Klaus Deimling

# Nonlinear Functional Analysis

With 35 Figures

Springer-Verlag
World Publishing Corporation

# Nonlinear Functional Analysis

With 35 Figures

Springer-Verlag
World Publishing Corporation

Klaus Deimling
Gesamthochschule Paderborn
Postfach 1621
D-4790 Paderborn
Federal Republic of Germany

Reprinted by World Publishing Corporation, Beijing, 1988
for distribution and sale in The People's Republic of China only

AMS Subject Classification (1980):
47Hxx, 58-01, 58C30, 58Cxx

ISBN 3-540-13928-1
Springer-Verlag Berlin Heidelberg New York Tokyo
ISBN 0-387-13928-1
Springer-Verlag New York Heidelberg Berlin Tokyo
ISBN 7-5062-0260-3

Library of Congress Cataloging in Publication Data.
Deimling, Klaus, 1943–. Nonlinear functional analysis.
Bibliography: p. Includes index.
1. Nonlinear functional analysis. I. Title.
QA320.D4 1985 515.7 84-26880
ISBN 0-387-13928-1 (U.S.)


Printed in Germany
Typesetting: Daten- und Lichtsatz-Service, Würzburg
Printing and binding: Graphischer Betrieb, Konrad Triltsch, Würzburg
2141/3140-543210

German intellect is an excellent thing, but when a German product is presented it must be analyzed. Most probably it is a combination of intellect ($I$) and tobacco-smoke ($T$). In many cases metaphysics ($M$) occurs and I hold that $I_a T_b M_c$ never occurs without $b + c > 2a$.

Augustus de Morgan

# Preface

Dear Reader,

The title tells you that this book deals with 'nonlinear functional analysis'. Roughly speaking, (linear) functional analysis is that mathematical discipline which is concerned with infinite-dimensional topological vector spaces, a fruitful combination of linear and topological structure, and the study of mappings between such spaces which respect these structures, i.e. linear maps that are somehow linked with the topologies of the spaces – continuous linear maps in the simplest case. Originally functional analysis could be understood as a unifying and abstract treatment of important aspects of linear mathematical models for problems in science, but the latter receded more and more into the background during the intensive theoretical investigations. It was clear from the start that most of the linear models are in fact only first approximations to models involving nonlinear maps. But given that some classes of linear topological spaces had already been basically understood, it was of course more natural to study linear maps, and this was further justified by the fact that not a few natural phenomena can be explained by linearization of nonlinear models. Thus, except for a fruitful period in the 1930s, the abstract treatment of the latter remained in the shade of the linear theory until a real boom started in the 1960s. Since then the existing methods, which had existed for thirty years or more, have been considerably extended, mainly motivated by new types of problems appearing also in nonclassical fields of application such as biology, chemistry or economics, and many new concepts and methods have been developed. Today some of these theories are well established and have almost reached their boundaries while others are still the subject of much activity. The purpose of the book is therefore to present a survey of the main elementary ideas, concepts and methods which constituted nonlinear functional analysis so far.

To explain what we understand by 'elementary', let us first remark that we have tried to present things in such a way that a graduate student can understand not only what is formally going on but also the spirit of the whole subject and its relations to adjacent parts of mathematics; so it is clear that one has to invest some more labour and time than for a conventional introduction to one of the special topics. However, only a modest preliminary knowledge is needed. In the first chapter, where we introduce an important topological concept, the so-called topological degree for continuous maps from subsets of $\mathbb{R}^n$ into $\mathbb{R}^n$, you need not know anything about functional analysis. Starting with Chapter 2, where infinite dimensions first appear, one should be familiar with the essential step of considering a sequence or a function of some sort as a point in the corresponding vector space of all such sequences or functions, whenever this abstraction is worthwhile. One should also work out the things which are proved in § 7 and accept certain basic principles of linear functional analysis quoted there for easier references, until they are applied in later chapters. In other words, even the 'completely linear' sections which we have included for your convenience serve only as a vehicle for progress in nonlinearity.

Another point that makes the text introductory is the use of an essentially uniform mathematical language and way of thinking, one which is no doubt familiar from elementary lectures in analysis that did not worry much about its connections with algebra and topology. Of course we shall use some elementary topological concepts, which may be new, but in fact only a few remarks here and there pertain to algebraic or differential topological concepts and methods. This will become clear as early as the first chapter (where an introduction, on the same level, of the basic concepts of algebraic topology needed for degree theory and some other ideas, would have taken at least as much space) but also in later chapters, say in § 27, where we deal with certain manifolds yet hardly use the language of the professionals in the field. This explains why we have described the topological concepts used as 'elementary', although we could have similarly described those ideas and concepts from algebraic or differential topology which have been used so far in nonlinear functional analysis, if we had chosen to begin with a different introductory chapter. We will come back to this remark in the epilogue.

Finally, let us mention a few things about 'examples' and 'applications'. As in the linear case, nonlinear functional analysis starts with the inspection of various types of equations or questions arising in nonlinear models for problems in, for example, natural science. Observing a phenomenon shown by such diverse problems, we may be led to introduce a certain class of nonlinear maps on a certain class of subsets of a certain class of Banach spaces. This class will then be studied by, say, analytical, topological or geometric means, first with regard to the phenomenon, but then also for purely theoretical reasons and for interest. Without saying more, it is clear that a book on this subject must contain examples of models, examples illustrating concepts and methods, and examples illustrating how the abstract results can be applied to the questions arising in a 'concrete' model or in other abstract contexts. In almost all cases we have deliberately chosen the simplest significant class of concrete equations or problems to which an abstract result applies.

Having explained for which reasons the book was written and what is needed to understand it, let us explain how it is organized. There are thirty sections arranged in ten groups called chapters. Every chapter has an introduction which explains what you will find there and how it is related to earlier chapters. It is necessary but of course not sufficient to read these introductions. Every section ends with final remarks and exercises. Some of these remarks will become clearer when you see them in the context of final remarks to later sections. The exercises range from almost obvious to by no means obvious. Only the major concepts are recorded in definitions, others can be rediscovered by means of the index. References are indicated by names followed by numbers in square brackets which you find in the bibliography. The latter contains most of the relevant books, lecture notes and survey articles up to date, but the selection of other research papers is more personal. The numbering of theorems etc. is evident: for example Theorem 15.8 means Theorem 8 in § 15.

Now knowing that writing a book is a waste of time unless somebody is going to publish it and that the long road from the first handwritten version to the final form of the manuscript could not be managed without considerable help from others, I have great pleasure in thanking the publishers for fruitful collaboration; Mr. Alan Whittle for his hard work in replacing a lot of Germanisms by (sometimes too) proper English; Mrs. Walburga Kropp for typing the manuscript even with enthusiasm and never grumbling at a lot of changes; my wife Brigitte for preparing the index and designing the bifurcation ghost (Fig. 29.1); Dipl. Math. Dieter Päschke for drawing the figures and reading proofs; colleagues who send me re- and preprints. I am especially grateful to Drs. Sönke Hansen, Harald Mönch and Jan Prüß for a lot of discussions and helpful suggestions which considerably improved the content of the book.

Paderborn, autumn 1984 Klaus Deimling

# Contents

Everything should be made as simple as possible,
but not simpler.

Albert Einstein

When a mathematician has no more ideas,
he pursues axiomatics.

Felix Klein

I hope, good luck lies in odd numbers ...
They say, there is divinity in odd numbers,
either in nativity, chance, or death.

William Shakespeare

# Chapter 1. Topological Degree in Finite Dimensions

In this basic chapter we shall study some basic problems concerning equations of the form $f(x) = y$, where $f$ is a continuous map from a subset $\Omega \subset \mathbb{R}^n$ into $\mathbb{R}^n$ and $y$ is a given point in $\mathbb{R}^n$. First of all we want to know whether such an equation has at least one solution $x \in \Omega$. If this is the case for some equation, we are then interested in the question of whether this solution is unique or not. We then also want to decide how the solutions are distributed in $\Omega$. Once we have some answers for a particular equation, we need also to study whether these answers remain the same or change drastically if we change $f$ and $y$ in some way. It is most probable that you have already been confronted, more or less explicitly, by all these questions at this stage in your mathematical development.

Let us review, for example, the problem of finding the zeros of a polynomial. First we learn that a real polynomial need not have a real zero. Then we are taught that a real polynomial of odd degree, say $p_{2m+1}(t) = t^{2m+1} + p_{2m}(t)$, has a real zero, and you will recall the simple proof which exploits the fact that $p_{2m}(t)$ is 'negligible' relative to $t^{2m-1}$ for large $t$, and therefore $p_{2m+1}(t) > 0$ for $t \geqq r$ and $p_{2m+1}(t) < 0$ for $t \leqq -r$ with $r$ sufficiently large, which in turn implies that $p_{2m+1}$ has a zero in $(-r, r)$, by Bolzano's intermediate value theorem. Next we learn that every polynomial of degree $m \geqq 1$ has at least one zero in the complex plane $\mathbb{C}$. Then we introduce the multiplicity of a zero $z_0$. If this is $k$, then $z_0$ is counted $k$ times, and by means of this concept the more precise statement is arrived at that every polynomial of degree $m \geqq 1$ has exactly $m$ zeros in $\mathbb{C}$. At this stage the problem of finding the zeros of a polynomial over $\mathbb{C}$ is solved for the pure algebraist and he will turn to the same question for more general functions over more general structures. The 'practical' man, if he is fair, will appreciate that the 'pure' fellows have proved a nice theorem, but it does not satisfy his needs. Suppose that he is led to investigate the behaviour as $t \to \infty$ of solutions of a linear system $x' = Ax$ of ordinary differential equations, where $A$ is an $n \times n$ matrix. Then the information that the characteristic polynomial of $A$ has exactly $n$ zeros in $\mathbb{C}$, the eigenvalues of $A$, is not enough for him since he has to know whether they are in the left or right half plane or on the imaginary axis. In another situation he may have obtained his polynomial by interpolation of certain experimental data which usually contain some hopefully small errors. Then he may need to know that the zeros of polynomials close to $p$ are close to the zeros of $p$.

Now, we want to construct a tool, the topological degree of $f$ with respect to $\Omega$ and $y$, which is very useful in the investigation of the problems mentioned at the beginning. To motivate the process, let us recall the winding number of plane curves and its connection with theorems on zeros of analytic functions. If you missed this topic in an elementary course in complex analysis, you may either consult Ahlfors [1], Dieudonné [1], Krasnoselskii et al. [1], or believe in what we are going to mention in the sequel, since we shall indicate in § 6.6 how the winding number is related to the degree in the case of $\mathbb{R}^2$.

Let $\Gamma \subset \mathbb{C}$ be an oriented closed curve with the continuously differentiable ($C^1$ for short) representation $z(t)$ ($t \in [0,1]$, $z(0) = z(1)$) and let $a \in \mathbb{C} \backslash \Gamma$. Then, the integer

$$(1) \qquad w(\Gamma, a) = \frac{1}{2\pi i} \int_\Gamma \frac{dz}{z-a} = \frac{1}{2\pi} \int_0^1 \frac{x(t)\, y'(t) - x'(t)\, y(t)}{x^2(t) + y^2(t)}\, dt$$
$$\text{for } z(t) = x(t) + i y(t) + a$$

is called the *winding number* (or *index*) of $\Gamma$ with respect to $a \in \mathbb{C} \backslash \Gamma$, since it tells us how many times $\Gamma$ winds around $a$, roughly speaking. If $\Gamma$ is only continuous then we can approximate $\Gamma$ as closely as we wish by $C^1$-curves, and it is easy to see that all these approximations have the same winding number provided that they are sufficiently close to $\Gamma$. More precisely, if $z_1(t)$ and $z_2(t)$ are $C^1$-representations of the closed curves $\Gamma_1$ and $\Gamma_2$ with the same orientation as $\Gamma$ and are such that

$$\max\{|z_j(t) - z(t)| : t \in [0,1]\} < \min\{|a - z(t)| : t \in [0,1]\} \qquad \text{for } j = 1, 2$$

then $w(\Gamma_1, a) = w(\Gamma_2, a)$. Therefore, we can define $w(\Gamma, a)$ to be $w(\Gamma_1, a)$ for any such $\Gamma_1$. Then we have defined

$$w\colon \{(\Gamma, a) : \Gamma \text{ closed continuous, } a \in \mathbb{C} \backslash \Gamma\} \to \mathbb{Z}$$

and it is not hard to see that this function $w$ has the following properties:

(a) $w$ is continuous in $(\Gamma, a)$, i.e. constant in some neighbourhood of $(\Gamma, a)$.
(b) $w(\Gamma, \cdot)$ is constant on every connected component of $\mathbb{C} \backslash \Gamma$ – in particular, equal to zero on the unbounded component.
(c) If the curves $\Gamma_0$ and $\Gamma_1$ are homotopic in $\mathbb{C} \backslash \{a\}$, then $w(\Gamma_0, a) = w(\Gamma_1, a)$. More explicitly, let $z_0(t)$ and $z_1(t)$ be representations of $\Gamma_0$ and $\Gamma_1$ such that there exists a continuous $h\colon [0,1] \times [0,1] \to \mathbb{C} \backslash \{a\}$ satisfying $h(0,t) = z_0(t)$ and $h(1,t) = z_1(t)$ in $[0,1]$ and $h(s,0) = h(s,1)$ for every $s \in [0,1]$; then $w(\Gamma_s, a)$ is the same integer for all $s \in [0,1]$, where $\Gamma_s$ is the closed curve represented by $h(s, \cdot)$.
(d) If $\Gamma^-$ denotes the curve $\Gamma$ with its orientation reversed, then $w(\Gamma^-, a) = -w(\Gamma, a)$.

Property (c) is the most important one, since it allows us for example to calculate the winding number of a complicated curve by means of the winding number of a possibly simpler homotopic curve. Furthermore, (a) and (b) are simple consequences of (c).

Now, let $G \subset \mathbb{C}$ be a simply connected region, $f: G \to \mathbb{C}$ be analytic and $\Gamma \subset G$ be a closed $C^1$-curve such that $f(z) \neq 0$ on $\Gamma$. Then the 'argument principle' tells us that

$$(2) \qquad w(f(\Gamma), 0) = \frac{1}{2\pi i} \int_{f(\Gamma)} \frac{dz}{z} = \frac{1}{2\pi i} \int_{\Gamma} \frac{f'(z)}{f(z)}\, dz = \sum_k w(\Gamma, z_k)\, \alpha_k,$$

where the $z_k$ are the zeros of $f$ in the regions enclosed by $\Gamma$ and the $\alpha_k$ are the corresponding multiplicities. If we assume in addition that $\Gamma$ has positive orientation and no intersection points, then we know from Jordan's curve theorem, which will be proved in this chapter, that there is exactly one region $G_0 \subset G$ enclosed by $\Gamma$, and $w(\Gamma, z_0) = 1$ for every $z_0 \in G_0$. Thus, (2) becomes

$$w(f(\Gamma), 0) = \sum_k \alpha_k,$$

i.e. the total number of zeros of $f$ in $G_0$ is obtained by calculating the winding number of the image curve $f(\Gamma)$ with respect to 0. In general, $w(\Gamma, z_k)$ can also be negative and then we can only conclude that $f$ has at least $|w(f(\Gamma), 0)|$ zeros in the regions enclosed by $\Gamma$.

In the more general case of continuous maps from subsets of $\mathbb{R}^n$ into $\mathbb{R}^n$ we shall imitate these ideas. We consider open bounded subsets $\Omega \subset \mathbb{R}^n$ instead of the regions enclosed by $\Gamma$, continuous maps $f: \bar{\Omega} \to \mathbb{R}^n$ and points $y \in \mathbb{R}^n$ which do not belong to the image $f(\partial\Omega)$ of the boundary of $\Omega$. With each such 'admissible' triple $(f, \Omega, y)$ we associate an integer $d(f, \Omega, y)$ such that the properties of the function $d$ allow us to give significant answers to the questions raised at the beginning. Of course, as in daily life, we cannot achieve everything, but the following minimal requirements and their useful consequences turn out to be a good compromise.

The first condition is simply a normalization. If $f = \mathrm{id}$, the identity map of $\mathbb{R}^n$ defined by $\mathrm{id}(x) = x$, then $f(x) = y \in \Omega$ has the unique solution $x = y$, and therefore we require

$$(\mathrm{d}1) \qquad d(\mathrm{id}, \Omega, y) = 1 \quad \text{for } y \in \Omega.$$

The second condition is a natural formulation of the desire that $d$ should yield information on the location of solutions. Suppose that $\Omega_1$ and $\Omega_2$ are disjoint open subsets of $\Omega$ and suppose that $f(x) = y$ has finitely many solutions in $\Omega_1 \cup \Omega_2$ but no solution in $\bar{\Omega} \backslash (\Omega_1 \cup \Omega_2)$. Then the number of solutions in $\Omega$ is the sum of the numbers of solutions in $\Omega_1$ and $\Omega_2$, and this suggests that $d$ should be additive in its argument $\Omega$, that is

(d2) $\quad d(f, \Omega, y) = d(f, \Omega_1, y) + d(f, \Omega_2, y)$ whenever $\Omega_1$ and $\Omega_2$ are disjoint open subsets of $\Omega$ such that $y \notin f(\bar{\Omega} \backslash (\Omega_1 \cup \Omega_2))$.

The third and last condition reflects the desire that for complicated $f$ the number $d(f, \Omega, y)$ can be calculated by $d(g, \Omega, y)$ with simpler $g$, at least if $f$ can be continuously deformed into $g$ such that at no stage of the deformation we get solutions on the boundary. This leads to

(d 3) $d(h(t,\cdot),\Omega,y(t))$ is independent of $t\in J=[0,1]$ whenever $h\colon J\times\bar{\Omega}\to\mathbb{R}^n$ and $y\colon J\to\mathbb{R}^n$ are continuous and $y(t)\notin h(t,\partial\Omega)$ for all $t\in J$.

There are essentially two different approaches to the construction of such a function $d$. The older one uses only concepts from algebraic topology, which is quite natural, since (d 1)–(d 3) involve only topological concepts such as open sets and continuous maps and a 'little bit' theory of groups like $\mathbb{Z}$; see, for example, Alexandroff and Hopf [1], Cronin [2], Dold [2], Dugundji and Granas [1].

We shall present the more recent second approach which is simpler for 'true' analysts, not worrying much about topology and algebra, since it uses only some basic analytical tools such as the approximation theorem of K. Weierstraß, the implicit function theorem and the so-called lemma of Sard (see § 2). Presentations still using topological arguments can be found in books on differential topology, for example, in Guillemin and Pollack [1], Hirsch [1] and Milnor [2], while purely analytical versions have been given by Nagumo [1] and Heinz [1] in the 1950s. An interesting mixture of the two methods has been given in Peitgen and Siegberg [1] – an outgrowth of recent efforts in finding numerical approximations to degrees and fixed points, based on the observation that the essential steps of the old method can be put into the form of algorithms.

In principle, it is an inessential question how we introduce degree theory, since there is only one $\mathbb{Z}$-valued function $d$ satisfying (d 1)–(d 3), as you will see in § 1, and since it are the properties of $d$ which count, as you will see throughout this chapter. Starting with the uniqueness of $d$, by exploiting (d 1)–(d 3) until we end up with the simplest case $f(x)=Ax$ with $\det A\neq 0$, has the advantage that the basic formula, which a purely analytical definition has to start with, does not fall from heaven – it is enough that the natural numbers do (according to L. Kronecker) – and that we are already motivated to introduce some prerequisites which we need anyway later on. However, you will keep in mind that choosing the analytical approach we lose topological insight to a considerable extent, while going through the mill of the elements of combinatorial topology you will hardly become aware of the fact that the same goal can be arrived at so simply by an analytical procedure. Thus, the essential question is why we introduce degree theory, but this has already been answered by the general remarks given in the foreword and the more special ones in this introduction which we are going to close by a few historical remarks.

The winding number is a very old concept. Its essentials can already be found in papers of C. F. Gauß and A. L. Cauchy at the beginning of the 19th century. Later on L. Kronecker, J. Hadamard, H. Poincaré and others extended formula (1) by consideration of integrals of differentiable maps over $\{x\in\mathbb{R}^n\colon |x|=1\}$. Finally, L. E. J. Brouwer established the degree for continuous maps in 1912. It is now tradition to speak of the Brouwer degree. The way towards an analytical definition was paved by A. Sard's investigation of the measure of the critical values of differentiable maps in 1942. You will find much more in the interesting papers of Siegberg [1], [2].

## § 1. Uniqueness of the Degree

In this section we shall show that there is only one function

$$d: \{(f, \Omega, y): \Omega \subset \mathbb{R}^n \text{ open and bounded, } f: \bar{\Omega} \to \mathbb{R}^n \text{ continuous,}$$
$$y \in \mathbb{R}^n \backslash f(\partial\Omega)\} \to \mathbb{Z}$$

satisfying

(d 1) $d(\mathrm{id}, \Omega, y) = 1$ for $y \in \Omega$

(d 2) $d(f, \Omega, y) = d(f, \Omega_1, y) + d(f, \Omega_2, y)$ whenever $\Omega_1, \Omega_2$ are disjoint open subsets of $\Omega$ such that $y \notin f(\bar{\Omega} \backslash (\Omega_1 \cup \Omega_2))$.

(d 3) $d(h(t, \cdot), \Omega, y(t))$ is independent of $t \in J = [0, 1]$ whenever $h: J \times \bar{\Omega} \to \mathbb{R}^n$ is continuous, $y: J \to \mathbb{R}^n$ is continuous and $y(t) \notin h(t, \partial\Omega)$ for all $t \in J$.

This will be done by reduction to more agreeable conditions, the final one being the case where $f$ is linear, i.e. $f(x) = Ax$ with $\det A \neq 0$. During the simplifying process we introduce basic tools which are also needed for the construction of the function $d$ in § 2, and you will see already here that the homotopy invariance $(d\,3)$ of $d$ is a very powerful property.

Let us start with some notation for the whole chapter.

**1.1 Notation.** We let $\mathbb{R}^n = \{x = (x_1, \dots, x_n): x_i \in \mathbb{R}$ for $i = 1, \dots, n\}$ with $|x| = \left(\sum_{i=1}^{n} x_i^2\right)^{1/2}$. For subsets $A \subset \mathbb{R}^n$ we use the usual symbols $\bar{A}$, $\partial A$ to denote the closure and the boundary of $A$, respectively. If also $B \subset \mathbb{R}^n$ then $B \backslash A = \{x \in B : x \notin A\}$, which may be the empty set $\emptyset$. The open and the closed ball of centre $x_0$ and radius $r > 0$ will be denoted by

$$B_r(x_0) = \{x \in \mathbb{R}^n: |x - x_0| < r\} = x_0 + B_r(0) \quad \text{and} \quad \bar{B}_r(x_0) = \overline{B_r(x_0)}.$$

Unless otherwise stated, $\Omega$ is always an open bounded subset of $\mathbb{R}^n$.

For maps $f: A \subset \mathbb{R}^n \to \mathbb{R}^n$ we let $f(A) = \{f(x): x \in A\}$ and $f^{-1}(y) = \{x \in A: f(x) = y\}$. The identity of $\mathbb{R}^n$ is denoted by id, i.e. $\mathrm{id}(x) = x$ for all $x \in \mathbb{R}^n$. Linear maps will be identified with their matrix $A = (a_{ij})$ and we write $\det A$ for the determinant of $A$. We shall also use *L. Kronecker's symbol* $\delta_{ij}$, defined by $\delta_{ij} = 1$ for $i = j$ and $\delta_{ij} = 0$ for $i \neq j$, so that $\mathrm{id} = (\delta_{ij})$. If $B \subset \mathbb{R}^n$ is compact, i.e. closed and bounded, then $C(B)$ is the space of continuous $f: B \to \mathbb{R}^n$, and we let $|f|_0 = \max_B |f(x)|$ for $f \in C(B)$. We shall write $f \in C(B; \mathbb{R}^m)$ to emphasize $f(B) \subset \mathbb{R}^m$, if necessary.

You will recall that $f: \Omega \to \mathbb{R}^n$ is said to be differentiable at $x_0$ if there is a matrix $f'(x_0)$ such that

$$f(x_0 + h) = f(x_0) + f'(x_0)\,h + \omega(h) \qquad \text{for } h \in \Omega - x_0 = \{x - x_0 : x \in \Omega\}$$

where the remainder $\omega(h)$ satisfies $|\omega(h)| \leq \varepsilon\,|h|$ for $|h| \leq \delta = \delta(\varepsilon, x_0)$. In this case $f'(x_0)_{ij} = \partial_j f_i(x_0) = \partial f_i(x_0)/\partial x_j$, the partial derivative of the $i$th component $f_i$ with respect to $x_j$.