CONTENTS

# PREFACE

This book is a revised and expanded version of a series of talks given in Hanoi at the Viện Toán học (Mathematical Institute) in July, 1978. The purpose of the book is the same as the purpose of the talks: to make certain recent applications of p-adic analysis to number theory accessible to graduate students and researchers in related fields. The emphasis is on new results and conjectures, or new interpretations of earlier results, which have come to light in the past couple of years and which indicate intriguing and as yet imperfectly understood new connections between algebraic number theory, algebraic geometry, and p-adic analysis.

I occasionally state without proof or assume some familiarity with facts or techniques of other fields: algebraic geometry (Chapter III), algebraic number theory (Chapter IV), analysis (the Appendix). But I include down-to-earth examples and words of motivation whenever possible, so that even a reader with little background in these areas should be able to see what's going on.

Chapter I contains the basic information about p-adic numbers and p-adic analysis needed for what follows. Chapter II describes the construction and properties of p-adic Dirichlet L-functions, including Leopoldt's formula for the value at 1, using the approach of p-adic integration. The p-adic gamma function and log gamma function are introduced, their properties are developed and compared with the identities satisfied by the classical gamma function, and two formulas relating them to the p-adic L-functions $L_p(s,\chi)$ are proved. The first formula--expressing $L_p'(0,\chi)$ in terms of special

values of log gamma--will be used later (Chapter IV) in the discussion of Gross' p-adic regulator. The other formula--a p-adic Stirling series for log gamma near infinity--will be a key motivating example for the p-adic Stieltjes transform, discussed in the Appendix.

Chapter III is devoted primarily to proving a p-adic formula for Gauss sums, which expresses them essentially as values of the p-adic gamma function. The approach emphasizes the analogy with the complex-analytic periods of differentials on certain special curves, and uses some algebraic geometry. The reader who is interested in a treatment that is more "elementary" and self-contained (but more computational rather than geometric) is referred to [62].

Chapter IV discusses two different types of p-adic regulators. One, due to Leopoldt, is connected with the behavior of $L_p(s,\chi)$ at $s=1$; the other, due to Gross, is connected with the behavior at $s=0$. Conjectures describing these connections between regulators and L-functions are explained and compared to the classical case. The conjectures are proved in the case of a one-dimensional character $\chi$ with base field $Q$ (the "abelian over $Q$" case). The proof of Gross' conjecture in this case combines the formula for $L_p'(0,\chi)$ in Chapter II and the p-adic formula for Gauss sums in Chapter III, together with a p-adic version of the linear independence over $\overline{Q}$ of logarithms of algebraic numbers (Baker's theorem). This proof provides the culmination of the main part of the book.

The Appendix concerns some general constructions in p-adic analysis: the Stieltjes transform and the Shnirelman integral. I first use the Stieltjes transform to highlight the analogy between the p-adic and classical log gamma functions. I then give a complete account of M. M. Vishik's p-adic spectral theorem. This material has been relegated to the Appendix because it has not yet led to new number theoretic or algebra-geometric facts, perhaps because Vishik's theory is not very well known.

I would like to thank N. M. Katz, whose Spring 1978 lectures at

Princeton provided the explanations of the algebraic geometry and p-adic cohomology given in Chapter III; R. Greenberg, whose seminar talks at the University of Washington in October 1979 and whose comments on the manuscript were of great help in writing Chapter IV; B. H. Gross, whose preprint [35] and correspondence were the basis for the second half of Chapter IV; and M. M. Vishik, whose preprint [95] is given in modified form in §§3–4 of the Appendix.

I am also grateful to Ju. I. Manin and A. A. Kirillov for the stimulation provided by their seminars on Diophantine geometry and p-adic analysis during my stays in Moscow in 1974–75 and in Spring 1978; and to the Vietnamese mathematicians, in particular Lê-văn-Thiêm, Hà-huy-Khoái, Vương-ngọc-Châu and Dỗ-ngọc-Diệp, for their hospitality, which contributed to a fruitful and enjoyable visit to Hanoi.

Seattle
April 1980

Neal Koblitz

*FRONTISPIECE*: Artist's conception of the construction of the 2-adic number system as an inverse limit. By Professor A. T. Fomenko of Moscow State University.

# I. BASICS

In some places in this chapter detailed proofs and computations are omitted, in order not to bore the reader before we get to the main subject matter. These details are readily available (see, for example, [53]).

## 1. History (very brief)

| | | |
|---|---|---|
| Kummer and Hensel..... | 1850-1900 | introduced p-adic numbers and developed their basic properties |
| Minkowski..... | 1884 | proved: an equation $a_1x_1^2+\ldots+a_nx_n^2 = 0$ ($a_i$ rational) is solvable in the rational numbers if and only if it is solvable in the reals and in the p-adic numbers for all primes $p$ (see [13,84]) |
| Tate..... | 1950 | Fourier analysis on p-adic groups; pointed toward interrelations between p-adic numbers and L-functions and representation theory (see [59]) |
| Dwork..... | 1960 | used p-adic analysis to prove rationality of the zeta-function of an algebraic variety defined over a finite field, part of the Weil conjectures (see [25,53] ) |
| (Kummer..... | 1851 | congruences for Bernoulli numbers--but he approached them in an ad hoc way, without p-adic numbers) |
| Kubota-Leopoldt | 1964 | interpretation of Kummer congruences for Bernoulli numbers using p-adic zeta-function |
| Iwasawa, Serre, Mazur, Manin, Katz, others | past 15 years | p-adic theories for many arithmetically interesting functions |

| | | |
|---|---|---|
| Dwork, Grothendieck and their students | past 15 years | p-adic differential equations, p-adic cohomology, crystals |

## 2. Basic concepts

Let $p$ be a prime number, fixed once and for all. The "p-adic numbers" are all expressions of the form

$$a_m p^m + a_{m+1}p^{m+1} + a_{m+2}p^{m+2} + \ldots,$$

where the $a_i \in \{0,1,2,\ldots,p-1\}$ are digits, and $m$ is any integer. These expressions form a field ($+$ and $\times$ are defined in the obvious way), which contains the nonnegative integers

$$n = a_0 + a_1 p + \ldots + a_r p^r \quad (\text{"}n \text{ written to the base } p\text{"}),$$

and hence contains the field of rational numbers $Q$. For example,

$$-1 = (p-1) + (p-1)p + (p-1)p^2 + \ldots;$$

$$\frac{-a_0}{p-1} = a_0 + a_0 p + a_0 p^2 + \ldots,$$

as is readily seen by adding $1$ to the first expression on the right and multiplying the second expression on the right by $1-p$.

An equivalent way to define the field $Q_p$ of p-adic numbers is as the completion of $Q$ under the "p-adic metric" determined by the norm $|\ |_p : Q \to$ nonnegative real numbers, defined by

$$\left|\frac{a}{b}\right|_p = p^{\mathrm{ord}_p b - \mathrm{ord}_p a}, \quad |0|_p = 0,$$

where $\mathrm{ord}_p$ of a nonzero integer is the highest power of $p$ dividing it. Under this norm, numbers highly divisible by $p$ are "small", while numbers with $p$ in the denominator are "large". For example, $|250|_5 = 1/125$, $|1/250|_5 = 125$. Clearly, $|\ |_p$ is multiplicative, because $\mathrm{ord}_p$ behaves like log:

$$\mathrm{ord}_p(xy) = \mathrm{ord}_p x + \mathrm{ord}_p y.$$

Also note that $|n|_p \leq 1$ for $n$ an integer.

It is not hard to verify that the completion of $Q$ under the p-adic metric can be identified with the set $Q_p$ of "p-adic expansions" $a_m p^m + a_{m+1}p^{m+1} + \ldots$ . The norm $|\ |_p$ is easy to

evaluate on an element of $Q_p$ written in its p-adic expansion: if $x = a_m p^m + a_{m+1} p^{m+1} + \ldots$ with $a_m \neq 0$, then $|x|_p = p^{-m}$.

Thus, $Q_p$ is obtained from $|\ |_p$ in the same way as the real number field $R$ is obtained from the usual absolute value $|\ |$: as the completion of $Q$. In fact, a theorem of Ostrowski (see [13] or [53]) says that any norm on $Q$ is equivalent to the usual $|\ |$ or to $|\ |_p$ for some $p$. Hence, together with $R$, the various $Q_p$ make up all possible completions of $Q$:

$$\begin{tikzcd} R & Q_2 \arrow[dl] & Q_3 \arrow[dl] & Q_5 \arrow[dl] & \cdots & Q_p \arrow[dl] \\ & & Q & & & \end{tikzcd}$$

Often, a situation can be studied more easily over $R$ and $Q_p$ than over $Q$; and then the information obtained can be put together to conclude something about the situation over $Q$. For example, one can readily show that a rational number has a square root in $Q$ if and only if it has a square root in $R$ and for all $p$ has a square root in $Q_p$. This assertion is a special case of the Hasse-Minkowski theorem (see §1 above).

In addition to multiplicativity, the other basic property of a norm $|\ |$ on a field is the "triangle inequality" $|x + y| \leq |x| + |y|$, so named because in the case of the complex numbers $C$ it says that in the complex plane one side of a triangle is less than or equal to the sum of the other two sides. The norm $|\ |_p$ on $Q_p$ satisfies a stronger inequality:

$$|x + y|_p \leq \max(|x|_p, |y|_p). \tag{2.1}$$

This is obvious if we recall how to evaluate $|x|_p$ for $x = a_m p^m + a_{m+1} p^{m+1} + \ldots$ (see above). A norm that satisfies (2.1) is called "non-Archimedean". Inequality (2.1) is sometimes called the "isosceles triangle principle", because it immediately implies that, among the three "sides" $|x|_p$, $|y|_p$ and $|x+y|_p$, at least two must be equal. Thus, in non-Archimedean geometry "all triangles are isosceles".

Here is another strange consequence of (2.1). In a field with a non-Archimedean norm $|\ |_p$, define

$$D_a(r) = \{x \mid |x-a|_p \le r\} \text{ ("closed" disc of radius } r \text{ centered at } a)$$
$$D_a(r^-) = \{x \mid |x-a|_p < r\} \text{ ("open" disc of radius } r \text{ centered at } a). \tag{2.2}$$

Then if $b \in D_a(r)$, it follows from (2.1) that $D_b(r) = D_a(r)$. (Also, if $b \in D_a(r^-)$, then $D_b(r^-) = D_a(r^-)$.) Thus, any point in a disc is its center! In particular, any point in a disc (or in its complement) has a neighborhood completely contained in the disc (resp., in its complement). Therefore, any disc is <u>both</u> open <u>and</u> closed in the topological sense. That is why the words "open" and "closed" in (2.2) are in quotation marks; these words are used <u>only</u> by analogy with classical geometry, and one should not be misled by them.

In $Q_p$, it is not hard to see that all discs of finite radius are compact. The most important such disc is

$$Z_p \overset{\text{def}}{=} D_0(1) = \{x \mid |x|_p \le 1\} = \{x = a_0 + a_1 p + a_2 p^2 + \ldots\}.$$

$Z_p$ is a ring, whose elements are called p-adic "integers". $Z_p$ is the closure of the ordinary integers $Z$ in $Q_p$. In $Q_p$, the other discs centered at $0$ are

$$p^m Z_p = \{x = a_m p^m + a_{m+1} p^{m+1} + \ldots\} \quad \text{for } m \in Z.$$

$Z_p$ is a local ring, i.e., it has a unique maximal ideal $pZ_p$, and its residue field $Z_p/pZ_p$ is the field of $p$ elements $F_p = Z/pZ$. The set of invertible elements in the ring $Z_p$ is

$$Z_p^* \overset{\text{def}}{=} Z_p - pZ_p = \{x \mid |x|_p = 1\}$$
$$= \{x = a_0 + a_1 p + a_2 p^2 + \ldots \mid a_0 \ne 0\}.$$

There are $p-1$ numbers in $Z_p^*$ which play a special role: the $(p-1)$-th roots of one. For each possible choice of $a_0 = 1, 2, \ldots, p-1$, there is a unique such root whose first digit is $a_0$; we denote it $\omega(a_0)$ and call it the <u>Teichmüller representative</u> of $a_0$. For example, for $p = 5$

$$\omega(1) = 1$$
$$\omega(2) = 2 + 1\cdot 5 + 2\cdot 5^2 + 1\cdot 5^3 + 3\cdot 5^4 + \ldots$$
$$\omega(3) = 3 + 3\cdot 5 + 2\cdot 5^2 + 3\cdot 5^3 + 1\cdot 5^4 + \ldots = -\omega(2)$$
$$\omega(4) = 4 + 4\cdot 5 + 4\cdot 5^2 + 4\cdot 5^3 + 4\cdot 5^4 + \ldots = -1.$$

Except for $\omega(\pm 1)$, the Teichmüller representatives are irrational, so their p-adic digits do not repeat, and can be expected to be just as random as, say, the decimal digits in $\sqrt{2}$.

If $x = a_0 + a_1 p + \ldots \in Z_p^*$, we set $\omega(x) = \omega(a_0)$. Any $x \in Q_p$ can be written as $x = p^{\mathrm{ord}_p x} x_0$ for $x_0 \in Z_p^*$. Then we write

$$x = p^{\mathrm{ord}_p x}\, \omega(x_0)\, \langle x_0 \rangle,$$

where $\langle x_0 \rangle \underset{\mathrm{def}}{=} x_0/\omega(x_0)$ is in $1 + pZ_p$, the set of $x$ such that $|x-1|_p < 1$.

The ring $Z_p$ is the inverse limit of the rings $Z/p^n Z$ with respect to the map "reduction mod $p^n$" from $Z/p^m Z$ to $Z/p^n Z$ for $m \geq n$. This suggests that, if we want to solve an equation $f(x) = 0$ for $x \in Z_p$, we should first solve it in $Z/pZ = F_p$, then in $Z/p^2 Z$, $Z/p^3 Z$, and so on. An important condition under which a solution in $F_p$ can be "lifted" to a solution in $Z_p$ is given by

Hensel's Lemma. <u>Suppose that</u> $f(x) \in Z_p[x]$, $f(a_0) \equiv 0 \pmod p$, <u>and</u> $f'(a_0) \not\equiv 0 \pmod p$ (<u>here</u> $f'$ <u>is the formal derivative of the polynomial</u> $f$). <u>Then there exists a unique</u> $x = a_0 + \ldots \in Z_p$ <u>such that</u> $f(x) = 0$.

Hensel's Lemma is proved by Newton's method for approximating roots (see [59,53]).

For example, when $f(x) = x^{p-1} - 1$, any $a_0 \in \{1,\ldots,p-1\}$ satisfies $f(a_0) \equiv 0 \pmod p$, while $f'(a_0) = (p-1)a_0^{p-2} \not\equiv 0 \pmod p$; so Hensel's Lemma tells us that $a_0$ has a unique Teichmüller representative $\omega(a_0) \in Z_p^*$.

Unlike in the case of $R$, whose algebraic closure $C$ is only a quadratic extension, $Q_p$ has algebraic extensions of arbitrary