|_p, |y - x|_p\} < r.$$

Therefore $w \in B_p(y, r)$ and $B_p(x, r) \subseteq B_p(y, r)$ which implies that $B_p(y, r) = B_p(x, r)$. □

This proposition gives us another fact about $\mathbb{Q}_p$: every open set is clopen.

Proposition 10. *Let $x \in \mathbb{Q}_p$ and $r \in \mathbb{R}_{\geq 0}$. Then $B_p(x, r)$ is clopen.*

Proof. Let $B = B_p(x, r)$ for $x \in \mathbb{Q}_p$ and $r \in \mathbb{R}_{\geq 0}$. We have by definition that B is open so it remains to show that B is closed. Let $cl(B)$ denote the closure of B. We will show that $cl(B) = B$. Let $y \in cl(B)$. Then for all $s \in \mathbb{R}_{\geq 0}$, $B_p(y, s) \cap B_p(x, r) \neq \emptyset$ by definition of closure. For $s = r$ then we have that there exists $z \in B_p(y, r) \cap B_p(x, r)$. So $z \in B_p(y, r)$ and $z \in B_p(x, r)$. By proposition 9 then,

$$B_p(y, r) = B_p(z, r) = B_p(x, r).$$

So $y \in B_p(y, r) = B_p(x, r)$ for all $y \in cl(B_p(x, r))$ implies that $cl(B_p(x, r)) \subseteq B_p(x, r)$. By definition of closure $B_p(x, r) \subseteq cl(B_p(x, r))$ and therefore $cl(B) = B$ as desired. □

These properties altogether describe the totally disconnected nature of $\mathbb{Q}_p$.

Proposition 11. *$\mathbb{Q}_p$ is a totally disconnected, space.*

Proof. Let $A \subseteq \mathbb{Q}_p$ such that $x, y \in A$ with $x \neq y$. Let $r = |x - y|_p$ and notice that $y \notin B_p(x, r) = \{z \in \mathbb{Q}_p \mid |x - z|_p < |x - y|_p\}$. By proposition 10, $B_p(x, r)$ is clopen. By definition we have that $\mathbb{Q}_p - B_p(x, r)$ is clopen. Let

$B = A \cap B_p(x, r)$ and $C = A \cap \{\mathbb{Q}_p - B_p(x, r)\}$ and notice that these are both open. Furthermore notice that $x \in B$, $y \in C$, and $A = B \cup C$. Then B and C are a separation of A and A is disconnected by definition. Therefore any subset with at least two elements in $\mathbb{Q}_p$ is disconnected which implies that the only connected sets in $\mathbb{Q}_p$ are the singleton sets. Thus $\mathbb{Q}_p$ is totally disconnected by definition. □

Furthermore, we can see that $\mathbb{Q}_p$ is also Hausdorff. Of course we have already shown that $\mathbb{Q}_p$ is a metric space which implies the Hausdorff property however if one is not familiar with the proof that all metric spaces are Hausdorff, we have provided one for the case of $\mathbb{Q}_p$ here.

Proposition 12. *$\mathbb{Q}_p$ is a Hausdorff space.*

Proof. Let $x, y \in \mathbb{Q}_p$ such that $x \neq y$ and let $r = |x - y|_p$. We will show that $B_p(x, r)$ and $B_p(y, r)$ are disjoint. We have that

$$B_p(x, r) = \{z \in \mathbb{Q}_p \mid |x - z| < r\}$$

and

$$B_p(y, r) = \{z \in \mathbb{Q}_p \mid |y - z| < r\}.$$

Therefore,

$$B_p(x, r) \cap B_p(y, r) = \{z \in \mathbb{Q}_p \mid |x - z| \text{ and } |y - z| < r\}.$$

By the non-Archimedean property of the p-adic absolute value we have that for all $z \in \mathbb{Q}_p$,

$$\max\{|x - z|_p, |y - z|_p\} \geq |x - y|_p = r.$$

Therefore, $|x-z|_p \geq r$ or $|y-z|_p \geq r$ for all $z \in \mathbb{Q}_p$. Thus,

$$B_p(x,r) \cap B_p(y,r) = \{z \in \mathbb{Q}_p \mid |x-z| \text{ and } |y-z| < r\} = \emptyset$$

and $\mathbb{Q}_p$ is Hausdorff. □

Clearly the topology on $\mathbb{Q}_p$ has some strange if not intriguing quirks. Sadly, we now move on from topology. In considering $\mathbb{Q}_p$ we sometimes pay special attention to the elements whose absolute value is less or equal to 1. These elements have positive valuations and we consider them to be the p-adic integers. Because they are all contained in a closed ball of radius 1, they are easy to define and study.

Definition 8. The ring of **p-adic integers**, or the valuation ring, denoted $\mathbb{Z}_p$, is defined as,

$$\mathbb{Z}_p = \{x \in \mathbb{Q}_p | \ |x|_p \leq 1\}.$$

The elements of $\mathbb{Z}_p$ will be those whose valuations is positive. Let us consider examples of elements in $\mathbb{Z}_p$ and outside of it.

Example 6. Let p=7 and $x = \frac{2}{7}$. Then,

$$|\frac{2}{7}|_7 = 7^{-v_7(\frac{2}{7})}.$$

Computing $v_7(\frac{2}{7})$ we get,

$$v_7(\frac{2}{7}) = v_7(2) - v_7(7) = -1.$$

Thus,

$$|\frac{2}{7}|_7 = 7^{-(-1)} = 7 > 1.$$

Therefore $\frac{2}{7} \notin \mathbb{Z}_7$.

Example 7. Let $p = 7$ again but this time let $x = \frac{7}{2}$. Computing the valuation we get,

$$v_7(\frac{7}{2}) = v_7(7) - v_7(2) = 1.$$

Thus,

$$|\frac{7}{2}|_7 = 7^{-(1)} = 1/7 \leq 1.$$

Therefore $\frac{7}{2} \in \mathbb{Z}_7$.

Equipped with a new way to talk about p-adic numbers we might want to ask what the relationship between $\mathbb{Q}_p$ and $\mathbb{Z}_p$ is. In particular is it true that every element in $\mathbb{Q}_p$ can be mapped to the smaller set $\mathbb{Z}_p$? This does turn out to be true and we can do it by multiplying an element of $\mathbb{Q}_p$ by some power of p in order to decrease its size.

Lemma 1. *For all $x \in \mathbb{Q}_p$, there exists an $n \in \mathbb{N} \cup \{0\}$ such that $p^n x \in \mathbb{Z}_p$.*

Proof. Let $x \in \mathbb{Q}_p$. If $|x|_p \leq 1$ then $x \in \mathbb{Z}_p$ by definition so assume that $|x|_p > 1$. Notice that $|x|_p > 1$ if and only if $v_p(x) < 0$ as,

$$|x|_p = p^{-v_p(x)} > 1$$

by definition. Then we have,

$$\log_p(p^{-v_p(x)}) = -v_p(x) > \log_p(1) = 0.$$

Thus $-v_p(x) > 0$ and $v_p(x) < 0$. So it suffices to show that there exists an $n \in \mathbb{N} \cup \{0\}$ such that $v_p(p^n x) \geq 0$. Let $n = -\lfloor v_p(x) \rfloor$. Notice that this is a

positive because $v_p(x)$ is negative. By Proposition 4 §2,

$$v_p(p^n x) = v_p(p^n) + v_p(x) = n + v_p(x) = v_p(x) - \lceil v_p(x) \rceil \geq 0$$

as desired. □

Having acquainted ourselves with $\mathbb{Q}_p$ and $\mathbb{Z}_p$ a natural next step would be to consider the field of polynomials with p-adic coefficients. This field we denote $\mathbb{Q}_p[x]$. If we wish to consider only coefficients that are p-adic integers we write $\mathbb{Z}_p[x]$. These fields have several interesting properties. To begin, one of the first questions we must consider when discussing polynomials in any field is if derivatives behave in the way we expect them to. Recall that in $\mathbb{R}[x]$, we can try Newton's method to find a root of any given polynomial. In many cases this tells us whether a given polynomial has roots in $\mathbb{R}[x]$. One must ask if the same is true for the p-adics. In his seminal paper introducing the p-adics, Kurt Hensel considered this [3]. With some slight adjustments, it turns out that it does. In $\mathbb{R}$, we can often determine whether a polynomial will have roots by looking at signs. For example, it is clear to see that $x^2 + 1$ has no roots. In the world of p-adics, we replace sign considerations with reduction modulo p.

Before diving into the proof however we must be careful in introducing derivatives. By only considering polynomials we can forgo the analysis required to construct a general theory of differentiation in $\mathbb{Q}_p$. Instead, we can simply define them algebraically.

Definition 9. Let

$$f(x) = a_0 + a_1x + a_2x^2 + ... + a_nx^n = \sum_{i=0}^{n} a_ix^i \in \mathbb{Z}_p[x].$$

Then the **derivative** of $f(x)$ denoted $f'(x)$, is given by,

$$f'(x) = a_1 + 2a_2x + ... + na_nx^{n-1} = \sum_{i=0}^{n} ia_ix^{i-1}.$$

This leads us right to Hensel's lemma. It should be noted that there are many logically equivalent formulations of this theorem that go by the same name. Here we present one such formulation.

Theorem 2 (Hensel's Lemma)**.** *Let $f(x) \in \mathbb{Z}_p[x]$ and $a \in \mathbb{Z}_p$. Suppose $|f(a)|_p < |f'(a)|_p^2$. Then there exists a unique $z \in \mathbb{Z}_p$ in the open ball around a with radius $\frac{|f(a)|_p}{|f'(a)|_p}$ such that $f(z) = 0$.*

In order to prove Hensel's lemma we rely on the Contraction Mapping theorem from analysis. For a proof see *Conrad*'s paper [4].

Theorem 3 (Contraction Mapping Theorem)**.** *Let (X, d) be a complete metric space and $f : X \to X$ be a contraction mapping, i.e., a map such that $d(f(x), f(y)) \leq c \cdot d(x, y)$ for some $0 \leq c < 1$ and any $x, y \in X$. Then f has a unique fixed point in X.*

Proof. (Hensel's lemma) Let $f(x) = c_ix^i \in \mathbb{Z}_p[x]$ and $a \in \mathbb{Z}_p$ such that $|f(a)|_p < |f'(a)|_p^2$. We will use the function

$$g(x) = x - \frac{f(x)}{f'(a)}$$

iteratively to construct a Cauchy sequence which converges to z. First however we must show that this is a contraction mapping of an open ball around a with radius less than 1. Let $r = |\frac{f(a)}{f'(a)}|_p$. We will show that $0 \leq r < 1$. By definition of absolute value, $r \geq 0$. By assumption,

$$r = \frac{|f(a)|_p}{|f'(a)|_p} < |f'(a)|_p$$

so it suffices to show that $|f'(a)|_p \leq 1$. By definition, $|f'(a)|_p = p^{-v_p(f'(a))}$ so it suffices to show that $v_p(f'(a)) \geq 0$. By definition,

$$v_p(f'(a)) = v_p(ic_i a^{i-1}).$$

Using Proposition 4 §2,

$$v_p(ic_i a^{i-1}) \geq \min\{v_p(ic_i a^{i-1}) \mid 0 \leq i \leq n\} = \min\{v_p(i)+v_p(c_i)+v_p(a)(i-1) \mid 0 \leq i \leq n\}.$$

Notice that if $i = 0$, then $v_p(i) = +\infty$ so

$$\min\{v_p(i) + v_p(c_i) + v_p(a)(i-1) \mid 1 \leq i \leq n\} \leq v_p(0) + v_p(c_0) - v_p(a) = +\infty$$

so we need only consider $i \in [1, n]$ for the minimum. Furthermore, notice that $c_i, a \in \mathbb{Z}_p$ implies that $v_p(c_i), v_p(a) \geq 0$ and that $i \in \mathbb{Z}$ implies that $v_p(i) \geq 0$ as well. Then,

$$\min\{v_p(i) + v_p(c_i) + v_p(a)(i-1) \mid 1 \leq i \leq n\} \geq 0.$$

Therefore, $r = \frac{|f(a)|_p}{|f'(a)|_p} < 1$. This gives us,

$$B_p(a, r) = \{x \in \mathbb{Q}_p \ : \ |x - a|_p < r\} \subseteq \mathbb{Z}_p$$

as $r < 1$. Let $c = |\frac{f(a)}{f'(a)^2}|_p$ and notice that $0 \leq c < 1$ by the assumption that $|f(a)|_p < |f'(a)|_p^2$. Then g maps $B_p(a, r)$ onto itself and forms a contraction. That is, for all $x, y \in B_p(a, r)$,

$$|g(x) - g(y)|_p = |x - y - \frac{f(x) - f(y)}{f'(a)}|_p < c|x - y|_p.$$

Consider the Cauchy sequence (a_n) where $a_1 = 0$ and $a_{i+1} = g(a_i)$ for all $i > 1$. As shown in section 2, $\mathbb{Q}_p$ is a complete metric space so by the Contraction mapping theorem g has a unique fixed point z which must be the limit of (a_n). That is, $g(z) = z$. Because $g(z) = z - \frac{f(z)}{f'(a)} = z$, we have that $f(z) = 0$. $\square$

After all that theory, we may lose sight of what Hensel's Lemma is actually doing. In order to ground ourselves, let us consider some examples of what using Hensel's lemma looks like.

Example 8. Let $p = 7$ and $f(x) = x^2 - 15$. Say we wanted to find a root of this polynomial in $\mathbb{Z}_7$. Taking a random guess, let us plug in $a = 1$. Then we have $|f(1)|_7 = |-14|_7 = 1/7$ and $|f'(1)|_7^2 = |2|_7^2 = 1$ so Hensel's Lemma applies in this case as $|f(a)|_p = 1/7 < 1 = |f'(a)|_p^2$. Therefore this function has a unique root which is congruent to 1 (mod 7). In other words, $\sqrt{15} \in \mathbb{Z}_5$. We can find the base 7 expression of $\sqrt{15}$ through the following method:

$$15 \equiv 1^2 \pmod 7$$

$$15 \equiv (1 + 7)^2 \pmod{7^2}$$

$$15 \equiv (1 + 7 + 3(7^2))^2 \pmod{7^3} \cdots$$

So $\sqrt{15} = 1 + 7 + 3(7^2) + \cdots$ in $\mathbb{Z}_p$.

Example 9. The previous example is actually a special case of a much larger class of solutions that we can obtain through Hensel's Lemma. Let p be a prime, and $n, m \in \mathbb{Z}$ such that $n > 0$, $p \nmid n$, and $m \equiv 1 \pmod{p}$. Let $f(x) = x^n - m \in \mathbb{Z}_p[x]$. Then for $a = 1$ again we have,

$$|f(a)|_p = |1 - m|_p < 1$$

by $m \equiv 1 \pmod{p}$. Furthermore,

$$|f'(a)|_p^2 = |n(1)^{n-1}|_p^2 = |n|_p^2 = 1$$

by $p \nmid n$. So $|f(a)|_p < 1 = |f'(a)|_p^2$ and Hensel's lemma applies in this case. What this tells us is that m is an n^{th} power in $\mathbb{Z}_p$. That is, there exists a unique $z \in \mathbb{Z}_p$ such that $z^n = m$ and $z \equiv 1 \pmod{p}$. Plugging in $p = 7$, $m = 15$, and $n = 2$, we get the result from example 8.

Hensel's lemma can be a very powerful problem solving tool for all polynomials in $\mathbb{Z}_p[x]$. However, we need not always consider the entirety of $\mathbb{Z}_p[x]$. Some polynomials have special properties that are particularily "nice". By narrowing our focus to a certain family of polynomials we can actually learn a lot about the field as a whole. For example we have the Eisenstein polynomials whose "nice" property is that they are easily identifiable as irreducible.

Definition 10. Let $f(x) = \sum_{i=0}^{n} a_i x^i \in \mathbb{Z}_p[x]$ such that,

(1) $|a_n|_p = 1$

(2) $|a_i|_p \leq 1$ for $0 \leq i < n$

(3) $|a_0|_p = 1/p$.

Then f is **Eisenstein**. If we want to say that f satisfies this criterion for a specific prime p we say that f is **p-Eisenstein**.

It can sometimes be helpful to write this definition in terms of p-adic valuations instead of absolute values. This gives us the following three conditions:

(1) $v_p(a_n) = 0$,

(2) $v_p(a_i) \geq 1$ for $0 \leq i < n$,

(3) $v_p(a_0) = 1$.

Of course we do not introduce these polynomials randomly. They have one property which will be very useful for us in the future.

Proposition 13. *Eisenstein polynomials are irreducible over $\mathbb{Q}_p$.*

A proof of this proposition by the methods available to us thus far exists however it will be much easier to prove using the methods of the next chapter so we wait until then.

4. Newton polygons

We now turn our attention to the study of Newton polygons. Newton polygons allow us to extract from a polynomial (or potentially a power series) much of the information that we are interested in without having to get bogged down. This does not mean that the Newton polygons are visual representation for the sake of visualizing; they allow us to perform certain maneuvers much more easily then we would be able to otherwise. In order to gain intuition, we will first informally describe the process of constructing a Newton polygon and then move on to a formal definition. For a given polynomial

$a_0 + a_1x + a_2x^2 + \cdots + a_nx^n = f(x) \in \mathbb{Q}_p[x]$ with p-adic coefficients, we construct the Newton polygon of $f(x)$ through the following process:

(1) On the Euclidean plane we plot the points $(i, v_p(a_i))$ for all $i \in [0, n]$. The only hiccup with this is that if $a_i = 0$ then $v_p(a_i) = +\infty$. In effect this just means we ignore the point.

(2) Take the part of the y-axis below the first point, the set $\{(0, y)|y \leq v_p(a_0)\}$, and rotate it counterclockwise until it hits one or more of the points we have plotted.

(3) Let the right most point that was hit be P_1. "Break" the line at P_1 and fix the line connecting the first point $P_0 = (a_0, v_p(a_0))$ and P_1, that is the line $\overline{P_0, P_1}$.

(4) Continue rotating counter clockwise until another point P_2 is hit. At this point fix the line $\overline{P_1, P_2}$.

(5) Repeat step (4) until all points have either been hit or lie above a portion of the polygon.

This explanation is serviceable in most circumstances but in order to have a more formal definition we must first introduce the concept of a convex hull. Intuitively, given a set of point S in $\mathbb{R}^2$, we can think of the convex hull of S as the shape that we get by stretching a rubber band around the points of S and letting it settle.

Definition 11. A set $S \subseteq \mathbb{R}^2$ is **convex** if for all points $x, y \in S$, the line $\overline{xy}$ connecting x and y is entirely contained in S. The **convex hull** of S, denoted $CH(S)$, is the intersection of all convex sets containing S. Let s_1 be the right most point of $CH(S)$ and s_2 the left most. Then the **lower convex hull** of S,

denoted $LCH(S)$, is the set of points in $CH(S)$ that are on or below the line $\overline{s_1 s_2}$ as well as all the points above the line $\overline{s_1 s_2}$ (not necessarily in $CH(S)$.

This concept of convexity allows us to write a formal definition for the process we outlined above.

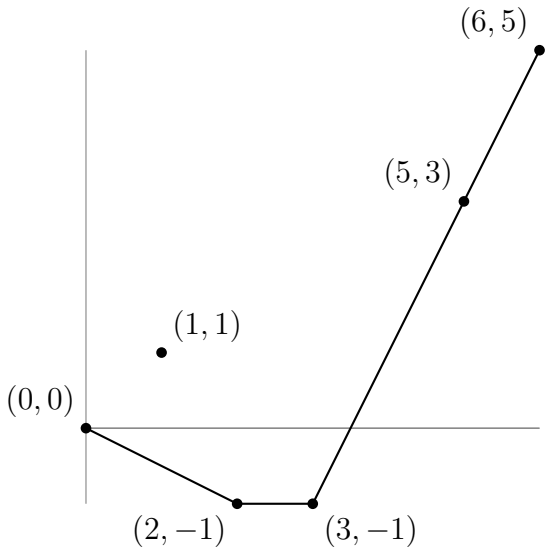
Definition 12. Let $a_0 + a_1 x + a_2 x^2 + \cdots + a_n x^n = f(x) \in \mathbb{Q}_p[x]$. Then the **Newton polygon** of $f(x)$, denoted $NP(f)$, is the lower convex hull of the set $\{(i, v_p(a_i)) | 0 \leq i \leq n\}$.

Let us consider an example to see what this looks like in practice.

Example 10. Let $p = 3$ and let $f(x) = 1 + 3x + \frac{1}{3}x^2 + \frac{1}{21}x^3 + 27x^5 + 486x^6$. Then the points we want to work with are:

$$\{i, v_3(a_i) | 0 \leq i \leq 6\} = \{(0,0), (1,1), (2,-1), (3,-1), (4,+\infty), (5,3), (6,5)\}.$$

Here we denote the coefficients of the i^{th} term of $f(x)$ as a_i so that $a_0 = 1, a_1 = 3, a_2 = \frac{1}{3}$ etc. Constructing the lower convex hull we ignore the point $(4, +\infty)$ as the lower convex hull cannot include it by definition. If this feels uncomfortable, think of the point as being "infinitely high" in the y direction. Then our Newton Polygon of $f(x)$ will look like this:



Equipped with an example we can begin to explore the information encoded in the graph. In this case, the slopes of the edges of the polygons are $-2, 0$ and 2 with lengths $2, 1$ and 3 respectively. Counter intuitively, by "length" here we mean the length of the segment projected onto the x-axis, not the length of the segment itself. Lastly, the breaks happen at $i = 0, 2, 3, 6$. These three features, (slope, length, and vertices) will help us reveal certain facts about the roots of $f(x)$.

A few things are worth pointing out at this point. Firstly, the sum of the lengths will always equal the degree of the f. Furthermore, $(a_n, v_p(a_n))$ will always be a vertex. Next notice that the slopes of the segments will always form an increasing sequence. This is clear from the rotating line construction. This leads us to our first theorem, and indeed the most famous theorem, about Newton polygons.

Theorem 4. *Let $f \in \mathbb{Q}_p[x]$ and let $NP(f)$ denote the Newton polygon of f. Then,*

(1) If $NP_p(f)$ has t segments of lengths $\{l_i \mid 1 \leq i \leq t\}$, then there exist $\{f_i \mid 1 \leq i \leq t\} \subset \mathbb{Q}_p[x]$ with $\deg(f_i) = l_i$ such that $f = \prod_{i=1}^{t} f_i$.

(2) If $NP_p(f)$ has a segment of length l and slope s, then f has exactly l roots (in $\overline{\mathbb{Q}}_p$) of valuation $-s$.

Here $\overline{\mathbb{Q}}_p$ denotes the *algebraic closure of* $\mathbb{Q}_p$. For a proof of the theorem see page 259 of Gouvea [1]

Some polynomials will produce a Newton polygon that consists of only one slope. These turn out to be quite useful so we have a special name for them: we say that they are *pure*.

Definition 13. Let $f(x) \in \mathbb{Q}_p[x]$. Then f is **pure of slope** m if $NP(f)$ has only one segment of slope m. If we do not wish to specify the slope, we simply say that f is **pure**.

Using the above theorem we can prove a very nice result about irreducability which uses Newton polygons in its proof.

Proposition 14. *Let $f(x) \in \mathbb{Z}_p[x]$ such that, $deg(f(x)) = n$, the valuation of its constant term is m, $\gcd(m, n) = 1$, and $f(x)$ is pure. Then f is irreducible.*

Proof. Let $f(x) \in \mathbb{Z}_p[x]$ such that, $deg(f(x)) = n$, the valuation of its constant term is m, $\gcd(m, n) = 1$, and $f(x)$ is pure. Suppose $f(x)$ is reducible in $\mathbb{Z}_p[x]$. Then there exist polynomials of degree at least 1 $g(x), h(x) \in \mathbb{Z}_p[x]$ such that

$$f(x) = g(x)h(x).$$

Let $A = \{\alpha_i | 1 \leq i \leq n\}$ be the roots of $f(x)$. Additionally, let
$B = \{\beta_i | 1 \leq i \leq b\}$ be the roots of $g(x)$. Notice that B is a proper subset of A

and thus $b < n$. By part 2 of Theorem 3, $v_p(\alpha_i) = \frac{m}{n}$ for all i and consequently $v_p(\beta_i) = \frac{m}{n}$. We have that,

$$g(x) = \prod_{i=1}^{b}(x - \beta_i) = x^b + \cdots + \prod_{i=1}^{b}\beta_i.$$

Because $g(x) \in \mathbb{Z}_p[x]$, the valuation of each of its coefficients must be an integer. Therefore,

$$v_p(\prod_{i=1}^{b}\beta_i) \in \mathbb{Z}.$$

Moreover,

$$v_p(\prod_{i=1}^{b}\beta_i) = \sum_{i=1}^{b} v_p(\beta_i) = \sum_{i=1}^{b}\frac{m}{n} = \frac{bm}{n} \in \mathbb{Z}.$$

However, we have that $\gcd(m, n) = 1$ and that $b < n$ so $\frac{bm}{n} \notin \mathbb{Z}$. Thus we have a contradiction and therefore $f(x)$ is irreducible. □

This result finally allows us to fulfill the promise of proving that Eisenstein polynomials are irreducible. However we must first prove a short lemma about the Newton polygons of Eisenstein polynomials.

Lemma 2. *Let $f(x) \in \mathbb{Z}_p[x]$ be Eisenstein with degree m. Then f is pure of slope $\frac{-1}{m}$.*

Proof. Because f is Eisenstein with degree m, $v_p(a_0) = 1$ and $v_p(a_m) = 0$. Then $NP(f)$ will have vertices at $(0, 1)$ and $(m, 0)$. By definition of Eisenstein, for all i except 0 or n, the vertex $(i, v_p(a_i))$ will lie above or on the line $y = 1$. Thus $NP(f)$ consists of a single segment with slope $\frac{-1}{m}$ and is therefore pure of slope $\frac{1}{m}$. □

This gives us everything we need.

Corollary 1. *Eisenstein polynomials are irreducible.*

Proof. Let $f(x) \in \mathbb{Z}_p[x]$ be Eisenstein. Let a_0 be the constant term of $f(x)$ and let $n = \deg(f)$. By definition then, $v_p(a_0) = 1$ and consequently $\gcd(1, n) = 1$. Furthermore, $f(x)$ is pure by Lemma 2. By proposition 12 then, $f(x)$ is irreducible. □

4.1. **Composition of Newton Polygons.** The next theorem is the product of our research.

Theorem 5. *Let $m, n \in \mathbb{Z}$ such that $m < n$. Let $f(x) \in \mathbb{Z}_p[x]$ have degree n and a Newton polygon with one segment of slope $-\frac{m}{n}$ which includes the point $(0, m)$ and let $g(x) \in \mathbb{Z}_p[x]$ have degree d and be Eisenstein. then the Newton polygon of $f(g(x))$ will have one segment of slope $-\frac{m}{nd}$ and contains the point $(0, m)$.*

To aid in the proof of this theorem we will first prove a few other results.

Proposition 15. *Let $f(x), g(x) \in \mathbb{Z}_p[x]$ be Eisenstein such that $NP(f)$ and $NP(g)$ have slope $-\frac{1}{m}$. Then $NP(fg)$ is pure of slope $-\frac{1}{m}$.*

Proof. Because of the slope of their Newton polygons we know that f and g both have degree m. Let

$$f(x) = \sum_{i=0}^{m} a_i x^i \text{ and } \ g(x) = \sum_{i=0}^{m} b_i x^i \text{ and } \ f(x)g(x) = \sum_{i=0}^{2m} c_i x^i.$$

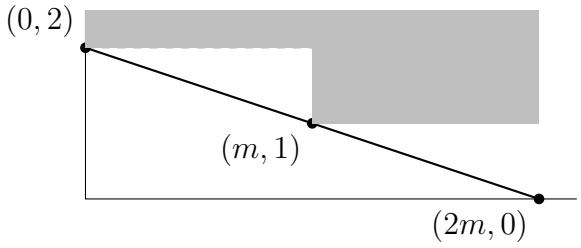
It is clear that, $c_0 = a_0 b_0$ and $c_{2m} = a_m b_m$. By proposition 4 § 2, we have,

$$v_p(c_0) = v_p(a_0 b_0) = v_p(a_0) + v_p(a_0) = 1 + 1 = 2$$

and,

$$v_p(c_{2m}) = v_p(a_m) + v_p(b_m) = 0 + 0 = 0.$$

Next we must show that all vertices of $NP(fg)$ lie above or on the line from $(0, 2)$ to $(2m, 0)$. In other words, we must show that For any n such that $0 < n < 2m$, $v_p(c_n) \geq 2 - \frac{n}{m}$. It will suffice to show that for $0 < n < m$, $v_p(c_n) \geq 2$ and that for $m \leq n \leq 2m$, $v_p(c_n) \geq 1$. That is, all vertices of $NP(fg)$ except $(0, 2)$ and $(2m, 0)$ will lie in the gray region shaded in the diagram below.



(1) Consider c_n such that $0 < n < m$. We have that

$$f(x)g(x) = a_0b_0+(a_1b_0+a_0+b_1)x+(a_2b_0+a_1b_1+a_0b_2)x^2+\cdots+(a_0b_m+\cdots+a_mb_0)x^m$$

$$+(a_mb_1 + \cdots + a_1b_m)x^{m+1} + \cdots + (a_mb_{m-1} + a_{m-1}b_m)x^{2m-1} + (a_mb_m)x^{2m}$$

$$= \sum_{n=0}^{m-1}(x^n(\sum_{i=0}^{n} a_ib_{n-i})) + \sum_{n=m}^{2m}(x^n(\sum_{i=0}^{2m-n} a_{m-i}b_{i-m+n})).$$

Therefore,

$$c_n = \begin{cases} \sum_{i=0}^{n} a_ib_{n-i} & \text{if } n < m \\ \sum_{i=0}^{2m-n} a_{m-i}b_{i-m+n} & \text{if } n \geq m \end{cases}.$$

In this case we need only consider the terms where $n < m$ so we can simplify to

$$c_n = \sum_{i=0}^{n} a_i b_{n-i}.$$

Using Proposition 4 § 2 we have,

$$v_p(c_n) = v_p(\sum_{i=0}^{n} a_i b_{n-1}) \geq \min\{v_p(a_i b_{n-i}) | 0 \leq i \leq n\} = \min\{v_p(a_i) + v_p(b_{n-i}) | 0 \leq i \leq n\}.$$

Recall that $f(x)$ and $g(x)$ are Eisenstein with degree m so $v_p(a_i)$ and $v_p(b_i)$ is greater than 1 for all i less than m. By assumption $n < m$ in this case so,

$$v_p(c_n) \geq \min\{v_p(a_i) + v_p(b_{n-i}) | 0 \leq i \leq n\} \geq 2$$

as desired.

(2) Next consider c_n with $m \leq n < 2m$. From the previous case we have that

$$c_n = \begin{cases} \sum_{i=0}^{n} a_i b_{n-1} & \text{if } n < m \\ \sum_{i=0}^{2m-n} a_{m-i} b_{i-m+n} & \text{if } n \geq m \end{cases}.$$

Using the assumption that $m \leq n < 2m$ we can simplify to

$$c_n = \sum_{i=0}^{2m-n} a_{m-i} b_{i-m+n}.$$

Again by Proposition 4 § 2 we have,

$$v_p(c_n) = v_p(\sum_{i=0}^{2m-n} a_{m-i} b_{i-m+n}) \geq \min\{v_p(a_{m-i} b_{i-m+n}) | 0 \leq i \leq 2m-n\}$$

$$= \min\{v_p(a_{m-i}) + v_p(b_{i-m+n}) | 0 \leq i \leq 2m-n\}.$$

If $i > 0$ then $v_p(a_{m-i}) \geq 1$ by the fact that $f(x)$ is Eisenstein. Furthermore, for all i, $v_p(b_i) \geq 0$ by the fact that $g(x)$ is Eisenstein. Therefore,

$$\min\{v_p(a_{m-i}) + v_p(b_{i-m+n}) | 1 \leq i \leq 2m-n\} \geq \min\{1+0\} = 1.$$

In the case that $i = 0$ we have,

$$v_p(a_{m-i}) + v_p(b_{i-m+n}) = v_p(a_m) + v_p(b_{-m+n}) = v_p(b_{-m+n})$$

by the fact that $f(x)$ is Eisenstein. From our assumption that $m \leq n < 2m$ we have that $-m+n < m$. From the fact that $g(x)$ is Eisenstein we have that $v_p(b_i) \geq 1$ if $i < m$. It follows that $v_p(b_{-m+n}) \geq 1$. Therefore,

$$v_p(c_n) = \min\{v_p(a_{m-i}) + v_p(b_{i-m+n}) | 0 \leq i \leq 2m-n\} \geq 1$$

as desired.

□

Corollary 2. *If $f(x) \in \mathbb{Z}_p[x]$ is Eisenstein with degree m then $(f(x))^n$ is pure of slope $-\frac{1}{m}$ for all $n \in \mathbb{N}$.*

Proof. We will proceed with induction. By construction, $(f(x))^1$ is Eisenstein with slope $-\frac{1}{m}$. Assume that for all $i \leq n-1$ that $(f(x))^i$ is Eisenstein with slope $-\frac{1}{m}$. Consider $(f(x))^n$. Clearly $(f(x))^n = (f(x))^{n-1}(f(x))$. Both $(f(x))^{n-1}$ and $(f(x))$ are Eisenstein with slope $-\frac{1}{m}$. By Proposition 14 then $(f(x))^n$ is Eisenstein with slope $-\frac{1}{m}$. □

Lemma 3. *Let $f(x) \in \mathbb{Z}_p[x]$ be pure of slope m and let c be a constant. Then $cf(x)$ is pure of slope m.*

Proof. Let $f(x) = \sum_{i=0}^{n} a_i x^i$. Then for all i, $v_p(ca_i) = v_p(c) + v_p(a_i)$. So $v_p(ca_0) = v_p(c) + v_p(a_0)$ and $v_p(ca_n) = v_p(c) + v_p(a_n)$. Furthermore, we have by construction that for all $i \neq 0, n$, $v_p(a_i) \geq v_p(a_0) + mi$. So every point $(i, v_p(c) + v_p(a_i))$ lies above the line $y = v_p(a_0) + v_p(c)(mi)$. Therefore $cf(x)$ is pure of slope m. □

Lemma 4. *Let $f(x), g(x), h(x) \in \mathbb{Z}_p[x]$ such that $f(x) = g(x) + h(x)$. Then $NP(f) \subseteq LCH(NP(g) \cup NP(h))$.*

Proof. Let,

$$g(x) = \sum_{i=0}^{n} a_i x^i \text{ and } \ h(x) = \sum_{i=0}^{n} b_i x^i$$

Where $n = \max\{\deg(g(x)), \deg(g(x))\}$. By definition then,

$$NP(g) \cup NP(h) = LCH\{(i, v_p(a_i))\}_{0 \leq i \leq n} \cup LCH\{(i, v_p(b_i))\}_{0 \leq i \leq n}$$

$$= LCH\{(i, \min\{v_p(a_i), v_p(b_i)\})\}_{0 \leq i \leq n}.$$

We also have,

$$f(x) = \sum_{i=0}^{n} (a_i + b_i) x^i.$$

Therefore,

$$NP(f) = LCH(\{(i, v_p(a_i + b_i)\}).$$

If $a_i \neq b_i$ then,

$$(i, v_p(a_i + b_i)) = (i, \min\{v_p(a_i), v_p(b_i)\}) \in LCH\{(i, \min\{v_p(a_i), v_p(b_i)\})\}.$$

If $a_i = b_i$ then,

$$(i, v_p(a_i + b_i)) \geq (i, \min\{v_p(a_i), v_p(b_i)\}).$$

Thus $(i, v_p(a_i + b_i))$ is above $(i, \min\{v_p(a_i), v_p(b_i))$ which implies that

$$(i, v_p(a_i + b_i)) \in LCH\{(i, \min\{v_p(a_i), v_p(b_i)\})\}.$$

So,

$$NP(f) \subseteq LCH\{(i, \min\{v_p(a_i), v_p(b_i)\})\} = LCH(NP(g) \cup NP(h))$$

□

Corollary 3. *Let $f(x) \in \mathbb{Z}_p[x]$ and let $\{f_i(x) | 0 \leq i \leq n\}$ be a finite collection of functions in $\mathbb{Z}_p[x]$ such that*

$$f(x) = \sum_{i=0}^{n} f_i(x).$$

Then,

$$NP(f) \subseteq LCH(\bigcup_{i=0}^{n} NP(f_i))$$

Proof. We proceed by induction. By lemma 4, if $n = 2$, then

$$NP(f) \subseteq LCH(NP(f_1) \cup NP(f_2)).$$

Assume that,

$$NP(f) \subseteq LCH(\bigcup_{i=0}^{n-1} NP(f_i)).$$

Then,

$$NP(f) \subseteq LCH(\bigcup_{i=0}^{n-1} NP(f_i)) \cup LCH(NP(f_n))$$

$$\subseteq LCH(\bigcup_{i=0}^{n} NP(f_i)).$$

□

Now we can finally move on to the proof of the theorem stated earlier.

Proof of Theorem 5. By construction we have that the degree of $f(x)$ is n and the degree of $g(x)$ is d. Thus the degree of $f(g(x))$ is nd. Let,

$$f(x) = \sum_{i=0}^{n} a_i x^i \text{ and } \; g(x) = \sum_{i=0}^{d} b_i x^i \text{ and } \; f(g(x)) = \sum_{i=0}^{nd} c_i x^i$$

We must prove the following 3 conditions:

(1) $v_p(c_0) = m$

(2) $v_p(c_{nd}) = 0$

(3) $v_p(c_i) \geq m - i\frac{m}{nd}$ for all i

For (1) we have,

$$c_0 = a_0 + a_1 b_0 + a_2 b_0^2 + \cdots + a_n b_0^n = \sum_{i=0}^{n} a_i b_0^i$$

Taking the valuation we have,

$$v_p(c_0) = v_p(\sum_{i=0}^{n} a_i b_0^i) = v_p(a_0 + \sum_{i=1}^{n} a_i b_0^i).$$

We know that $v_p(a_0) = m$ and that $v_p(c_0) = \min\{v_p(a_0), v_p(\sum_{i=1}^{n} a_i b_0^i)\}$ if $v_p(a_0) \neq v_p(\sum_{i=1}^{n} a_i b_0^i)$ so it suffices to show that $v_p(\sum_{i=1}^{n} a_i b_0^i) > m$. Using the same additive property of valuations as above we have,

$$v_p(\sum_{i=1}^{n} a_i b_0^i) \geq \min\{v_p(a_i b_0^i) | 1 \leq i \leq n\} = \min\{v_p(a_i) + \sum_{k=0}^{i} v_p(b_0) | 1 \leq i \leq n\}$$

$$= \min\{v_p(a_i) + \sum_{k=0}^{i} 1 | 1 \leq i \leq n\} = \min\{v_p(a_i) + i | 1 \leq i \leq n\}.$$

Because $f(x)$ is pure of slope $-\frac{m}{n}$, we have that $v_p(a_i) \geq m - i\frac{m}{n}$ for all i. Thus,

$$v_p(a_i) + i \geq m - i\frac{m}{n} + i = m + i(1 - \frac{m}{n}).$$

Notice that because $m < n$, we have that for all $i \geq 1$, $m + i(1 - \frac{m}{n}) > m$. Therefore,

$$v_p(\sum_{i=1}^{n} a_i b_0^i) \geq \min\{v_p(a_i) + i | 1 \leq i \leq n\} \geq \min\{m + i(1 - \frac{m}{n}) | 1 \leq i \leq n\} > m$$

as desired.

For (2) we have $c_{nd} = a_n b_d^n$ so,

$$v_p(c_{nd}) = v_p(a_n b_d^n) = v_p(a_n) + \sum_{k=0}^{n} v_p(b_d) = v_p(a_n) + n v_p(b_d).$$

By definition we have $v_p(a_n) = 0$. Also, by the fact that $g(x)$ is Eisenstein we have that $v_p(b_d) = 0$. Thus,

$$v_p(c_{nd}) = v_p(a_n) + n v_p(b_d) = 0.$$

For (3) Let $h_i(x) = a_i(g(x))^i$ so that

$$f(g(x)) = \sum_{i=0}^{n} h_i(x).$$

We will consider the Newton polygons of each $h_i(x)$ individually. The first vertex of $NP(h_i(x))$ is determined by $v_p(h_i(0))$. We can compute this value

by using the properties of p-adic valuations:

$$v_p(h_i(0)) = v_p(a_i g(0)^i) = v_p(a_i b_0^i) = v_p(a_i) + v_p(b_0^i)$$

$$= v_p(a_i) + \sum_{k=0}^{i} v_p(b_0) = v_p(a_i) \sum_{k=0}^{i} 1 = v_p(a_i) + i.$$

By definition $v_p(a_i) \geq -i\frac{m}{n} + m$ for all i. So we have,

$$v_p(h_i(0)) = v_p(a_i) + i \geq -i\frac{m}{n} + m + i.$$

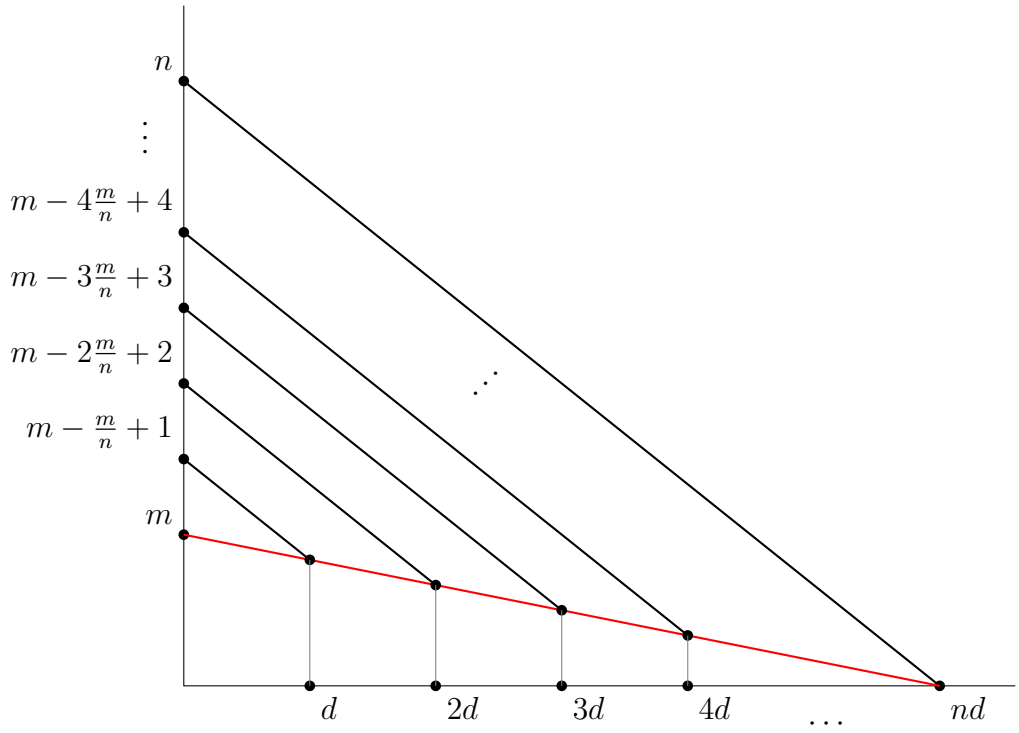
Therefore, $(0, v_p(a_i) + i) \in NP(h_i)$ and $v_p(a_i) + i \geq m - i\frac{m}{n} + i$. By the fact that $g(x)$ is Eisenstein and lemma 2, we know that $g(x)$ is pure of slope $\frac{-1}{d}$. Furthermore, by Corollary 2 we have that $g(x)^i$ is pure of slope $\frac{-1}{d}$. Then by lemma 3, this implies that $a_i g(x)^i = h_i(x)$ is pure of slope $\frac{-1}{d}$. Notice that $\deg(h_i(x)) = \deg(a_i g(x)^i) = i \deg(a_i g(x)) = id$. Thus, the last vertex of $NP(h_i)$ will have x-value id. Using the fact that $h_i(x)$ is pure of slope $\frac{-1}{d}$ and that $(0, v_p(a_i) + i) \in NP(h_i)$ we can calculate this vertex,

$$(id, v_p(a_i) + i - \frac{id}{d}) = (id, v_p(a_i)) \in NP(h_i).$$

Furthermore, we can use the above inequality to get

$$v_p(a_i) \geq m - i\frac{m}{n}.$$

Let $L_i = LCH((0, m - i\frac{m}{n} + i), (id, m - i\frac{m}{n}))$ and notice that $NP(h_i) \subseteq L_i$. Graphing all L_i's we see that they all have the same slope and are separated by regular intervals. Furthermore, we notice that taking the lowermost point of each one gives us a line which gives the Newton polygon of $h(x)$ by Corollary 3.



More formally we have that,

$$LCH(\bigcup_{i=0}^{n} NP(h_i)) \subseteq LCH(\bigcup_{i=0}^{n} L_i) = LCH(\{(id, m - i\frac{m}{n}) | 0 \leq i \leq n\}).$$

In addition, the lower convex hull of this line is the lower convex hull of the its endpoints,

$$LCH(\{(id, m - i\frac{m}{n}) | 0 \leq i \leq n\}) = LCH(\{(0, m), (nd, 0)\})$$

Recall that $f(g(x)) = \sum_{i=0}^{n} h_i(x)$. Then we have by Corollary 3 that the Newton polygon of $f(g(x))$ is contained within $LCH(\{(0, m), (nd, 0)\})$.

To summarize,

$$NP(f \circ g) \subseteq LCH(\bigcup_{i=0}^{n} NP(h_i)) \subseteq LCH(\bigcup_{i=0}^{n} L_i)$$

$$= LCH(\{(id, m - i\frac{m}{n}) | 0 \leq i \leq n\}) = LCH(\{(0, m), (nd, 0)\}).$$

This implies that all vertices of $NP(f \circ g)$ lie above or on the line from $(0, m)$ to $(nd, 0)$. This line has slope $-\frac{m}{nd}$ and a vertex at $(0, m)$ so we finally have that $v_p(c_i) \geq -i\frac{m}{nd} + m$ for all i which is (3). $\square$

So we have that composing a single segment function f, with an Eisenstein function g results in a a new function $f \circ g$, whose Newton Polygon is the Newton polygon of f "stretched" by the degree g. In order to achieve this result one of the conditions we placed on f is that the valuation of f's constant term, which we denoted m, must be less than or equal to the degree of f which we denoted n. This seems minor and it would be nice to do away with it. After all, we only made use of it in the proof to fix the first vertex of $NP(fg)$. One must wonder if there is a clever way to expunge the condition. Unfortunately it is indeed necessary. To see this, let us consider a case where $m > n$.

Example 11. Let $p = 5$, $f(x) = 375 + x + x^2$ and $g(x) = 5 + x^3$. This setup fulfills all the requirements of Theorem 5 *except* that $m = 3 > 2 = n$ in this case. Let us try composing the two functions and see what happens. We have,

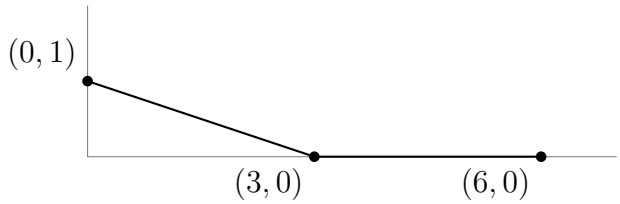
$$f(g(x)) = 375 + (5 + x^3) + (5 + x^3)^2 = x^6 + 11x^3 + 405.$$

From this we have that the vertices of $NP(f \circ g)$ are

$$\{(i, v_5(a_i(f(g(x))) | 0 \leq i \leq 6\}$$

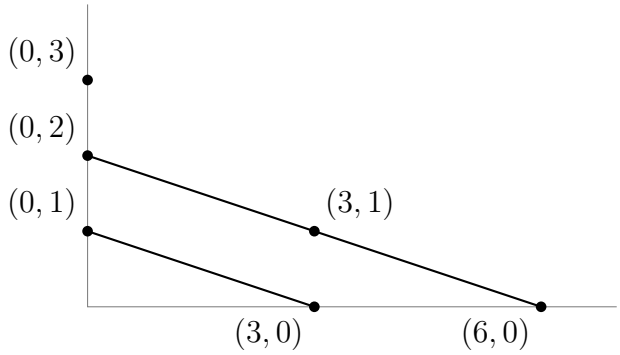
$$= \{(0, 1), (1, +\infty), (2, +\infty), (3, 0), (4, +\infty), (5, +\infty), (6, 0)\}.$$

This gives us an additional break at $(3, 0)$ which makes it clear that Theorem 5 does not apply in these cases.



Newton Polygon of $f(g(x))$

In fact, taking the approach that we used in the proof above and considering each $h_i(x) = a_i g(x)^i$ individually we get 3 h_i's: $h_0(x) = 375$, $h_1(x) = 5 + x^3$ and $h_2(x) = 25 + 10x^3 + x^6$. Graphing the Newton polygons of these h_i's we get



Here the top line is the Newton Polygon of $h_2(x)$, the bottom line is the Newton Polygon of $h_1(x)$, and the Newton polygon of $h_0(x)$ is just the point $(3, 0)$. Contrary to the picture we relied on in the proof, in this case the Newton polygons stack downwards. Hence, we cannot consider them subsets of the Newton polygon of the line from $(0, m)$ to $(nd, 0)$. Accordingly, the assumption that $m < n$ is quite necessary.

Another nice consequence of this theorem is that it is a more general case of *Odoni's theorem* [2].

Corollary 4 (Odoni 1985)**.** *Let* $f(x), g(x) \in \mathbb{Z}_p[x]$ *be Eisenstein with* $\deg(f(x)) \geq 2$*. Then* $f(g(x))$ *is Eisenstein.*

Proof. Let $\deg(f(x)) = n$ and $\deg(g(x)) = d$. By definition of Eisenstein, $(0,1) \in NP(f)$ and $NP(f)$ has pure slope $-\frac{1}{n}$. Clearly $1 < 2 \leq n$. By theorem 5 then $NP(f \circ g)$ is pure of slope $-\frac{1}{nd}$ and contains the point $(0,1)$. Thus, for $f(g(x)) = \sum_{i=0}^{nd} a_i x^i$,

(1) $v_p(a_0) = 1$,

(2) $v_p(a_{nd}) = 0$,

(3) and $v_p(a_i) \geq 1$ for all $i \neq nd$.

By definition then $f(g(x))$ is Eisenstein. □

Naturally we can induct on this process to show that any finite number of compositions of Eisenstein functions yields an Eisenstein function.

Corollary 5 (Odoni 1985)**.** *Let* $F = \{f_i(x) | 0 \leq i \leq n\}$ *be a finite set of Eisenstein functions in* $\mathbb{Z}_p[x]$ *with* $deg(f_i(x)) \geq 2$ *for all* i*. Then,*

$$f_n \circ f_{n-1} \circ \cdots \circ f_1 \circ f_0$$

is Eisenstein.

Proof. For ease of notation let,

$$g_i(x) = \begin{cases} f_0(x) \text{ if } i = 0 \\ f_i(g_{i-1}(x)) \text{ if } i > 0 \end{cases}.$$

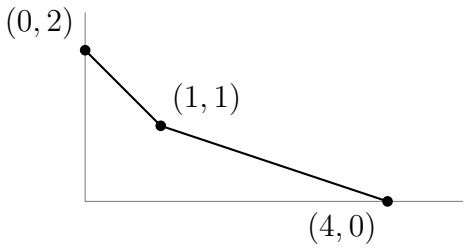
That is, $g_i(x)$ is the composition of the first i functions in F. By Odoni's theorem if $n = 2$ then $g_2(x)$ is Eisenstein. Assume that $g_{n-1}(x)$ is Eisenstein. Then $f_n(g_{n-1}(x)) = g_n(x)$ is a composition of 2 Eisenstein functions and is therefore Eisenstein by Odoni's theorem. By induction then $g_n(x)$ is Eisenstein for all $n \in \mathbb{N}$. □

This theorem has yielded some nice results already but it is limited in that it only applies to functions of pure slope. Ideally we would like to make a similar argument for functions with any number of breaks. This leads us to the following conjecture.

Conjecture 1. *Let $m, n \in \mathbb{Z}$ such that $m < n$. Let $f(x) = \sum_{i=0}^{n} a_i x^i \in \mathbb{Z}_p[x]$ have a Newton polygon with breaks at $\{(b_j, v_p(a_{b_j}) | 1 \leq j \leq r\}$ with $(0, v_p(a_0)) = (0, m)$. Furthermore let $g(x) \in \mathbb{Z}_p[x]$ have degree d and be Eisenstein. then the Newton polygon of $f(g(x))$ will have breaks at $\{(b_j d, v_p(a_{b_j}) | 1 \leq j \leq r\}$.*

In plain language we want to say that composing $f(x)$ with $g(x)$ "stretches" the newton polygon of $f(x)$ by a factor of d, the degree of $g(x)$, just as in the case where $f(x)$ was pure. Furthermore, the number of breaks, r, is preserved under this type of composition. Let us consider some examples.

Example 12. Let $p = 3$ and $f(x) = 18 + 3x + 4x^4$ and let $g(x) = 3 + 5x^2$. This example fulfills all the requirements of the above conjecture and we have that the breaks of $f(x)$ are $\{(0, 2), (1, 1), (0, 4)\}$. If the conjecture is true then we should have that the breaks of $f(g(x))$ are $\{(0, 2), (2, 1), (8, 0)\}$ as $d = 2$. Notice that f is not pure so this example is not covered by theorem 5.

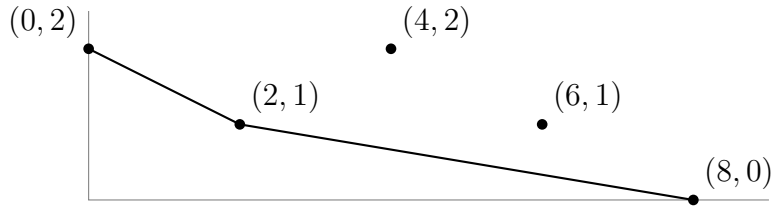


Newton Polygon of $f(x)$

Despite this, we can check if the conjecture holds by calculating $f(g(x))$. This gives us $f(g(x)) = 2500x^8 + 6000x^6 + 5400x^4 + 2175x^2 + 351$. Then $NP(f \circ g)$ will contain the following points:

$$\{(0,2), (1,+\infty), (2,1), (3,+\infty), (4,2), (5,+\infty), (6,1), (7,+\infty), (8,0)\}.$$

Then the breaks of $f(g(x))$ will occur at $\{(0,2), (2,1), (8,0)\}$ as predicted.



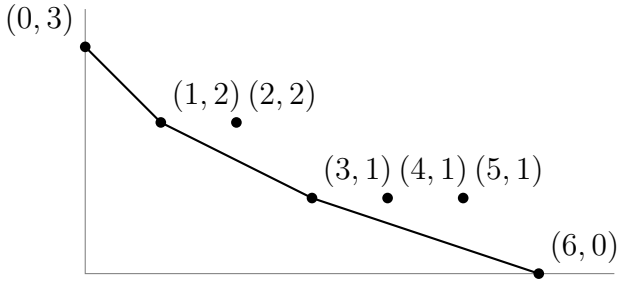
Newton Polygon of $f(g(x))$

Indeed the Newton polygon of $f(x)$ gets "stretched" to produce the Newton polygon of $f(g(x))$ as predicted.

We can also notice that, despite not being on the boundary of the Newton polygon, the points $\{(4,2), (6,1), (8,0)\}$ are colinear on the line $y = -2x + 4$. All other points that are not on the boundary are "at infinity" so to speak. Showing that all points aside from the breaks are either colinear or at infinity

will be helpful in proving the conjecture. Thus let us consider a situation with no points at infinity in the hope of detecting a general pattern.

Example 13. Let $p = 3$, $f(x) = 27 + 9x + 9x^2 + 3x^3 + 3x^4 + 3x^5 + x^6$ and let $g(x) = 3 + 3x + x^2$.



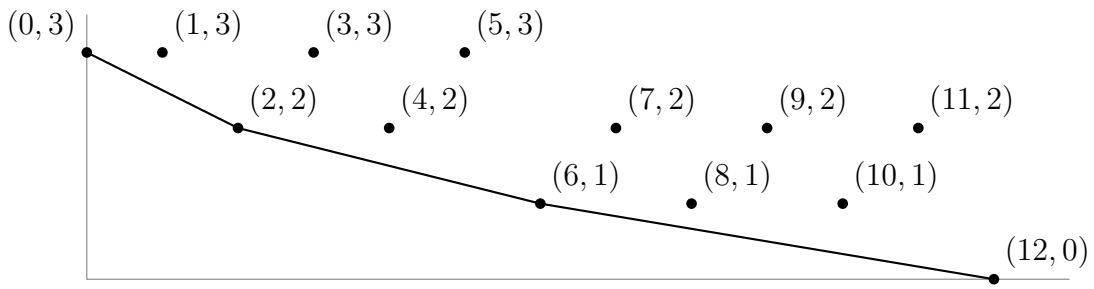
Newton Polygon of $f(x)$

This set up will not have any points at infinity as $f(g(x))$ will have no zero coefficients. Computing $f(g(x))$ we get,

$$x^{12} + 18x^{11} + 156x^{10} + 855x^9 + 3288x^8 + 9324x^7 + 19965x^6 + 32508x^5$$

$$+39969x^4 + 36261x^3 + 23148x^2 + 9423x + 1917.$$

We would expect the breaks to occur at $\{(0, 3), (2, 2), (6, 1), (12, 0)\}$ and indeed this is true.



Newton Polygon of $f(g(x))$

So there is some emperical evidence to support this conjecture. Furthermore we can see that in more complicated examples, some patterns may be emmerging. In terms of proof strategy, we can take a similar approach to the proof of Theorem 5. For $f(x) = \sum_{i=0}^{n} a_i x^i$ and $f(g(x)) = \sum_{i=0}^{nd} c_i x^i$, we would need to prove the following conditions:

(1) $v_p(c_0) = m$,

(2) $v_p(c_{nd}) = 0$,

(3) If there is a break at $(i, v_p(a_i))$ in $NP(f)$, then there is a break at $(id, v_p(a_i))$ in $NP(f \circ g)$. And,

(4) For any 2 sequential breaks at $x = b_j$ and $x = b_{j+1}$ in $NP(f \circ g)$, for all k such that $b_j < k < b_{j+1}$, the point $(k, v_p(c_k))$ is above the line from $(b_j, v_p(c_{b_j}))$ to $(b_{j+1}, v_p(c_{b_{j+1}}))$.

The first and second conditions should be almost identical to the case with f being pure. The third condition, which in plain language states that all breaks of $NP(f)$ get their x-values multiplied by a factor of d in $NP(f \circ g)$, is a little more tricky but possibly could be cracked by a method similar to the h_i method in the previous proof. In the picture of the L_i's we would expect

some L_i's to dip below the line from $(0, m)$ to $(nd, 0)$ in exactly the places we would like them to. This would result in breaks occuring at appropriate spots. Keep in mind that we need not prove that these are breaks of $NP(f \circ g)$ for condition 3, that is what the fourth condition is for. However, the h_i method might be able to resolve this one as well.

In conclusion, the p-adic numbers are a fascinating field of study with some beautiful results and methods lurking behind what at first may seem like very intimidating algebra. At first glance they seem to almost resist visualization, we did not discuss any attempt to visualize them in this paper though these attempts do exist. See page 85 of Gouvea [1]. Despite this seeming reluctance, we found that in considering polynomials with p-adic coefficients we can make use of the Newton polygons not only to help with visual intuition, but also to do some very nice math. There is always more to be said but for now we end here.

References

[1] Gouvea, Fernando Q. P-Adic Numbers: An Introduction. Springer, 2020.

[2] Odoni, R. W. "The Galois theory of iterates and composites of polynomials." Proceedings of the London Mathematical Society, vol. s3-51, no. 3, Nov. 1985, pp. 385–414, https://doi.org/10.1112/plms/s3-51.3.385.

[3] Kurt Hensel. Uber eine neue Begr¨undung der Theorie der algebraischen Za ¨ hlen. Jahresbericht der Deutschen Mathematiker-Vereinigung 6 (1897), 83–88

[4] Keith Conrad. The Contraction Mapping Theorem. https://kconrad.math.uconn.edu/blurbs /analysis/contraction.pdf