Nonlinear Functional Analysis

Gerald Teschl

Gerald Teschl
Fakultät für Mathematik
Nordbergstraße 15
Universität Wien
1090 Wien, Austria

E-mail address: `Gerald.Teschl@univie.ac.at`
URL: `http://www.mat.univie.ac.at/~gerald/`

1991 Mathematics subject classification. 46-01, 47H10, 47H11, 58Fxx, 76D05

Abstract. This manuscript provides a brief introduction to nonlinear functional analysis.

We start out with calculus in Banach spaces, review differentiation and integration, derive the implicit function theorem (using the uniform contraction principle) and apply the result to prove existence and uniqueness of solutions for ordinary differential equations in Banach spaces.

Next we introduce the mapping degree in both finite (Brouwer degree) and infinite dimensional (Leray-Schauder degree) Banach spaces. Several applications to game theory, integral equations, and ordinary differential equations are discussed.

As an application we consider partial differential equations and prove existence and uniqueness for solutions of the stationary Navier-Stokes equation.

Finally, we give a brief discussion of monotone operators.

Keywords and phrases. Mapping degree, fixed-point theorems, differential equations, Navier–Stokes equation.

Typeset by LaTeX and Makeindex.
Version: October 13, 2005
Copyright © 1998-2004 by Gerald Teschl

Preface

The present manuscript was written for my course *Nonlinear Functional Analysis* held at the University of Vienna in Summer 1998 and 2001. It is supposed to give a brief introduction to the field of Nonlinear Functional Analysis with emphasis on applications and examples. The material covered is highly selective and many important and interesting topics are not covered.

It is available from

`http://www.mat.univie.ac.at/~gerald/ftp/book-nlfa/`

Acknowledgments

I'd like to thank Volker Enß for making his lecture notes available to me and Matthias Hammerl for pointing out errors in previous versions.

Gerald Teschl

Vienna, Austria
February 2001

Contents

Preface **iii**

1 Analysis in Banach spaces **1**
- 1.1 Differentiation and integration in Banach spaces 1
- 1.2 Contraction principles 5
- 1.3 Ordinary differential equations 8

2 The Brouwer mapping degree **11**
- 2.1 Introduction 11
- 2.2 Definition of the mapping degree and the determinant formula . . . 13
- 2.3 Extension of the determinant formula 17
- 2.4 The Brouwer fixed-point theorem 24
- 2.5 Kakutani's fixed-point theorem and applications to game theory . . 25
- 2.6 Further properties of the degree 29
- 2.7 The Jordan curve theorem 31

3 The Leray–Schauder mapping degree **33**
- 3.1 The mapping degree on finite dimensional Banach spaces 33
- 3.2 Compact operators 34
- 3.3 The Leray–Schauder mapping degree 35
- 3.4 The Leray–Schauder principle and the Schauder fixed-point theorem 37
- 3.5 Applications to integral and differential equations 39

4 The stationary Navier–Stokes equation **43**
- 4.1 Introduction and motivation 43
- 4.2 An insert on Sobolev spaces 44
- 4.3 Existence and uniqueness of solutions 50

5 Monotone operators 53
5.1 Monotone operators 53
5.2 The nonlinear Lax–Milgram theorem 55
5.3 The main theorem of monotone operators 57

Bibliography 61

Glossary of notations 63

Index 65

Chapter 1

Analysis in Banach spaces

1.1 Differentiation and integration in Banach spaces

We first review some basic facts from calculus in Banach spaces.

Let X and Y be two Banach spaces and denote by $C(X,Y)$ the set of continuous functions from X to Y and by $\mathcal{L}(X,Y) \subset C(X,Y)$ the set of (bounded) linear functions. Let U be an open subset of X. Then a function $F: U \to Y$ is called differentiable at $x \in U$ if there exists a linear function $dF(x) \in \mathcal{L}(X,Y)$ such that

$$F(x+u) = F(x) + dF(x)\, u + o(u), \tag{1.1}$$

where o, O are the Landau symbols. The linear map $dF(x)$ is called derivative of F at x. If F is differentiable for all $x \in U$ we call F differentiable. In this case we get a map

$$\begin{array}{rccc} dF: & U & \to & \mathcal{L}(X,Y) \\ & x & \mapsto & dF(x) \end{array} . \tag{1.2}$$

If dF is continuous, we call F continuously differentiable and write $F \in C^1(U,Y)$.

Let $Y = \prod_{j=1}^m Y_j$ and let $F: X \to Y$ be given by $F = (F_1, \dots, F_m)$ with $F_j : X \to Y_i$. Then $F \in C^1(X,Y)$ if and only if $F_j \in C^1(X, Y_j)$, $1 \le j \le m$, and in this case $dF = (dF_1, \dots, dF_m)$. Similarly, if $X = \prod_{i=1}^m X_i$, then one can define the partial derivative $\partial_i F \in \mathcal{L}(X_i, Y)$, which is the derivative of F considered as a function of the i-th variable alone (the other variables being fixed). We have $dF\, v = \sum_{i=1}^n \partial_i F\, v_i$, $v = (v_1, \dots, v_n) \in X$, and $F \in C^1(X,Y)$ if and only if all partial derivatives exist and are continuous.

In the case of $X = \mathbb{R}^m$ and $Y = \mathbb{R}^n$,the matrix representation of dF with respect to the canonical basis in $\mathbb{R}^m$ and $\mathbb{R}^n$ is given by the partial derivatives $\partial_i F_j(x)$ and is called Jacobi matrix of F at x.

We can iterate the procedure of differentiation and write $F \in C^r(U,Y)$, $r \geq 1$, if the r-th derivative of F, $d^r F$ (i.e., the derivative of the $(r-1)$-th derivative of F), exists and is continuous. Finally, we set $C^\infty(U,Y) = \bigcap_{r\in\mathbb{N}} C^r(U,Y)$ and, for notational convenience, $C^0(U,Y) = C(U,Y)$ and $d^0 F = F$.

It is often necessary to equip $C^r(U,Y)$ with a norm. A suitable choice is

$$|F| = \max_{0\leq j\leq r} \sup_{x\in U} |d^j F(x)|. \tag{1.3}$$

The set of all r times continuously differentiable functions for which this norm is finite forms a Banach space which is denoted by $C_b^r(U,Y)$.

If F is bijective and F, F^{-1} are both of class C^r, $r \geq 1$, then F is called a diffeomorphism of class C^r.

Note that if $F \in \mathcal{L}(X,Y)$, then $dF(x) = F$ (independent of x) and $d^r F(x) = 0$, $r > 1$.

For the composition of mappings we note the following result (which is easy to prove).

Lemma 1.1 (Chain rule) *Let $F \in C^r(X,Y)$ and $G \in C^r(Y,Z)$, $r \geq 1$. Then $G \circ F \in C^r(X,Z)$ and*

$$d(G \circ F)(x) = dG(F(x)) \circ dF(x), \qquad x \in X. \tag{1.4}$$

In particular, if $\lambda \in Y^*$ is a linear functional, then $d(\lambda \circ F) = d\lambda \circ dF = \lambda \circ dF$. In addition, we have the following mean value theorem.

Theorem 1.2 (Mean value) *Suppose $U \subseteq X$ and $F \in C^1(U,Y)$. If U is convex, then*

$$|F(x) - F(y)| \leq M|x-y|, \qquad M = \max_{0\leq t\leq 1} |dF((1-t)x + ty)|. \tag{1.5}$$

Conversely, (for any open U) if

$$|F(x) - F(y)| \leq M|x-y|, \qquad x,y \in U, \tag{1.6}$$

then

$$\sup_{x\in U} |dF(x)| \leq M. \tag{1.7}$$

Proof. Abbreviate $f(t) = F((1-t)x + ty)$, $0 \leq t \leq 1$, and hence $df(t) = dF((1-t)x + ty)(y-x)$ implying $|df(t)| \leq \tilde{M} = M|x-y|$. For the first part it suffices to show

$$\phi(t) = |f(t) - f(0)| - (\tilde{M} + \delta)t \leq 0 \tag{1.8}$$

for any $\delta > 0$. Let $t_0 = \max\{t \in [0,1] | \phi(t) \leq 0\}$. If $t_0 < 1$ then

$$\begin{aligned} \phi(t_0 + \varepsilon) &= |f(t_0+\varepsilon) - f(t_0) + f(t_0) - f(0)| - (\tilde{M}+\delta)(t_0+\varepsilon) \\ &\leq |f(t_0+\varepsilon) - f(t_0)| - (\tilde{M}+\delta)\varepsilon + \phi(t_0) \\ &\leq |df(t_0)\varepsilon + o(\varepsilon)| - (\tilde{M}+\delta)\varepsilon \\ &\leq (\tilde{M} + o(1) - \tilde{M} - \delta)\varepsilon = (-\delta + o(1))\varepsilon \leq 0, \end{aligned} \tag{1.9}$$

for $\varepsilon \geq 0$, small enough. Thus $t_0 = 1$.

To prove the second claim suppose there is an $x_0 \in U$ such that $|dF(x_0)| = M + \delta$, $\delta > 0$. Then we can find an $e \in X$, $|e| = 1$ such that $|dF(x_0)e| = M + \delta$ and hence

$$\begin{aligned} M\varepsilon &\geq |F(x_0 + \varepsilon e) - F(x_0)| = |dF(x_0)(\varepsilon e) + o(\varepsilon)| \\ &\geq (M+\delta)\varepsilon - |o(\varepsilon)| > M\varepsilon \end{aligned} \tag{1.10}$$

since we can assume $|o(\varepsilon)| < \varepsilon\delta$ for $\varepsilon > 0$ small enough, a contradiction. □

As an immediate consequence we obtain

Corollary 1.3 *Suppose U is a connected subset of a Banach space X. A mapping $F \in C^1(U,Y)$ is constant if and only if $dF = 0$. In addition, if $F_{1,2} \in C^1(U,Y)$ and $dF_1 = dF_2$, then F_1 and F_2 differ only by a constant.*

Next we want to look at higher derivatives more closely. Let $X = \prod_{i=1}^m X_i$, then $F : X \to Y$ is called multilinear if it is linear with respect to each argument.

It is not hard to see that F is continuous if and only if

$$|F| = \sup_{x:\prod_{i=1}^m |x_i|=1} |F(x_1, \dots, x_m)| < \infty. \tag{1.11}$$

If we take n copies of the same space, the set of multilinear functions $F : X^n \to Y$ will be denoted by $\mathcal{L}^n(X,Y)$. A multilinear function is called symmetric provided its value remains unchanged if any two arguments are switched. With the norm from above it is a Banach space and in fact there is a canonical isometric isomorphism between $\mathcal{L}^n(X,Y)$ and $\mathcal{L}(X, \mathcal{L}^{n-1}(X,Y))$ given by $F : (x_1, \dots, x_n) \mapsto$

$F(x_1, \dots, x_n)$ maps to $x_1 \mapsto F(x_1, .)$. In addition, note that to each $F \in \mathcal{L}^n(X, Y)$ we can assign its polar form $F \in C(X, Y)$ using $F(x) = F(x, \dots, x)$, $x \in X$. If F is symmetric it can be reconstructed from its polar form using

$$F(x_1, \dots, x_n) = \frac{1}{n!} \partial_{t_1} \cdots \partial_{t_n} F(\sum_{i=1}^{n} t_i x_i)|_{t_1 = \cdots = t_n = 0}. \tag{1.12}$$

Moreover, the r-th derivative of $F \in C^r(X, Y)$ is symmetric since,

$$d^r F_x(v_1, \dots, v_r) = \partial_{t_1} \cdots \partial_{t_r} F(x + \sum_{i=1}^{r} t_i v_i)|_{t_1 = \cdots = t_r = 0}, \tag{1.13}$$

where the order of the partial derivatives can be shown to be irrelevant.

Now we turn to integration. We will only consider the case of mappings $f : I \to X$ where $I = [a, b] \subset \mathbb{R}$ is a compact interval and X is a Banach space. A function $f : I \to X$ is called simple if the image of f is finite, $f(I) = \{x_i\}_{i=1}^n$, and if each inverse image $f^{-1}(x_i)$, $1 \le i \le n$ is a Borel set. The set of simple functions $S(I, X)$ forms a linear space and can be equipped with the sup norm. The corresponding Banach space obtained after completion is called the set of regulated functions $R(I, X)$.

Observe that $C(I, X) \subset R(I, X)$. In fact, consider $f_n = \sum_{i=0}^{n-1} f(t_i) \chi_{[t_i, t_{i+1})} \in S(I, X)$, where $t_i = a + i\frac{b-a}{n}$ and χ is the characteristic function. Since $f \in C(I, X)$ is uniformly continuous, we infer that f_n converges uniformly to f.

For $f \in S(I, X)$ we can define a linear map $\int : S(I, X) \to X$ by

$$\int_a^b f(t) dt = \sum_{i=1}^{n} x_i \mu(f^{-1}(x_i)), \tag{1.14}$$

where μ denotes the Lebesgue measure on I. This map satisfies

$$\int_a^b f(t) dt \le |f|(b - a). \tag{1.15}$$

and hence it can be extended uniquely to a linear map $\int : R(I, X) \to X$ with the same norm $(b - a)$. We even have

$$\int_a^b f(t) dt \le \int_a^b |f(t)| dt. \tag{1.16}$$

In addition, if $\lambda \in X^*$ is a continuous linear functional, then

$$\lambda(\int_a^b f(t)dt) = \int_a^b \lambda(f(t))dt, \qquad f \in R(I,X). \tag{1.17}$$

We use the usual conventions $\int_{t_1}^{t_2} f(s)ds = \int_a^b \chi_{(t_1,t_2)}(s)f(s)ds$ and $\int_{t_2}^{t_1} f(s)ds = -\int_{t_1}^{t_2} f(s)ds$.

If $I \subseteq \mathbb{R}$, we have an isomorphism $\mathcal{L}(I,X) \equiv X$ and if $F : I \to X$ we will write $\dot{F}(t)$ in stead of $dF(t)$ if we regard $dF(t)$ as an element of X. In particular, if $f \in C(I,X)$, then $F(t) = \int_a^t f(s)ds \in C^1(I,X)$ and $\dot{F}(t) = f(t)$ as can be seen from

$$|\int_a^{t+\varepsilon} f(s)ds - \int_a^t f(s)ds - f(t)\varepsilon| = |\int_t^{t+\varepsilon}(f(s) - f(t))ds| \le |\varepsilon| \sup_{s\in[t,t+\varepsilon]} |f(s) - f(t)|. \tag{1.18}$$

This even shows that $F(t) = F(a) + \int_a^t (\dot{F}(s))ds$ for any $F \in C^1(I,X)$.

1.2 Contraction principles

A fixed point of a mapping $F : C \subseteq X \to C$ is an element $x \in C$ such that $F(x) = x$. Moreover, F is called a contraction if there is a contraction constant $\theta \in [0,1)$ such that

$$|F(x) - F(\tilde{x})| \le \theta|x - \tilde{x}|, \qquad x, \tilde{x} \in C. \tag{1.19}$$

Note that a contraction is continuous. We also recall the notation $F^n(x) = F(F^{n-1}(x))$, $F^0(x) = x$.

Theorem 1.4 (Contraction principle) *Let C be a closed subset of a Banach space X and let $F : C \to C$ be a contraction, then F has a unique fixed point $\overline{x} \in C$ such that*

$$|F^n(x) - \overline{x}| \le \frac{\theta^n}{1-\theta}|F(x) - x|, \qquad x \in C. \tag{1.20}$$

Proof. If $x = F(x)$ and $\tilde{x} = F(\tilde{x})$, then $|x - \tilde{x}| = |F(x) - F(\tilde{x})| \le \theta|x - \tilde{x}|$ shows that there can be at most one fixed point.

Concerning existence, fix $x_0 \in C$ and consider the sequence $x_n = F^n(x_0)$. We have

$$|x_{n+1} - x_n| \le \theta|x_n - x_{n-1}| \le \cdots \le \theta^n|x_1 - x_0| \tag{1.21}$$

and hence by the triangle inequality (for $n > m$)

$$\begin{aligned} |x_n - x_m| &\le \sum_{j=m+1}^{n} |x_j - x_{j-1}| \le \theta^m \sum_{j=0}^{n-m-1} \theta^j |x_1 - x_0| \\ &\le \frac{\theta^m}{1-\theta} |x_1 - x_0|. \end{aligned} \tag{1.22}$$

Thus x_n is Cauchy and tends to a limit $\overline{x}$. Moreover,

$$|F(\overline{x}) - \overline{x}| = \lim_{n\to\infty} |x_{n+1} - x_n| = 0 \tag{1.23}$$

shows that $\overline{x}$ is a fixed point and the estimate (1.20) follows after taking the limit $n \to \infty$ in (1.22). □

Next, we want to investigate how fixed points of contractions vary with respect to a parameter. Let $U \subseteq X$, $V \subseteq Y$ be open and consider $F : \overline{U} \times V \to U$. The mapping F is called a uniform contraction if there is a $\theta \in [0, 1)$ such that

$$|F(x, y) - F(\tilde{x}, y)| \le \theta |x - \tilde{x}|, \quad x, \tilde{x} \in \overline{U},\ y \in V. \tag{1.24}$$

Theorem 1.5 (Uniform contraction principle) *Let U, V be open subsets of Banach spaces X, Y, respectively. Let $F : \overline{U} \times V \to U$ be a uniform contraction and denote by $\overline{x}(y) \in U$ the unique fixed point of $F(., y)$. If $F \in C^r(U \times V, U)$, $r \ge 0$, then $\overline{x}(.) \in C^r(V, U)$.*

Proof. Let us first show that $\overline{x}(y)$ is continuous. From

$$\begin{aligned} |\overline{x}(y+v) - \overline{x}(y)| &= |F(\overline{x}(y+v), y+v) - F(\overline{x}(y), y+v) \\ &\quad + F(\overline{x}(y), y+v) - F(\overline{x}(y), y)| \\ &\le \theta |\overline{x}(y+v) - \overline{x}(y)| + |F(\overline{x}(y), y+v) - F(\overline{x}(y), y)| \end{aligned} \tag{1.25}$$

we infer

$$|\overline{x}(y+v) - \overline{x}(y)| \le \frac{1}{1-\theta} |F(\overline{x}(y), y+v) - F(\overline{x}(y), y)| \tag{1.26}$$

and hence $\overline{x}(y) \in C(V, U)$. Now let $r = 1$ and let us formally differentiate $\overline{x}(y) = F(\overline{x}(y), y)$ with respect to y,

$$d\,\overline{x}(y) = \partial_x F(\overline{x}(y), y) d\,\overline{x}(y) + \partial_y F(\overline{x}(y), y). \tag{1.27}$$

Considering this as a fixed point equation $T(x', y) = x'$, where $T(., y) : \mathcal{L}(Y, X) \to \mathcal{L}(Y, X)$, $x' \mapsto \partial_x F(\overline{x}(y), y) x' + \partial_y F(\overline{x}(y), y)$ is a uniform contraction since we have

$|\partial_x F(\overline{x}(y), y)| \le \theta$ by Theorem 1.2. Hence we get a unique continuous solution $\overline{x}'(y)$. It remains to show

$$\overline{x}(y+v) - \overline{x}(y) - \overline{x}'(y)v = o(v). \tag{1.28}$$

Let us abbreviate $u = \overline{x}(y+v) - \overline{x}(y)$, then using (1.27) and the fixed point property of $\overline{x}(y)$ we see

$$\begin{aligned}(1 - \partial_x F(\overline{x}(y), y))(u - \overline{x}'(y)v) =& \\ = F(\overline{x}(y) + u, y + v) - F(\overline{x}(y), y) - \partial_x F(\overline{x}(y), y)u - \partial_y F(\overline{x}(y), y)v& \\ = o(u) + o(v)&\end{aligned} \tag{1.29}$$

since $F \in C^1(U \times V, U)$ by assumption. Moreover, $|(1 - \partial_x F(\overline{x}(y), y))^{-1}| \le (1-\theta)^{-1}$ and $u = O(v)$ (by (1.26)) implying $u - \overline{x}'(y)v = o(v)$ as desired.

Finally, suppose that the result holds for some $r - 1 \ge 1$. Thus, if F is C^r, then $\overline{x}(y)$ is at least C^{r-1} and the fact that $d\,\overline{x}(y)$ satisfies (1.27) implies $\overline{x}(y) \in C^r(V, U)$. □

As an important consequence we obtain the implicit function theorem.

Theorem 1.6 (Implicit function) *Let X, Y, and Z be Banach spaces and let U, V be open subsets of X, Y, respectively. Let $F \in C^r(U \times V, Z)$, $r \ge 1$, and fix $(x_0, y_0) \in U \times V$. Suppose $\partial_x F(x_0, y_0) \in \mathcal{L}(X, Z)$ is an isomorphism. Then there exists an open neighborhood $U_1 \times V_1 \subseteq U \times V$ of (x_0, y_0) such that for each $y \in V_1$ there exists a unique point $(\xi(y), y) \in U_1 \times V_1$ satisfying $F(\xi(y), y) = F(x_0, y_0)$. Moreover, the map ξ is in $C^r(V_1, Z)$ and fulfills*

$$d\xi(y) = -(\partial_x F(\xi(y), y))^{-1} \circ \partial_y F(\xi(y), y). \tag{1.30}$$

Proof. Using the shift $F \to F - F(x_0, y_0)$ we can assume $F(x_0, y_0) = 0$. Next, the fixed points of $G(x, y) = x - (\partial_x F(x_0, y_0))^{-1} F(x, y)$ are the solutions of $F(x, y) = 0$. The function G has the same smoothness properties as F and since $|\partial_x G(x_0, y_0)| = 0$, we can find balls U_1 and V_1 around x_0 and y_0 such that $|\partial_x G(x, y)| \le \theta < 1$. Thus $G(., y)$ is a uniform contraction and in particular, $G(U_1, y) \subset U_1$, that is, $G : U_1 \times V_1 \to U_1$. The rest follows from the uniform contraction principle. Formula (1.30) follows from differentiating $F(\xi(y), y) = 0$ using the chain rule. □

Note that our proof is constructive, since it shows that the solution $\xi(y)$ can be obtained by iterating $x - (\partial_x F(x_0, y_0))^{-1} F(x, y)$.

Moreover, as a corollary of the implicit function theorem we also obtain the inverse function theorem.

Theorem 1.7 (Inverse function) *Suppose* $F \in C^r(U,Y)$, $U \subseteq X$, *and let* $dF(x_0)$ *be an isomorphism for some* $x_0 \in U$. *Then there are neighborhoods* U_1, V_1 *of* x_0, $F(x_0)$, *respectively, such that* $F \in C^r(U_1, V_1)$ *is a diffeomorphism.*

Proof. Apply the implicit function theorem to $G(x,y) = y - F(x)$. □

1.3 Ordinary differential equations

As a first application of the implicit function theorem, we prove (local) existence and uniqueness for solutions of ordinary differential equations in Banach spaces.

The following lemma will be needed in the proof.

Lemma 1.8 *Suppose* $I \subseteq \mathbb{R}$ *is a compact interval and* $f \in C^r(U,Y)$. *Then* $f_* \in C^r(C_b(I,U), C_b(I,Y))$, *where*

$$(f_* x)(t) = f(x(t)). \tag{1.31}$$

Proof. Fix $x_0 \in C_b(I,U)$ and $\varepsilon > 0$. For each $t \in I$ we have a $\delta(t) > 0$ such that $|f(x) - f(x_0(t))| \le \varepsilon/2$ for all $x \in U$ with $|x - x_0(t)| \le 2\delta(t)$. The balls $B_{\delta(t)}(x_0(t))$, $t \in I$, cover the set $\{x_0(t)\}_{t\in I}$ and since I is compact, there is a finite subcover $B_{\delta(t_j)}(x_0(t_j))$, $1 \le j \le n$. Let $|x - x_0| \le \delta = \min_{1\le j\le n} \delta(t_j)$. Then for each $t \in I$ there is t_i such that $|x_0(t) - x_0(t_j)| \le \delta(t_j)$ and hence $|f(x(t)) - f(x_0(t))| \le |f(x(t)) - f(x_0(t_j))| + |f(x_0(t_j)) - f(x_0(t))| \le \varepsilon$ since $|x(t) - x_0(t_j)| \le |x(t) - x_0(t)| + |x_0(t) - x_0(t_j)| \le 2\delta(t_j)$. This settles the case $r = 0$.

Next let us turn to $r = 1$. We claim that df_* is given by $(df_*(x_0)x)(t) = df(x_0(t))x(t)$. Hence we need to show that for each $\varepsilon > 0$ we can find a $\delta > 0$ such that

$$\sup_{t\in I} |f_*(x_0(t) + x(t)) - f_*(x_0(t)) - df(x_0(t))x(t)| \le \varepsilon\delta \tag{1.32}$$

whenever $|x - x_0| \le \delta$. By assumption we have

$$|f_*(x_0(t) + x(t)) - f_*(x_0(t)) - df(x_0(t))x(t)| \le \varepsilon\delta(t) \tag{1.33}$$

whenever $|x(t) - x_0(t)| \le \delta(t)$. Now argue as before. It remains to show that df_* is continuous. To see this we use the linear map

$$\begin{array}{rccl} \lambda: & C_b(I,\mathcal{L}(X,Y)) & \to & \mathcal{L}(C_b(I,X), C_b(I,Y)) \;, \\ & T & \mapsto & T_* x \end{array} \tag{1.34}$$

where $(T_* x)(t) = T(t)x(t)$. Since we have

$$|T_* x| = \sup_{t\in I} |T(t)x(t)| \le \sup_{t\in I} |T(t)||x(t)| \le |T||x|, \tag{1.35}$$

we infer $|\lambda| \le 1$ and hence λ is continuous. Now observe $df_* = \lambda \circ (df)_*$.

The general case $r > 1$ follows from induction. □

Now we come to our existence and uniqueness result for the initial value problem in Banach spaces.

Theorem 1.9 *Let I be an open interval, U an open subset of a Banach space X and Λ an open subset of another Banach space. Suppose $F \in C^r(I \times U \times \Lambda, X)$, then the initial value problem*

$$\dot{x}(t) = F(t, x, \lambda), \qquad x(t_0) = x_0, \quad (t_0, x_0, \lambda) \in I \times U \times \Lambda, \tag{1.36}$$

has a unique solution $x(t, t_0, x_0, \lambda) \in C^r(I_1 \times I_2 \times U_1 \times \Lambda_1, X)$, where $I_{1,2}$, U_1, and Λ_1 are open subsets of I, U, and Λ, respectively. The sets I_2, U_1, and Λ_1 can be chosen to contain any point $t_0 \in I$, $x_0 \in U$, and $\lambda_0 \in \Lambda$, respectively.

Proof. If we shift $t \to t - t_0$, $x \to x - x_0$, and hence $F \to F(. + t_0, . + x_0, \lambda)$, we see that it is no restriction to assume $x_0 = 0$, $t_0 = 0$ and to consider (t_0, x_0) as part of the parameter λ (i.e., $\lambda \to (t_0, x_0, \lambda)$). Moreover, using the standard transformation $\dot{x} = F(\tau, x, \lambda)$, $\dot{\tau} = 1$, we can even assume that F is independent of t. We will also replace U by a smaller (bounded) subset such that F is uniformly continuous with respect to x on this subset.

Our goal is to invoke the implicit function theorem. In order to do this we introduce an additional parameter $\varepsilon \in \mathbb{R}$ and consider

$$\dot{x} = \varepsilon F(x, \lambda), \quad x \in D^{r+1} = \{x \in C_b^{r+1}((-1,1), U) | x(0) = 0\}, \tag{1.37}$$

such that we know the solution for $\varepsilon = 0$. The implicit function theorem will show that solutions still exist as long as ε remains small. At first sight this doesn't seem to be good enough for us since our original problem corresponds to $\varepsilon = 1$. But since ε corresponds to a scaling $t \to \varepsilon t$, the solution for one $\varepsilon > 0$ suffices. Now let us turn to the details.

Our problem (1.37) is equivalent to looking for zeros of the function

$$\begin{array}{llll} G: & D^{r+1} \times \Lambda \times (-\varepsilon_0, \varepsilon_0) & \to & C_b^r((-1,1), X) \\ & (x, \lambda, \varepsilon) & \mapsto & \dot{x} - \varepsilon F(x, \lambda) \end{array} . \tag{1.38}$$

Lemma 1.8 ensures that this function is C^r. Now fix λ_0, then $G(0,\lambda_0,0) = 0$ and $\partial_x G(0,\lambda_0,0) = T$, where $Tx = \dot{x}$. Since $(T^{-1}x)(t) = \int_0^t x(s)ds$ we can apply the implicit function theorem to conclude that there is a unique solution $x(\lambda,\varepsilon) \in C^r(\Lambda_1 \times (-\varepsilon_0,\varepsilon_0), D^{r+1})$. In particular, the map $(\lambda,t) \mapsto x(\lambda,\varepsilon)(t/\varepsilon)$ is in $C^r(\Lambda_1, C^{r+1}((-\varepsilon,\varepsilon),X)) \hookrightarrow C^r(\Lambda \times (-\varepsilon,\varepsilon),X)$. Hence it is the desired solution of our original problem. $\square$

Chapter 2

The Brouwer mapping degree

2.1 Introduction

Many applications lead to the problem of finding all zeros of a mapping $f : U \subseteq X \to X$, where X is some (real) Banach space. That is, we are interested in the solutions of

$$f(x) = 0, \quad x \in U. \tag{2.1}$$

In most cases it turns out that this is too much to ask for, since determining the zeros analytically is in general impossible.

Hence one has to ask some weaker questions and hope to find answers for them. One such question would be "Are there any solutions, respectively, how many are there?". Luckily, this questions allows some progress.

To see how, lets consider the case $f \in \mathcal{H}(\mathbb{C})$, where $\mathcal{H}(\mathbb{C})$ denotes the set of holomorphic functions on a domain $U \subset \mathbb{C}$. Recall the concept of the winding number from complex analysis. The winding number of a path $\gamma : [0,1] \to \mathbb{C}$ around a point $z_0 \in \mathbb{C}$ is defined by

$$n(\gamma, z_0) = \frac{1}{2\pi\mathrm{i}} \int_\gamma \frac{dz}{z - z_0} \in \mathbb{Z}. \tag{2.2}$$

It gives the number of times γ encircles z_0 taking orientation into account. That is, encirclings in opposite directions are counted with opposite signs.

In particular, if we pick $f \in \mathcal{H}(\mathbb{C})$ one computes (assuming $0 \notin f(\gamma)$)

$$n(f(\gamma), 0) = \frac{1}{2\pi\mathrm{i}} \int_\gamma \frac{f'(z)}{f(z)} dz = \sum_k n(\gamma, z_k)\alpha_k, \tag{2.3}$$

where z_k denote the zeros of f and α_k their respective multiplicity. Moreover, if γ is a Jordan curve encircling a simply connected domain $U \subset \mathbb{C}$, then $n(\gamma, z_k) = 0$ if $z_k \notin U$ and $n(\gamma, z_k) = 1$ if $z_k \in U$. Hence $n(f(\gamma), 0)$ counts the number of zeros inside U.

However, this result is useless unless we have an efficient way of computing $n(f(\gamma), 0)$ (which does not involve the knowledge of the zeros z_k). This is our next task.

Now, lets recall how one would compute complex integrals along complicated paths. Clearly, one would use homotopy invariance and look for a simpler path along which the integral can be computed and which is homotopic to the original one. In particular, if $f : \gamma \to \mathbb{C}\backslash\{0\}$ and $g : \gamma \to \mathbb{C}\backslash\{0\}$ are homotopic, we have $n(f(\gamma), 0) = n(g(\gamma), 0)$ (which is known as Rouchés theorem).

More explicitly, we need to find a mapping g for which $n(g(\gamma), 0)$ can be computed and a homotopy $H : [0, 1] \times \gamma \to \mathbb{C}\backslash\{0\}$ such that $H(0, z) = f(z)$ and $H(1, z) = g(z)$ for $z \in \gamma$. For example, how many zeros of $f(z) = \frac{1}{2}z^6 + z - \frac{1}{3}$ lie inside the unit circle? Consider $g(z) = z$, then $H(t, z) = (1-t)f(z) + t\,g(z)$ is the required homotopy since $|f(z) - g(z)| < |g(z)|$, $|z| = 1$, implying $H(t, z) \neq 0$ on $[0, 1] \times \gamma$. Hence $f(z)$ has one zero inside the unit circle.

Summarizing, given a (sufficiently smooth) domain U with enclosing Jordan curve ∂U, we have defined a degree $\deg(f, U, z_0) = n(f(\partial U), z_0) = n(f(\partial U) - z_0, 0) \in \mathbb{Z}$ which counts the number of solutions of $f(z) = z_0$ inside U. The invariance of this degree with respect to certain deformations of f allowed us to explicitly compute $\deg(f, U, z_0)$ even in nontrivial cases.

Our ultimate goal is to extend this approach to continuous functions $f : \mathbb{R}^n \to \mathbb{R}^n$. However, such a generalization runs into several problems. First of all, it is unclear how one should define the multiplicity of a zero. But even more severe is the fact, that the number of zeros is unstable with respect to small perturbations. For example, consider $f_\varepsilon : [-1, 2] \to \mathbb{R}$, $x \mapsto x^2 - \varepsilon$. Then f_ε has no zeros for $\varepsilon < 0$, one zero for $\varepsilon = 0$, two zeros for $0 < \varepsilon \leq 1$, one for $1 < \varepsilon \leq \sqrt{2}$, and none for $\varepsilon > \sqrt{2}$. This shows the following facts.

1. Zeros with $f' \neq 0$ are stable under small perturbations.
2. The number of zeros can change if two zeros with opposite sign change (i.e., opposite signs of f') run into each other.
3. The number of zeros can change if a zero drops over the boundary.

Hence we see that we cannot expect too much from our degree. In addition, since it is unclear how it should be defined, we will first require some basic properties a degree should have and then we will look for functions satisfying these properties.

2.2 Definition of the mapping degree and the determinant formula

To begin with, let us introduce some useful notation. Throughout this section U will be a bounded open subset of $\mathbb{R}^n$. For $f \in C^1(U, \mathbb{R}^n)$ the Jacobi matrix of f at $x \in U$ is $f'(x) = (\partial_{x_i} f_j(x))_{1 \le i,j \le n}$ and the Jacobi determinant of f at $x \in U$ is

$$J_f(x) = \det f'(x). \tag{2.4}$$

The set of regular values is

$$\mathrm{RV}(f) = \{y \in \mathbb{R}^n | \forall x \in f^{-1}(y) : J_f(x) \neq 0\}. \tag{2.5}$$

Its complement $\mathrm{CV}(f) = \mathbb{R}^n \backslash \mathrm{RV}(f)$ is called the set of critical values. We set $C^r(\overline{U}, \mathbb{R}^n) = \{f \in C^r(U, \mathbb{R}^n) | d^j f \in C(\overline{U}, \mathbb{R}^n), 0 \le j \le r\}$ and

$$D_y^r(\overline{U}, \mathbb{R}^n) = \{f \in C^r(\overline{U}, \mathbb{R}^n) | y \notin f(\partial U)\}, \quad D_y(\overline{U}, \mathbb{R}^n) = D_y^0(\overline{U}, \mathbb{R}^n), \quad y \in \mathbb{R}^n. \tag{2.6}$$

Now that these things are out of the way, we come to the formulation of the requirements for our degree.

A function deg which assigns each $f \in D_y(\overline{U}, \mathbb{R}^n)$, $y \in \mathbb{R}^n$, a real number $\deg(f, U, y)$ will be called degree if it satisfies the following conditions.

(D1). $\deg(f, U, y) = \deg(f - y, U, 0)$ (*translation invariance*).

(D2). $\deg(\mathbb{1}, U, y) = 1$ if $y \in U$ (*normalization*).

(D3). If $U_{1,2}$ are open, disjoint subsets of U such that $y \notin f(\overline{U} \backslash (U_1 \cup U_2))$, then $\deg(f, U, y) = \deg(f, U_1, y) + \deg(f, U_2, y)$ (*additivity*).

(D4). If $H(t) = (1-t)f + tg \in D_y(\overline{U}, \mathbb{R}^n)$, $t \in [0,1]$, then $\deg(f, U, y) = \deg(g, U, y)$ (*homotopy invariance*).

Before we draw some first conclusions form this definition, let us discuss the properties (D1)–(D4) first. (D1) is natural since $\deg(f, U, y)$ should have something to do with the solutions of $f(x) = y$, $x \in U$, which is the same as the solutions of $f(x) - y = 0$, $x \in U$. $(D2)$ is a normalization since any multiple of deg would also satisfy the other requirements. (D3) is also quite natural since it requires deg to be additive with respect to components. In addition, it implies that sets where $f \neq y$ do not contribute. (D4) is not that natural since it already rules out the case where deg is the cardinality of $f^{-1}(U)$. On the other hand it will give us the ability to compute $\deg(f, U, y)$ in several cases.

Theorem 2.1 *Suppose* deg *satisfies (D1)–(D4) and let* $f, g \in D_y(\overline{U}, \mathbb{R}^n)$, *then the following statements hold.*

(i). We have $\deg(f, \emptyset, y) = 0$. *Moreover, if* U_i, $1 \le i \le N$, *are disjoint open subsets of* U *such that* $y \notin f(\overline{U} \setminus \bigcup_{i=1}^N U_i)$, *then* $\deg(f, U, y) = \sum_{i=1}^N \deg(f, U_i, y)$.

(ii). If $y \notin f(U)$, *then* $\deg(f, U, y) = 0$ *(but not the other way round). Equivalently, if* $\deg(f, U, y) \neq 0$, *then* $y \in f(U)$.

(iii). If $|f(x) - g(x)| < \operatorname{dist}(y, f(\partial U))$, $x \in \partial U$, *then* $\deg(f, U, y) = \deg(g, U, y)$. *In particular, this is true if* $f(x) = g(x)$ *for* $x \in \partial U$.

Proof. For the first part of (i) use (D3) with $U_1 = U$ and $U_2 = \emptyset$. For the second part use $U_2 = \emptyset$ in (D3) if $i = 1$ and the rest follows from induction. For (ii) use $i = 1$ and $U_1 = \emptyset$ in (ii). For (iii) note that $H(t, x) = (1 - t) f(x) + t\, g(x)$ satisfies $|H(t, x) - y| \ge \operatorname{dist}(y, f(\partial U)) - |f(x) - g(x)|$ for x on the boundary. □

Next we show that (D.4) implies several at first sight much stronger looking facts.

Theorem 2.2 *We have that* $\deg(., U, y)$ *and* $\deg(f, U, .)$ *are both continuous. In fact, we even have*

(i). $\deg(., U, y)$ *is constant on each component of* $D_y(\overline{U}, \mathbb{R}^n)$.

(ii). $\deg(f, U, .)$ *is constant on each component of* $\mathbb{R}^n \setminus f(\partial U)$.

Moreover, if $H : [0, 1] \times \overline{U} \to \mathbb{R}^n$ *and* $y : [0, 1] \to \mathbb{R}^n$ *are both continuous such that* $H(t) \in D_{y(t)}(U, \mathbb{R}^n)$, $t \in [0, 1]$, *then* $\deg(H(0), U, y(0)) = \deg(H(1), U, y(1))$.

Proof. For (i) let C be a component of $D_y(\overline{U}, \mathbb{R}^n)$ and let $d_0 \in \deg(C, U, y)$. It suffices to show that $\deg(., U, y)$ is locally constant. But if $|g - f| < \operatorname{dist}(y, f(\partial U))$, then $\deg(f, U, y) = \deg(g, U, y)$ by (D.4) since $|H(t) - y| \ge |f - y| - |g - f| > 0$, $H(t) = (1 - t) f + t\, g$. The proof of (ii) is similar. For the remaining part observe, that if $H : [0, 1] \times \overline{U} \to \mathbb{R}^n$, $(t, x) \mapsto H(t, x)$, is continuous, then so is $H : [0,1] \to C(\overline{U}, \mathbb{R}^n)$, $t \mapsto H(t)$, since $\overline{U}$ is compact. Hence, if in addition $H(t) \in D_y(\overline{U}, \mathbb{R}^n)$, then $\deg(H(t), U, y)$ is independent of t and if $y = y(t)$ we can use $\deg(H(0), U, y(0)) = \deg(H(t) - y(t), U, 0) = \deg(H(1), U, y(1))$. □

Note that this result also shows why $\deg(f, U, y)$ cannot be defined meaningful for $y \in f(\partial D)$. Indeed, approaching y from within different components of $\mathbb{R}^n \backslash f(\partial U)$ will result in different limits in general!

In addition, note that if Q is a closed subset of a locally pathwise connected space X, then the components of $X \backslash Q$ are open (in the topology of X) and pathwise connected (the set of points for which a path to a fixed point x_0 exists is both open and closed).

Now let us try to compute deg using its properties. Lets start with a simple case and suppose $f \in C^1(U, \mathbb{R}^n)$ and $y \notin \mathrm{CV}(f) \cup f(\partial U)$. Without restriction we consider $y = 0$. In addition, we avoid the trivial case $f^{-1}(y) = \emptyset$. Since the points of $f^{-1}(0)$ inside U are isolated (use $J_f(x) \neq 0$ and the inverse function theorem) they can only cluster at the boundary ∂U. But this is also impossible since f would equal y at the limit point on the boundary by continuity. Hence $f^{-1}(0) = \{x^i\}_{i=1}^N$. Picking sufficiently small neighborhoods $U(x^i)$ around x^i we consequently get

$$\deg(f, U, 0) = \sum_{i=1}^{N} \deg(f, U(x^i), 0). \tag{2.7}$$

It suffices to consider one of the zeros, say x^1. Moreover, we can even assume $x^1 = 0$ and $U(x^1) = B_\delta(0)$. Next we replace f by its linear approximation around 0. By the definition of the derivative we have

$$f(x) = f'(0)x + |x| r(x), \qquad r \in C(B_\delta(0), \mathbb{R}^n), \quad r(0) = 0. \tag{2.8}$$

Now consider the homotopy $H(t, x) = f'(0)x + (1-t)|x|r(x)$. In order to conclude $\deg(f, B_\delta(0), 0) = \deg(f'(0), B_\delta(0), 0)$ we need to show $0 \notin H(t, \partial B_\delta(0))$. Since $J_f(0) \neq 0$ we can find a constant λ such that $|f'(0)x| \geq \lambda |x|$ and since $r(0) = 0$ we can decrease δ such that $|r| < \lambda$. This implies $|H(t,x)| \geq ||f'(0)x| - (1-t)|x||r(x)|| \geq \lambda\delta - \delta|r| > 0$ for $x \in \partial B_\delta(0)$ as desired.

In order to compute the degree of a nonsingular matrix we need the following lemma.

Lemma 2.3 *Two nonsingular matrices* $M_{1,2} \in \mathrm{GL}(n)$ *are homotopic in* $\mathrm{GL}(n)$ *if and only if* $\operatorname{sgn}\det M_1 = \operatorname{sgn}\det M_2$.

Proof. We will show that any given nonsingular matrix M is homotopic to $\mathrm{diag}(\mathrm{sgn}\det M, 1, \dots, 1)$, where $\mathrm{diag}(m_1, \dots, m_n)$ denotes a diagonal matrix with diagonal entries m_i.

In fact, note that adding one row to another and multiplying a row by a positive constant can be realized by continuous deformations such that all intermediate matrices are nonsingular. Hence we can reduce M to a diagonal matrix $\mathrm{diag}(m_1, \dots, m_n)$ with $(m_i)^2 = 1$. Next,

$$\begin{pmatrix} \pm\cos(\pi t) & \mp\sin(\pi t) \\ sin(\pi t) & \cos(\pi t), \end{pmatrix} \tag{2.9}$$

shows that $\mathrm{diag}(\pm 1, 1)$ and $\mathrm{diag}(\mp 1, -1)$ are homotopic. Now we apply this result to all two by two subblocks as follows. For each i starting from n and going down to 2 transform the subblock $\mathrm{diag}(m_{i-1}, m_i)$ into $\mathrm{diag}(1, 1)$ respectively $\mathrm{diag}(-1, 1)$. The result is the desired form for M.

To conclude the proof note that a continuous deformation within $\mathrm{GL}(n)$ cannot change the sign of the determinant since otherwise the determinant would have to vanish somewhere in between (i.e., we would leave $\mathrm{GL}(n)$). □

Using this lemma we can now show the main result of this section.

Theorem 2.4 *Suppose* $f \in D^1_y(\overline{U}, \mathbb{R}^n)$ *and* $y \notin \mathrm{CV}(f)$, *then a degree satisfying (D1)–(D4) satisfies*

$$\deg(f, U, y) = \sum_{x \in f^{-1}(y)} \mathrm{sgn}\, J_f(x), \tag{2.10}$$

where the sum is finite and we agree to set $\sum_{x\in\emptyset} = 0$.

Proof. By the previous lemma we obtain

$$\deg(f'(0), B_\delta(0), 0) = \deg(\mathrm{diag}(\mathrm{sgn} J_f(0), 1, \dots, 1), B_\delta(0), 0) \tag{2.11}$$

since $\det M \neq 0$ is equivalent to $Mx \neq 0$ for $x \in \partial B_\delta(0)$. Hence it remains to show $\deg(f'(0), B_\delta(0), 0) = \mathrm{sgn} J_f(0)$.

If $\mathrm{sgn} J_f(0) = 1$ this is true by (D2). Otherwise we can replace $f'(0)$ by $M_- = \mathrm{diag}(-1, 1, \dots, 1)$.

Now let $U_1 = \{x \in \mathbb{R}^n | |x_i| < 1, 1 \le i \le n\}$, $U_2 = \{x \in \mathbb{R}^n | 1 < x_1 < 3, |x_i| < 1, 2 \le i \le n\}$, $U = \{x \in \mathbb{R}^n | -1 < x_1 < 3, |x_i| < 1, 2 \le i \le n\}$, and abbreviate $y_0 = (2, 0, \dots, 0)$. On U consider two continuous mappings $M_{1,2} : U \to \mathbb{R}^n$ such that $M_1(x) = M_-$ if $x \in U_1$, $M_1(x) = \mathbb{1} - y_0$ if $x \in U_2$, and $M_2(x) = (1, x_2, \dots, x_n)$.

Since $M_1(x) = M_2(x)$ for $x \in \partial U$ we infer $\deg(M_1, U, 0) = \deg(M_2, U, 0) = 0$. Moreover, we have $\deg(M_1, U, 0) = \deg(M_1, U_1, 0) + \deg(M_1, U_2, 0)$ and hence $\deg(M_-, U_1, 0) = -\deg(\mathbb{1} - y_0, U_2, 0) = -\deg(\mathbb{1}, U_2, y_0) = -1$ as claimed. □

Up to this point we have only shown that a degree (provided there is one at all) necessarily satisfies (2.10). Once we have shown that regular values are dense, it will follow that the degree is uniquely determined by (2.10) since the remaining values follow from point (iv) of Theorem 2.1. On the other hand, we don't even know whether a degree exists. Hence we need to show that (2.10) can be extended to $f \in D_y(\overline{U}, \mathbb{R}^n)$ and that this extension satisfies our requirements (D1)–(D4).

2.3 Extension of the determinant formula

Our present objective is to show that the determinant formula (2.10) can be extended to all $f \in D_y(\overline{U}, \mathbb{R}^n)$. This will be done in two steps, where we will show that $\deg(f, U, y)$ as defined in (2.10) is locally constant with respect to both y (step one) and f (step two).

Before we work out the technical details for these two steps, we prove that the set of regular values is dense as a warm up. This is a consequence of a special case of Sard's theorem which says that $\mathrm{CV}(f)$ has zero measure.

Lemma 2.5 (Sard) *Suppose $f \in C^1(U, \mathbb{R}^n)$, then the Lebesgue measure of $\mathrm{CV}(f)$ is zero.*

Proof. Since the claim is easy for linear mappings our strategy is as follows. We divide U into sufficiently small subsets. Then we replace f by its linear approximation in each subset and estimate the error.

Let $\mathrm{CP}(f) = \{x \in U | J_f(x) = 0\}$ be the set of critical points of f. We first pass to cubes which are easier to divide. Let $\{Q_i\}_{i \in \mathbb{N}}$ be a countable cover for U consisting of open cubes such that $\overline{Q_i} \subset U$. Then it suffices to prove that $f(\mathrm{CP}(f) \cap Q_i)$ has zero measure since $\mathrm{CV}(f) = f(\mathrm{CP}(f)) = \bigcup_i f(\mathrm{CP}(f) \cap Q_i)$ (the Q_i's are a cover).

Let Q be any of these cubes and denote by ρ the length of its edges. Fix $\varepsilon > 0$ and divide Q into N^n cubes Q_i of length ρ/N. Since $f'(x)$ is uniformly continuous on Q we can find an N (independent of i) such that

$$|f(x) - f(\tilde{x}) - f'(\tilde{x})(x - \tilde{x})| \leq \int_0^1 |f'(\tilde{x} + t(x - \tilde{x})) - f'(\tilde{x})||\tilde{x} - x| dt \leq \frac{\varepsilon\rho}{N} \quad (2.12)$$

for $\tilde{x}, x \in Q_i$. Now pick a Q_i which contains a critical point $\tilde{x}_i \in \mathrm{CP}(f)$. Without restriction we assume $\tilde{x}_i = 0$, $f(\tilde{x}_i) = 0$ and set $M = f'(\tilde{x}_i)$. By $\det M = 0$ there is an orthonormal basis $\{b^i\}_{1\le i\le n}$ of $\mathbb{R}^n$ such that b^n is orthogonal to the image of M. In addition, there is a constant C_1 such that $Q_i \subseteq \{\sum_{i=1}^{n-1} \lambda_i b^i \,|\, |\lambda_i| \le C_1 \frac{\rho}{N}\}$ (e.g., $C_1 = n2^{(n/2)}$) and hence there is a second constant (again independent of i) such that

$$MQ_i \subseteq \{\sum_{i=1}^{n-1} \lambda_i b^i \,|\, |\lambda_i| \le C_2 \frac{\rho}{N}\} \tag{2.13}$$

(e.g., $C_2 = nC_1 \max_{x\in\overline{Q}} |f'(x)|$). Next, by our estimate (2.12) we even have

$$f(Q_i) \subseteq \{\sum_{i=1}^{n} \lambda_i b^i \,|\, |\lambda_i| \le (C_2+\varepsilon)\frac{\rho}{N}, |\lambda_n| \le \varepsilon\frac{\rho}{N}\} \tag{2.14}$$

and hence the measure of $f(Q_i)$ is smaller than $\frac{C_3\varepsilon}{N^n}$. Since there are at most N^n such Q_i's, we see that the measure of $f(Q)$ is smaller than $C_3\varepsilon$. □

Having this result out of the way we can come to step one and two from above.

Step 1: Admitting critical values

By (v) of Theorem 2.1, $\deg(f, U, y)$ should be constant on each component of $\mathbb{R}^n \backslash f(\partial U)$. Unfortunately, if we connect y and a nearby regular value $\tilde{y}$ by a path, then there might be some critical values in between. To overcome this problem we need a definition for deg which works for critical values as well. Let us try to look for an integral representation. Formally (2.10) can be written as $\deg(f, U, y) = \int_U \delta_y(f(x)) J_f(x) dx$, where $\delta_y(.)$ is the Dirac distribution at y. But since we don't want to mess with distributions, we replace $\delta_y(.)$ by $\phi_\varepsilon(. - y)$, where $\{\phi_\varepsilon\}_{\varepsilon>0}$ is a family of functions such that ϕ_ε is supported on the ball $B_\varepsilon(0)$ of radius ε around 0 and satisfies $\int_{\mathbb{R}^n} \phi_\varepsilon(x) dx = 1$.

Lemma 2.6 *Let $f \in D^1_y(\overline{U}, \mathbb{R}^n)$, $y \notin \mathrm{CV}(f)$. Then*

$$\deg(f, U, y) = \int_U \phi_\varepsilon(f(x) - y) J_f(x) dx \tag{2.15}$$

for all positive ε smaller than a certain ε_0 depending on f and y. Moreover, $\mathrm{supp}(\phi_\varepsilon(f(.) - y)) \subset U$ *for $\varepsilon < \mathrm{dist}(y, f(\partial U))$.*

Proof. If $f^{-1}(y) = \emptyset$, we can set $\varepsilon_0 = \operatorname{dist}(y, f(\overline{U}))$, implying $\phi_\varepsilon(f(x) - y) = 0$ for $x \in \overline{U}$.

If $f^{-1}(y) = \{x^i\}_{1 \le i \le N}$, we can find an $\varepsilon_0 > 0$ such that $f^{-1}(B_{\varepsilon_0}(y))$ is a union of disjoint neighborhoods $U(x^i)$ of x^i by the inverse function theorem. Moreover, after possibly decreasing ε_0 we can assume that $f|_{U(x^i)}$ is a bijection and that $J_f(x)$ is nonzero on $U(x^i)$. Again $\phi_\varepsilon(f(x) - y) = 0$ for $x \in \overline{U} \setminus \bigcup_{i=1}^N U(x^i)$ and hence

$$\begin{aligned} \int_U \phi_\varepsilon(f(x) - y) J_f(x) dx &= \sum_{i=1}^N \int_{U(x^i)} \phi_\varepsilon(f(x) - y) J_f(x) dx \\ &= \sum_{i=1}^N \operatorname{sgn}(J_f(x)) \int_{B_{\varepsilon_0}(0)} \phi_\varepsilon(\tilde{x}) d\tilde{x} = \deg(f, U, y), \end{aligned} \tag{2.16}$$

where we have used the change of variables $\tilde{x} = f(x)$ in the second step. □

Our new integral representation makes sense even for critical values. But since ε depends on y, continuity with respect to y is not clear. This will be shown next at the expense of requiring $f \in C^2$ rather than $f \in C^1$.

The key idea is to rewrite $\deg(f, U, y^2) - \deg(f, U, y^1)$ as an integral over a divergence (here we will need $f \in C^2$) supported in U and then apply Stokes theorem. For this purpose the following result will be used.

Lemma 2.7 *Suppose $f \in C^2(U, \mathbb{R}^n)$ and $u \in C^1(\mathbb{R}^n, \mathbb{R}^n)$, then*

$$(\operatorname{div} u)(f) J_f = \operatorname{div} D_f(u), \tag{2.17}$$

where $D_f(u)_j$ is the determinant of the matrix obtained from f' by replacing the j-th column by $u(f)$.

Proof. We compute

$$\operatorname{div} D_f(u) = \sum_{j=1}^n \partial_{x_j} D_f(u)_j = \sum_{j,k=1}^n D_f(u)_{j,k}, \tag{2.18}$$

where $D_f(u)_{j,k}$ is the determinant of the matrix obtained from the matrix associated with $D_f(u)_j$ by applying ∂_{x_j} to the k-th column. Since $\partial_{x_j} \partial_{x_k} f = \partial_{x_k} \partial_{x_j} f$ we infer $D_f(u)_{j,k} = -D_f(u)_{k,j}$, $j \neq k$, by exchanging the k-th and the j-th column. Hence

$$\operatorname{div} D_f(u) = \sum_{i=1}^n D_f(u)_{i,i}. \tag{2.19}$$

Now let $J_f^{(i,j)}(x)$ denote the (i,j) minor of $f'(x)$ and recall $\sum_{i=1}^n J_f^{(i,j)} \partial_{x_i} f_k = \delta_{j,k} J_f$. Using this to expand the determinant $D_f(u)_{i,i}$ along the i-th column shows

$$\begin{aligned} \mathrm{div} D_f(u) &= \sum_{i,j=1}^n J_f^{(i,j)} \partial_{x_i} u_j(f) = \sum_{i,j=1}^n J_f^{(i,j)} \sum_{k=1}^n (\partial_{x_k} u_j)(f) \partial_{x_i} f_k \\ &= \sum_{j,k=1}^n (\partial_{x_k} u_j)(f) \sum_{i=1}^n J_f^{(i,j)} \partial_{x_j} f_k = \sum_{j=1}^n (\partial_{x_j} u_j)(f) J_f \end{aligned} \tag{2.20}$$

as required. □

Now we can prove

Lemma 2.8 *Suppose* $f \in C^2(\overline{U}, \mathbb{R}^n)$. *Then* $\deg(f,U,.)$ *is constant in each ball contained in* $\mathbb{R}^n \backslash f(\partial U)$, *whenever defined.*

Proof. Fix $\tilde{y} \in \mathbb{R}^n \backslash f(\partial U)$ and consider the largest ball $B_\rho(\tilde{y})$, $\rho = \mathrm{dist}(\tilde{y}, f(\partial U))$ around $\tilde{y}$ contained in $\mathbb{R}^n \backslash f(\partial U)$. Pick $y^i \in B_\rho(\tilde{y}) \cap \mathrm{RV}(f)$ and consider

$$\deg(f,U,y^2) - \deg(f,U,y^1) = \int_U (\phi_\varepsilon(f(x) - y^2) - \phi_\varepsilon(f(x) - y^1)) J_f(x) dx \tag{2.21}$$

for suitable $\phi_\varepsilon \in C^2(\mathbb{R}^n, \mathbb{R})$ and suitable $\varepsilon > 0$. Now observe

$$\begin{aligned} (\mathrm{div}\, u)(y) &= \int_0^1 z_j \partial_{y_j} \phi(y + tz) dt \\ &= \int_0^1 (\frac{d}{dt} \phi(y + t\,z)) dt = \phi_\varepsilon(y - y^2) - \phi_\varepsilon(y - y^1), \end{aligned} \tag{2.22}$$

where

$$u(y) = z \int_0^1 \phi(y + t\,z) dt, \quad \phi(y) = \phi_\varepsilon(y - y^1),\ z = y^2 - y^1, \tag{2.23}$$

and apply the previous lemma to rewrite the integral as $\int_U \mathrm{div} D_f(u) dx$. Since the integrand vanishes in a neighborhood of ∂U it is no restriction to assume that ∂U is smooth such that we can apply Stokes theorem. Hence we have $\int_U \mathrm{div} D_f(u) dx = \int_{\partial U} D_f(u) dF = 0$ since u is supported inside $B_\rho(\tilde{y})$ provided ε is small enough (e.g., $\varepsilon < \rho - \max\{|y^i - \tilde{y}|\}_{i=1,2}$). □

As a consequence we can define

$$\deg(f,U,y) = \deg(f,U,\tilde{y}), \qquad y \notin f(\partial U), \quad f \in C^2(\overline{U},\mathbb{R}^n), \tag{2.24}$$

where $\tilde{y}$ is a regular value of f with $|\tilde{y} - y| < \operatorname{dist}(y, f(\partial U))$.

Remark 2.9 *Let me remark a different approach due to Kronecker. For U with sufficiently smooth boundary we have*

$$\deg(f,U,0) = \frac{1}{|S^{n-1}|}\int_{\partial U} D_{\tilde{f}}(x)dF = \frac{1}{|S^n|}\int_{\partial U} \frac{1}{|f|^n} D_f(x)dF, \quad \tilde{f} = \frac{f}{|f|}, \tag{2.25}$$

for $f \in C^2_y(\overline{U},\mathbb{R}^n)$. Explicitly we have

$$\deg(f,U,0) = \frac{1}{|S^{n-1}|}\int_{\partial U} \sum_{j=1}^{n} (-1)^{j-1} \frac{f_j}{|f|^n} df_1 \wedge \cdots \wedge df_{j-1} \wedge df_{j+1} \wedge \cdots \wedge df_n. \tag{2.26}$$

Since $\tilde{f} : \partial U \to S^{n-1}$ the integrand can also be written as the pull back $\tilde{f}^* dS$ of the canonical surface element dS on S^{n-1}.

This coincides with the boundary value approach for complex functions (note that holomorphic functions are orientation preserving).

Step 2: Admitting continuous functions

Our final step is to remove the condition $f \in C^2$. As before we want the degree to be constant in each ball contained in $D_y(\overline{U},\mathbb{R}^n)$. For example, fix $f \in D_y(\overline{U},\mathbb{R}^n)$ and set $\rho = \operatorname{dist}(y, f(\partial U)) > 0$. Choose $f^i \in C^2(\overline{U},\mathbb{R}^n)$ such that $|f^i - f| < \rho$, implying $f^i \in D_y(\overline{U},\mathbb{R}^n)$. Then $H(t,x) = (1-t)f^1(x) + tf^2(x) \in D_y(\overline{U},\mathbb{R}^n) \cap C^2(U,\mathbb{R}^n)$, $t \in [0,1]$, and $|H(t) - f| < \rho$. If we can show that $\deg(H(t),U,y)$ is locally constant with respect to t, then it is continuous with respect to t and hence constant (since $[0,1]$ is connected). Consequently we can define

$$\deg(f,U,y) = \deg(\tilde{f},U,y), \qquad f \in D_y(\overline{U},\mathbb{R}^n), \tag{2.27}$$

where $\tilde{f} \in C^2(\overline{U},\mathbb{R}^n)$ with $|\tilde{f} - f| < \operatorname{dist}(y, f(\partial U))$.

It remains to show that $t \mapsto \deg(H(t),U,y)$ is locally constant.

Lemma 2.10 *Suppose $f \in C^2_y(\overline{U},\mathbb{R}^n)$. Then for each $\tilde{f} \in C^2(\overline{U},\mathbb{R}^n)$ there is an $\varepsilon > 0$ such that* $\deg(f + t\tilde{f}, U, y) = \deg(f,U,y)$ *for all* $t \in (-\varepsilon,\varepsilon)$.

Proof. If $f^{-1}(y) = \emptyset$ the same is true for $f + t\,g$ if $|t| < \mathrm{dist}(y, f(\overline{U}))/|g|$. Hence we can exclude this case. For the remaining case we use our usual strategy of considering $y \in \mathrm{RV}(f)$ first and then approximating general y by regular ones.

Suppose $y \in \mathrm{RV}(f)$ and let $f^{-1}(y) = \{x^i\}_{j=1}^N$. By the implicit function theorem we can find disjoint neighborhoods $U(x^i)$ such that there exists a unique solution $x^i(t) \in U(x^i)$ of $(f + t\,g)(x) = y$ for $|t| < \varepsilon_1$. By reducing $U(x^i)$ if necessary, we can even assume that the sign of $J_{f+t\,g}$ is constant on $U(x^i)$. Finally, let $\varepsilon_2 = \mathrm{dist}(y, f(U \backslash \bigcup_{i=1}^N U(x^i)))/|g|$. Then $|f + t\,g| > 0$ for $|t| < \varepsilon_2$ and $\varepsilon = \min(\varepsilon_1, \varepsilon_2)$ is the quantity we are looking for.

It remains to consider the case $y \in \mathrm{CV}(f)$. pick a regular value $\tilde{y} \in B_{\rho/3}(y)$, where $\rho = \mathrm{dist}(y, f(\partial U))$, implying $\deg(f, U, y) = \deg(f, U, \tilde{y})$. Then we can find an $\tilde{\varepsilon} > 0$ such that $\deg(f, U, \tilde{y}) = \deg(f + t\,g, U, \tilde{y})$ for $|t| < \tilde{\varepsilon}$. Setting $\varepsilon = \min(\tilde{\varepsilon}, \rho/(3|g|))$ we infer $\tilde{y} - (f + t\,g)(x) \geq \rho/3$ for $x \in \partial U$, that is $|\tilde{y} - y| < \mathrm{dist}(\tilde{y}, (f + t\,g)(\partial U))$, and thus $\deg(f + t\,g, U, \tilde{y}) = \deg(f + t\,g, U, y)$. Putting it all together implies $\deg(f, U, y) = \deg(f + t\,g, U, y)$ for $|t| < \varepsilon$ as required. □

Now we can finally prove our main theorem.

Theorem 2.11 *There is a unique degree* deg *satisfying (D1)-(D4). Moreover,* $\deg(., U, y) : D_y(\overline{U}, \mathbb{R}^n) \to \mathbb{Z}$ *is constant on each component and given* $f \in D_y(\overline{U}, \mathbb{R}^n)$ *we have*

$$\deg(f, U, y) = \sum_{x \in \tilde{f}^{-1}(y)} \mathrm{sgn}\, J_{\tilde{f}}(x) \tag{2.28}$$

where $\tilde{f} \in D_y^2(\overline{U}, \mathbb{R}^n)$ *is in the same component of* $D_y(\overline{U}, \mathbb{R}^n)$, *say* $|f - \tilde{f}| < \mathrm{dist}(y, f(\partial U))$, *such that* $y \in \mathrm{RV}(\tilde{f})$.

Proof. Our previous considerations show that deg is well-defined and locally constant with respect to the first argument by construction. Hence $\deg(., U, y) : D_y(\overline{U}, \mathbb{R}^n) \to \mathbb{Z}$ is continuous and thus necessarily constant on components since $\mathbb{Z}$ is discrete. (D2) is clear and (D1) is satisfied since it holds for $\tilde{f}$ by construction. Similarly, taking $U_{1,2}$ as in (D3) we can require $|f - \tilde{f}| < \mathrm{dist}(y, f(\overline{U} \backslash (U_1 \cup U_2))$. Then (D3) is satisfied since it also holds for $\tilde{f}$ by construction. Finally, (D4) is a consequence of continuity. □

To conclude this section, let us give a few simple examples illustrating the use of the Brouwer degree.

First, let's investigate the zeros of

$$f(x_1, x_2) = (x_1 - 2x_2 + \cos(x_1 + x_2), x_2 + 2x_1 + \sin(x_1 + x_2)). \tag{2.29}$$

Denote the linear part by

$$g(x_1, x_2) = (x_1 - 2x_2, x_2 + 2x_1). \tag{2.30}$$

Then we have $|g(x)| = \sqrt{5}|x|$ and $|f(x) - g(x)| = 1$ and hence $h(t) = (1-t)g + t\,f = g + t(f - g)$ satisfies $|h(t)| \geq |g| - t|f - g| > 0$ for $|x| > 1/\sqrt{5}$ implying

$$\deg(f, B_5(0), 0) = \deg(g, B_5(0), 0) = 1. \tag{2.31}$$

Moreover, since $J_f(x) = 5 + 3\cos(x_1 + x_2) + \sin(x_1 + x_2) > 1$ we see that $f(x) = 0$ has a unique solution in $\mathbb{R}^2$. This solution has even to lie on the circle $|x| = 1/\sqrt{5}$ since $f(x) = 0$ implies $1 = |f(x) - g(x)| = |g(x)| = \sqrt{5}|x|$.

Next let us prove the following result which implies the hairy ball (or hedgehog) theorem.

Theorem 2.12 *Suppose U contains the origin and let $f : \partial U \to \mathbb{R}^n\backslash\{0\}$ be continuous. If n is odd, then there exists a $x \in \partial U$ and a $\lambda \neq 0$ such that $f(x) = \lambda x$.*

Proof. By Theorem 2.15 we can assume $f \in C(\overline{U}, \mathbb{R}^n)$ and since n is odd we have $\deg(-\mathbb{1}, U, 0) = -1$. Now if $\deg(f, U, 0) \neq -1$, then $H(t, x) = (1-t)f(x) - tx$ must have a zero $(t_0, x_0) \in (0, 1) \times \partial U$ and hence $f(x_0) = \frac{t_0}{1-t_0} x_0$. Otherwise, if $\deg(f, U, 0) = -1$ we can apply the same argument to $H(t, x) = (1-t)f(x) + tx$. □

In particular this result implies that a continuous tangent vector field on the unit sphere $f : S^{n-1} \to \mathbb{R}^n$ (with $f(x)x = 0$ for all $x \in S^n$) must vanish somewhere if n is odd. Or, for $n = 3$, you cannot smoothly comb a hedgehog without leaving a bald spot or making a parting. It is however possible to comb the hair smoothly on a torus and that is why the magnetic containers in nuclear fusion are toroidal.

Another simple consequence is the fact that a vector field on $\mathbb{R}^n$, which points outwards (or inwards) on a sphere, must vanish somewhere inside the sphere.

Theorem 2.13 *Suppose $f : B_R(0) \to \mathbb{R}^n$ is continuous and satisfies*

$$f(x)x > 0, \qquad |x| = R. \tag{2.32}$$

Then $f(x)$ vanishes somewhere inside $B_R(0)$.

Proof. If f does not vanish, then $H(t, x) = (1 - t)x + tf(x)$ must vanish at some point $(t_0, x_0) \in (0, 1) \times \partial B_R(0)$ and thus

$$0 = H(t_0, x_0)x_0 = (1 - t_0)R^2 + t_0 f(x_0)x_0. \tag{2.33}$$

But the last part is positive by assumption, a contradiction. □

2.4 The Brouwer fixed-point theorem

Now we can show that the famous Brouwer fixed-point theorem is a simple consequence of the properties of our degree.

Theorem 2.14 (Brouwer fixed point) *Let K be a topological space homeomorphic to a compact, convex subset of $\mathbb{R}^n$ and let $f \in C(K, K)$, then f has at least one fixed point.*

Proof. Clearly we can assume $K \subset \mathbb{R}^n$ since homeomorphisms preserve fixed points. Now lets assume $K = B_r(0)$. If there is a fixed-point on the boundary $\partial B_r(0))$ we are done. Otherwise $H(t, x) = x - t\, f(x)$ satisfies $0 \notin H(t, \partial B_r(0))$ since $|H(t, x)| \ge |x| - t|f(x)| \ge (1 - t)r > 0$, $0 \le t < 1$. And the claim follows from $\deg(x - f(x), B_r(0), 0) = \deg(x, B_r(0), 0) = 1$.

Now let K be convex. Then $K \subseteq B_\rho(0)$ and, by Theorem 2.15 below, we can find a continuous retraction $R : \mathbb{R}^n \to K$ (i.e., $R(x) = x$ for $x \in K$) and consider $\tilde{f} = f \circ R \in \mathcal{C}(\overline{B_\rho(0)}, \overline{B_\rho(0)})$. By our previous analysis, there is a fixed point $x = \tilde{f}(x) \in \mathrm{conv}(f(K)) \subseteq K$. □

Note that any compact, convex subset of a finite dimensional Banach space (complex or real) is isomorphic to a compact, convex subset of $\mathbb{R}^n$ since linear transformations preserve both properties. In addition, observe that all assumptions are needed. For example, the map $f : \mathbb{R} \to \mathbb{R}$, $x \mapsto x + 1$, has no fixed point ($\mathbb{R}$ is homeomorphic to a bounded set but not to a compact one). The same is true for the map $f : \partial B_1(0) \to \partial B_1(0)$, $x \mapsto -x$ ($\partial B_1(0) \subset \mathbb{R}^n$ is simply connected for $n \ge 3$ but not homeomorphic to a convex set).

It remains to prove the result from topology needed in the proof of the Brouwer fixed-point theorem.

Theorem 2.15 *Let X and Y be Banach spaces and let K be a closed subset of X. Then $F \in C(K, Y)$ has a continuous extension $F \in C(X, Y)$ such that $F(X) \subseteq \mathrm{conv}(F(K))$.*

Proof. Consider the open cover $\{B_{\rho(x)}(x)\}_{x \in X \setminus K}$ for $X \setminus K$, where $\rho(x) = \mathrm{dist}(x, X \setminus K)/2$. Choose a (locally finite) partition of unity $\{\phi_\lambda\}_{\lambda \in \Lambda}$ subordinate to this cover and set

$$F(x) = \sum_{\lambda \in \Lambda} \phi_\lambda(x) F(x_\lambda) \text{ for } x \in X \setminus K, \tag{2.34}$$

where $x_\lambda \in K$ satisfies $\mathrm{dist}(x_\lambda, \mathrm{supp}\phi_\lambda) \le 2\mathrm{dist}(K, \mathrm{supp}\phi_\lambda)$. By construction, F is continuous except for possibly at the boundary of K. Fix $x_0 \in \partial K$, $\varepsilon > 0$ and choose $\delta > 0$ such that $|F(x) - F(x_0)| \le \varepsilon$ for all $x \in K$ with $|x - x_0| < 4\delta$. We will show that $|F(x) - F(x_0)| \le \varepsilon$ for all $x \in X$ with $|x - x_0| < \delta$. Suppose $x \notin K$, then $|F(x) - F(x_0)| \le \sum_{\lambda\in\Lambda} \phi_\lambda(x)|F(x_\lambda) - F(x_0)|$. By our construction, x_λ should be close to x for all λ with $x \in \mathrm{supp}\phi_\lambda$ since x is close to K. In fact, if $x \in \mathrm{supp}\phi_\lambda$ we have

$$|x - x_\lambda| \le \mathrm{dist}(x_\lambda, \mathrm{supp}\phi_\lambda) + d(\mathrm{supp}\phi_\lambda) \le 2\mathrm{dist}(K, \mathrm{supp}\phi_\lambda) + d(\mathrm{supp}\phi_\lambda), \quad (2.35)$$

where $d(\mathrm{supp}\phi_\lambda) = \sup_{x,y\in\mathrm{supp}\phi_\lambda} |x-y|$. Since our partition of unity is subordinate to the cover $\{B_{\rho(x)}(x)\}_{x\in X\backslash K}$ we can find a $\tilde{x} \in X\backslash K$ such that $\mathrm{supp}\phi_\lambda \subset B_{\rho(\tilde{x})}(\tilde{x})$ and hence $d(\mathrm{supp}\phi_\lambda) \le \rho(\tilde{x}) \le \mathrm{dist}(K, \mathrm{supp}\phi_\lambda)$. Putting it all together we have $|x - x_\lambda| \le 3\mathrm{dist}(x_\lambda, \mathrm{supp}\phi_\lambda)$ and hence

$$|x_0 - x_\lambda| \le |x_0 - x| + |x - x_\lambda| \le 4\mathrm{dist}(x_\lambda, \mathrm{supp}\phi_\lambda) \le 4|x - x_0| \le 4\delta \qquad (2.36)$$

as expected. By our choice of δ we have $|F(x_\lambda) - F(x_0)| \le \varepsilon$ for all λ with $\phi_\lambda(x) \ne 0$. Hence $|F(x) - F(x_0)| \le \varepsilon$ whenever $|x - x_0| \le \delta$ and we are done. □

Note that the same proof works if X is only a metric space.

Finally, let me remark that the Brouwer fixed point theorem is equivalent to the fact that there is no continuous retraction $R : B_1(0) \to \partial B_1(0)$ (with $R(x) = x$ for $x \in \partial B_1(0)$) from the unit ball to the unit sphere in $\mathbb{R}^n$.

In fact, if R would be such a retraction, $-R$ would have a fixed point $x_0 \in \partial B_1(0)$ by Brouwer's theorem. But then $x_0 = -f(x_0) = -x_0$ which is impossible. Conversely, if a continuous function $f : B_1(0) \to B_1(0)$ has no fixed point we can define a retraction $R(x) = f(x) + t(x)(x - f(x))$, where $t(x) \ge 0$ is chosen such that $|R(x)|^2 = 1$ (i.e., $R(x)$ lies on the intersection of the line spanned by x, $f(x)$ with the unit sphere).

Using this equivalence the Brouwer fixed point theorem can also be derived easily by showing that the homology groups of the unit ball $B_1(0)$ and its boundary (the unit sphere) differ (see, e.g., [9] for details).

2.5 Kakutani's fixed-point theorem and applications to game theory

In this section we want to apply Brouwer's fixed-point theorem to show the existence of Nash equilibria for n-person games. As a preparation we extend Brouwer's fixed-point theorem to set valued functions. This generalization will be more suitable for our purpose.

Denote by $\mathrm{CS}(K)$ the set of all nonempty convex subsets of K.

Theorem 2.16 (Kakutani) *Suppose K is a compact convex subset of $\mathbb{R}^n$ and $f: K \to \mathrm{CS}(K)$. If the set*

$$\Gamma = \{(x,y) | y \in f(x)\} \subseteq K^2 \tag{2.37}$$

is closed, then there is a point $x \in K$ such that $x \in f(x)$.

Proof. Our strategy is to apply Brouwer's theorem, hence we need a function related to f. For this purpose it is convenient to assume that K is a simplex

$$K = \langle v_1, \dots, v_m \rangle, \qquad m \le n, \tag{2.38}$$

where v_i are the vertices. If we pick $y_i \in f(v_i)$ we could set

$$f^1(x) = \sum_{i=1}^{m} \lambda_i y_i, \tag{2.39}$$

where λ_i are the barycentric coordinates of x (i.e., $\lambda_i \ge 0$, $\sum_{i=1}^m \lambda_i = 1$ and $x = \sum_{i=1}^n \lambda_i v_i$). By construction, $f^1 \in C(K,K)$ and there is a fixed point x^1. But unless x^1 is one of the vertices, this doesn't help us too much. So lets choose a better function as follows. Consider the k-th barycentric subdivision and for each vertex v_i in this subdivision pick an element $y_i \in f(v_i)$. Now define $f^k(v_i) = y_i$ and extend f^k to the interior of each subsimplex as before. Hence $f^k \in C(K,K)$ and there is a fixed point

$$x^k = \sum_{i=1}^{m} \lambda_i^k v_i^k = \sum_{i=1}^{m} \lambda_i^k y_i^k, \quad y_i^k = f^k(v_i^k), \tag{2.40}$$

in the subsimplex $\langle v_1^k, \dots, v_m^k \rangle$. Since $(x^k, \lambda_1^k, \dots, \lambda_m^k, y_1^k, \dots, y_m^k) \in K^{2m+1}$ we can assume that this sequence converges to $(x^0, \lambda_1^0, \dots, \lambda_m^0, y_1^0, \dots, y_m^0)$ after passing to a subsequence. Since the subsimplices shrink to a point, this implies $v_i^k \to x^0$ and hence $y_i^0 \in f(x^0)$ since $(v_i^k, y_i^k) \in \Gamma \to (v_i^0, y_i^0) \in \Gamma$ by the closedness assumption. Now (2.40) tells us

$$x^0 = \sum_{i=1}^{m} \lambda_i^k y_i^k \in f(x^0) \tag{2.41}$$

since $f(x^0)$ is convex and the claim holds if K is a simplex.

If K is not a simplex, we can pick a simplex S containing K and proceed as in the proof of the Brouwer theorem. □

If $f(x)$ contains precisely one point for all x, then Kakutani's theorem reduces to the Brouwer's theorem.

Now we want to see how this applies to game theory.

An n-person game consists of n players who have m_i possible actions to choose from. The set of all possible actions for the i-th player will be denoted by $\Phi_i = \{1, \dots, m_i\}$. An element $\varphi_i \in \Phi_i$ is also called a pure strategy for reasons to become clear in a moment. Once all players have chosen their move φ_i, the payoff for each player is given by the payoff function

$$R_i(\varphi), \qquad \varphi = (\varphi_1, \dots, \varphi_n) \in \Phi = \prod_{i=1}^{n} \Phi_i \tag{2.42}$$

of the i-th player. We will consider the case where the game is repeated a large number of times and where in each step the players choose their action according to a fixed strategy. Here a strategy s_i for the i-th player is a probability distribution on Φ_i, that is, $s_i = (s_i^1, \dots, s_i^{m_i})$ such that $s_i^k \ge 0$ and $\sum_{k=1}^{m_i} s_i^k = 1$. The set of all possible strategies for the i-th player is denoted by S_i. The number s_i^k is the probability for the k-th pure strategy to be chosen. Consequently, if $s = (s_1, \dots, s_n) \in S = \prod_{i=1}^n S_i$ is a collection of strategies, then the probability that a given collection of pure strategies gets chosen is

$$s(\varphi) = \prod_{i=1}^{n} s_i(\varphi), \qquad s_i(\varphi) = s_i^{k_i},\ \varphi = (k_1, \dots, k_n) \in \Phi \tag{2.43}$$

(assuming all players make their choice independently) and the expected payoff for player i is

$$R_i(s) = \sum_{\varphi \in \Phi} s(\varphi) R_i(\varphi). \tag{2.44}$$

By construction, $R_i(s)$ is continuous.

The question is of course, what is an optimal strategy for a player? If the other strategies are known, a best reply of player i against s would be a strategy $\overline{s_i}$ satisfying

$$R_i(s \backslash \overline{s}_i) = \max_{\tilde{s}_i \in S_i} R_i(s \backslash \tilde{s}_i) \tag{2.45}$$

Here $s\backslash\tilde{s}_i$ denotes the strategy combination obtained from s by replacing s_i by $\tilde{s}_i$. The set of all best replies against s for the i-th player is denoted by $B_i(s)$. Explicitly, $\overline{s}_i \in B(s)$ if and only if $\overline{s}_i^k = 0$ whenever $R_i(s\backslash k) < \max_{1\le l\le m_i} R_i(s\backslash l)$ (in particular $B_i(s) \neq \emptyset$).

Let $s, \overline{s} \in S$, we call $\overline{s}$ a best reply against s if $\overline{s}_i$ is a best reply against s for all i. The set of all best replies against s is $B(s) = \prod_{i=1}^n B_i(s)$.

A strategy combination $\overline{s} \in S$ is a Nash equilibrium for the game if it is a best reply against itself, that is,

$$\overline{s} \in B(\overline{s}). \tag{2.46}$$

Or, put differently, $\overline{s}$ is a Nash equilibrium if no player can increase his payoff by changing his strategy as long as all others stick to their respective strategies. In addition, if a player sticks to his equilibrium strategy, he is assured that his payoff will not decrease no matter what the others do.

To illustrate these concepts, let us consider the famous *prisoners dilemma*. Here we have two players which can choose to defect or cooperate. The payoff is symmetric for both players and given by the following diagram

$$\begin{array}{c|cc} R_1 & d_2 & c_2 \\ \hline d_1 & 0 & 2 \\ c_1 & -1 & 1 \end{array} \qquad \begin{array}{c|cc} R_2 & d_2 & c_2 \\ \hline d_1 & 0 & -1 \\ c_1 & 2 & 1 \end{array} \tag{2.47}$$

where c_i or d_i means that player i cooperates or defects, respectively. It is easy to see that the (pure) strategy pair (d_1, d_2) is the only Nash equilibrium for this game and that the expected payoff is 0 for both players. Of course, both players could get the payoff 1 if they both agree to cooperate. But if one would break this agreement in order to increase his payoff, the other one would get less. Hence it might be safer to defect.

Now that we have seen that Nash equilibria are a useful concept, we want to know when such an equilibrium exists. Luckily we have the following result.

Theorem 2.17 (Nash) *Every n-person game has at least one Nash equilibrium.*

Proof. The definition of a Nash equilibrium begs us to apply Kakutani's theorem to the set valued function $s \mapsto B(s)$. First of all, S is compact and convex and so are the sets $B(s)$. Next, observe that the closedness condition of Kakutani's theorem is satisfied since if $s^m \in S$ and $\overline{s}^m \in B(s^n)$ both converge to s and $\overline{s}$, respectively, then (2.45) for s^m, $\overline{s}^m$

$$R_i(s^m\backslash\tilde{s}_i) \le R_i(s^m\backslash\overline{s}_i^m), \qquad \tilde{s}_i \in S_i,\ 1 \le i \le n, \tag{2.48}$$

implies (2.45) for the limits s, $\overline{s}$

$$R_i(s\backslash\tilde{s}_i) \le R_i(s\backslash\overline{s}_i), \qquad \tilde{s}_i \in S_i,\ 1 \le i \le n, \tag{2.49}$$

by continuity of $R_i(s)$. □

2.6 Further properties of the degree

We now prove some additional properties of the mapping degree. The first one will relate the degree in $\mathbb{R}^n$ with the degree in $\mathbb{R}^m$. It will be needed later on to extend the definition of degree to infinite dimensional spaces. By virtue of the canonical embedding $\mathbb{R}^m \hookrightarrow \mathbb{R}^m \times \{0\} \subset \mathbb{R}^n$ we can consider $\mathbb{R}^m$ as a subspace of $\mathbb{R}^n$.

Theorem 2.18 (Reduction property) *Let $f \in C(\overline{U}, \mathbb{R}^m)$ and $y \in \mathbb{R}^m\backslash(\mathbb{1} + f)(\partial U)$, then*

$$\deg(\mathbb{1} + f, U, y) = \deg(\mathbb{1} + f_m, U_m, y), \tag{2.50}$$

where $f_m = f|_{U_m}$, where $U - M$ is the projection of U to $\mathbb{R}^m$.

Proof. Choose a $\tilde{f} \in C^2(U, \mathbb{R}^m)$ sufficiently close to f such that $y \in \mathrm{RV}(\tilde{f})$. Let $x \in (\mathbb{1}+\tilde{f})^{-1}(y)$, then $x = y - f(x) \in \mathbb{R}^m$ implies $(\mathbb{1}+\tilde{f})^{-1}(y) = (\mathbb{1}+\tilde{f}_m)^{-1}(y)$. Moreover,

$$\begin{aligned} J_{\mathbf{1}+\tilde{f}}(x) &= \det(\mathbb{1} + \tilde{f}')(x) = \det \begin{pmatrix} \delta_{ij} + \partial_j \tilde{f}_i(x) & \partial_j \tilde{f}_j(x) \\ 0 & \delta_{ij} \end{pmatrix} \\ &= \det(\delta_{ij} + \partial_j \tilde{f}_i) = J_{\mathbf{1}+\tilde{f}_m}(x) \end{aligned} \tag{2.51}$$

shows $\deg(\mathbb{1}+f, U, y) = \deg(\mathbb{1}+\tilde{f}, U, y) = \deg(\mathbb{1}+\tilde{f}_m, U_m, y) = \deg(\mathbb{1}+f_m, U_m, y)$ as desired. □

Let $U \subseteq \mathbb{R}^n$ and $f \in C(\overline{U}, \mathbb{R}^n)$ be as usual. By Theorem 2.2 we know that $\deg(f, U, y)$ is the same for every y in a connected component of $\mathbb{R}^n\backslash f(\partial U)$. We will denote these components by K_j and write $\deg(f, U, y) = \deg(f, U, K_j)$ if $y \in K_j$.

Theorem 2.19 (Product formula) *Let $U \subseteq \mathbb{R}^n$ be a bounded and open set and denote by G_j the connected components of $\mathbb{R}^n\backslash f(\partial U)$. If $g \circ f \in D_y(U, \mathbb{R}^n)$, then*

$$\deg(g \circ f, U, y) = \sum_j \deg(f, U, G_j) \deg(g, G_j, y), \tag{2.52}$$

where only finitely many terms in the sum are nonzero.

Proof. Since $f(\overline{U})$ is is compact, we can find an $r > 0$ such that $f(\overline{U}) \subseteq B_r(0)$. Moreover, since $g^{-1}(y)$ is closed, $g^{-1}(y) \cap B_r(0)$ is compact and hence can be covered by finitely many components $\{G_j\}_{j=1}^m$. In particular, the others will have $\deg(f, U, G_k) = 0$ and hence only finitely many terms in the above sum are nonzero.

We begin by computing $\deg(g \circ f, U, y)$ in the case where $f, g \in C^1$ and $y \notin \mathrm{CV}(g \circ f)$. Since $(g \circ f)'(x) = g'(f(x))f'(x)$ the claim is a straightforward calculation

$$\begin{aligned}
\deg(g \circ f, U, y) &= \sum_{x \in (g \circ f)^{-1}(y)} \mathrm{sgn}(J_{g \circ f}(x)) \\
&= \sum_{x \in (g \circ f)^{-1}(y)} \mathrm{sgn}(J_g(f(x)))\mathrm{sgn}(J_f(x)) \\
&= \sum_{z \in g^{-1}(y)} \mathrm{sgn}(J_g(z)) \sum_{x \in f^{-1}(z)} \mathrm{sgn}(J_f(x)) \\
&= \sum_{z \in g^{-1}(y)} \mathrm{sgn}(J_g(z)) \deg(f, U, z)
\end{aligned}$$

and, using our cover $\{G_j\}_{j=1}^m$,

$$\begin{aligned}
\deg(g \circ f, U, y) &= \sum_{j=1}^m \sum_{z \in g^{-1}(y) \cap G_j} \mathrm{sgn}(J_g(z)) \deg(f, U, z) \\
&= \sum_{j=1}^m \deg(f, U, G_j) \sum_{z \in g^{-1}(y) \cap G_j} \mathrm{sgn}(J_g(z)) && (2.53) \\
&= \sum_{j=1}^m \deg(f, U, G_j) \deg(g, G_j, y). && (2.54)
\end{aligned}$$

Moreover, this formula still holds for $y \in \mathrm{CV}(g \circ f)$ and for $g \in C$ by construction of the Brouwer degree. However, the case $f \in C$ will need a closer investigation since the sets G_j depend on f. To overcome this problem we will introduce the sets

$$L_l = \{z \in \mathbb{R}^n \backslash f(\partial U) | \deg(f, U, z) = l\}. \tag{2.55}$$

Observe that L_l, $l > 0$, must be a union of some sets of $\{G_j\}_{j=1}^m$.

Now choose $\tilde{f} \in C^1$ such that $|f(x) - \tilde{f}(x)| < 2^{-1}\mathrm{dist}(g^{-1}(y), f(\partial U))$ for $x \in \overline{U}$ and define $\tilde{K}_j$, $\tilde{L}_l$ accordingly. Then we have $U_l \cap g^{-1}(y) = \tilde{U}_l \cap g^{-1}(y)$ by

Theorem 2.1 (iii). Moreover,

$$\begin{aligned}\deg(f\circ g,U,y) &= \deg(\tilde{f}\circ g,U,y) = \sum_j \deg(f,U,\tilde{K}_j)\deg(g,\tilde{K}_j,y)\\ &= \sum_{l>0} l\deg(g,\tilde{U}_l,y) = \sum_{l>0} l\deg(g,U_l,y)\\ &= \sum_j \deg(f,U,G_j)\deg(g,G_j,y) \qquad (2.56)\end{aligned}$$

which proves the claim. □

2.7 The Jordan curve theorem

In this section we want to show how the product formula (2.52) for the Brouwer degree can be used to prove the famous Jordan curve theorem which states that a homeomorphic image of the circle dissects $\mathbb{R}^2$ into two components (which necessarily have the image of the circle as common boundary). In fact, we will even prove a slightly more general result.

Theorem 2.20 *Let $C_j \subset \mathbb{R}^n$, $j = 1,2$, be homeomorphic compact sets. Then $\mathbb{R}^n \backslash C_1$ and $\mathbb{R}^n \backslash C_2$ have the same number of connected components.*

Proof. Denote the components of $\mathbb{R}^n\backslash C_1$ by H_j and those of $\mathbb{R}^n\backslash C_2$ by K_j. Let $h : C_1 \to C_2$ be a homeomorphism with inverse $k : C_2 \to C_1$. By Theorem 2.15 we can extend both to $\mathbb{R}^n$. Then Theorem 2.1 (iii) and the product formula imply

$$1 = \deg(k\circ h, H_j, y) = \sum_l \deg(h,H_j,G_l)\deg(k,G_l,y) \qquad (2.57)$$

for any $y \in H_j$. Now we have

$$\bigcup_i K_i = \mathbb{R}^n\backslash C_2 \subseteq \mathbb{R}^n\backslash h(\partial H_j) = \bigcup_l G_l \qquad (2.58)$$

and hence fore every i we have $K_i \subseteq G_l$ for some l since components are maximal connected sets. Let $N_l = \{i | K_i \subseteq G_l\}$ and observe that we have $\deg(k,G_l,y) =$

$\sum_{i\in N_l}\deg(k,K_i,y)$ and $\deg(h,H_j,G_l)=\deg(h,H_j,K_i)$ for every $i\in N_l$. Therefore,

$$1=\sum_l\sum_{i\in N_l}\deg(h,H_j,K_i)\deg(k,K_i,y)=\sum_i\deg(h,H_j,K_i)\deg(k,K_i,H_j) \tag{2.59}$$

By reversing the role of C_1 and C_2, the same formula holds with H_j and K_i interchanged.

Hence

$$\sum_i 1=\sum_i\sum_j\deg(h,H_j,K_i)\deg(k,K_i,H_j)=\sum_j 1 \tag{2.60}$$

shows that if the number of components of $\mathbb{R}^n\backslash C_1$ or $\mathbb{R}^n\backslash C_2$ is finite, then so is the other and both are equal. Otherwise there is nothing to prove. □

Chapter 3

The Leray–Schauder mapping degree

3.1 The mapping degree on finite dimensional Banach spaces

The objective of this section is to extend the mapping degree from $\mathbb{R}^n$ to general Banach spaces. Naturally, we will first consider the finite dimensional case.

Let X be a (real) Banach space of dimension n and let ϕ be any isomorphism between X and $\mathbb{R}^n$. Then, for $f \in D_y(\overline{U}, X)$, $U \subset X$ open, $y \in X$, we can define

$$\deg(f, U, y) = \deg(\phi \circ f \circ \phi^{-1}, \phi(U), \phi(y)) \tag{3.1}$$

provided this definition is independent of the basis chosen. To see this let ψ be a second isomorphism. Then $A = \psi \circ \phi^{-1} \in \mathrm{GL}(n)$. Abbreviate $f^* = \phi \circ f \circ \phi^{-1}$, $y^* = \phi(y)$ and pick $\tilde{f}^* \in C^1_y(\phi(\overline{U}), \mathbb{R}^n)$ in the same component of $D_y(\phi(\overline{U}), \mathbb{R}^n)$ as f^* such that $y^* \in \mathrm{RV}(f^*)$. Then $A \circ \tilde{f}^* \circ A^{-1} \in C^1_y(\psi(U), \mathbb{R}^n)$ is the same component of $D_y(\psi(\overline{U}), \mathbb{R}^n)$ as $A \circ f^* \circ A^{-1} = \psi \circ f \circ \psi^{-1}$ (since A is also a homeomorphism) and

$$J_{A\circ\tilde{f}^*\circ A^{-1}}(Ay^*) = \det(A) J_{\tilde{f}^*}(y^*) \det(A^{-1}) = J_{\tilde{f}^*}(y^*) \tag{3.2}$$

by the chain rule. Thus we have $\deg(\psi \circ f \circ \psi^{-1}, \psi(U), \psi(y)) = \deg(\phi \circ f \circ \phi^{-1}, \phi(U), \phi(y))$ and our definition is independent of the basis chosen. In addition, it inherits all properties from the mapping degree in $\mathbb{R}^n$. Note also that the reduction property holds if $\mathbb{R}^m$ is replaced by an arbitrary subspace X_1 since we can always choose $\phi : X \to \mathbb{R}^n$ such that $\phi(X_1) = \mathbb{R}^m$.

Our next aim is to tackle the infinite dimensional case. The general idea is to approximate F by finite dimensional operators (in the same spirit as we approximated continuous f by smooth functions). To do this we need to know which operators can be approximated by finite dimensional operators. Hence we have to recall some basic facts first.

3.2 Compact operators

Let X, Y be Banach spaces and $U \subset X$. An operator $F : U \subset X \to Y$ is called finite dimensional if its range is finite dimensional. In addition, it is called compact if it is continuous and maps bounded sets into relatively compact ones. The set of all compact operators is denoted by $\mathcal{C}(U, Y)$ and the set of all compact, finite dimensional operators is denoted by $\mathcal{F}(U, Y)$. Both sets are normed linear spaces and we have $\mathcal{F}(U, Y) \subseteq \mathcal{C}(U, Y) \subseteq C(U, Y)$.

If U is compact, then $\mathcal{C}(U, Y) = C(U, Y)$ (since the continuous image of a compact set is compact) and if $\dim(Y) < \infty$, then $\mathcal{F}(U, Y) = \mathcal{C}(U, Y)$. In particular, if $U \subset \mathbb{R}^n$ is bounded, then $\mathcal{F}(\overline{U}, \mathbb{R}^n) = \mathcal{C}(\overline{U}, \mathbb{R}^n) = C(\overline{U}, \mathbb{R}^n)$.

Now let us collect some results to be needed in the sequel.

Lemma 3.1 *If $K \subset X$ is compact, then for every $\varepsilon > 0$ there is a finite dimensional subspace $X_\varepsilon \subseteq X$ and a continuous map $P_\varepsilon : K \to X_\varepsilon$ such that $|P_\varepsilon(x) - x| \le \varepsilon$ for all $x \in K$.*

Proof. Pick $\{x_i\}_{i=1}^n \subseteq K$ such that $\bigcup_{i=1}^n B_\varepsilon(x_i)$ covers K. Let $\{\phi_i\}_{i=1}^n$ be a partition of unity (restricted to K) subordinate to $\{B_\varepsilon(x_i)\}_{i=1}^n$, that is, $\phi_i \in C(K, [0, 1])$ with $\mathrm{supp}(\phi_i) \subset B_\varepsilon(x_i)$ and $\sum_{i=1}^n \phi_i(x) = 1$, $x \in K$. Set

$$P_\varepsilon(x) = \sum_{i=1}^n \phi_i(x) x_i, \tag{3.3}$$

then

$$\begin{aligned} |P_\varepsilon(x) - x| &= |\sum_{i=1}^n \phi_i(x) x - \sum_{i=1}^n \phi_i(x) x_i| \\ &\le \sum_{i=1}^n \phi_i(x) |x - x_i| \le \varepsilon. \end{aligned} \tag{3.4}$$

□

This lemma enables us to prove the following important result.

Theorem 3.2 *Let U be bounded, then the closure of $\mathcal{F}(U,Y)$ in $C(U,Y)$ is $\mathcal{C}(U,Y)$.*

Proof. Suppose $F_N \in \mathcal{C}(U,Y)$ converges to F. If $F \not\in \mathcal{C}(U,Y)$ then we can find a sequence $x_n \in U$ such that $|F(x_n) - F(x_m)| \geq \rho > 0$ for $n \neq m$. If N is so large that $|F - F_N| \leq \rho/4$, then

$$\begin{aligned} |F_N(x_n) - F_N(x_m)| &\geq |F(x_n) - F(x_m)| - |F_N(x_n) - F(x_n)| - |F_N(x_m) - F(x_m)| \\ &\geq \rho - 2\frac{\rho}{4} = \frac{\rho}{2} \end{aligned} \tag{3.5}$$

This contradiction shows $\overline{\mathcal{F}(U,y)} \subseteq \mathcal{C}(U,Y)$. Conversely, let $K = \overline{F(U)}$ and choose P_ε according to Lemma 3.1, then $F_\varepsilon = P_\varepsilon \circ F \in \mathcal{F}(U,Y)$ converges to F. Hence $\mathcal{C}(U,Y) \subseteq \overline{\mathcal{F}(U,y)}$ and we are done. □

Finally, let us show some interesting properties of mappings $\mathbb{1} + F$, where $F \in \mathcal{C}(U,Y)$.

Lemma 3.3 *Let U be bounded and closed. Suppose $F \in \mathcal{C}(U,Y)$, then $\mathbb{1} + F$ is proper (i.e., inverse images of compact sets are compact) and maps closed subsets to closed subsets.*

Proof. Let $A \subseteq U$ be closed and $y_n = (\mathbb{1} + F)(x_n) \in (\mathbb{1} + F)(A)$. Since $\{y_n - x_n\} \subset F^{-1}(\{y_n\})$ we can assume that $y_n - x_n \to z$ after passing to a subsequence and hence $x_n \to x = y - z \in A$. Since $y = x + F(x) \in (\mathbb{1} + F)(A)$, $(\mathbb{1} + F)(A)$ is closed.

Next, let U be closed and $K \subset X$ be compact. Let $\{x_n\} \subseteq (\mathbb{1} + F)^{-1}(K)$. Then we can pass to a subsequence $y_{n_m} = x_{n_m} + F(x_{n_m})$ such that $y_{n_m} \to y$. As before this implies $x_{n_m} \to x$ and thus $(\mathbb{1} + F)^{-1}(K)$ is compact. □

Now we are all set for the definition of the Leray–Schauder degree, that is, for the extension of our degree to infinite dimensional Banach spaces.

3.3 The Leray–Schauder mapping degree

For $U \subset X$ we set $\mathcal{D}_y(\overline{U}, X) = \{F \in \mathcal{C}(\overline{U}, X) | y \not\in (\mathbb{1} + F)(\partial U)\}$ and $\mathcal{F}_y(\overline{U}, X) = \{F \in \mathcal{F}(\overline{U}, X) | y \not\in (\mathbb{1} + F)(\partial U)\}$. Note that for $F \in \mathcal{D}_y(\overline{U}, X)$ we have $\operatorname{dist}(y, (\mathbb{1} + F)(\partial U)) > 0$ since $\mathbb{1} + F$ maps closed sets to closed sets.

Abbreviate $\rho = \operatorname{dist}(y, (\mathbb{1} + F)(\partial U))$ and pick $F_1 \in \mathcal{F}(\overline{U}, X)$ such that $|F - F_1| < \rho$ implying $F_1 \in \mathcal{F}_y(\overline{U}, X)$. Next, let X_1 be a finite dimensional subspace of X such that $F_1(U) \subset X_1$, $y \in X_1$ and set $U_1 = U \cap X_1$. Then we have $F_1 \in \mathcal{F}_y(\overline{U_1}, X_1)$ and might define

$$\deg(\mathbb{1} + F, U, y) = \deg(\mathbb{1} + F_1, U_1, y) \tag{3.6}$$

provided we show that this definition is independent of F_1 and X_1 (as above). Pick another operator $F_2 \in \mathcal{F}(\overline{U}, X)$ such that $|F - F_2| < \rho$ and let X_2 be a corresponding finite dimensional subspace as above. Consider $X_0 = X_1 + X_2$, $U_0 = U \cap X_0$, then $F_i \in \mathcal{F}_y(\overline{U_0}, X_0)$, $i = 1, 2$, and

$$\deg(\mathbb{1} + F_i, U_0, y) = \deg(\mathbb{1} + F_i, U_i, y), \quad i = 1, 2, \tag{3.7}$$

by the reduction property. Moreover, set $H(t) = \mathbb{1} + (1-t)F_1 + t\,F_2$ implying $H(t) \in$, $t \in [0,1]$, since $|H(t) - (\mathbb{1} + F)| < \rho$ for $t \in [0,1]$. Hence homotopy invariance

$$\deg(\mathbb{1} + F_1, U_0, y) = \deg(\mathbb{1} + F_2, U_0, y) \tag{3.8}$$

shows that (3.6) is independent of F_1, X_1.

Theorem 3.4 *Let U be a bounded open subset of a (real) Banach space X and let $F \in \mathcal{D}_y(\overline{U}, X)$, $y \in X$. Then the following hold true.*

(i). $\deg(\mathbb{1} + F, U, y) = \deg(\mathbb{1} + F - y, U, 0)$.

(ii). $\deg(\mathbb{1}, U, y) = 1$ *if* $y \in U$.

(iii). If $U_{1,2}$ are open, disjoint subsets of U such that $y \notin f(\overline{U} \backslash (U_1 \cup U_2))$, then $\deg(\mathbb{1} + F, U, y) = \deg(\mathbb{1} + F, U_1, y) + \deg(\mathbb{1} + F, U_2, y)$.

(iv). If $H : [0,1] \times \overline{U} \to X$ and $y : [0,1] \to X$ are both continuous such that $H(t) \in D_{y(t)}(U, \mathbb{R}^n)$, $t \in [0,1]$, then $\deg(\mathbb{1} + H(0), U, y(0)) = \deg(\mathbb{1} + H(1), U, y(1))$.

Proof. Except for (iv) all statements follow easily from the definition of the degree and the corresponding property for the degree in finite dimensional spaces. Considering $H(t,x) - y(t)$, we can assume $y(t) = 0$ by (i). Since $H([0,1], \partial U)$ is compact, we have $\rho = \text{dist}(y, H([0,1], \partial U) > 0$. By Theorem 3.2 we can pick $H_1 \in \mathcal{F}([0,1] \times U, X)$ such that $|H(t) - H_1(t)| < \rho$, $t \in [0,1]$. this implies $\deg(\mathbb{1} + H(t), U, 0) = \deg(\mathbb{1} + H_1(t), U, 0)$ and the rest follows from Theorem 2.2. □

In addition, Theorem 2.1 and Theorem 2.2 hold for the new situation as well (no changes are needed in the proofs).

Theorem 3.5 *Let $F, G \in \mathcal{D}_y(U, X)$, then the following statements hold.*

(i). We have $\deg(\mathbb{1} + F, \emptyset, y) = 0$. *Moreover, if U_i, $1 \le i \le N$, are disjoint open subsets of U such that $y \notin (\mathbb{1}+F)(\overline{U} \setminus \bigcup_{i=1}^N U_i)$, then* $\deg(\mathbb{1}+F, U, y) = \sum_{i=1}^N \deg(\mathbb{1}+F, U_i, y)$.

(ii). If $y \notin (\mathbb{1}+F)(U)$, then $\deg(\mathbb{1}+F, U, y) = 0$ *(but not the other way round). Equivalently, if* $\deg(\mathbb{1}+F, U, y) \ne 0$, *then $y \in (\mathbb{1}+F)(U)$.*

(iii). If $|f(x) - g(x)| < \mathrm{dist}(y, f(\partial U))$, $x \in \partial U$, then $\deg(f, U, y) = \deg(g, U, y)$. *In particular, this is true if $f(x) = g(x)$ for $x \in \partial U$.*

(iv). $\deg(\mathbb{1} + ., U, y)$ *is constant on each component of $D_y(\overline{U}, X)$.*

(v). $\deg(\mathbb{1} + F, U, .)$ *is constant on each component of $X \setminus f(\partial U)$.*

3.4 The Leray–Schauder principle and the Schauder fixed-point theorem

As a first consequence we note the Leray–Schauder principle which says that a priori estimates yield existence.

Theorem 3.6 (Leray–Schauder principle) *Suppose $F \in \mathcal{C}(X, X)$ and any solution x of $x = tF(x)$, $t \in [0,1]$ satisfies the a priori bound $|x| \le M$ for some $M > 0$, then F has a fixed point.*

Proof. Pick $\rho > M$ and observe $\deg(\mathbb{1} + F, B_\rho(0), 0) = \deg(\mathbb{1}, B_\rho(0), 0) = 1$ using the compact homotopy $H(t,x) = tF(x)$. Here $0 \notin H(t, \partial B_\rho(0))$ due to the a priori bound. □

Now we can extend the Brouwer fixed-point theorem to infinite dimensional spaces as well.

Theorem 3.7 (Schauder fixed point) *Let K be a closed, convex, and bounded subset of a Banach space X. If $F \in \mathcal{C}(K, K)$, then F has at least one fixed point. The result remains valid if K is only homeomorphic to a closed, convex, and bounded subset.*

Proof. Since K is bounded, there is a $\rho > 0$ such that $K \subseteq B_\rho(0)$. By Theorem 2.15 we can find a continuous retraction $R : X \to K$ (i.e., $R(x) = x$ for $x \in K$) and consider $\tilde{F} = F \circ R \in \mathcal{C}(\overline{B_\rho(0)}, \overline{B_\rho(0)})$. The compact homotopy $H(t, x) = t\tilde{F}(x)$ shows that $\deg(\mathbb{1} + \tilde{F}, B_\rho(0), 0) = \deg(\mathbb{1}, B_\rho(0), 0) = 1$. Hence there is a point $x_0 = \tilde{F}(x_0) \in K$. Since $\tilde{F}(x_0) = F(x_0)$ for $x_0 \in K$ we are done. □

Finally, let us prove another fixed-point theorem which covers several others as special cases.

Theorem 3.8 *Let $U \subset X$ be open and bounded and let $F \in \mathcal{C}(\overline{U}, X)$. Suppose there is an $x_0 \in U$ such that*

$$F(x) - x_0 \neq \alpha(x - x_0), \qquad x \in \partial U,\ \alpha \in (1, \infty). \tag{3.9}$$

Then F has a fixed point.

Proof. Consider $H(t, x) = x - x_0 - t(F(x) - x_0)$, then we have $H(t, x) \neq 0$ for $x \in \partial U$ and $t \in [0, 1]$ by assumption. If $H(1, x) = 0$ for some $x \in \partial U$, then x is a fixed point and we are done. Otherwise we have $\deg(\mathbb{1} - F, U, 0) = \deg(\mathbb{1} - x_0, U, 0) = \deg(\mathbb{1}, U, x_0) = 1$ and hence F has a fixed point. □

Now we come to the anticipated corollaries.

Corollary 3.9 *Let $U \subset X$ be open and bounded and let $F \in \mathcal{C}(\overline{U}, X)$. Then F has a fixed point if one of the following conditions holds.*

1. *$U = B_\rho(0)$ and $F(\partial U) \subseteq \overline{U}$ (Rothe).*
2. *$U = B_\rho(0)$ and $|F(x) - x|^2 \geq |F(x)|^2 - |x|^2$ for $x \in \partial U$ (Altman).*
3. *X is a Hilbert space, $U = B_\rho(0)$ and $\langle F(x), x\rangle \leq |x|^2$ for $x \in \partial U$ (Krasnosel'skii).*

Proof. (1). $F(\partial U) \subseteq \overline{U}$ and $F(x) = \alpha x$ for $|x| = \rho$ implies $|\alpha|\rho \leq \rho$ and hence (3.9) holds. (2). $F(x) = \alpha x$ for $|x| = \rho$ implies $(\alpha - 1)^2\rho^2 \geq (\alpha^2 - 1)\rho^2$ and hence $\alpha \leq 0$. (3). Special case of (2) since $|F(x) - x|^2 = |F(x)|^2 - 2\langle F(x), x\rangle + |x|^2$. □

3.5 Applications to integral and differential equations

In this section we want to show how our results can be applied to integral and differential equations. To be able to apply our results we will need to know that certain integral operators are compact.

Lemma 3.10 *Suppose* $I = [a, b] \subset \mathbb{R}$ *and* $f \in C(I \times I \times \mathbb{R}^n, \mathbb{R}^n)$, $\tau \in C(I, I)$, *then*

$$\begin{array}{rlcl} F: & C(I, \mathbb{R}^n) & \to & C(I, \mathbb{R}^n) \\ & x(t) & \mapsto & F(x)(t) = \int_a^{\tau(t)} f(t, s, x(s)) ds \end{array} \tag{3.10}$$

is compact.

Proof. We first need to prove that F is continuous. Fix $x_0 \in C(I, \mathbb{R}^n)$ and $\varepsilon > 0$. Set $\rho = |x_0| + 1$ and abbreviate $\overline{B} = \overline{B_\rho(0)} \subset \mathbb{R}^n$. The function f is uniformly continuous on $Q = I \times I \times \overline{B}$ since Q is compact. Hence for $\varepsilon_1 = \varepsilon/(b-a)$ we can find a $\delta \in (0, 1]$ such that $|f(t, s, x) - f(t, s, y)| \le \varepsilon_1$ for $|x - y| < \delta$. But this implies

$$\begin{aligned} |F(x) - F(x_0)| &= \sup_{t \in I} \left| \int_a^{\tau(t)} f(t, s, x(s)) - f(t, s, x_0(s)) ds \right| \\ &\le \sup_{t \in I} \int_a^{\tau(t)} |f(t, s, x(s)) - f(t, s, x_0(s))| ds \\ &\le \sup_{t \in I} (b-a)\varepsilon_1 = \varepsilon, \end{aligned} \tag{3.11}$$

for $|x - x_0| < \delta$. In other words, F is continuous. Next we note that if $U \subset C(I, \mathbb{R}^n)$ is bounded, say $|U| < \rho$, then

$$|F(U)| \le \sup_{x \in U} \left| \int_a^{\tau(t)} f(t, s, x(s)) ds \right| \le (b-a)M, \tag{3.12}$$

where $M = \max |f(I, I, \overline{B_\rho(0)})|$. Moreover, the family $F(U)$ is equicontinuous. Fix ε and $\varepsilon_1 = \varepsilon/(2(b-a))$, $\varepsilon_2 = \varepsilon/(2M)$. Since f and τ are uniformly continuous on $I \times I \times \overline{B_\rho(0)}$ and I, respectively, we can find a $\delta > 0$ such that $|f(t, s, x) -$

$f(t_0,s,x)| \le \varepsilon_1$ and $|\tau(t)-\tau(t_0)| \le \varepsilon_2$ for $|t-t_0| < \delta$. Hence we infer for $|t-t_0| < \delta$

$$\begin{aligned}|F(x)(t) - F(x)(t_0)| &= \left| \int_a^{\tau(t)} f(t,s,x(s))ds - \int_a^{\tau(t_0)} f(t_0,s,x(s))ds \right| \\ &\le \int_a^{\tau(t_0)} |f(t,s,x(s)) - f(t_0,s,x(s))|ds + \left| \int_{\tau(t_0)}^{\tau(t)} |f(t,s,x(s))|ds \right| \\ &\le (b-a)\varepsilon_1 + \varepsilon_2 M = \varepsilon. \end{aligned} \tag{3.13}$$

This implies that $F(U)$ is relatively compact by the Arzelà-Ascoli theorem. Thus F is compact. □

As a first application we use this result to show existence of solutions to integral equations.

Theorem 3.11 *Let F be as in the previous lemma. Then the integral equation*

$$x - \lambda F(x) = y, \qquad \lambda \in \mathbb{R}, y \in C(I, \mathbb{R}^n) \tag{3.14}$$

has at least one solution $x \in C(I,\mathbb{R}^n)$ if $|\lambda| \le \rho/M(\rho)$, where $M(\rho) = (b-a)\max_{(s,t,x)\in I\times I\times\overline{B_\rho(0)}} |f(s,t,x-y(s))|$ and $\rho > 0$ is arbitrary.

Proof. Note that, by our assumption on λ, λF maps $B_\rho(y)$ into itself. Now apply the Schauder fixed-point theorem. □

This result immediately gives the Peano theorem for ordinary differential equations.

Theorem 3.12 (Peano) *Consider the initial value problem*

$$\dot{x} = f(t,x), \quad x(t_0) = x_0, \tag{3.15}$$

where $f \in C(I,\mathbb{R}^n)$ and $I \subset \mathbb{R}$ is an interval containing t_0. Then (3.15) has at least one local solution $x \in C^1([t_0-\varepsilon, t_0+\varepsilon], \mathbb{R}^n)$, $\varepsilon > 0$. For example, any ε satisfying $\varepsilon M(\varepsilon,\rho) \le \rho$, $\rho > 0$ with $M(\varepsilon,\rho) = \max|f([t_0-\varepsilon,t_0+\varepsilon],\overline{B_\rho(x_0)})|$ works. In addition, if $M(\varepsilon,\rho) \le \tilde{M}(\varepsilon)(1+\rho)$, then there exists a global solution.

Proof. For notational simplicity we make the shift $t \to t-t_0$, $x \to x-x_0$, $f(t,x) \to f(t+t_0, x+t_0)$ and assume $t_0 = 0$, $x_0 = 0$. In addition, it suffices to consider $t \ge 0$ since $t \to -t$ amounts to $f \to -f$.

Now observe, that (3.15) is equivalent to

$$x(t) - \int_0^t f(s, x(s))ds, \qquad x \in C([-\varepsilon, \varepsilon], \mathbb{R}^n) \tag{3.16}$$

and the first part follows from our previous theorem. To show the second, fix $\varepsilon > 0$ and assume $M(\varepsilon, \rho) \leq \tilde{M}(\varepsilon)(1 + \rho)$. Then

$$|x(t)| \leq \int_0^t |f(s, x(s))|ds \leq \tilde{M}(\varepsilon) \int_0^t (1 + |x(s)|)ds \tag{3.17}$$

implies $|x(t)| \leq \exp(\tilde{M}(\varepsilon)\varepsilon)$ by Gronwall's inequality. Hence we have an a priori bound which implies existence by the Leary–Schauder principle. Since ε was arbitrary we are done. □

Chapter 4

The stationary Navier–Stokes equation

4.1 Introduction and motivation

In this chapter we turn to partial differential equations. In fact, we will only consider one example, namely the stationary Navier–Stokes equation. Our goal is to use the Leray–Schauder principle to prove an existence and uniqueness result for solutions.

Let U $(\neq \emptyset)$ be an open, bounded, and connected subset of $\mathbb{R}^3$. We assume that U is filled with an incompressible fluid described by its velocity field $v_j(t, x)$ and its pressure $p(t, x)$, $(t, x) \in \mathbb{R} \times U$. The requirement that our fluid is incompressible implies $\partial_j v_j = 0$ (we sum over two equal indices from 1 to 3), which follows from the Gauss theorem since the flux trough any closed surface must be zero.

Rather than just writing down the equation, let me give a short physical motivation. To obtain the equation which governs such a fluid we consider the forces acting on a small cube spanned by the points (x_1, x_2, x_3) and $(x_1 + \Delta x_1, x_2 + \Delta x_2, x_3 + \Delta x_3)$. We have three contributions from outer forces, pressure differences, and viscosity.

The outer force density (force per volume) will be denoted by K_j and we assume that it is known (e.g. gravity).

The force from pressure acting on the surface through (x_1, x_2, x_3) normal to the x_1-direction is $p\Delta x_2 \Delta x_3 \delta_{1j}$. The force from pressure acting on the opposite surface is $-(p + \partial_1 p \Delta x_1)\Delta x_2 \Delta x_3 \delta_{1j}$. In summary, we obtain

$$-(\partial_j p)\Delta V, \tag{4.1}$$

where $\Delta V = \Delta x_1 \Delta x_2 \Delta x_3$.

The viscosity acting on the surface through (x_1, x_2, x_3) normal to the x_1-direction is $-\eta \Delta x_2 \Delta x_3 \partial_1 v_j$ by some physical law. Here $\eta > 0$ is the viscosity constant of the fluid. On the opposite surface we have $\eta \Delta x_2 \Delta x_3 \partial_1 (v_j + \partial_1 v_j \Delta x_1)$. Adding up the contributions of all surface we end up with

$$\eta \Delta V \partial_i \partial_i v_j. \tag{4.2}$$

Putting it all together we obtain from Newton's law

$$\rho \Delta V \frac{d}{dt} v_j(t, x(t)) = \eta \Delta V\, \partial_i \partial_i v_j(t, x(t)) - (\partial_j p(t, x(t)) + \Delta V\, K_j(t, x(t)), \tag{4.3}$$

where $\rho > 0$ is the density of the fluid. Dividing by ΔV and using the chain rule yields the Navier–Stokes equation

$$\rho \partial_t v_j = \eta \partial_i \partial_i v_j - \rho (v_i \partial_i) v_j - \partial_j p + K_j. \tag{4.4}$$

Note that it is no restriction to assume $\rho = 1$.

In what follows we will only consider the stationary Navier–Stokes equation

$$0 = \eta \partial_i \partial_i v_j - (v_i \partial_i) v_j - \partial_j p + K_j. \tag{4.5}$$

In addition to the incompressibility condition $\partial_j v_j = 0$ we also require the boundary condition $v|_{\partial U} = 0$, which follows from experimental observations.

In summary, we consider the problem (4.5) for v in (e.g.) $X = \{v \in C^2(\overline{U}, \mathbb{R}^3) | \partial_j v_j = 0 \text{ and } v|_{\partial U} = 0\}$.

Our strategy is to rewrite the stationary Navier–Stokes equation in integral form, which is more suitable for our further analysis. For this purpose we need to introduce some function spaces first.

4.2 An insert on Sobolev spaces

Let U be a bounded open subset of $\mathbb{R}^n$ and let $L^p(U, \mathbb{R})$ denote the Lebesgue spaces of p integrable functions with norm

$$|u|_p = \left(\int_U |u(x)|^p dx \right)^{1/p}. \tag{4.6}$$

In the case $p = 2$ we even have a scalar product

$$\langle u, v\rangle_2 = \int_U u(x)v(x)dx \tag{4.7}$$

and our aim is to extend this case to include derivatives.

Given the set $C^1(U, \mathbb{R})$ we can consider the scalar product

$$\langle u, v\rangle_{2,1} = \int_U u(x)v(x)dx + \int_U (\partial_j u)(x)(\partial_j v)(x)dx. \tag{4.8}$$

Taking the completion with respect to the associated norm we obtain the Sobolev space $H^1(U, \mathbb{R})$. Similarly, taking the completion of $C_0^1(U, \mathbb{R})$ with respect to the same norm, we obtain the Sobolev space $H_0^1(U, \mathbb{R})$. Here $C_0^r(U, Y)$ denotes the set of functions in $C^r(U, Y)$ with compact support. This construction of $H^1(U, \mathbb{R})$ implies that a sequence u_k in $C^1(U, \mathbb{R})$ converges to $u \in H^1(U, \mathbb{R})$ if and only if u_k and all its first order derivatives $\partial_j u_k$ converge in $L^2(U, \mathbb{R})$. Hence we can assign each $u \in H^1(U, \mathbb{R})$ its first order derivatives $\partial_j u$ by taking the limits from above. In order to show that this is a useful generalization of the ordinary derivative, we need to show that the derivative depends only on the limiting function $u \in L^2(U, \mathbb{R})$. To see this we need the following lemma.

Lemma 4.1 (Integration by parts) *Suppose $u \in H_0^1(U, \mathbb{R})$ and $v \in H^1(U, \mathbb{R})$, then*

$$\int_U u(\partial_j v)dx = -\int_U (\partial_j u)v\, dx. \tag{4.9}$$

Proof. By continuity it is no restriction to assume $u \in C_0^1(U, \mathbb{R})$ and $v \in C^1(U, \mathbb{R})$. Moreover, we can find a function $\phi \in C_0^1(U, \mathbb{R})$ which is 1 on the support of u. Hence by considering ϕv we can even assume $v \in C_0^1(U, \mathbb{R})$.

Moreover, we can replace U by a rectangle K containing U and extend u, v to K by setting it 0 outside U. Now use integration by parts with respect to the j-th coordinate. □

In particular, this lemma says that if $u \in H^1(U, \mathbb{R})$, then

$$\int_U (\partial_j u)\phi dx = -\int_U u(\partial_j \phi)\, dx, \qquad \phi \in C_0^\infty(U, \mathbb{R}). \tag{4.10}$$

And since $C_0^\infty(U, \mathbb{R})$ is dense in $L^2(U, \mathbb{R})$, the derivatives are uniquely determined by $u \in L^2(U, \mathbb{R})$ alone. Moreover, if $u \in C^1(U, \mathbb{R})$, then the derivative in the

Sobolev space corresponds to the usual derivative. In summary, $H^1(U,\mathbb{R})$ is the space of all functions $u \in L^2(U,\mathbb{R})$ which have first order derivatives (in the sense of distributions, i.e., (4.10)) in $L^2(U,\mathbb{R})$.

Next, we want to consider some additional properties which will be used later on. First of all, the Poincaré-Friedrichs inequality.

Lemma 4.2 (Poincaré-Friedrichs inequality) *Suppose* $u \in H_0^1(U,\mathbb{R})$*, then*

$$\int_U u^2 dx \le d_j^2 \int_U (\partial_j u)^2 dx, \tag{4.11}$$

where $d_j = \sup\{(x_j - y_j)^2 | (x_1,\dots,x_n),(y_1,\dots,y_n) \in U\}$.

Proof. Again we can assume $u \in C_0^1(U,\mathbb{R})$ and we assume $j = 1$ for notational convenience. Replace U by a set $K = [a,b] \times \tilde{K}$ containing U and extend u to K by setting it 0 outside U. Then we have

$$\begin{aligned} u(x_1,x_2,\dots,x_n)^2 &= \left(\int_a^{x_1} 1 \cdot (\partial_1 u)(\xi,x_2,\dots,x_n) d\xi\right)^2 \\ &\le (b-a)\int_a^b (\partial_1 u)^2(\xi,x_2,\dots,x_n) d\xi, \end{aligned} \tag{4.12}$$

where we have used the Cauchy-Schwarz inequality. Integrating this result over $[a,b]$ gives

$$\int_a^b u^2(\xi,x_2,\dots,x_n) d\xi \le (b-a)^2 \int_a^b (\partial_1 u)^2(\xi,x_2,\dots,x_n) d\xi \tag{4.13}$$

and integrating over $\tilde{K}$ finishes the proof. □

Hence, from the view point of Banach spaces, we could also equip $H_0^1(U,\mathbb{R})$ with the scalar product

$$\langle u,v\rangle = \int_U (\partial_j u)(x)(\partial_j v)(x) dx. \tag{4.14}$$

This scalar product will be more convenient for our purpose and hence we will use it from now on. (However, all results stated will hold in either case.) The norm corresponding to this scalar product will be denoted by $|.|$.

Next, we want to consider the embedding $H^1_0(U,\mathbb{R}) \hookrightarrow L^2(U,\mathbb{R})$ a little closer. This embedding is clearly continuous since by the Poincaré-Friedrichs inequality we have

$$|u|_2 \le \frac{d(U)}{\sqrt{n}} |u|, \qquad d(U) = \sup\{|x-y| \,|x, y \in U\}. \tag{4.15}$$

Moreover, by a famous result of Rellich, it is even compact. To see this we first prove the following inequality.

Lemma 4.3 (Poincaré inequality) *Let $Q \subset \mathbb{R}^n$ be a cube with edge length ρ. Then*

$$\int_Q u^2 dx \le \frac{1}{\rho^n} \left(\int_Q u dx \right)^2 + \frac{n\rho^2}{2} \int_Q (\partial_k u)(\partial_k u) dx \tag{4.16}$$

for all $u \in H^1(Q,\mathbb{R})$.

Proof. After a scaling we can assume $Q = (0,1)^n$. Moreover, it suffices to consider $u \in C^1(Q,\mathbb{R})$.

Now observe

$$u(x) - u(\tilde{x}) = \sum_{i=1}^n \int_{x^{i-1}}^{x^i} (\partial_i u) dx_i, \tag{4.17}$$

where $x^i = (\tilde{x}_1, \dots, \tilde{x}_i, x_{i+1}, \dots, x_n)$. Squaring this equation and using Cauchy–Schwarz on the right hand side we obtain

$$\begin{aligned} u(x)^2 - 2u(x)u(\tilde{x}) + u(\tilde{x})^2 &\le \left(\sum_{i=1}^n \int_0^1 |\partial_i u| dx_i \right)^2 \le n \sum_{i=1}^n \left(\int_0^1 |\partial_i u| dx_i \right)^2 \\ &\le n \sum_{i=1}^n \int_0^1 (\partial_i u)^2 dx_i. \end{aligned} \tag{4.18}$$

Now we integrate over x and $\tilde{x}$, which gives

$$2 \int_Q u^2 dx - 2 \left(\int_Q u\, dx \right)^2 \le n \int_Q (\partial_i u)(\partial_i u) dx \tag{4.19}$$

and finishes the proof. □

Now we are ready to show Rellich's compactness theorem.

Theorem 4.4 (Rellich's compactness theorem) *Let U be a bounded open subset of $\mathbb{R}^n$. Then the embedding*

$$H_0^1(U,\mathbb{R}) \hookrightarrow L^2(U,\mathbb{R}) \tag{4.20}$$

is compact.

Proof. Pick a cube Q (with edge length ρ) containing U and a bounded sequence $u^k \in H_0^1(U,\mathbb{R})$. Since bounded sets are weakly compact, it is no restriction to assume that u^k is weakly convergent in $L^2(U,\mathbb{R})$. By setting $u^k(x) = 0$ for $x \notin U$ we can also assume $u^k \in H^1(Q,\mathbb{R})$. Next, subdivide Q into N subcubes Q_i with edge lengths ρ/N. On each subcube (4.16) holds and hence

$$\int_U u^2 dx = \int_Q u^2 dx = \sum_{i=1}^{N^n} \frac{N}{\rho} \left(\int_{Q_i} u dx \right)^2 + \frac{n\rho^2}{2N^2} \int_U (\partial_k u)(\partial_k u) dx \tag{4.21}$$

for all $u \in H^1(U,\mathbb{R})$. Hence we infer

$$|u^k - u^\ell|_2^2 \le \sum_{i=1}^{N^n} \frac{N}{\rho} \left(\int_{Q_i} (u^k - u^\ell) dx \right)^2 + \frac{n\rho^2}{2N^2} |u^k - u^\ell|^2. \tag{4.22}$$

The last term can be made arbitrarily small by picking N large. The first term converges to 0 since u^k converges weakly and each summand contains the L^2 scalar product of $u^k - u^\ell$ and χ_{Q_i} (the characteristic function of Q_i). □

In addition to this result we will also need the following interpolation inequality.

Lemma 4.5 (Ladyzhenskaya inequality) *Let $U \subset \mathbb{R}^3$. For all $u \in H_0^1(U,\mathbb{R})$ we have*

$$|u|_4 \le \sqrt[4]{8} |u|_2^{1/4} |u|^{3/4}. \tag{4.23}$$

Proof. We first prove the case where $u \in C_0^1(U,\mathbb{R})$. The key idea is to start with $U \subset \mathbb{R}^1$ and then work ones way up to $U \subset \mathbb{R}^2$ and $U \subset \mathbb{R}^3$.

If $U \subset \mathbb{R}^1$ we have

$$u(x)^2 = \int^x \partial_1 u^2(x_1) dx_1 \le 2 \int |u \partial_1 u| dx_1 \tag{4.24}$$

and hence

$$\max_{x \in U} u(x)^2 \le 2 \int |u \partial_1 u| dx_1. \tag{4.25}$$

Here, if an integration limit is missing, it means that the integral is taken over the whole support of the function.

If $U \subset \mathbb{R}^2$ we have

$$\begin{aligned}
\iint u^4 dx_1 dx_2 &\leq \int \max_x u(x,x_2)^2 dx_2 \int \max_y u(x_1,y)^2 dx_1 \\
&\leq 4 \iint |u\partial_1 u| dx_1 dx_2 \iint |u\partial_2 u| dx_1 dx_2 \\
&\leq 4 \Big(\iint u^2 dx_1 dx_2 \Big)^{2/2} \Big(\iint (\partial_1 u)^2 dx_1 dx_2 \Big)^{1/2} \Big(\iint (\partial_2 u)^2 dx_1 dx_2 \Big)^{1/2} \\
&\leq 4 \iint u^2 dx_1 dx_2 \iint ((\partial_1 u)^2 + (\partial_2 u)^2) dx_1 dx_2 \qquad (4.26)
\end{aligned}$$

Now let $U \subset \mathbb{R}^3$, then

$$\begin{aligned}
\iiint u^4 dx_1 dx_2 dx_3 &\leq 4 \int dx_3 \iint u^2 dx_1 dx_2 \iint ((\partial_1 u)^2 + (\partial_2 u)^2) dx_1 dx_2 \\
&\leq 4 \iint \max_z u(x_1,x_2,z)^2 dx_1 dx_2 \iiint ((\partial_1 u)^2 + (\partial_2 u)^2) dx_1 dx_2 dx_3 \\
&\leq 8 \iiint |u\partial_3 u| dx_1 dx_2 dx_3 \iiint ((\partial_1 u)^2 + (\partial_2 u)^2) dx_1 dx_2 dx_3 \qquad (4.27)
\end{aligned}$$

and applying Cauchy–Schwarz finishes the proof for $u \in C_0^1(U,\mathbb{R})$.

If $u \in H_0^1(U,\mathbb{R})$ pick a sequence u_k in $C_0^1(U,\mathbb{R})$ which converges to u in $H_0^1(U,\mathbb{R})$ and hence in $L^2(U,\mathbb{R})$. By our inequality, this sequence is Cauchy in $L^4(U,\mathbb{R})$ and converges to a limit $v \in L^4(U,\mathbb{R})$. Since $|u|_2 \leq \sqrt[4]{|U|}|u|_4$ ($\int 1 \cdot u^2 dx \leq \sqrt{\int 1\, dx \int u^4 dx}$), u_k converges to v in $L^2(U,\mathbb{R})$ as well and hence $u = v$. Now take the limit in the inequality for u_k. □

As a consequence we obtain

$$|u|_4 \leq \left(\frac{8d(U)}{\sqrt{3}} \right)^{1/4} |u|, \qquad U \subset \mathbb{R}^3, \tag{4.28}$$

and

Corollary 4.6 *The embedding*

$$H_0^1(U,\mathbb{R}) \hookrightarrow L^4(U,\mathbb{R}), \qquad U \subset \mathbb{R}^3, \tag{4.29}$$

is compact.

Proof. Let u_k be a bounded sequence in $H^1_0(U,\mathbb{R})$. By Rellich's theorem there is a subsequence converging in $L^2(U,\mathbb{R})$. By the Ladyzhenskaya inequality this subsequence converges in $L^4(U,\mathbb{R})$. □

Our analysis clearly extends to functions with values in $\mathbb{R}^n$ since $H^1_0(U,\mathbb{R}^n) = \oplus_{j=1}^n H^1_0(U,\mathbb{R})$.

4.3 Existence and uniqueness of solutions

Now we come to the reformulation of our original problem (4.5). We pick as underlying Hilbert space $H^1_0(U,\mathbb{R}^3)$ with scalar product

$$\langle u, v\rangle = \int_U (\partial_j u_i)(\partial_j v_i) dx. \tag{4.30}$$

Let $\mathcal{X}$ be the closure of X in $H^1_0(U,\mathbb{R}^3)$, that is,

$$\mathcal{X} = \overline{\{v \in C^2(\overline{U},\mathbb{R}^3) | \partial_j v_j = 0 \text{ and } v|_{\partial U} = 0\}} = \{v \in H^1_0(U,\mathbb{R}^3) | \partial_j v_j = 0\}. \tag{4.31}$$

Now we multiply (4.5) by $w \in X$ and integrate over U

$$\int_U \Big(\eta\partial_k\partial_k v_j - (v_k\partial_k)v_j + K_j\Big) w_j\, dx = \int_U (\partial_j p) w_j\, dx = 0. \tag{4.32}$$

Using integration by parts this can be rewritten as

$$\int_U \Big(\eta(\partial_k v_j)(\partial_k w_j) - v_k v_j(\partial_k w_j) - K_j w_j\Big) dx = 0. \tag{4.33}$$

Hence if v is a solution of the Navier-Stokes equation, then it is also a solution of

$$\eta\langle v, w\rangle - a(v, v, w) - \int_U K w\, dx = 0, \quad \text{for all } w \in \mathcal{X}, \tag{4.34}$$

where

$$a(u, v, w) = \int_U u_k v_j (\partial_k w_j)\, dx. \tag{4.35}$$

In other words, (4.34) represents a necessary solubility condition for the Navier-Stokes equations. A solution of (4.34) will also be called a **weak solution** of the Navier-Stokes equations. If we can show that a weak solution is in C^2, then we can read our argument backwards and it will be also a classical solution. However, in general this might not be true and it will only solve the Navier-Stokes equations in the sense of distributions. But let us try to show existence of solutions for (4.34) first.

For later use we note

$$\begin{aligned} a(v,v,v) &= \int_U v_k v_j (\partial_k v_j)\, dx = \frac{1}{2} \int_U v_k \partial_k (v_j v_j)\, dx \\ &= -\frac{1}{2} \int_U (v_j v_j) \partial_k v_k\, dx = 0, \qquad v \in \mathcal{X}. \end{aligned} \tag{4.36}$$

We proceed by studying (4.34). Let $K \in L^2(U,\mathbb{R}^3)$, then $\int_U Kw\, dx$ is a linear functional on $\mathcal{X}$ and hence there is a $\tilde{K} \in \mathcal{X}$ such that

$$\int_U Kw\, dx = \langle \tilde{K}, w \rangle, \qquad w \in \mathcal{X}. \tag{4.37}$$

Moreover, the same is true for the map $a(u,v,.)$, $u,v \in \mathcal{X}$, and hence there is an element $B(u,v) \in \mathcal{X}$ such that

$$a(u,v,w) = \langle B(u,v), w \rangle, \qquad w \in \mathcal{X}. \tag{4.38}$$

In addition, the map $B : \mathcal{X}^2 \to \mathcal{X}$ is bilinear. In summary we obtain

$$\langle \eta v - B(v,v) - \tilde{K}, w \rangle = 0, \qquad w \in \mathcal{X}, \tag{4.39}$$

and hence

$$\eta v - B(v,v) = \tilde{K}. \tag{4.40}$$

So in order to apply the theory from our previous chapter, we need a Banach space Y such that $\mathcal{X} \hookrightarrow Y$ is compact.

Let us pick $Y = L^4(U,\mathbb{R}^3)$. Then, applying the Cauchy-Schwarz inequality twice to each summand in $a(u,v,w)$ we see

$$\begin{aligned} |a(u,v,w)| &\le \sum_{j,k} \Big(\int_U (u_k v_j)^2 dx \Big)^{1/2} \Big(\int_U (\partial_k w_j)^2 dx \Big)^{1/2} \\ &\le |w| \sum_{j,k} \Big(\int_U (u_k)^4 dx \Big)^{1/4} \Big(\int_U (v_j)^4 dx \Big)^{1/4} = |u|_4 |v|_4 |w|. \end{aligned} \tag{4.41}$$

Moreover, by Corollary 4.6 the embedding $\mathcal{X} \hookrightarrow Y$ is compact as required.

Motivated by this analysis we formulate the following theorem.

Theorem 4.7 *Let $\mathcal{X}$ be a Hilbert space, Y a Banach space, and suppose there is a compact embedding $\mathcal{X} \hookrightarrow Y$. In particular, $|u|_Y \le \beta |u|$. Let $a : \mathcal{X}^3 \to \mathbb{R}$ be a multilinear form such that*

$$|a(u,v,w)| \le \alpha |u|_Y |v|_Y |w| \tag{4.42}$$

and $a(v,v,v) = 0$. Then for any $\tilde{K} \in \mathcal{X}$, $\eta > 0$ we have a solution $v \in \mathcal{X}$ to the problem

$$\eta \langle v, w\rangle - a(v,v,w) = \langle \tilde{K}, w\rangle, \qquad w \in \mathcal{X}. \tag{4.43}$$

Moreover, if $2\alpha\beta|\tilde{K}| < \eta^2$ this solution is unique.

Proof. It is no loss to set $\eta = 1$. Arguing as before we see that our equation is equivalent to

$$v - B(v,v) + \tilde{K} = 0, \tag{4.44}$$

where our assumption (4.42) implies

$$|B(u,v)| \le \alpha |u|_Y |v|_Y \le \alpha\beta^2 |u||v| \tag{4.45}$$

Here the second equality follows since the embedding $\mathcal{X} \hookrightarrow Y$ is continuous.

Abbreviate $F(v) = B(v,v)$. Observe that F is locally Lipschitz continuous since if $|u|, |v| \le \rho$ we have

$$|F(u) - F(v)| = |B(u-v,u) - B(v,u-v)| \le 2\alpha\,\rho\,|u-v|_Y \le 2\alpha\beta^2\,\rho|u-v|. \tag{4.46}$$

Moreover, let v_n be a bounded sequence in $\mathcal{X}$. After passing to a subsequence we can assume that v_n is Cauchy in Y and hence $F(v_n)$ is Cauchy in $\mathcal{X}$ by $|F(u) - F(v)| \le 2\alpha\,\rho|u-v|_Y$. Thus $F : \mathcal{X} \to \mathcal{X}$ is compact.

Hence all we need to apply the Leray-Schauder principle is an a priori estimate. Suppose v solves $v = tF(v) + t\tilde{K}$, $t \in [0,1]$, then

$$\langle v, v\rangle = t\,a(v,v,v) + t\langle \tilde{K}, v\rangle = t\langle \tilde{K}, v\rangle. \tag{4.47}$$

Hence $|v| \le |\tilde{K}|$ is the desired estimate and the Leray-Schauder principle yields existence of a solution.

Now suppose there are two solutions v_i, $i = 1,2$. By our estimate they satisfy $|v_i| \le |\tilde{K}|$ and hence $|v_1 - v_2| = |F(v_1) - F(v_2)| \le 2\alpha\beta^2\,|\tilde{K}||v_1 - v_2|$ which is a contradiction if $2\alpha\beta^2\,|\tilde{K}| < 1$. □

Hence we have found a solution v to the generalized problem (4.34). This solution is unique if $2(\frac{2d(U)}{\sqrt{3}})^{3/2}|K|_2 < \eta^2$. Under suitable additional conditions on the outer forces and the domain, it can be shown that weak solutions are C^2 and thus also classical solutions. However, this is beyond the scope of this introductory text.

Chapter 5

Monotone operators

5.1 Monotone operators

The Leray–Schauder theory can only be applied to compact perturbations of the identity. If F is not compact, we need different tools. In this section we briefly present another class of operators, namely monotone ones, which allow some progress.

If $F : \mathbb{R} \to \mathbb{R}$ is continuous and we want $F(x) = y$ to have a unique solution for every $y \in \mathbb{R}$, then f should clearly be strictly monotone increasing (or decreasing) and satisfy $\lim_{x\to\pm\infty} F(x) = \pm\infty$. Rewriting these conditions slightly such that they make sense for vector valued functions the analogous result holds.

Lemma 5.1 *Suppose $F : \mathbb{R}^n \to \mathbb{R}^n$ is continuous and satisfies*

$$\lim_{|x|\to\infty} \frac{F(x)x}{|x|} = \infty. \tag{5.1}$$

Then the equation

$$F(x) = y \tag{5.2}$$

has a solution for every $y \in \mathbb{R}^n$. If F is strictly monotone

$$(F(x) - F(y))(x - y) > 0, \qquad x \neq y, \tag{5.3}$$

then this solution is unique.

Proof. Our first assumption implies that $G(x) = F(x) - y$ satisfies $G(x)x = F(x)x - yx > 0$ for $|x|$ sufficiently large. Hence the first claim follows from Theorem 2.13. The second claim is trivial. □

Now we want to generalize this result to infinite dimensional spaces. Throughout this chapter, X will be a Hilbert space with scalar product $\langle ., .. \rangle$. An operator $F : X \to X$ is called **monotone** if

$$\langle F(x) - F(y), x - y \rangle \geq 0, \qquad x, y \in X, \tag{5.4}$$

strictly monotone if

$$\langle F(x) - F(y), x - y \rangle > 0, \qquad x \neq y \in X, \tag{5.5}$$

and finally **strongly monotone** if there is a constant $C > 0$ such that

$$\langle F(x) - F(y), x - y \rangle \geq C|x - y|^2, \qquad x, y \in X. \tag{5.6}$$

Note that the same definitions can be made if X is a Banach space and $F : X \to X^*$.

Observe that if F is strongly monotone, then it automatically satisfies

$$\lim_{|x| \to \infty} \frac{\langle F(x), x \rangle}{|x|} = \infty. \tag{5.7}$$

(Just take $y = 0$ in the definition of strong monotonicity.) Hence the following result is not surprising.

Theorem 5.2 (Zarantonello) *Suppose $F \in C(X, X)$ is (globally) Lipschitz continuous and strongly monotone. Then, for each $y \in X$ the equation*

$$F(x) = y \tag{5.8}$$

has a unique solution $x \in X$.

Proof. Set

$$G(x) = x - t(F(x) - y), \qquad t > 0, \tag{5.9}$$

then $F(x) = y$ is equivalent to the fixed point equation

$$G(x) = x. \tag{5.10}$$

It remains to show that G is a contraction. We compute

$$\begin{aligned} |G(x) - G(\tilde{x})|^2 &= |x - \tilde{x}|^2 - 2t\langle F(x) - F(\tilde{x}), x - \tilde{x} \rangle + t^2 |F(x) - F(\tilde{x})|^2 \\ &\leq (1 - 2\frac{C}{L}(Lt) + (Lt)^2)|x - \tilde{x}|^2, \end{aligned} \tag{5.11}$$

where L is a Lipschitz constant for F (i.e., $|F(x)-F(\tilde{x})| \le L|x-\tilde{x}|$). Thus, if $t \in (0, \frac{2C}{L})$, G is a contraction and the rest follows from the contraction principle. □

Again observe that our proof is constructive. In fact, the best choice for t is clearly $t = \frac{C}{L}$ such that the contraction constant $\theta = 1 - (\frac{C}{L})^2$ is minimal. Then the sequence

$$x_{n+1} = x_n - (1 - (\frac{C}{L})^2)(F(x_n) - y), \quad x_0 = x, \tag{5.12}$$

converges to the solution.

5.2 The nonlinear Lax–Milgram theorem

As a consequence of the last theorem we obtain a nonlinear version of the Lax–Milgram theorem. We want to investigate the following problem:

$$a(x,y) = b(y), \qquad \text{for all } y \in X, \tag{5.13}$$

where $a : X^2 \to \mathbb{R}$ and $b : X \to \mathbb{R}$. For this equation the following result holds.

Theorem 5.3 (Nonlinear Lax–Milgram theorem) *Suppose $b \in \mathcal{L}(X,\mathbb{R})$ and $a(x,.) \in \mathcal{L}(X,\mathbb{R})$, $x \in X$, are linear functionals such that there are positive constants L and C such that for all $x, y, z \in X$ we have*

$$a(x, x-y) - a(y, x-y) \ge C|x-y|^2 \tag{5.14}$$

and

$$|a(x,z) - a(y,z)| \le L|z||x-y|. \tag{5.15}$$

Then there is a unique $x \in X$ such that (5.13) holds.

Proof. By the Riez theorem there are elements $F(x) \in X$ and $z \in X$ such that $a(x,y) = b(y)$ is equivalent to $\langle F(x) - z, y\rangle = 0$, $y \in X$, and hence to

$$F(x) = z. \tag{5.16}$$

By (5.14) the operator F is strongly monotone. Moreover, by (5.15) we infer

$$|F(x) - F(y)| = \sup_{\tilde{x} \in X, |\tilde{x}|=1} |\langle F(x) - F(y), \tilde{x}\rangle| \le L|x-y| \tag{5.17}$$

that F is Lipschitz continuous. Now apply Theorem 5.2. □

The special case where $a \in \mathcal{L}^2(X, \mathbb{R})$ is a bounded bilinear form which is strongly continuous, that is,

$$a(x,x) \geq C|x|^2, \qquad x \in X, \tag{5.18}$$

is usually known as (linear) Lax–Milgram theorem.

The typical application of this theorem is the existence of a unique weak solution of the Dirichlet problem for **elliptic equations**

$$\begin{aligned} \partial_i A_{ij}(x)\partial_j u(x) + b_j(x)\partial_j u(x) + c(x)u(x) = f(x), &\qquad x \in U, \\ u(x) = 0, &\qquad x \in \partial U, \end{aligned} \tag{5.19}$$

where U is a bounded open subset of $\mathbb{R}^n$. By elliptic we mean that all coefficients A, b, c plus the right hand side f are bounded and $a_0 > 0$, where

$$a_0 = \inf_{e \in S^n, x \in U} e_i A_{ij}(x) e_j, \quad b_0 = -\inf_{x \in U} b(x), \quad c_0 = \inf_{x \in U} c(x). \tag{5.20}$$

As in Section 4.3 we pick $H_0^1(U, \mathbb{R})$ with scalar product

$$\langle u, v \rangle = \int_U (\partial_j u)(\partial_j v) dx \tag{5.21}$$

as underlying Hilbert space. Next we multiply (5.19) by $v \in H_0^1$ and integrate over U

$$\int_U \Big(\partial_i A_{ij}(x)\partial_j u(x) + b_j(x)\partial_j u(x) + c(x)u(x)\Big) v(x)\, dx = \int_U f(x)v(x)\, dx. \tag{5.22}$$

After a partial integration we can write this equation as

$$a(v,u) = f(v), \qquad v \in H_0^1, \tag{5.23}$$

where

$$\begin{aligned} a(v,u) &= \int_U \Big(\partial_i v(x) A_{ij}(x)\partial_j u(x) + b_j(x)v(x)\partial_j u(x) + c(x)v(x)u(x)\Big) dx \\ f(v) &= \int_U f(x)v(x)\, dx, \end{aligned} \tag{5.24}$$

We call a solution of (5.23) a **weak solution** of the elliptic Dirichlet problem (5.19).

By a simple use of the Cauchy-Schwarz and Poincaré-Friedrichs inequalities we see that the bilinear form $a(u,v)$ is bounded. To be able to apply the (linear) Lax–Milgram theorem we need to show that it satisfies $a(u,u) \geq \int |\partial_j u|^2 dx$.

Using (5.20) we have

$$a(u,u) \geq \int_U \Big(a_0 |\partial_j u|^2 - b_0 |u||\partial_j u| + c_0 |u|^2 \Big), \tag{5.25}$$

where $-b_0 = \inf b(x)$, $c_0 = \inf c(x)$ and we need to control the middle term. If $b_0 \leq 0$ there is nothing to do and it suffices to require $c_0 \geq 0$.

If $b_0 > 0$ we distribute the middle term by means of the elementary inequality

$$|u||\partial_j u| \leq \frac{\varepsilon}{2}|u|^2 + \frac{1}{2\varepsilon}|\partial_j u|^2 \tag{5.26}$$

which gives

$$a(u,u) \geq \int_U \Big((a_0 - \frac{b_0}{2\varepsilon})|\partial_j u|^2 + (c_0 - \frac{\varepsilon b_0}{2})|u|^2 \Big). \tag{5.27}$$

Since we need $a_0 - \frac{b_0}{2\varepsilon} > 0$ and $c_0 - \frac{\varepsilon b_0}{2} \geq 0$, or equivalently $\frac{2c_0}{b_0} \geq \varepsilon > \frac{b_0}{2a_0}$, we see that we can apply the Lax–Milgram theorem if $4a_0c_0 > b_0^2$. In summary, we have proven

Theorem 5.4 *The elliptic Dirichlet problem (5.19) has a unique weak solution* $u \in H_0^1(U,\mathbb{R})$ *if* $a_0 > 0$, $b_0 \leq 0$, $c_0 \geq 0$ *or* $4a_0c_0 > b_0^2$.

5.3 The main theorem of monotone operators

Now we return to the investigation of $F(x) = y$ and weaken the conditions of Theorem 5.2. We will assume that X is a separable Hilbert space and that $F : X \to X$ is a continuous monotone operator satisfying

$$\lim_{|x|\to\infty} \frac{\langle F(x), x\rangle}{|x|} = \infty. \tag{5.28}$$

In fact, if suffices to assume that F is **weakly continuous**

$$\lim_{n\to\infty} \langle F(x_n), y\rangle = \langle F(x), y\rangle, \qquad \text{for all } y \in X \tag{5.29}$$

whenever $x_n \to x$.

The idea is as follows: Start with a finite dimensional subspace $X_n \subset X$ and project the equation $F(x) = y$ to X_n resulting in an equation

$$F_n(x_n) = y_n, \qquad x_n, y_n \in X_n. \tag{5.30}$$

More precisely, let P_n be the (linear) projection onto X_n and set $F_n(x_n) = P_n F(x_n)$, $y_n = P_n y$ (verify that F_n is continuous and monotone!).

Now Lemma 5.1 ensures that there exists a solution u_n. Now chose the subspaces X_n such that $X_n \to X$ (i.e., $X_n \subset X_{n+1}$ and $\bigcup_{n=1}^{\infty} X_n$ is dense). Then our hope is that u_n converges to a solution u.

This approach is quite common when solving equations in infinite dimensional spaces and is known as **Galerkin approximation**. It can often be used for numerical computations and the right choice of the spaces X_n will have a significant impact on the quality of the approximation.

So how should we show that x_n converges? First of all observe that our construction of x_n shows that x_n lies in some ball with radius R_n, which is chosen such that

$$\langle F_n(x), x\rangle > |y_n||x|, \qquad |x| \geq R_n,\ x \in X_n. \tag{5.31}$$

Since $\langle F_n(x), x\rangle = \langle P_n F(x), x\rangle = \langle F(x), P_n x\rangle = \langle F(x), x\rangle$ for $x \in X_n$ we can drop all n's to obtain a constant R which works for all n. So the sequence x_n is uniformly bounded

$$|x_n| \leq R. \tag{5.32}$$

Now by a well-known result there exists a weakly convergent subsequence. That is, after dropping some terms, we can assume that there is some x such that $x_n \rightharpoonup x$, that is,

$$\langle x_n, z\rangle \to \langle x, z\rangle, \qquad \text{for every } z \in X. \tag{5.33}$$

And it remains to show that x is indeed a solution. This follows from

Lemma 5.5 *Suppose $F : X \to X$ is weakly continuous and monotone, then*

$$\langle y - F(z), x - z\rangle \geq 0 \quad \textit{for every } z \in X \tag{5.34}$$

implies $F(x) = y$.

Proof. Choose $z = x \pm tw$, then $\mp\langle y - F(x \pm tw), w\rangle \geq 0$ and by continuity $\mp\langle y - F(x), w\rangle \geq 0$. Thus $\langle y - F(x), w\rangle = 0$ for every w implying $y - F(x) = 0$. □

Now we can show

Theorem 5.6 (Browder, Minty) *Suppose $F : X \to X$ is weakly continuous, monotone, and satisfies*

$$\lim_{|x|\to\infty} \frac{\langle F(x), x\rangle}{|x|} = \infty. \tag{5.35}$$

Then the equation

$$F(x) = y \tag{5.36}$$

has a solution for every $y \in X$. If F is strictly monotone then this solution is unique.

Proof. Abbreviate $y_n = F(x_n)$, then we have $\langle y - F(z), x_n - z\rangle = \langle y_n - F_n(z), x_n - z\rangle \geq 0$ for $z \in X_n$. Taking the limit implies $\langle y - F(z), x - z\rangle \geq 0$ for every $z \in X_\infty = \bigcup_{n=1}^\infty X_n$. Since X_∞ is dense, $\langle y - F(z), x - z\rangle \geq 0$ for every $z \in X$ by continuity and hence $F(x) = y$ by our lemma. □

Note that in the infinite dimensional case we need monotonicity even to show existence. Moreover, this result can be further generalized in two more ways. First of all, the Hilbert space X can be replaced by a reflexive Banach space if $F : X \to X^*$. The proof is almost identical. Secondly, it suffices if

$$t \mapsto \langle F(x + ty), z\rangle \tag{5.37}$$

is continuous for $t \in [0, 1]$ and all $x, y, z \in X$, since this condition together with monotonicity can be shown to imply weak continuity.

Bibliography

[1] M. Berger and M. Berger, *Perspectives in Nonlinearity*, Benjamin, New York, 1968.

[2] L. C. Evans, *Weak Convergence Methods for nonlinear Partial Differential Equations*, CBMS 74, American Mathematical Society, Providence, 1990.

[3] S.-N. Chow and J. K. Hale, *Methods of Bifurcation Theory*, Springer, New York, 1982.

[4] K. Deimling, *Nichtlineare Gleichungen und Abbildungsgrade*, Springer, Berlin, 1974.

[5] K. Deimling, *Nonlinear Functional Analysis*, Springer, Berlin, 1985.

[6] J. Franklin, *Methods of Mathematical Economics*, Springer, New York 1980.

[7] O.A. Ladyzhenskaya, *The Boundary Values Problems of Mathematical Physics*, Springer, New York, 1985.

[8] N. Lloyd, *Degree Theory*, Cambridge University Press, London, 1978.

[9] J.J. Rotman, *Introduction to Algebraic Topology*, Springer, New York, 1988.

[10] M. Růžička, *Nichtlineare Funktionalanalysis*, Springer, Berlin, 2004.

[11] E. Zeidler, *Applied Functional Analysis: Applications to Mathematical Physics*, Springer, New York 1995.

[12] E. Zeidler, *Applied Functional Analysis: Main Principles and Their Applications*, Springer, New York 1995.

Glossary of notations

$B_\rho(x)$	...ball of radius ρ around x
$\mathrm{conv}(.)$	...convex hull
$C(U,Y)$	...set of continuous functions from U to Y, 1
$C^r(U,Y)$	...set of r times continuously differentiable functions, 2
$C_0^r(U,Y)$	...functions in C^r with compact support, 45
$\mathcal{C}(U,Y)$	...set of compact functions from U to Y, 34
$\mathrm{CP}(f)$	...critical points of f, 13
$\mathrm{CS}(K)$	...nonempty convex subsets of K, 26
$\mathrm{CV}(f)$	...critical values of f, 13
$\deg(D,f,y)$	...mapping degree, 13, 22
$\det$	...determinant
$\dim$	...dimension of a linear space
div	...divergence
$\mathrm{dist}(U,V)$	$= \inf_{(x,y)\in U\times V} \lvert x-y\rvert$ distance of two sets
$D_y^r(U,Y)$	...functions in $C^r(\overline{U},Y)$ which do not attain y on the boundary.
dF	...derivative of F, 1
$\mathcal{F}(X,Y)$	...set of compact finite dimensional functions, 34
$\mathrm{GL}(n)$	...general linear group in n dimensions
$\mathcal{H}(\mathbb{C})$	...set of holomorphic functions, 11
$H^1(U,\mathbb{R}^n)$	...Sobolev space, 45
$H_0^1(U,\mathbb{R}^n)$	...Sobolev space, 45
$\inf$	...infimum
$J_f(x)$	$= \det f'(x)$ Jacobi determinant of f at x, 13
$\mathcal{L}(X,Y)$	...set of bounded linear functions, 1
$L^p(U,\mathbb{R}^n)$	...Lebesgue space of p integrable functions, 44
$\max$	...maximum
$n(\gamma,z_0)$	...winding number
$O(.)$	...Landau symbol, $f=O(g)$ iff $\limsup_{x\to x_0} \lvert f(x)/g(x)\rvert < \infty$
$o(.)$	...Landau symbol, $f=o(g)$ iff $\lim_{x\to x_0} \lvert f(x)/g(x)\rvert = 0$

∂U	. . . boundary of the set U
$\partial_x F(x,y)$	. . . partial derivative with respect to x, 1
$\mathrm{RV}(f)$	. . . regular values of f, 13
$R(I,X)$	. . . set of regulated functions, 4
$S(I,X)$	. . . set of simple functions, 4
sgn	. . . sign of a number
sup	. . . supremum
supp	. . . support of a functions

Index

Arzelà-Ascoli theorem, 40

Best reply, 27
Brouwer fixed-point theorem, 24

Chain rule, 2
Characteristic function, 4
Compact operator, 34
Contraction principle, 5
Critical values, 13

Derivative, 1
  partial, 1
Diffeomorphism, 2
Differentiable, 1
Differential equations, 8
Distribution, 46

Elliptic equation, 56
Embedding, 48
Equilibrium
  Nash, 28

Finite dimensional operator, 34
Fixed-point theorem
  Altman, 38
  Brouwer, 24
  contraction principle, 5
  Kakutani, 26
  Krasnosel'skii, 38
  Rothe, 38
  Schauder, 37
Functional, linear, 5

Galerkin approximation, 58
Gronwall's inequality, 41

Holomorphic function, 11
Homotopy, 12
Homotopy invariance, 13

Implicit function theorem, 7
Integral, 4
Integration by parts, 45
Inverse function theorem, 8

Jordan curve theorem, 31

Kakutani's fixed-point theorem, 26

Ladyzhenskaya inequality, 48
Landau symbols, 1
Lax–Milgram theorem, 55
Leray–Schauder principle, 37

Mean value theorem, 2
monotone, 54
  operator, 53
  strictly, 54
  strongly, 54
Multilinear function, 3

Nash equilibrium, 28
Nash theorem, 28
Navier–Stokes equation, 44

  stationary, 44
n-person game, 27

Payoff, 27
Peano theorem, 40
Poincaré inequality, 47
Poincaré-Friedrichs inequality, 46
Prisoners dilemma, 28
Proper, 35

Reduction property, 29
Regular values, 13
Regulated function, 4
Rellich's compactness theorem, 48
Rouchés theorem, 12

Sard's theorem, 17
Simple function, 4
Stokes theorem, 19
Strategy, 27
Symmetric multilinear function, 3

Uniform contraction principle, 6

Weak solution, 50, 56
Winding number, 11