

Introduction to p -adic Numbers

by Rebecca A. Herb

The p -adic numbers were introduced as a tool in number theory in the late 19th century. Today p -adic numbers are important in number theory, algebraic geometry, representation theory, and even physics. (Disclaimer: I know nothing about physics and will not mention it in this minicourse. Try doing a Google search on p -adic physics if you're curious about it.) In the first two lectures of this minicourse I will give an elementary introduction to the p -adic numbers, starting with the construction of p -adic numbers as a completion of the rational numbers with respect to the p -adic norm. This process parallels the more familiar construction of the real numbers using Cauchy sequences, but has some interesting differences. I will then discuss some of the interesting properties of p -adic numbers.

References

Complete proofs for most of the results in these notes are in

1. N. Koblitz, p -adic numbers, p -adic analysis, and zeta functions, 2nd edition, Springer-Verlag, 1984.

Some other possible sources are:

2. G. Bachman, Introduction to p -adic numbers and valuation theory, Academic Press, 1964.

3. A.M. Robert, A course in p -adic analysis, Springer-Verlag, 2000.

In addition to the above books, useful lecture notes are available on the web. For example,

4. A. J. Baker, An introduction to p -adic numbers and p -adic analysis,

<http://www.maths.gla.ac.uk/~ajb>

Finally, try doing a Google search on p -adic numbers for many more possibilities.

§1. Norms and Completions.

Let \mathbf{Z} , \mathbf{Q} , \mathbf{R} denote the sets of integers, rational numbers, and real numbers respectively.

Definition. A function $N : \mathbf{Q} \rightarrow \mathbf{R}$ is called a norm on \mathbf{Q} if for all $x, y \in \mathbf{Q}$,

(N1) $N(x) \geq 0$ and $N(x) = 0$ if and only if $x = 0$;

(N2) $N(xy) = N(x)N(y)$;

(N3) $N(x + y) \leq N(x) + N(y)$.

Let N be a norm on \mathbf{Q} . We can define a distance by $d(x, y) = N(x - y)$, $x, y \in \mathbf{Q}$.

Lemma. For all $x, y, z \in \mathbf{Q}$,

(D1) $d(x, y) \geq 0$ and $d(x, y) = 0$ if and only if $x = y$;

(D2) $d(x, y) = d(y, x)$;

(D3) $d(x, y) \leq d(x, z) + d(z, y)$. (triangle inequality)

Let N be a norm on \mathbf{Q} with distance d . A sequence $\{x_n\}$ is called a Cauchy sequence (with respect to N) if given any $\epsilon > 0$ there is $M > 0$ so that $d(x_n, x_m) < \epsilon$ for all $n, m \geq M$. Let $x \in \mathbf{Q}$. Then we say the sequence $\{x_n\}$ converges to x if given any $\epsilon > 0$ there is $M > 0$ so that $d(x_n, x) < \epsilon$ for all $n \geq M$. It is easy to prove that every convergent sequence is Cauchy. We say \mathbf{Q} is complete (with respect to N) if every Cauchy sequence converges to an element of \mathbf{Q} .

Definition. A field K with norm N_K is called a completion on \mathbf{Q} with N if:

(C1) $\mathbf{Q} \subset K$ and $N_K(x) = N(x)$ for all $x \in \mathbf{Q}$;

(C2) K is complete with respect to N_K . That is, every Cauchy sequence in K converges to an element of K ;

(C3) \mathbf{Q} is dense in K . That is, every element of K is the limit of a sequence of elements of \mathbf{Q} .

Example 1. The trivial norm is given by $N(x) = 1, x \neq 0, N(0) = 0$. With the trivial norm, $d(x, y) = 1, x \neq y$, and $d(x, x) = 0$. A sequence $\{x_n\}$ is Cauchy just in case it is constant after a finite number of terms, that is there are $x \in \mathbf{Q}$ and M such that $x_n = x$ for all $n \geq M$. In this case $\{x_n\}$ converges to x . Thus \mathbf{Q} is complete with respect to the trivial norm.

Example 2. The usual absolute value gives a norm $N(x) = |x|$ with the usual distance $d(x, y) = |x - y|$. The field of real numbers $K = \mathbf{R}$ with the usual absolute value $N_K(x) = |x|$ is a completion of \mathbf{Q} with N .

The real numbers can be rigorously defined as equivalence classes of Cauchy sequences of rational numbers. This procedure can be used in general to produce a completion of \mathbf{Q} with respect to any norm N . This completion is essentially unique.

Example 3. Fix a prime number p . Define $N_p(0) = 0$. If $x \neq 0 \in \mathbf{Q}$, there is a unique integer k such that $x = p^k(a/b)$ where a and b are integers not divisible by p . We define $N_p(x) = p^{-k}$. The field of p -adic numbers \mathbf{Q}_p is the completion of \mathbf{Q} with respect to this norm. We will study this example in detail in §2.

We say two norms N_1 and N_2 on \mathbf{Q} are equivalent if a sequence $\{x_n\}$ is Cauchy with

respect to N_1 if and only if it is Cauchy with respect to N_2 . Thus equivalent norms will yield the same completion of \mathbf{Q} .

Theorem (Ostrowski) Every nontrivial norm on \mathbf{Q} is equivalent to the usual absolute value or to N_p for a unique prime number p .

§2. The p -adic Numbers.

Fix a prime number p . If $x \neq 0 \in \mathbf{Q}$, there is a unique integer k such that $x = p^k(a/b)$ where a and b are integers not divisible by p . We define $\nu_p(x) = k$. Define $\nu_p(0) = +\infty$.

Lemma. Let $x, y \in \mathbf{Q}$. Then

- (1) $\nu_p(x) = +\infty$ if and only if $x = 0$;
- (2) $\nu_p(xy) = \nu_p(x) + \nu_p(y)$;
- (3) $\nu_p(x + y) \geq \min\{\nu_p(x), \nu_p(y)\}$ and $\nu_p(x + y) = \min\{\nu_p(x), \nu_p(y)\}$ if $\nu_p(x) \neq \nu_p(y)$.

Proof. (1) and (2) are easy, and so is (3) when $x = 0$ or $y = 0$. Assume that x and y are both non-zero. Let $r = \nu_p(x)$ and $s = \nu_p(y)$. We may as well assume that $s \geq r$. Then we can write $x = p^r(a/b)$ and $y = p^s(c/d)$ where a, b, c, d are not divisible by p . Then

$$x + y = p^r \cdot \frac{ad + p^{s-r}cb}{bd}.$$

Now bd is not divisible by p so that $\nu_p((ad + p^{s-r}cb)/bd) \geq 0$. Thus

$$\nu_p(x + y) = r + \nu_p((ad + p^{s-r}cb)/bd) \geq r = \min\{\nu_p(x), \nu_p(y)\}.$$

Suppose that $s > r$. Then p divides $p^{s-r}cb$ but p does not divide ad so that p does not divide $ad + p^{s-r}cb$. Thus in this case $\nu_p((ad + p^{s-r}cb)/bd) = 0$ so that $\nu_p(x + y) = r = \min\{\nu_p(x), \nu_p(y)\}$.

For $x \in \mathbf{Q}$ we define $|x|_p = p^{-\nu_p(x)}$ where $|0|_p = p^{-\infty} = 0$. Then the following is an easy consequence of the above lemma.

Proposition. Let $x, y \in \mathbf{Q}$. Then

- (N1) $|x|_p \geq 0$, and $|x|_p = 0$ if and only if $x = 0$;
- (N2) $|xy|_p = |x|_p|y|_p$;
- (N3') $|x + y|_p \leq \max\{|x|_p, |y|_p\}$, and $|x + y|_p = \max\{|x|_p, |y|_p\}$ if $|x|_p \neq |y|_p$.

The proposition says that $N(x) = |x|_p$ is a norm on \mathbf{Q} since for $x, y \in \mathbf{Q}$,

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

That is, (N3') implies (N3). Note that (N3) does not imply (N3'). For example the usual absolute value satisfies (N3) but not (N3'). Any norm on \mathbf{Q} satisfying the stronger condition (N3') is called a non-Archimedean norm.

Let $n \in \mathbf{Z}$. Then $\nu_p(n) \geq 0$ so that $|n|_p \leq 1$. $|n|_p$ is small if n is divisible by a high power of p . If $x = n/m$ is written in reduced form, $|x|_p < 1$ if p divides n and $|x|_p > 1$ if p divides m . If both n and m are not divisible by p , then $|x|_p = 1$. For example

$$|100/13|_5 = |100|_5 = 10^{-2}, \quad |13/100|_5 = 10^2, \quad |99/4|_5 = 1 = |4/99|_5.$$

As before we can define a p -adic distance function on \mathbf{Q} by $d(x, y) = |x - y|_p, x, y \in \mathbf{Q}$. Let $x, y, z \in \mathbf{Q}$. Then using (N3') we have

(D3') $d(x, y) \leq \max\{d(x, z), d(z, y)\}$, with $d(x, y) = \max\{d(x, z), d(z, y)\}$ if $d(x, z) \neq d(z, y)$.

If we think of $d(x, y), d(x, z), d(z, y)$ as being three sides of a triangle, two of the sides must be equal, for either $d(x, z) = d(z, y)$ or $d(x, y) = \max\{d(x, z), d(z, y)\}$. Thus every p -adic "triangle" is isosceles.

As another example of the consequences of (D3') is the following. For $a \in \mathbf{Q}$ and $r > 0$ define the open ball of radius r centered at a by $B_r(a) = \{x \in \mathbf{Q} : d(x, a) < r\}$.

Lemma. Let $b \in B_r(a)$. Then $B_r(a) = B_r(b)$. That is every element of $B_r(a)$ is a center.

Proof. Let $x \in B_r(a)$. Then $d(x, b) \leq \max\{d(x, a), d(b, a)\} < r$ since $x, b \in B_r(a)$. Thus $x \in B_r(b)$. The reverse inclusion is the same.

The formal construction of \mathbf{Q}_p , the completion of \mathbf{Q} with respect to the p -adic norm, using equivalence classes of Cauchy sequences parallels that of the construction of \mathbf{R} , the completion of \mathbf{Q} with respect to the ordinary absolute value. There is one important difference. Let $\{x_n\}$ be a Cauchy sequence with respect to a norm N . Then the Cauchy sequence is assigned norm $N(\{x_n\}) = \lim_n N(x_n)$. When N is the ordinary absolute value, these limits give all possible non-negative real numbers. When N is a p -adic absolute value, this is not the case.

To see this, let $\{x_n\}$ be a Cauchy sequence in \mathbf{Q} with respect to the p -adic norm. If $\lim_{n \rightarrow \infty} |x_n|_p = 0$, then $|\{x_n\}|_p = 0$. Otherwise, there is $\epsilon > 0$ such that for every $M > 0$ there is $k_M > M$ such that $|x_{k_M}|_p > \epsilon$. Since $\{x_n\}$ is Cauchy, we can pick $M > 0$ so that $|x_j - x_k|_p < \epsilon$ for all $j, k \geq M$. But now, for all $j \geq k_M > M$, $x_j = (x_j - x_{k_M}) + x_{k_M}$ where $|x_{k_M}|_p > \epsilon > |x_j - x_{k_M}|_p$ so that $|x_j|_p = |x_{k_M}|_p$ by the strong version of the triangle inequality (N3'). Thus the sequence $\{|x_n|_p\}$ is eventually constant, and $|\{x_n\}|_p = \lim_{n \rightarrow \infty} |x_n|_p = |x_{k_M}|_p$. That is, the possible values of $|x|_p, x \in \mathbf{Q}_p$ is not enlarged, but is still $\{0\} \cup \{p^k : k \in \mathbf{Z}\}$.

The extension of $|\cdot|_p$ to \mathbf{Q}_p still satisfies conditions (N1), (N2), (N3'). An interesting consequence of (N3') is the fact that if $x_n \in \mathbf{Q}_p, n \geq 0$, then $\sum_{n=0}^{\infty} x_n$ is a convergent

series if and only if $\lim_{n \rightarrow \infty} x_n = 0$. That is in p -adic calculus class we have nothing like the divergent series $1 + 1/2 + 1/3 + \dots$.

Most of us don't picture a real number as an equivalence class of Cauchy sequences. Just as real numbers can be described using a possibly infinite decimal expansion, elements of \mathbf{Q}_p can be described using a p -adic digit expansion.

Fix $m \in \mathbf{Z}$ and let $\alpha_k \in \{0, 1, 2, \dots, p-1\}, k \geq m$. For $n \geq m$, write $x_n = \sum_{k=m}^n \alpha_k p^k$. Then $x_n \in \mathbf{Q}$. We claim that $\{x_n\}$ is a Cauchy sequence. For all $k \geq m$, $|\alpha_k p^k|_p = p^{-k} |\alpha_k|_p \leq p^{-k}$. Thus for $r < s$,

$$|x_s - x_r|_p = \left| \sum_{k=r+1}^s \alpha_k p^k \right|_p \leq \max\{|\alpha_k p^k|_p : r+1 \leq k \leq s\} \leq p^{-r-1}.$$

Thus the sequence is Cauchy and so it represents an element x of \mathbf{Q}_p which we denote by $x = \lim x_n = \sum_{k=m}^{\infty} \alpha_k p^k$. Moreover, if $x \neq 0$, we may as well assume that m was chosen to be the first non-zero term, so that $\alpha_m \neq 0$. Then $1 \leq \alpha_m \leq p-1$ so that p does not divide α_m . Thus $|\alpha_m p^m|_p = p^{-m}$. Further, for $k > m$, $|\alpha_k p^k|_p \leq p^{-k} < p^{-m}$. Thus for all $n \geq m$, $|x_n|_p = p^{-m}$, and $|x|_p = \lim_n |x_n|_p = p^{-m}$. The following theorem which we will not prove asserts that every equivalence class of Cauchy sequences has a representative of this type.

Theorem. Every $x \in \mathbf{Q}_p$ has a unique convergent p -adic expansion $x = \sum_{k=m}^{\infty} \alpha_k p^k$ with $\alpha_k \in \{0, 1, 2, \dots, p-1\}$ for all k . If $\alpha_m \neq 0$, then $|x|_p = p^{-m}$.

This p -adic expansion is the analog of the decimal expansion of a real number which expresses a real number as $x = \sum_{k=-\infty}^m \alpha_k 10^k$ with digits $\alpha_i \in \{0, 1, \dots, 9\}$. Since $|10^k| = 10^k$ is large for $k > 0$ and converges to zero as $k \rightarrow -\infty$, these series have finitely many positive powers, but are allowed to have infinitely many negative powers. In contrast, $|p^k|_p = p^{-k}$ is large for $k < 0$ and converges to zero as $k \rightarrow \infty$, so p -adic decimal expansions have finitely many negative powers, but are allowed to have infinitely many positive powers. In addition, unlike decimal expansions of real numbers, where $1.0000\bar{0} = .99999\bar{9}$, the p -adic expansion is unique.

The mechanics of adding, subtracting, multiplying, and dividing p -adic numbers using the p -adic decimal expansion are like those for ordinary decimals using carrying, borrowing, long multiplication, and long division. The following two properties of p -adic expansions are also similar to those for decimal expansions.

1. The p -adic expansion of $x \in \mathbf{Q}_p$ is finite if and only if x is a positive rational number whose denominator is a power of p .

2. The p -adic expansion of $x \in \mathbf{Q}_p$ has repeating digits from some point on (ie. $\alpha_{i+r} = \alpha_i$ for some r and all $i \geq N$) if and only if $\alpha \in \mathbf{Q}$.

§3. Some More Properties of p -adic Numbers.

Elements of the set $\mathbf{Z}_p = \{x \in \mathbf{Q}_p : |x|_p \leq 1\}$ are called p -adic integers. Thus \mathbf{Z}_p is the set of all $x \in \mathbf{Q}_p$ with p -adic expansions $x = \sum_{k=0}^{\infty} \alpha_k p^k$. The ordinary integers $\mathbf{Z} \subset \mathbf{Z}_p$ and every $x = \sum_{k=0}^{\infty} \alpha_k p^k \in \mathbf{Z}_p$ is the limit of the $x_n = \sum_{k=0}^n \alpha_k p^k \in \mathbf{Z}$. That is \mathbf{Z}_p is the closure in \mathbf{Q}_p of \mathbf{Z} . (Note that \mathbf{Z} is closed in \mathbf{R} so the real integers are just the ordinary integers.)

\mathbf{Z}_p is a subring of \mathbf{Q}_p because for $x, y \in \mathbf{Z}_p$, $|x \pm y|_p \leq \max\{|x|_p, |y|_p\} \leq 1$ and $|xy|_p = |x|_p |y|_p \leq 1$. Let $\mathbf{Z}_p^\times = \{x \in \mathbf{Q}_p : |x|_p = 1\}$. It is the set of $x \in \mathbf{Z}_p$ such that $x^{-1} \in \mathbf{Z}_p$, and is a group with multiplication called the group of units of \mathbf{Z}_p . Units have expansions $x = \sum_{k=0}^{\infty} \alpha_k p^k$ with $\alpha_0 \neq 0$. Let $x \neq 0 \in \mathbf{Q}_p$ with $|x|_p = p^{-k}$. Then $|p^{-k}x|_p = 1$ so that $p^{-k}x = u \in \mathbf{Z}_p^\times$. That is every nonzero $x \in \mathbf{Q}_p$ can be written (uniquely) as $x = p^k u$ where $k \in \mathbf{Z}$ and $u \in \mathbf{Z}_p^\times$. Let $M_p = \{x \in \mathbf{Q}_p : |x|_p < 1\}$. It is the unique maximal ideal of the ring \mathbf{Z}_p and \mathbf{Z}_p/M_p is isomorphic to the finite field F_p with p elements since M_p consists of the elements $x = \sum_{k=0}^{\infty} \alpha_k p^k$ with $\alpha_0 = 0$.

When we complete the rational numbers to obtain the real numbers, we get a bonus. Not only do all Cauchy sequences converge, but most polynomial equations can be solved. However, we still cannot solve equations such as $x^2 + 1 = 0$, so we introduce $i = \sqrt{-1}$ to obtain the complex numbers $\mathbf{C} = \{x + iy : x, y \in \mathbf{R}\}$. This field is algebraically closed, that is all polynomial equations with coefficients in \mathbf{C} have solutions in \mathbf{C} . It is also complete with respect to the norm $|x + iy| = \sqrt{x^2 + y^2}$. We can think of \mathbf{C} as a real vector space with basis $\{1, i\}$, and say \mathbf{C} is an extension of \mathbf{R} of degree 2.

Things aren't so simple for \mathbf{Q}_p , since \mathbf{Q}_p is very far from being algebraically closed. For example, look at equations of the form $x^n - p = 0$. A solution $a \in \mathbf{Q}_p$ would satisfy $a^n = p$ so that $|a|_p^n = |p|_p = p^{-1}$. Thus we would need $|a|_p = p^{-1/n}$. But for $n \geq 2$, there is no $a \in \mathbf{Q}_p$ with $|a|_p = p^{-1/n}$. We must adjoin infinitely many solutions of polynomial equations to obtain an algebraically closed field. That is, $\overline{\mathbf{Q}_p}$, the smallest algebraically closed field containing \mathbf{Q}_p , is an infinite extension of \mathbf{Q}_p . Moreover, $\overline{\mathbf{Q}_p}$ is no longer complete, so we need to complete it to form an even bigger field Ω_p . Fortunately Ω_p is now both algebraically closed and complete and so is the analog of \mathbf{C} . For details of this construction see Chapter III of Koblitz.

The topology on \mathbf{Q}_p is quite different from that of \mathbf{R} . Both \mathbf{Q}_p and \mathbf{R} are examples

of metric spaces, that is sets with a distance function d satisfying (D1),(D2),(D3). Let X be a metric space with distance function d . Then given $x \in X, r > 0$ we can define the open ball $B_r(x) = \{y \in X : d(x, y) < r\}$. Now $U \subset X$ is called open if for all $x \in U$ there is $r > 0$ such that $B_r(x) \subset U$. Further $C \subset X$ is called closed if its complement is open, that is if for $x \notin C$, there is $r > 0$ so that $B_r(x) \cap C = \emptyset$. It is easy to check that open balls are open sets and that $\overline{B}_r(x) = \{y \in X : d(x, y) \leq r\}$ is a closed set.

When $X = \mathbf{R}$, $B_r(x) = (x - r, x + r)$ and $\overline{B}_r(x) = [x - r, x + r]$ are the open and closed intervals centered at x with radius r . Let $X = \mathbf{Q}_p$. Then the only possible values of $d(x, y) = |x - y|_p$ are 0 and $p^k, k \in \mathbf{Z}$. Thus if $p^{-k} < r \leq p^{-k+1}$ we see that $d(x, y) < r$ if and only if $d(x, y) \leq p^{-k}$. That is $B_r(x) = \overline{B}_{p^{-k}}(x)$ is both open and closed. In contrast, the only subsets of \mathbf{R} which are both open and closed are \emptyset and \mathbf{R} .

A subset Y of X is called disconnected if there are open sets U_1, U_2 of X such that Y is the disjoint union $Y = (Y \cap U_1) \cup (Y \cap U_2)$ where $Y \cap U_i$ is non-empty for both $i = 1, 2$. Y is called connected if it is not disconnected. For example, the set $\{x\}$ consisting of a single point is always connected. When $X = \mathbf{R}$, a set is connected if and only if it is an interval. However, let Y be a subset of \mathbf{Q}_p containing two distinct elements x_1, x_2 . Then $d(x_1, x_2) = r > 0$. Let $U_1 = B_r(x_1)$ and let U_2 be the complement of $B_r(x_1)$. Since $B_r(x_1)$ is both open and closed, U_1 and U_2 are both open. Since $U_1 \cap U_2 = \emptyset$ and $U_1 \cup U_2 = \mathbf{Q}_p$, we have $Y = (Y \cap U_1) \cup (Y \cap U_2)$, disjoint union. Finally, $x_1 \in Y \cap U_1$ and $x_2 \in Y \cap U_2$ so both are non-empty. Thus Y is disconnected. That is, the only connected subsets of \mathbf{Q}_p are single points. For this reason \mathbf{Q}_p is called totally disconnected.