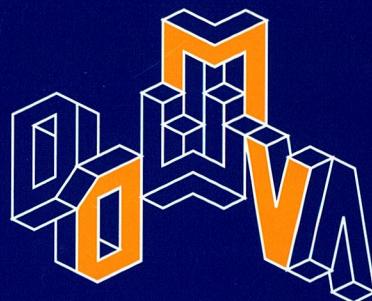


107. Band Heft 2, Juni 2005

D 20577

24/6/05



# Jahresbericht

## der Deutschen Mathematiker-Vereinigung

2 – 2005

Herausgegeben von K. Hulek  
unter Mitwirkung von  
U. Gather, H.-Ch. Grunau, H. Lange,  
J. Rambau, A. Schied, Th. Sonar



# Jahresbericht

der Deutschen Mathematiker-Vereinigung

Der „Jahresbericht“ ist das offizielle Veröffentlichungsorgan der Deutschen Mathematiker-Vereinigung, für dessen inhaltliche Gestaltung im Auftrag des Präsidiums der jeweilige Herausgeber zuständig ist. Im „Jahresbericht“ sollen vornehmlich Überblicksartikel über Teilgebiete der reinen und angewandten Mathematik, Nachrufe sowie historische Artikel, Berichte aus der Forschung und Buchbesprechungen veröffentlicht werden.

## Manuskripte:

Alle für die Schriftleitung des Jahresberichts bestimmten Briefe und Manuskripte sind an *Prof. Dr. K. Hulek* zu richten. Für Buchbesprechungen ist *Prof. Dr. H. Lange* zuständig. Bücher, von denen eine Besprechung erfolgen soll, werden bei den Verlagen angefordert.

Die Autoren werden gebeten, für Manuskripte und Buchbesprechungen die **Standard-LATEX-Klasse article mit 10pt (default), \textwidth139mm, \textheight205mm** zu benutzen. Sollen Illustrationen in die Arbeiten integriert werden, können diese auch in das Satzsystem übernommen werden. Dazu ist es erforderlich, dass die Bilddaten der Abbildungen nochmals in separaten Dateien einzeln abgespeichert werden. Ein Foto des Autors sollte als Bilddatei in einem der gängigen Grafikformate (am unproblematischsten: TIFF-Format; Graustufenbild mit einer Auflösung von mindestens 300 dpi) oder als normaler Papier-Fotoabzug zum Einscanen mitgeschickt werden. Als Datenträger sind ZIP-Disketten, CD-ROM bzw. Syquest (88 oder 200 MB) möglich.

Grundsätzlich sollen nur solche Manuskripte eingereicht werden, die nicht gleichzeitig an anderer Stelle zur Veröffentlichung eingereicht oder bereits veröffentlicht worden sind. Mit der Annahme zur Veröffentlichung erwirbt der Verlag das Verlagsrecht zur Vervielfältigung und Verbreitung sowie das Recht der Übersetzung in andere Sprachen.

Weitere Informationen zum „Jahresbericht“ finden Sie unter  
<http://www.mathematik.uni-bielefeld.de/DMV/jb/index.html>

## Verlag:

B. G. Teubner Verlag/GWV Fachverlage GmbH  
Abraham-Lincoln-Straße 46  
65189 Wiesbaden  
<http://www.teubner.de>  
<http://www.gwv-fachverlage.de>

*Geschäftsführer:* Andreas Kösters  
*Verlagsleitung:* Dr. Heinz Weinheimer  
*Gesamtleitung Anzeigen:* Thomas Werner  
*Gesamtleitung Produktion:* Reinhard van den Hövel  
*Gesamtleitung Vertrieb:* Gabriel Göttlinger

## Marketing/Sonderdrucke:

Eva Brechtel-Wahl  
Telefon: (06 11) 78 78-3 79  
Fax: (06 11) 78 78-4 39  
E-Mail: [eva.brechitel-wahl@gvv-fachverlage.de](mailto:eva.brechitel-wahl@gvv-fachverlage.de)

## Abonentenverwaltung:

(Änderungen von Adressen und Bankverbindung, Rückfragen zu Rechnungen oder Mahnung)  
VVA-Zeitschriftenservice, Abt. D6F6 / Jahresbericht der Deutschen Mathematiker-Vereinigung,  
Postfach 7777, 33310 Gütersloh  
Ursula Müller

Telefon: (0 52 41) 80-19 65  
Fax: (0 52 41) 80-96 20  
E-Mail: [ursula.mueller@bertelsmann.de](mailto:ursula.mueller@bertelsmann.de)

## Bezugsbedingungen:

Die Zeitschrift erscheint 4mal jährlich zum Jahresabonnementpreis von € 107,- (172,90 sF o. MwSt.) inkl. Versandkosten. Der Bezug von Einzelheften ist nicht möglich. Schriftliche Kündigung des Abonnements spätestens sechs Wochen vor Ablauf des Bezugsjahrs.

Für persönliche Mitglieder der DMV, die den Jahresbericht zu beziehen wünschen, ist der zwischen DMV und Verlag vereinbarte Bezugspreis maßgebend, der im Rahmen des Mitgliedsbeitrags erhoben wird.

## Copyright ©

B. G. Teubner Verlag/GWV Fachverlage GmbH, Wiesbaden 2005. Printed in Germany. Der Verlag B. G. Teubner ist ein Unternehmen von Springer Science+Business Media. Alle Rechte vorbehalten. Kein Teil dieser Zeitschrift darf ohne schriftliche Genehmigung des Verlages vervielfältigt oder verbreitet werden. Unter dieses Verbot fällt insbesondere die gewerbliche Vervielfältigung per Kopie, die Aufnahme in elektronischen Datenbanken und die Vervielfältigung auf CD-ROM und allen anderen elektronischen Datenträgern.

<b>Vorwort</b> . . . . .	55
--------------------------	----

Übersichtsartikel	Historische Beiträge	Berichte aus der Forschung	Buchbesprechungen
-------------------	----------------------	----------------------------	-------------------

**Don't shed tears over breaks**

G. Winkler et al. . . . .	57
---------------------------	----

**Der Multiplikationssatz der Mengenlehre**

O. Deiser . . . . .	88
---------------------	----

Übersichtsartikel	Historische Beiträge	Berichte aus der Forschung	Buchbesprechungen
-------------------	----------------------	----------------------------	-------------------

**Elmar Thoma zum Gedächtnis 1926–2002**

E. Kaniuth und G. Schlichting . . . . .	110
---	-----

Übersichtsartikel	Historische Beiträge	Berichte aus der Forschung	Buchbesprechungen
-------------------	----------------------	----------------------------	-------------------

**G. Aubert, P. Kornprobst: Mathematical Problems in Image Processing,  
Partial Differential Equations and the Calculus of Variations**

P. Maaß . . . . .	11
-------------------	----

**J. L. Nazareth: Differentiable Optimization and Equation Solving,  
A Treatise on Algorithmic Science and the Karmarkar Revolution**

C. Kanzow . . . . .	13
---------------------	----

**I. Elshakoff and Y. Ren: Finite Element methods for Structures  
with Large Stochastic Variations**

P. Kloeden . . . . .	14
----------------------	----

## In den nächsten Heften erscheinende Arbeiten

**P. Roquette:** From FLT to Finite Groups. The remarkable career of Otto Grün

**M. Groetschel:** MATHEON: Introducing the DFG Research Center „Mathematics for key technologies“

**B. Eick:** Computational Group Theory

---

## Anschriften der Herausgeber

Prof. Dr. K. Hulek, Institut für Mathematik, Universität Hannover,  
Welfengarten 1, 30167 Hannover  
E-Mail: hulek@math.uni-hannover.de

Prof. Dr. Ursula Gather, Lehrstuhl für Mathematische Statistik und industrielle Anwendungen, Universität Dortmund, 44221 Dortmund  
E-Mail: gather@statistik.uni-dortmund.de

Prof. Dr. H.-Ch. Grunau, Institut für Analysis und Numerik, Otto-von-Guericke-Universität Magdeburg, Postfach 4120, 39016 Magdeburg  
E-Mail: hans-christoph.grunau@mathematik.uni-magdeburg.de

Prof. Dr. Herbert Lange, Mathematisches Institut, Friedrich-Alexander-Universität Erlangen-Nürnberg, Bismarckstraße 1a, 91054 Erlangen  
E-Mail: lange@mi.uni-erlangen.de

Prof. Dr. J. Rambau, Fakultät für Mathematik, Physik und Informatik,  
Universität Bayreuth, 95440 Bayreuth  
E-Mail: rambau@zib.de

Prof. Dr. A. Schied, Institut für Mathematik, Technische Universität Berlin,  
Straße des 17. Juni 136, 10623 Berlin  
E-Mail: schied@math.tu-berlin.de

Prof. Dr. Th. Sonar, Institut für Analysis, Technische Universität Braunschweig,  
Pockelsstraße 14, 38106 Braunschweig  
E-Mail: t.sonar@tu-bs.de

## Bezugshinweis

Früher erschienene Bände (ab Band 68) des „Jahresberichts der Deutschen Mathematiker-Vereinigung“ können durch den Buchhandel oder den Verlag bezogen werden.

Nachdruck der Bände 41 bis 67 liefert: Swets & Zeitlinger, Heereweg 347 b, POB 810,  
NL-2160 SZ Lisse/Holland

## Vorwort

Das vorliegende Heft enthält drei Arbeiten. Die behandelten Themen zeigen dabei die Spannweite, die der Jahresbericht abdecken möchte. Auf der anwendungsbezogenen Seite steht die Arbeit „Don't shed tears over breaks“ von G. Winkler, O. Wittich, V. Liebscher und A. Kempe. Hier wird die Frage behandelt: wie kann man große Datenmengen, etwa aus dem Bereich der „Life Sciences“, sinnvoll interpretieren? Ausgangspunkt sind zwei konkrete Beispiele, eines aus der Hirnforschung, das andere aus dem Bereich der DNA-Analyse. Die Autoren zeigen in ihrer Arbeit, wie mathematische Methoden eingesetzt werden können, um solche Datenmengen so aufzubereiten, dass eine möglichst zuverlässige Interpretation möglich wird. Solche Methoden sind essentiell für eine erfolgreiche Forschung in Biologie, Medizin und auf anderen Gebieten.

Der Beitrag von O. Deiser über den Multiplikationssatz der Mengenlehre behandelt auf der anderen Seite ein Thema, das die Grundlagen der Mathematik betrifft. Anhand der Geschichte dieses Satzes diskutiert der Autor wesentliche Entwicklungslinien der Mengenlehre in der Zeit von ca. 1875–1945.

E. Kaniuth und G. Schlichting würdigen in ihrem Nachruf auf E. Thoma Leben und Werk dieses Mathematikers.

Wie stets wird das Heft durch eine Reihe aktueller Buchbesprechungen abgerundet.

K. Hulek

# Eine lebensnahe und praktische Einführung in die Mathematik für Ingenieure



Klaus Dürrschnabel  
**Mathematik für Ingenieure**  
*Eine Einführung  
mit Anwendungs- und  
Alltagsbeispielen*  
2004. 627 S. Br. EUR 34,90  
ISBN 3-519-00456-9

## Inhalt

Zahlenbereiche - Funktionen - Lineare Gleichungssysteme - Vektorrechnung - Analytische Geometrie - Matrizen - Grenzwerte - Differentialrechnung - Integralrechnung - Potenzreihen - Fourier-Reihen und Fourier-Transformation - Extrema bei Funktionen mehrerer Veränderlicher - Allgemeine Kurven - Differentialgleichungen

## Das Buch

Anschaulich und praktisch werden die grundlegenden mathematischen Kenntnisse für Studierende der Ingenieurwissenschaften vermittelt. Viele der Beispiele wurden gezielt aus dem vertrauten alltäglichen Leben gewählt. Damit erschließen sich komplexe mathematische Sachverhalte überraschend einfach. Mit mehr als 400 Aufgaben und Lösungen überwindet der Leser die Anfangsschwierigkeiten ganz leicht.

Teubner Lehrbücher:  
einfach clever



Abraham-Lincoln-Str. 46  
65189 Wiesbaden  
Fax 0611.7878-420  
[www.teubner.de](http://www.teubner.de)



G. Winkler



O. Wittich



V. Liebscher



A. Kempe\*

## Don't Shed Tears over Breaks

### Abstract

- Keywords and Phrases: edge preserving smoothing, variational problems, epiconvergence, wavelet shrinkage, regularization, nonlinear filters, Potts model, Mumford-Shah functional, segmentation, time-series, fractionation curves, functional magnetic resonance imaging
- Mathematical Subject Classification: 93 E 14, 62 G 08, 68 T 45, 49 M 20, 90C 31

This essay deals with ‘discontinuous phenomena’ in time-series. It is an introduction to, and a brief survey of aspects concerning the concepts of segmentation into ‘smooth’ pieces on the one hand, and the complementary notion of the identification of jumps, on the other hand. We restrict ourselves to variational approaches, both in discrete, and in continuous time. They will define ‘filters’, with data as ‘inputs’ and minimizers of functionals as ‘outputs’.

The main example are complexity penalized sums of squared deviations from data. We will argue that it is an appropriate tool for the extraction of the simplest and most basic morphological features from data. This is an attempt to *interpret* data from a well-defined point of view. It is *in contrast to restoration of a true signal* – perhaps distorted and degraded by noise – which is not in the main focus of this paper. The discussion proceeds along two real-world data sets, one from brain mapping, and one from

---

\* Partially supported by DFG Graduate Programme ‘Applied Algorithmic Mathematics’ at the TU München and DFG grant SFB 386 at the LMU München

Eingegangen: 19.07.2004, in überarbeiteter Form am 23.12.2004

Institute of Biomathematics and Biometry, GSF – National Research Center for Environment and Health, Postfach 1129, D-85758 Oberschleißheim, Germany; {gwinkler,wittich,liebscher,kempe}@gsf.de, http://ibb.gsf.de

**DMV**

**JAHRESBERICHT  
DER DMV**

© B. G. Teubner 2005

functional genomics. These are typical examples, where little or no ground truth is available. In view of the indistinct nature of such data, the whole procedure should follow the principle of *parsimony* or sparsity.

In edge-preserving smoothing, various formally similar variational models appear in the literature. We aim at an integral view; to this end we embed them into a parametric family, and study continuous dependence on parameters and sample size.

## 1 Prologue

In this essay, we reflect on the extraction of morphological features from data in the natural and in the life sciences. Partially, it is a survey of variational methods and their use in statistics; partially, it should be understood as a programme how to proceed further in this direction, both in theory and practice. The final aim is to support the work of natural scientists by sound interpretations of data, as well as the falsification or modification of present, and the generation of new hypotheses.

In the classical framework, models should be built on knowledge about the mechanisms of, and information gained by, experiments. In the examples we have in mind, focus is on data for which there is little or no ground truth. Let us only think of living organisms or other complex biological systems. It is by far not evident whether it is reasonable to believe in any (tractable) model of the underlying mechanism. Therefore, we must go beyond the established theory of estimation and test theory. *What* we can do in such cases is to ask nature simple questions, trying to extract specific features from data. And then we can endeavour to formulate the answer in a precise, transparent, and well-structured way as a basis for the discussion with natural scientists.

Here, we will restrict ourselves to features of time-series like persistent, steady, or smooth behaviour on the one hand, and rapid changes, or breaks on the other hand. For sake of brevity, the latter will be called *jumps*. The term jump implies that between subsequent jumps, or inside a boundary, the signal behaves ‘smoothly’, i.e. that the relevant characteristics do not change, or vary only moderately. Jump detection and identification of smooth regions are nonlocal and complementary concepts.

In this text, we will give examples of time-series from brain mapping and gene expression, and we will argue that jumps are the relevant features. But if this is our belief, then it is natural to examine and develop filters or estimators which transform data into *representations* based on jumps and segmentation. This is closely related to our second aim. We will discuss and illustrate by way of the mentioned data and some probabilistic models what we mean by parsimonious modelling and statistics. We will argue that parsimony is an important concept for the explanation of such data.

### 1.1 Two Data Sets from Life Sciences

In many situations, we are confronted with experiments, where the (stochastic) mechanism generating data essentially is unknown. This is particularly the case in a conglomeration of classical and new scientific disciplines like medicine, ecology, and biology, or genomics, proteomics, metabolomics, cellomics etc., for which in the last years the fash-

ionable name ‘life sciences’ was created. Compared to physics, which certainly is the discipline with the highest degree of mathematization among the natural sciences, there are seldom hard mechanistic models explaining the chains and loops of causes and effects, and nearly all disciplines in the life sciences are – due to the complexity of their questions and problems – still far from stringent mathematical formalization. Nevertheless there is a desire for precise mathematical formulations and arguments. In such cases, the only way to associate data to some hidden reality is to verify or falsify rough and basic criteria which characterize the event in question. Such criteria frequently are based on primitive signal features. In images these may be boundaries between regions of different intensity or texture, in time-series they may be morphological features like modes or ‘ups and downs’, domains of monotony, or plateaus where the signal is constant. A list of past work on such and related topics can be found in Section 1.4 of P. L. Davies and A. Kovac (2001).

We start the discussion with two one-dimensional data sets, one from brain mapping and one from functional genomics. In these examples, we expect that the observation period can be partitioned into intervals where the underlying signal *can reasonably be represented* by a constant. This is a very simple morphological feature, but the resulting representations by step functions allow for sound biological interpretations.

The first set of data consists of time-series from functional magnetic resonance imaging (fMRI) of the human brain. Please note that we selected an almost ideal example for the illustrations in order to make our intentions clearly visible. The second type of data are melting or fractionation curves for spots on a cDNA microchip.

**Example 1.1 (fMRI Brain Data: Identification of Response Regions)** In this example, the final aim is to identify regions of increased activity in the human brain in response to outer stimuli. Frequently, such stimuli are boxcar shaped as indicated in Fig. 1. They may represent ‘light or sound on and off’, i.e. visual or acoustic stimuli, or tactile ones like finger tipping on a desk. *Functional magnetic resonance imaging* (fMRI) exploits the *blood oxygenation level dependent* (BOLD) effect which basically is a change of paramagnetic properties caused by an increase of blood flow in response to the demand of activated neurons for more oxygen. The degradation mechanism along the path ‘(very complex) eye – (highly complex) visual cortex – (complicated) measuring device’ is only partially known. Moreover, measurement is indirect, since the recorded BOLD effect is a physiological quantity related to a local increase of blood flow and not a direct function of cortical activation.

In the present example, fMRI records for each brain voxel the BOLD effect along 70 time points. The voxel size is about  $3 \times 3 \times 5 \text{ mm}^3$ . Typical time-series are displayed in Fig 1. Each single value of a time-series is interpreted as a measure for the mean activation of the about 2 or 3 Millions of neurons inside the voxel. The final aim is to decide on the basis of these time-series, whether a majority of neurons in the voxel answers to the external stimulus or not. We claim, that a minimal requirement for the classification of a voxel as a response to the boxcar stimulus is that it shows the same number of plateaus or ‘ups and downs’ as the stimulus, irrespective of time shift, trend, scale, ‘random noise’ etc. Hence a parsimonious approach based on significant plateaus should be appropriate.

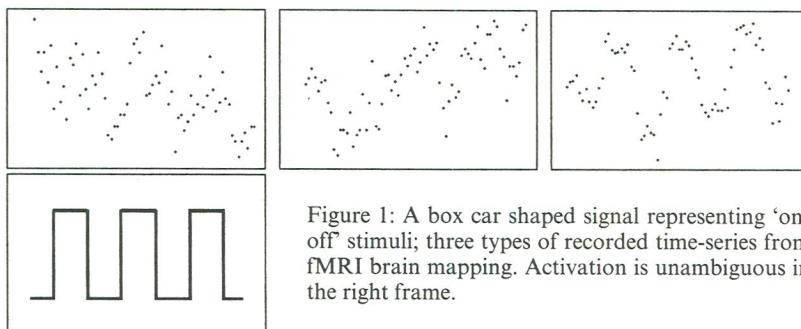


Figure 1: A box car shaped signal representing ‘on-off’ stimuli; three types of recorded time-series from fMRI brain mapping. Activation is unambiguous in the right frame.

**Example 1.2 (Gene Expression Experiments: Fractionation Curves)** The aim of this experiment is to judge the quality of cDNA chips in order to increase the reliability of conventional microarray experiments.

Single stranded sections of *known* cDNA are put on spots of a microchip; one chip typically consists of about 20.000 spots. Each cDNA section is characterized by a finite chain of nucleic acids, which are coded by the letters A(denine), C(ytosine), G(uanine), and T(hymine). If further – *unknown* – nucleic acids are added then they tend to bind to the target nucleic acids, where T binds to A, and G binds to C.

Hence sections of single stranded unknown cDNA tend to pair with known cDNA of similar sequences. The binding energy is maximal for perfect matches like

A	C	T	A	C	A	G	T	A	C	C	C	A
T	G	A	T	G	T	C	A	T	G	G	G	T

and such a perfect match has high stability. In this case, the unknown sequence could be identified perfectly. A main problem is *cross-hybridization*, which means that DNA sections also pair with DNA of nearly, but not precisely complementary sequence, for example

A	C	T	A	C	A	G	T	A	C	C	C	A
T	G	A	T	T	T	C	A	T	G	A	G	T

A new and innovative biochemical experiment, designed to improve the quality control, provides data which hopefully will allow to identify mismatch dissociation at low stringency. It is called ‘Specificity Assessment From Fractionation Experiments’ or in short-hand notation ‘SAFE’, see A. L. Drobyshev et al. (2003). Washing the chip repeatedly with *formamide* solutions of increasing concentration removes pieces of single stranded DNA with higher and higher binding energy. The chips are washed 29 times and *fractionation curves* like in Fig. 2 are recorded. Basically, they represent the fraction of binding DNA present on the spot at the different formamide concentrations. It is clear that ideal fractionation curves decrease. The aim of the statistical analysis is to identify locations and heights of abrupt decreases. Locations of jumps correspond to melting temperatures and indicate that a certain type of cross-hybridizing cDNA was washed away. There is still no satisfactory model which captures the relevant mechanisms.

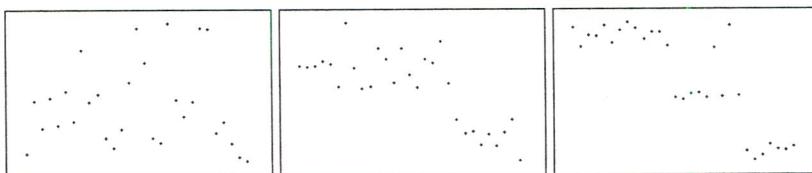


Figure 2: Typical fractionation curves of single spots: no specific binding, specific binding, cross-hybridization. Data are normalized to have the same range.

In view of such data, one may doubt about too ‘specific’ methods or too detailed models for their analysis, and in fact we do so. Estimates rely sensitively on the validity of the models, and therefore are not robust against even slight changes in the assumptions. Fitting too many parameters introduces more variance despite slight decrease in bias.

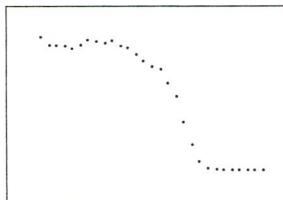


Figure 3: Median plot of fractionation curves (from the red channel): The pointwise median of 20 000 fractionation curves derived from a DNA-microchip which were robustly normalized by forcing the 3/29-quantile (approximately the median of the last 7 time points) to be 0 and the 24/29-quantile (approximately the median of the first 9 time points) to be 1.

A way out of this misery is to try a parsimonious approach. The principle of parsimony is a philosophical matter and will be addressed in the next section. The relevant features in Example 1.1 are successions of high and low plateaus and in Example 1.2, the positions of rapid decreases and their height. The form of typical signal representations in the second example is suggested by physical arguments, but also by statistical ones like the median plot in Fig. 3. It displays for each time-point the median over data in the about 20.000 spots on a microarray. It clearly indicates that typical features of typical signals are sharp decreases.

In summary, in both examples there is ample evidence for well-defined, and characteristic, morphological features. Therefore, we try to fit piecewise constant functions to these data. We will argue in Section 2.1 that minimizers of the Potts functional (4) with proper hyperparameters extract step functions with few jumps. This nourishes the hope that they will be appropriate for the explanation of data like those in the Examples 1.1 and 1.2.

Example 1.2 conceals another interesting aspect. Theory predicts that ideal fractionation curves are (decreasing) superpositions of sigmoids (with sharper turns on the lower right than on the upper left). Nevertheless, we want to see sharp breaks for the identification of cross-hybridization, insert jumps where the underlying signal is steep. This will be discussed in Section 1.3 below.

## 1.2 Why Parsimony?

The principle of parsimony is often called Ockham's Razor, named after the medieval philosopher Wilhelm von Ockham, who was born about 1285 at Ockham, Surrey, and who died about 1349 at München, Bavaria. He advanced the theory with phrases like "What can be explained by the assumption of fewer things is vainly explained by the assumption of more things." Another formulation reads "*Entia non sunt multiplicanda sine necessitate.*" (The being should not be multiplied without necessity). It remains unclear from these passages just what these 'things' *really are*. This becomes also evident in the words of B. Russel (1953): "What science does, in fact, is to select the simplest formula that will fit the facts. But this, quite obviously, is merely a methodological precept, not a law of nature. If the simplest formula ceases, after a time, to be applicable, the simplest formula that remains applicable is selected." There remains the question what 'simple' means. We quote from P. L. Davies (1995): "... it will often be possible to find a model which almost reproduces the data. The simplest example is that of linear regression with a large number of explanatory variables. In such a case the interesting question is therefore not what is the 'correct' number of explanatory variables but what is the smallest number required to give an adequate model. In this case parsimony is measured by the number of coefficients but it is also possible to use other measures such as continuity or the number of local extreme values."

Adapting the paradigm of parsimony *to the data we have in mind*, the rules of the game are as follows: Explain data by a minimal number of smooth pieces. The one who needs least, is the winner. In other cases, one may aim at the minimal number of modes, monotonous pieces etc. Explaining data includes a proper balance between complexity and fidelity to data.

## 1.3 Jumps and Segmentations

For the present, let a signal be a real function on the unit interval  $U = [0, 1]$ , or on a discrete subset of  $U$ . We are interested in rapid changes of this signal, or in *jumps*. This is a rather intricate matter, and we will shortly argue that naive concepts can be misleading.

For a finite stepfunction like in Fig. 4 (a), the discontinuities are natural candidates, and it certainly does not make sense to locate jumps inside the plateaus. Therefore it is reasonable to require the necessary condition that *a step function is discontinuous at each jump location*. How to formulate a sufficient condition is – even for step functions – less obvious. The naive definition as analytical discontinuities in general is problematic. In fact, the tiny jump in Fig. 4 (a) has a quite different quality as the large one. A continuous function does not have any analytical discontinuities at all.

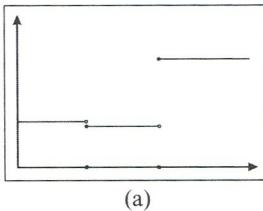
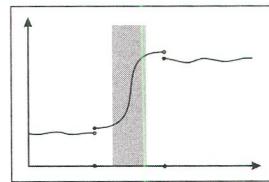
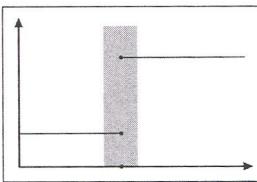
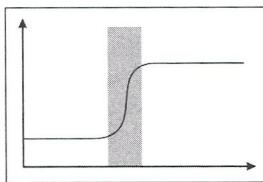


Figure 4: Analytical discontinuities versus an intuitive notion of jump.



Nevertheless, one may want to interpret the steep ascent of the sigmoid function  $f$  from Fig. 4 (b) as a jump. This is reasonable since one cannot tell  $f$  from a step function like that in (c) from sparse samples. A natural choice for time-continuous signals are càdlàg functions which in each point are right-continuous and have limits from the left. These can also have analytical discontinuities which are very small compared to the variation of the signal in continuous parts like the shadowed one in Fig. 4 (d). In such situations the identification of jumps as analytical discontinuities can be severely misleading and not desirable. This is underlined by the data from Example 1.2, in particular Fig. 3.

In summary, a notion of jump should be non-local: a time-point should be identified as a jump if the variation of the signal in a neighbourhood exceeds the variation in adjacent segments remarkably. This implies that a notion of smoothness is required if one wants to define what a jump is. In order to capture the antagonism of jumps and smoothness, we will write signals in terms of *segmentations*.

Let us start with continuous time. The first ingredient are finite subsets  $\{t_0 = 0 < t_1 < \dots < t_r < t_{r+1} = 1\} \subset U$ . They are the edges of the closed subintervals  $I_j = [t_{j-1}, t_j] \subset U$  between the  $t_j$  and of their interiors  $I_j^\circ$ . By slight abuse of notation, we call the set  $\mathcal{P} = \{I_j : 1 \leq j \leq r+1\}$  a *partition* of  $U$ . The second ingredient are tuples  $f_{\mathcal{P}} = (f_1, \dots, f_{r+1})$  of functions  $f_j$  on the  $I_j$  or on the interiors  $I_j^\circ$ . The single pieces  $f_j$  will be required to be in some ‘smoothness class’, for example in a space of polynomials or in a Sobolev space  $\mathcal{W}^{1,2}(I_j^\circ)$ . Each *segmentation*  $(\mathcal{P}, f_{\mathcal{P}})$  induces a function  $f$  on  $U$ , which coincides on  $I_j^\circ$  with  $f_j$ . The set of segmentations will be denoted by  $\mathfrak{S}$ . In general,  $f$  is not defined on the  $t_j$ ; this will not bother us, since we will work in the  $L^p$ -context anyway. Let us call functions admitting a representation by a segmentation *admissible*.

The same idea works in the discrete case. We simply partition the time-set  $\{1, 2, \dots, n\}$  or  $\{1/n, 2/n, \dots, 1\}$  into mutually disjoint discrete intervals  $I_j$  and specify the smoothness classes. A typical example are the subspaces of  $\mathbb{R}^{I_j}$  with  $u_k = u_{k+1}$  whenever the neighbours  $k, k+1$  both are in  $I_j$ . Other examples of smoothness classes are given by  $|u_{k+1} - u_k| \leq \delta$  for some  $\delta > 0$  whenever the neighbours  $k, k+1$  both are in  $I_j$ . We will pursue this aspect in the next sections.

## 2 Variational Approaches

Our concern is the decomposition of signals into smooth pieces, and simultaneously, the identification of boundaries between the regions of smoothness. Moreover, data have to be explained in a parsimonious way; in the present context this means by a minimal number of jumps.

We are now going to report on variational approaches to this problem. The general paradigm is to design a real-valued functional of signals and data, which captures and rates the relevant signal features and, simultaneously, measures the fit of signals to data. Usually, such functionals are a sum of a penalty or regularization term  $G(f)$  for the signal  $f$ , and a term  $D(g,f)$  measuring fidelity of the signal  $f$  to the recorded data  $g$ . The generic functional has the form

$$(1) \quad F(f) = G(f) + D(f,g).$$

Then – given data  $g$  – one selects a minimizer  $f^*$  of this functional, and decides that  $f^*$  is the desired ‘filter output’, or representation.

The functionals in this text will have a more special structure. Recall from Section 1.3 that each admissible signal has a representation as a (minimal) segmentation  $(\mathcal{P}, f_{\mathcal{P}})$ . We may and will identify admissible functions with their segmentations and write  $f_{\mathcal{P}}$  for both. All the functionals will have a penalty for the size of the partition  $\mathcal{P}$ , penalties for the roughness of each  $f_I$ , and penalties for their deviation from the data  $g|I$  on  $I$ . In summary, we will consider functionals of the form

$$(2) \quad F^{\Xi}(\mathcal{P}, f_{\mathcal{P}}) = \gamma(|\mathcal{P}| - 1) + \frac{1}{\mu^2} \sum_{I \in \mathcal{P}} \Psi_I(f_I) + \sum_{I \in \mathcal{P}} \varrho_I(g|I, f_I),$$

with control parameters  $\gamma > 0$  and  $\mu \geq 0$ . If we agree that the jumps are located at the endpoints of the intervals  $I$  of smoothness, then the first penalty is simply the weighted number of jumps. This formulation allows a significant simplification of the optimization problem, which reads

$$(3) \quad \min_{(\mathcal{P}, f_{\mathcal{P}})} F^{\Xi}(\mathcal{P}, f_{\mathcal{P}}) = \min_{\mathcal{P}} \left( \gamma(|\mathcal{P}| - 1) + \frac{1}{\mu^2} \sum_{I \in \mathcal{P}} \min_{f_I} (\Psi_I(f_I) + \varrho_I(g|I, f_I)) \right).$$

Recall that  $F^{\Xi}$  is defined for admissible segmentations only, whereas pairs  $(\mathcal{P}, (f_I)_{I \in \mathcal{P}})$  on the right-hand side need not be admissible. We shall therefore require  $\Psi_{I \cup J} \leq \Psi_I + \Psi_J$  and  $\varrho_{I \cup J}(g|I \cup J, f_{I \cup J}) \leq \varrho_I(g|I, f_I) + \varrho_J(g|J, f_J)$  for all adjacent  $I, J \in \mathcal{P}$  which are both not maximal. Then the minimum cannot be attained for non-admissible segmentations, since the fusion of non-maximal intervals decreases  $|\mathcal{P}|$  and does not change the other terms.

Frequently, solutions of the inner optimization problem can be determined analytically, or computed with reasonable effort. Then minimization boils down to that on the space of partitions which is considerably smaller than the original search-space. In the discrete-time setting the problem is even reduced to the *finite set of finite partitions*, which is easy to handle by dynamic programming.

Two further remarks are in order here: (1) Such functionals formally are similar to posterior or penalized log-likelihood functions from statistics, and there is a close con-

nection to maximum posterior estimation in Bayesian signal and image analysis. (2) The set of jumps may be interpreted as a ‘boundary’ between smooth regions. In dimension one, such a boundary consists of a number of isolated points. In higher dimension, boundaries are discrete or continuous ‘manifolds’ with an own structure. One example is regularity in real-world pictures at large scale, called ‘continuity of discontinuities’ in D. Marr (1982).

## 2.1 Time-Discrete Functionals

Let us start with time-discrete models. Suppose that data  $v_1, \dots, v_n$  with a sample size  $n \in \mathbb{N}$  and values  $v_k \in \mathbb{R}$  is recorded. The candidates for a representation then are time-series  $u_1, \dots, u_n \in \mathbb{R}$ . Let us mention in advance, that in the examples *all*  $u \in \mathbb{R}^n$  are admissible.

The simplest instance of an edge detecting functional counts each time-point  $k$  with  $u_k \neq u_{k+1}$ ,  $k = 1, \dots, n - 1$ , as a *jump location*. Let  $J(u)$  denote the set of these jump locations, and  $|J(u)|$  their number. The distance of a  $u$  to data  $v$  may be measured by various distances, for example by the sum of squared deviations. In this case the functional has the form

$$(4) \quad P_{\gamma, v} : \mathbb{R}^n \rightarrow \mathbb{R}, \quad u = (u_1, \dots, u_n) \mapsto \gamma|J(u)| + \|u - v\|^2$$

with Euclidean norm  $\|\cdot\|$  on  $\mathbb{R}^n$ . The first term penalizes the number of jumps of the representation  $u$ , and the second term rates fidelity of the representation  $u$  to data  $v$ . We call (4) the one-dimensional *time-discrete real-valued Potts functional*. The original Potts model is a Gibbs field introduced in R.B. Potts (1952) as a generalization of the well-known Ising model from E. Ising (1925), for a binary spin system, to more than two states. The three concepts discussed above are incorporated in this model:

- (i) Jumps: Changes in the signal. In the Potts functional there are jumps where the values in two subsequent time-points differ.
- (ii) Smoothness: The behaviour between two subsequent jumps. It is a consequence of (i) that in the Potts functional a signal is constant there.
- (iii) Fidelity to data: A distance between data and signal.

The item (ii) is a fairly strict notion of smoothness: in fact, the signal is ‘smooth’ only in regions where it is constant. Note that this is parsimony at its best, since the number of values in a smooth part is one and hence minimal. The first term in (4) penalizes the number of jumps irrespective of their size, and the parameter  $\gamma > 0$  controls the degree of smoothness.

For the general considerations, the explicit form of the data term addressed in (iii) is irrelevant. The sum of squares could be replaced by other distances, for instance by the sum of absolute deviations. Given the set of jumps, this term determines the constant value of the signal between subsequent jumps. For the sum of squares, it is the empirical mean of data values inbetween; for the sum of absolute deviations, it is an empirical median. For the mathematical and the numerical analysis, on the other hand, the special form of the data term plays an important role.

**Remark 2.1** Each minimizer  $u_\gamma^*$  of the Potts functional (4) is the representation with the least number of jumps among all representations not farther away from data  $v$  than  $u_\gamma^*$ :

$$u_\gamma^* \in \operatorname{argmin}_{u \in \mathbb{R}^n} |J(u)|, \text{ subject to } \|u - v\|^2 \leq \|u_\gamma^* - v\|^2.$$

In view of the tradeoff between penalty and data term this is obvious. This constraint optimization problem depends on  $\gamma$  only through the map  $\gamma \mapsto u_\gamma^*$ . Note that this is another indication for parsimony of the Potts functional.

Let us illustrate the concept of segmentations by way of this simple example. Let us start from a signal  $u$ . It induces a partition of  $\{1, \dots, n\}$  into intervals  $I$  with  $u_k = \mu_I$  for some  $\mu_I \in \mathbb{R}$  and each  $k \in I$ . The correspondence is one-to-one and onto if the intervals are maximal. If we take for each  $I$  the diagonal in  $\mathbb{R}^I$ , consisting of the constant vectors  $c \cdot (1, \dots, 1)$ , as the smoothness class then the correspondence between signals and segmentations is one-to-one and onto as well. In terms of such segmentations, and with  $v_I = (v_k)_{k \in I}$ , the functional (4) reads

$$P_{\gamma, v}^\Xi : \mathfrak{S} \longrightarrow \mathbb{R}, (\mathcal{P}, (\mu_I)_{I \in \mathcal{P}}) \longmapsto \gamma(|\mathcal{P}| - 1) + \sum_{I \in \mathcal{P}} \|v_I - \mu_I\|_I^2,$$

where  $\|\cdot\|_I$  is Euclidean norm on  $\mathbb{R}^I$ . This is (2) with  $\psi_I \equiv 0$ .

We consider the Potts functional as the simplest variational approach to the theoretical study of edge preserving smoothing, and simultaneously as a parsimonious statistical model for signals, in cases where there is little ‘ground truth’.

Let us give a second – closely related – example. It is a reformulation of the elasticity model, discussed in detail in the article A. Blake (1983) and in the monograph A. Blake and A. Zisserman (1987); for a comprehensive view see also G. Winkler et al. (1999), G. Winkler and V. Liebscher (2002), and the monograph G. Winkler (2003). Like the Potts functional, this *Blake-Zisserman functional* can be written in several forms. The one closest to (4) reads

$$(5) \quad B_{\gamma, \mu, v} : \mathbb{R}^n \longrightarrow \mathbb{R}, u = (u_1, \dots, u_n) \longmapsto \sum_{k=1}^{n-1} \varphi(u_{k+1} - u_k) + \|u - v\|^2$$

with the truncated square function

$$(6) \quad \varphi(\Delta) = \varphi_{\mu, \gamma}(\Delta) = \min\{\Delta^2/\mu^2, \gamma\}, \gamma > 0.$$

Here a difference  $\Delta = |u_{k+1} - u_k|$  is considered to be a jump if it is greater than  $\delta = \gamma^{1/2}\mu$  and then  $\varphi(\Delta) = \gamma$ . It treats values  $\Delta$  greater than  $\delta$  like the Potts functional, again irrespective of their size. At low values, this functional prefers signals  $u$  which are smooth between two subsequent jump locations in the  $\ell^2$ -sense  $\varphi(\Delta) = \Delta^2/\mu^2$ .

Formally, we get the Potts functional, if we let  $\varphi(\Delta) = 0$  if  $\Delta = 0$  and  $\varphi(\Delta) = \gamma$  if  $\Delta \neq 0$  in (5). This shows that Blake-Zisserman functionals converge pointwise to the Potts functionals as  $\delta \rightarrow 0$ . They also converge in the epi- or  $\Gamma$ -topology which even implies convergence of the respective minimizers. This aspect will be addressed in Section 5.2.

All this becomes more transparent in the language of segmentations. The discussion is similar to that for the Potts functional. The smoothness class for each interval  $I$  is

$$\{(u_k)_{k \in I} \in \mathbb{R}^I : |u_{k+1} - u_k| \leq \delta, \text{ whenever } k, k+1 \in I\}.$$

If we denote the generic element by  $u_I$  and the restriction of  $v$  to  $I$  by  $v_I$ , the functional becomes

$$(7) \quad B_{\gamma, \mu, v}^{\mathfrak{S}} : \mathfrak{S} \longrightarrow \mathbb{R},$$

$$(\mathcal{P}, (u_I)_{I \in \mathcal{P}}) \longmapsto \gamma(|\mathcal{P}| - 1) + \frac{1}{\mu^2} \sum_{I \in \mathcal{P}} \sum_{k, k+1 \in I} (u_{k+1} - u_k)^2 + \sum_{I \in \mathcal{P}} \|v_I - u_I\|_I^2.)$$

To put this into the general context of (2) we observe that the data term is the same as in the Potts model and  $\Psi_I(u_I) = \sum_{I \in \mathcal{P}} \sum_{k, k+1 \in I} (u_{k+1} - u_k)^2$ . Respective segmentations of the same signal are displayed in Fig. 5.

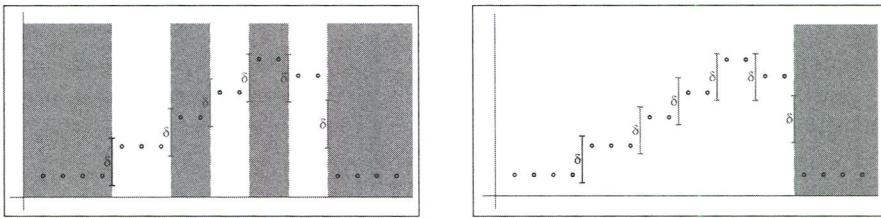


Figure 5: A Potts and a Blake-Zisserman segmentation

We can easily convince ourselves, that (5) and (7) are equivalent. Equivalence of the penalty terms is read off from the following identity. Let  $u$  be a signal and  $(\mathcal{P}, u_{\mathcal{P}})$  the associated minimal segmentation and note that  $(u_{k+1} - u_k)^2 / \mu^2 \leq \gamma$  if and only if  $|u_{k+1} - u_k| \leq \delta$ . Then

$$\begin{aligned} & \sum_{k=1}^{n-1} \min\{(u_{k+1} - u_k)^2 / \mu^2, \gamma\} \\ &= \gamma |\{k : |u_{k+1} - u_k| > \delta\}| + \frac{1}{\mu^2} \sum_{|u_{k+1} - u_k| \leq \delta} (u_{k+1} - u_k)^2. \end{aligned}$$

It is now clear that the right sum can be split into sums over maximal discrete intervals on which  $|u_{k+1} - u_k| \leq \delta$ . They define a minimal partition  $\mathcal{P}$  with  $|\mathcal{P}| - 1 = |\{k : |u_{k+1} - u_k| > \delta\}|$ . This completes the proof. We will later refer to arguments like this as the reduction principle.

The functional (7) can be interpreted as a discretisation of the Mumford-Shah functional (10) which lives on continuous spaces. It will be addressed in Section 2.2.

A. Blake and A. Zisserman (1987) started their discussion from the functional (8) below which is very close to (7). They work with the original signal  $u$  but introduce auxiliary or co-variables indicating the presence or absence of a jump. In one dimension these are binary variables  $b_k \in \{0, 1\}$ ,  $k = 1, \dots, n - 1$ ; the functional depends on the two variables  $u = (u_k)$  and  $b = (b_k)$  in the following way

$$(8) \quad ((u_1, \dots, u_n), (b_1, \dots, b_{n-1})) \longmapsto$$

$$G_{\gamma, \mu, v}(u, b) = \sum_{k=1}^{n-1} \left( (u_{k+1} - u_k)^2 (1 - b_k) / \mu^2 + \gamma b_k \right) + \|u - v\|^2.$$

The variables  $b_k$  are indicators of jumps; there is a jump between  $k$  and  $k + 1$  if and only if  $b_k = 1$ . The configurations  $b \in \{0, 1\}^{n-1}$  represent boundaries. In one dimension they correspond to our  $t_j$ ; in higher dimension they live on dual grids. It is clear from the above discussion that (8) is equivalent to (5). The explicit representation of boundaries was introduced in A. Blake and A. Zisserman (1987) since these authors considered (8) mainly as an edge detector. The form (5) for them was a tool to develop the *Graduated Non-Convexity algorithm* for minimization.

S. Geman and D. Geman (1984) started from similar functionals, but they augmented them by penalty terms for selected edge configurations; in one dimension these could be ‘double boundaries’ like  $b_{k-1} = 1, b_k = 0, b_{k+1} = 1$ . Most other penalized local edge configurations from S. Geman and D. Geman (1984) – like crosses or dead ends – do not appear in dimension one. These authors work in the Bayesian framework and let the variables  $u_k$  take values in finite discrete spaces. They also experimented with various smooth cup-shaped functions instead of the square penalty. These are conceptually much harder to grasp.

There is a close relation between edge preservation and robustness. In fact, the penalty in (5) can be interpreted as a robustification of an  $\ell^2$ -penalty – where  $\Delta^2$  is used instead of  $\varphi(\Delta)$  – in the sense of F.R. Hampel et al. (1986). A closely related robustification is the popular  $\ell^1$ -penalty with  $|\Delta|$ . The *taut-string* algorithm allows fast computations for the corresponding functionals, see P.L. Davies and A. Kovac (2001). A theoretical study is E. Mammen and S. van de Geer (1997). A mixture of  $\ell^1$ - and  $\ell^2$ -penalties results from functions  $\psi$  of the (convex) Huber type

$$(9) \quad \psi(\Delta) = \begin{cases} \Delta^2 / \mu^2 & \text{if } |\Delta| < \delta \\ 2\sqrt{\gamma}|\Delta|/\mu - \gamma & \text{if } |\Delta| \geq \delta \end{cases}, \quad \delta = \sqrt{\gamma}\mu,$$

see P.J. Huber (1981). The above discussion shows that only a strict truncation like in  $\varphi$  allows the unambiguous identification of jump locations.

The performances of some of the methods are illustrated and compared in Fig. 6.

## 2.2 Continuous Time and the Mumford-Shah Functional

Let us now turn to continuous time. D. Mumford and J. Shah (1985, 1989) introduced a time-continuous functional in  $\mathbb{R}^2$ , similar in spirit to the time-discrete functionals above.

In the time-continuous setting, ‘data’ are functions  $g$  on the unit interval  $U = [0, 1]$ . We continue with notation from Section 1.3. If  $(\mathcal{P}, f_{\mathcal{P}})$  is a segmentation, we denote the components of  $f_{\mathcal{P}}$  by  $f_I$ . The smoothness classes are Sobolev spaces  $\mathcal{W}^{1,2}(I^\circ)$ , and hence each  $f_I$  has a square integrable generalized derivative. Then, apart from technical details, one version of the Mumford-Shah functional, for parameters  $\gamma, \mu > 0$  and a signal  $g \in L^2(U)$ , may be written in the form

$$(10) \quad E_{\gamma, \mu}^{\mathfrak{S}} : \mathfrak{S} \longrightarrow \mathbb{R},$$

$$(\mathcal{P}, f_{\mathcal{P}}) \longmapsto \gamma(|\mathcal{P}| - 1) + \frac{1}{\mu^2} \sum_{I \in \mathcal{P}} \int_I |f'_I(t)|^2 dt + \sum_{I \in \mathcal{P}} \int_I (f_I(t) - g_I(t))^2 dt,$$

where  $f'_I$  denotes the Sobolev derivative of  $f_I$  on  $I^\circ$ .

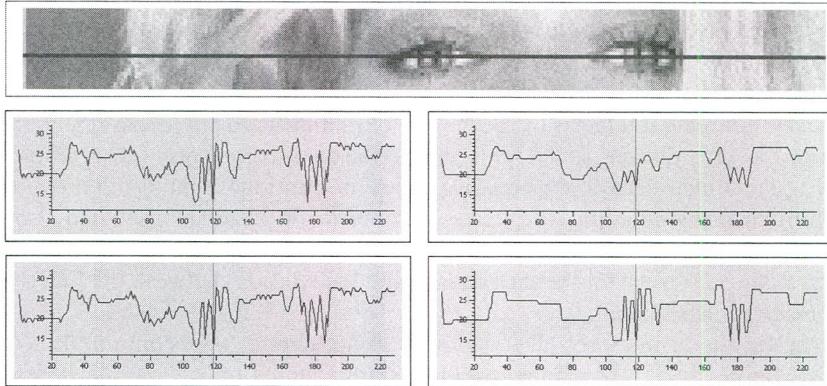


Figure 6: Intensity profiles along the horizontal line through the middle of the pupils in a portrait: original image (upper left); typical minimizers of three functionals: sum of squares penalty (upper right), Blake-Zisserman functional (lower left), and Potts functional (lower right). Jumps are indicated by the almost vertical lines. Compare the characteristic pupil shapes of the profiles, from 108 to 119 for the right eye and 175 to 185 for the left eye.

In analogy to (4) and (5), it will be convenient to define this and other functionals on signals rather than on segmentations. Albeit the natural ranges of definition of the various functionals will be different, they all can be embedded into the space  $L^2(U)$  of square integrable functions. It contains for example all signals which are admissible for the above segmentations. We will set the functionals to  $\infty$  on the nonadmissible functions. We will argue in Section 4.1 that the subspace  $SBV_2(U)$  of  $L^2(U)$  is the natural range of definition of the Mumford-Shah functional. It consists of functions which are the sum of a stepfunction and of an absolutely continuous function with square-integrable derivative. We set

$$(11) \quad E_{\gamma, \mu} : L^2(U) \longrightarrow \mathbb{R} \cup \{\infty\},$$

$$f \longmapsto \begin{cases} \gamma|J(f)| + \frac{1}{\mu^2} \int_U |f'(t)|^2 + \int_U (f(t) - g(t))^2 dt & \text{if } f \in SBV_2(U) \\ \infty & \text{otherwise} \end{cases}.$$

A more detailed discussion of the Mumford-Shah functional and the space  $SBV_2(U)$  will follow in Section 4. It will also reveal, that each function which minimizes (10) is a representative of an equivalence class in  $L^2(U)$  minimizing (11) and vice versa.

## 2.3 All Together Now

In the last two sections, we discussed typical functionals for jump detection and jump preserving smoothing of time-series, both in discrete and in continuous time. All these models depend on hyperparameters – which we denoted by  $\gamma$  and  $\mu$ . By means of them we can control the number of jumps and the degree of smoothness. Time-discrete functionals depend also on the sample size  $n$ . We aim at an integral view of all these models, in order to understand their mutual interrelations. The first step towards this aim simply is to draw up a stock list of the functionals; next we will embed them into a common setting; and finally, we will study the space of all these functionals from a topological point of view. Of main interest is the behaviour of minimizers as hyperparameters vary, in particular as time-discrete functionals tend to the continuous ones. For example, the Mumford-Shah functionals, and its relatives, are time-continuous limits of time-discrete models for increasing sampling rate, in a sense to be made precise. This would also be a prerequisite for the numerical optimization of the latter. In this section, we just give an overview. Rigorous results on the asymptotics and the relations between the functionals will be presented in Section 5.

Our strategy is to interpret all functionals in a common time-continuous setting. Since the Mumford-Shah functional (11) is already defined on  $L^2(U)$ , this space is a natural candidate. We will first embed time-discrete signals into  $L^2(U)$  as finite step-functions, and this way carry the time-discrete functionals over to functionals on  $L^2(U)$ . To fix the problem of notation once and for all let us agree about the following convention:

*Let  $\mathcal{F}$  denote a space of functions on  $U$ . If each  $f \in \mathcal{F}$  is the unique representative of some equivalence class  $[f]$ , say in some  $L^p(U)$ , then we write  $\mathcal{F} \subset L^p(U)$  and say that  $f \in \mathcal{F}$  is an element of  $L^p(U)$ .*

The mutual interrelation between the spaces  $\mathbb{R}^n$  and  $L^2(U)$  is described by the map

$$\mathbb{R}^n \longrightarrow L^2(U), (u_1, \dots, u_n) \longmapsto \left[ \sum_{k=1}^n u_k \mathbf{I}_{\left[\frac{k-1}{n}, \frac{k}{n}\right)} \right]_{L^2(U)},$$

where  $[\cdot]$  denotes equivalence classes, and its left inverse

$$L^2(U) \longrightarrow \mathbb{R}^n, g \longmapsto \left( g_k^n = (1/n)^{-1} \int_{(k-1)/n}^{k/n} g(t) dt : 1 \leq k \leq n \right).$$

These maps are isomorphisms between the Euclidean spaces  $\mathbb{R}^n$  endowed with scaled Euclidean norm  $n^{-1/2} \|\cdot\|$ , and the subspaces  $\mathcal{T}_n$  of  $L^2(U)$  spanned by the indicator functions  $\mathbf{I}_{[k-1/n, k/n)}$ ,  $k = 1, \dots, n-1$ , and  $\mathbf{I}_U$ . For each  $g \in L^2(U)$ , we define finite step-functions

$$g^n = \sum_{k=1}^{n-1} g_k^n \mathbf{I}_{[(k-1)/n, k/n)} + g_n^n \mathbf{I}_{[(n-1)/n, 1]}.$$

Note that the  $g^n$  are càdlàg versions of the conditional expectation of  $g$  with respect to the  $\sigma$ -algebras  $\mathcal{A}_n$  generated by the intervals  $[(k-1)/n, k/n)$ ,  $k = 1, \dots, n-1$ , and

$[(n-1)/n, 1]$  on  $[0, 1]$ , and as such they are unique. The definition is in accordance with the assumption that observations consist of means of the underlying signals over short time-periods.

It is now straightforward to define the time-continuous version of the Blake-Zisserman functional. Basically, it makes sense only on  $\mathcal{A}_n$ -measurable functions. Since each equivalence class has a càdlàg representative  $f$  in  $\mathcal{T}_n$ , we define for each  $f \in \mathcal{T}_n$  and  $g \in L^2(U)$  the functional

$$G_{1/n, \gamma, \mu}(f) = \sum_{k=1}^{n-1} \min \left\{ n(f(k/n) - f((k-1)/n)^2/\mu^2, \gamma) \right\} + \frac{1}{n} \sum_{k=0}^{n-1} (f - g^n)^2(k/n).$$

To simplify notation, let us write  $\varepsilon = 1/n$ . To extend the functional to all of  $L^2(U)$  we define for each  $g \in L^2(U)$  and  $\mu > 0$  the functional as

$$(12) \quad F_{\varepsilon, \gamma, \mu} : L^2(U) \longrightarrow \mathbb{R} \cup \{\infty\}, \quad f \mapsto \begin{cases} G_{\varepsilon, \gamma, \mu}(f) & \text{if } f \in \mathcal{T}_n \\ \infty & \text{if } f \in L^2(U) \setminus \mathcal{T}_n. \end{cases}$$

Note that we introduced the sampling rate  $1/n$  as an additional parameter. We call (12) the *scaled time-continuous Blake-Zisserman functional*. This is the time-continuous counterpart of (5). The degenerate version in the limit  $\mu \rightarrow 0$  is the *scaled time-continuous discrete Potts functional*. Let  $J(f)$  be the set of discontinuities of a stepfunction  $f \in \mathcal{T}_n$ . Then the functional reads

$$(13) \quad F_{\varepsilon, \gamma, 0} : L^2(U) \longrightarrow \mathbb{R} \cup \{\infty\}, \quad f \mapsto \begin{cases} \gamma|J(f)| + \frac{1}{n} \sum_{k=0}^{n-1} (f - g^n)^2(k/n) & \text{if } f \in \mathcal{T}_n \\ \infty & \text{if } f \in L^2(U) \setminus \mathcal{T}_n \end{cases}.$$

Let us write  $\varepsilon = 0$  if we are in continuous time. Then the Mumford-Shah functional becomes  $F_{0, \gamma, \mu}$ . It will turn out that, as  $\mu \rightarrow 0$ , it tends to the (*time-continuous*) *continuous Potts functional*. If  $\mathcal{T}$  denotes the space of *all* finite càdlàg step functions

$$f = \sum_{k=1}^{r-1} a_k \mathbf{1}_{[(k-1)/r, k/r)} + a_r \mathbf{1}_{[(r-1)/r, 1]}$$

then it is given by

$$(14) \quad F_{0, \gamma, 0} : L^2(U) \longrightarrow \mathbb{R} \cup \{\infty\}, \quad f \mapsto \begin{cases} \gamma|J(f)| + \sum_{k=1}^r \int_{(k-1)/r}^{k/r} (a_k - g(t))^2 dt & \text{if } f \in \mathcal{T} \\ \infty & \text{if } f \in L^2(U) \setminus \mathcal{T} \end{cases}.$$

We give now an overview of all these functionals (and another two). Once and for all, we fix a signal  $g \in L^2(U)$ . The parameters  $\gamma > 0$ ,  $\mu > 0$  and  $\varepsilon > 0$  form a subset of  $\mathbb{R}^3$  with closure

$$\mathcal{C} = \{(\varepsilon, \gamma, \mu) : \varepsilon \in \{1/n : n \in \mathbb{N}\} \cup \{0\}, \gamma, \mu \geq 0\}.$$

In the first component it is discrete with an accumulation point at zero.  $C$  is a vertical pile of two-dimensional cones pinned at the positions  $1/n$  and 0. To most of the parameters  $\gamma, \mu, \varepsilon \sim 1/n$  we associated functionals  $F_{\varepsilon, \gamma, \mu}$ . Recall that we reserved  $\varepsilon = 0$  for time-continuous functionals. To some combinations of parameters we did not yet associate functionals; this will be done shortly to complete the picture.

The various functionals correspond to parameters according to the following list, illustrated in Fig. 7:

- for  $\gamma > 0, \mu > 0, \varepsilon = 0$  (the interior of the top-face of the cube) we have the Mumford-Shah functional  $E_{\gamma, \mu} = F_{0, \gamma, \mu}$  from (11),
- for  $\gamma > 0, \mu > 0, \varepsilon \sim 1/n > 0$  (the interior of the cube) we have the scaled Blake-Zisserman functional  $F_{\varepsilon, \gamma, \mu}$  from (12),
- for  $\gamma > 0, \mu = 0, \varepsilon \sim 1/n > 0$  (the interior of the right front face) we have the discrete scaled Potts functional  $F_{\varepsilon, \gamma, 0}$  from (13),
- for  $\gamma > 0, \mu = 0, \varepsilon = 0$  (the intersection of the right front and the top face excluding the origin) we have the continuous scaled Potts functional  $F_{0, \gamma, 0}$  from (14)
- for the interior of the left front face including the right boundary, but excluding the origin, we have the discretized  $L^2$ -distance in the sense

$$F_{\varepsilon, 0, 0}(f) = \frac{1}{n} \sum_{k=1}^n (f_k^n - g_k^n)^2,$$

- for the upper boundary of the left front face including the origin, we simply have the  $L^2$ -distance

$$F_{0, 0, \mu}(f) = \|g - f\|^2.$$

Note that the last functional does not depend on  $\mu$ . We will argue in Section 5 that the map  $(\varepsilon, \gamma, \mu) \mapsto F_{\varepsilon, \gamma, \mu}$  is continuous, in a suitable topology, on the whole range of parameters, including the extreme cases. This is not only a beautiful mathematical result, but of practical importance: Denote the set of minimizers of  $F_{\varepsilon, \gamma, \mu}$  by  $\text{ArgMin}(\varepsilon, \gamma, \mu)$ . This set is in general not a singleton. We can show that the map  $(\varepsilon, \gamma, \mu) \mapsto \text{ArgMin}(\varepsilon, \gamma, \mu)$  is continuous as well.

### 3 Time-Discrete Potts Functionals

We will now describe an example for the statistical analysis and segmentation of data like those presented in Section 1.1. We adopt the simplest variational approach based on Potts functionals. We will first report on some – from the mathematical point of view – typical rigorous results, including existence and uniqueness theorems, a real-time algorithm for the computation of minimizers, results about continuous dependence on parameters, and – as a straightforward consequence – first consistency results. Then we discuss the crucial problem of model choice, or choice of hyperparameters; we start with the classical context, and then suggest new criteria like ‘interval estimators’. Finally, we will illustrate the performance of the classical and of the new criteria in practice.

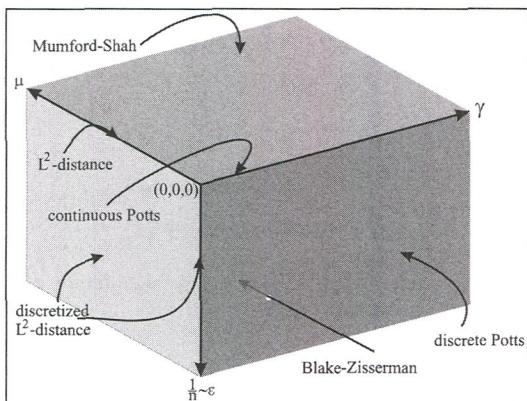


Figure 7: Symbolic display of model classes in terms of their hyperparameters  $\mu$  and  $\gamma$ , and the sampling rate  $\epsilon \sim 1/n$

### 3.1 Rigorous Results

We present now some rigorous results for the Potts filter to be defined shortly. To be consistent, we formulate them for one time dimension only. Some results like the Theorems 3.1 and 3.2 hold *mutatis mutandis* in all finite dimensions. Given data  $v \in \mathbb{R}^n$ , and a hyperparameter  $\gamma > 0$ , a ‘filter output’  $\Phi_\gamma(v)$  is defined as a signal which minimizes the Potts functional  $P_{\gamma,v}$  from (4). The next two theorems are taken from A. Kempe (2004) and O. Wittich et al. (2005). Let us start with the basic analytical result. In general,  $\Phi_\gamma(v)$  is not unique; but fortunately the following result guarantees uniqueness almost surely:

**Theorem 3.1** *For each  $v \in \mathbb{R}^n$ , a minimizer of the functional  $P_{\gamma,v}$  in (4) exists. For each single  $\gamma$ , it is unique outside a Lebesgue nullset of signals  $v$ . In particular, the filter output  $\Phi_\gamma(v)$  is well-defined Lebesgue almost surely, and for each Borel measure on  $\mathbb{R}^n$  with a Lebesgue density, and each  $\gamma > 0$ , it is unique for almost all data  $v$ .*

The exceptional set of data is a finite union of manifolds of codimension one, and hence it is of first Baire category too.

In view of Theorem 3.1, a *Potts estimator* or *Potts filter* can be defined for Lebesgue almost all  $v$  by

$$(15) \quad \Phi_\gamma(v) = \operatorname{argmin}_x P_{\gamma,v}(x).$$

We are going now to report on some of its essential properties. The first one states that the range of hyperparameters  $\gamma$  can be partitioned into intervals, on which  $\Phi_\gamma(v)$  does not change. In the following, we suppress the dependence of  $v$  on the hyperparameters  $\gamma$ . Dependence on hyperparameters is illustrated in Fig. 8.

**Theorem 3.2** *For Lebesgue almost all data  $v$  there are an integer  $k(v)$ ,  $1 \leq k(v) \leq n - 2$ , and hyperparameters  $0 = \gamma_0 < \gamma_1(v) < \dots < \gamma_{k(v)}(v) < \gamma_{k+1} = \infty$  such that*

- (i)  $\Phi_\gamma(v)$  is unique for and does not depend on the hyperparameters  $\gamma_j < \gamma < \gamma_{j+1}$ .
- (ii)  $\Phi_\gamma(v) = v$  for  $\gamma < \gamma_1$ .

- (iii)  $\Phi_\gamma(v)$  is a constant signal for each  $\gamma > \gamma_k$ .
- (iv) The number of jumps of  $\Phi_\gamma(v)$  on the intervals  $(\gamma_j, \gamma_{j+1})$  decreases in  $j$ .
- (v) For each  $0 < j \leq k$ , the functional  $P_{\gamma_j, v}$  has precisely the two minimizers belonging to the  $\gamma$ -intervals adjacent to  $\gamma_j$ .

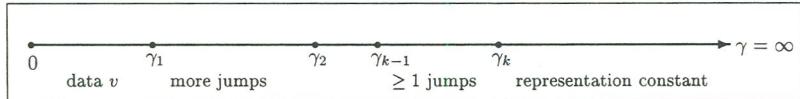


Figure 8: Intervals on which  $\Phi_\gamma(v)$  does not change. Terms like ‘more jumps’ refer to the approach from right to left.

Relevant simulations can only be carried out with exact optimization. Algorithms were sketched in G. Winkler and V. Liebscher (2002).

**Theorem 3.3** For each  $\gamma > 0$ , there is an algorithm to compute a minimizer of  $P_{\gamma, v}$  in time complexity  $O(n^2)$ . There is an algorithm to compute minimizers of  $P_{\gamma, v}$ , for all  $\gamma \in (0, \infty)$  simultaneously, in time complexity  $O(n^3)$ .

The Potts filter has pleasant properties. For example, it locates jumps precisely, and does not insert additional ones. Let  $J(v)$  be the set of all  $k$  with  $v_k \neq v_{k+1}$ .

**Theorem 3.4**  $J(\Phi_\gamma(v)) \subset J(v)$  for all data  $v$  and all hyperparameters  $\gamma$ .

Iteration of the Potts filter stops after one single step which means that the filter is *idempotent*, or more precisely, that  $\Phi_\gamma \circ \Phi_\gamma = \Phi_\gamma$ .

**Theorem 3.5** The Potts filter is idempotent. In particular,  $u$  is a fixed point of  $\Phi_\gamma$  if and only if  $u = \Phi_\gamma(v)$  for some  $v \in \mathbb{R}^n$ .

$\Phi_\gamma$  shares this property with the morphological filters from J. Serra (1982, 1988), see G. Winkler and V. Liebscher (2002). It is not a morphological filter in the strict sense of J. Serra (1982, 1988), since it is not order-preserving. Let us finish this section with some results on continuity and stability. Suppose that for a pair  $(\gamma^*, v^*)$  the filter output  $\Phi_{\gamma^*}(v^*)$  is unique. Then Theorem 3.2 tells us that  $\gamma \mapsto \Phi_\gamma(v^*)$  does not change in a whole open interval around  $\gamma^*$ . We even have joint continuity in  $\gamma$  and  $v$ .

**Theorem 3.6** Suppose that for a pair  $(\gamma^*, v^*)$  the filter output  $\Phi_{\gamma^*}(v^*)$  is unique. Then there is a neighbourhood of  $(\gamma^*, v^*)$  on which  $\Phi_\gamma(v)$  exists and is unique, and moreover, the map  $(\gamma, v) \mapsto \Phi_\gamma(v)$  is continuous.

This has a consequence important in practice.

**Theorem 3.7** Suppose that for a pair  $(\gamma^*, v^*)$  the filter output  $\Phi_{\gamma^*}(v^*)$  is unique. Suppose further that  $\gamma_m \rightarrow \gamma_\infty$ , and that  $Y_m$  and  $Y_\infty$  are random signals with  $Y_m \rightarrow Y_\infty$  in probability, or almost surely. Then  $\Phi_{\gamma_m}(Y_m) \rightarrow \Phi_{\gamma_\infty}(Y_\infty)$  in probability, and almost surely, respectively.

We are not interested in a ‘restoration’ but in feature extraction. This is reflected by the theorem since we recover  $\Phi_{\gamma_\infty}(Y_\infty)$  – and not  $Y_\infty$  – in the limit.

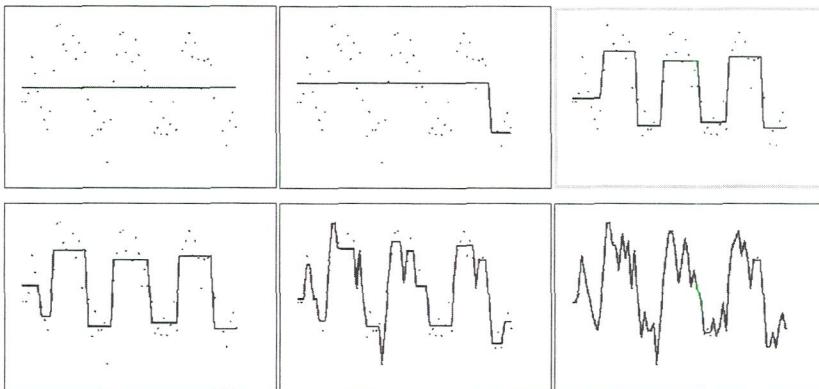


Figure 9: The first four and two later steps of a scan through  $\gamma \mapsto \Phi_\gamma(v)$  for fMRI brain data along decreasing hyperparameters  $\gamma$ . Dots indicate data  $v$ . Upper right is desired, cf. the display of stimulus in Fig. 1.

**Example 3.1** If  $u_\infty$  is degraded by random noise  $\varepsilon_m$  according to  $Y_m = u_\infty + \varepsilon_m$ , and noise  $\varepsilon_m$  tends to zero, then  $\Phi_\gamma(Y_m)$  tends to  $\Phi_\gamma(u_\infty)$ .

Statistical consistency of the estimators defined by the minimizers of Potts functionals is studied in L. Boysen et al. (2005). The Theorems 3.4 and 3.5 are from G. Winkler and V. Liebscher (2002).

### 3.2 Equivariance

Scanning the filter outputs  $\Phi_\gamma(v)$  along the hyperparameter  $\gamma$ , in view of Theorem 3.2, is illustrated in the Figures 9 and 10 for the brain and gene expression data. Such a scan contains the complete potential of the Potts functional for data  $v$ . There remains the crucial problem to decide on the right hyperparameter. Visual inspection of the plots reveals clearly the proper segmentations. On the other hand, it is hard to find a universal unsupervised method for the identification of hyperparameters.

Before we discuss examples of such methods, let us comment on invariance of jumps. The notion of a jump is genuinely non-local, since it compares the variation of a signal in one region with the variation in a larger one. We want to identify a time-point as a jump or change-point whenever the change in intensity near the point is considerably larger than the variation in a larger surrounding. This relation should not be affected by a change of scale since it increases or decreases both, the variation around and off the jump point, by the same factor. Thus, a *jump map*, which assigns to each signal  $f$  a set  $J(f) \subset U$  of jump locations should be invariant with respect to a change of scale. The latter is induced by the action of the affine group

$$\mathbf{A} = \{A_{a,c} : \mathbb{R} \longrightarrow \mathbb{R} : A_{a,c}x = ax + c, a \in \mathbb{R} \setminus \{0\}, c \in \mathbb{R}\}.$$

On signals  $f$ , it is pointwise given by  $A_{a,c}f(t) := af(t) + c$ . A jump map  $J$  will be called *invariant under affine scale changes*, if  $J(Af) = J(f)$  for all signals  $f$  and every  $A \in \mathbb{A}$ .

**Remark 3.1** Scale invariance over the whole range of data is not always appropriate in practice. For example, very weak recorded signals may correspond to various failures in the experiment: in a microarray experiment the spotter may miss a spot, evaluation may be wrong, or there is no binding of cDNA. Then no meaningful information can be extracted. An example of such a signal is contrasted to meaningful ones on different scales in Fig. 11. It should be treated in a different way than those of larger range. One can single out such data by preprocessing, and on the rest equivariant jump detection makes sense.

For situations like those sketched in Section 1.1, equivariance definitely is necessary. There is an enormous number of different time-series – about  $10^5$  for the brain data and about  $2 \times 10^4$  for microarrays – with a wide range of scales. Because of this high throughput, automatic evaluation must rely on the same scheme for all time-series to be processed.

In summary, it is worthwhile to reflect on scale invariant jump maps. We illustrate the concept by the following simple example:

**Example 3.2** Let  $f$  be a càdlàg function and set

$$\Delta(t) = |f(t) - f(t-)|.$$

On the space of finite step functions the jump map  $J_0(f) = \{t \in U : \Delta(t) > 0\}$  is  $\mathbb{A}$ -invariant, whereas  $J_\tau(f) = \{t \in U : \Delta(t) > \tau\}$  with a global threshold  $\tau > 0$  is not. The adaptive threshold

$$J_{\tau,V}(f) = \{t \in U : \Delta(t) > \tau V(f)\}, \quad V(f) = \sup_{t, t' \in U} |f(t) - f(t')|,$$

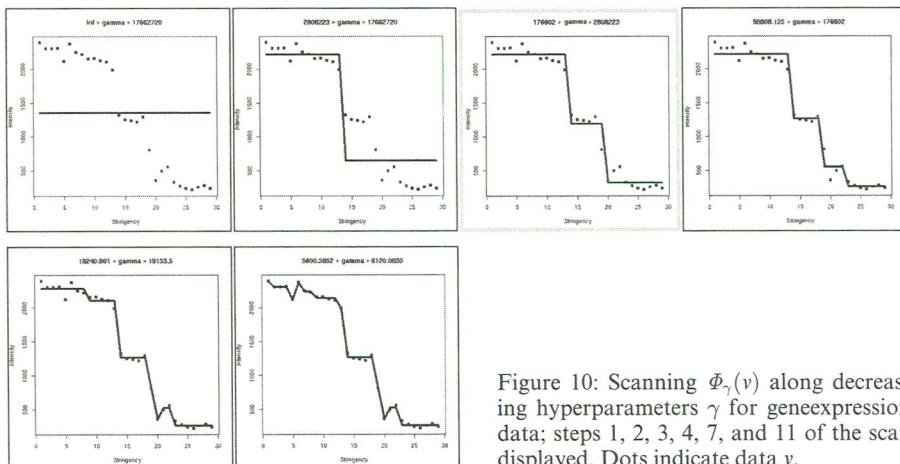


Figure 10: Scanning  $\Phi_\gamma(v)$  along decreasing hyperparameters  $\gamma$  for gene expression data; steps 1, 2, 3, 4, 7, and 11 of the scan displayed. Dots indicate data  $v$ .

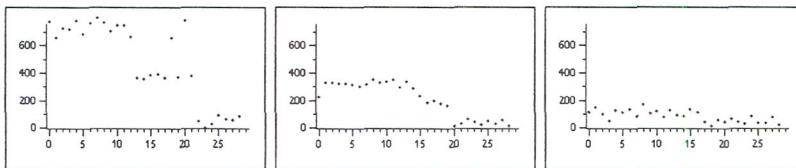


Figure 11: Strong and weak signals from gene expression: cross-hybridization (two jumps) on large and small scale, ‘noise’ on very small scale.

depends on the global range of the signal and is  $\mathbb{A}$ -invariant. This implies in particular that

$$J_{\tau,V}(f) = J_\tau \left( \frac{f - m(f)}{V(f)} \right)$$

where  $m(f)$  may be some normalization like the *midrange*  $m = (\max_x f(x) + \min_x f(x))/2$  or a mean is an invariant jump map. Hence application of  $J_{\tau,V}$  amounts to a standardization of the signal, followed by hard thresholding  $J_\tau$ .

It is evident, that step functions will play a special role. Since they are constant between consecutive points of discontinuity, jumps can only be located at analytical discontinuities.

In view of the Potts functional, let us consider the simple case where *each* analytical discontinuity of a step function defines a jump. In terms of Example 3.2 this corresponds to the jump map  $J_0(t) = \text{disc}(t)$ , where  $t$  is a finite step function, and  $\text{disc}(t)$  the set of its discontinuities. Then for each estimator  $\Phi$  with values in the space of finite step functions the jump map  $J(f) = J_0(\Phi(f))$  is  $\mathbb{A}$ -invariant, if the filter is  $\mathbb{A}$ -equivariant, i.e.  $\Phi(Af) = A\Phi(f)$  for all  $A \in \mathbb{A}$  and all signals  $f$ . This follows readily from

$$J(Af) = J_0(\Phi(Af)) = \text{disc}(A\Phi(f)) = J_0(\Phi(f)) = J(f).$$

This suggests that in this case we should watch out for filters which transform signals to step functions, and which are  $\mathbb{A}$ -equivariant.

By way of example, let us discuss the Potts filters  $v \mapsto \Phi_\gamma(v)$  for parameters  $\gamma \geq 0$ , introduced in (15). These maps are not  $\mathbb{A}$ -equivariant; in fact, a simple calculation shows

$$(16) \quad \Phi_\gamma(av + c) = a\Phi_{\gamma/a^2}(v) + c.$$

*Normalization* is a standard procedure to enforce equivariance of filters.

**Example 3.3** Choose  $\gamma_0 > 0$  and let  $O(f) := \{Af : A \in \mathbb{A}\}$  be the *orbit* of a signal  $f$  under the  $\mathbb{A}$ -action. Then the section

$$p(f) = \frac{f - m(f)}{V(f)} \in O(f)$$

through the orbits, with  $m$  and  $V$  from Example 3.2 leads to the equivariant filter

$$\Phi(f) := V(f)(\Phi_{\gamma_0}(p(f)) + m(f)).$$

In this example nothing was said about  $\gamma_0$ . It has still to be determined properly on the basis of external criteria. This means that the crucial problem is just shifted but not solved.

**Remark 3.2** By external criteria we mean something like testing the residuals for being noise, like the run- or wavelet-coefficient criteria in P.L. Davies and A. Kovac (2001). Such criteria may be meaningful, for example in denoising, where the residuals definitely should be noise. For the data we discuss in this essay they are not. In fact, if we want to extract a Heaviside function from a noisy sigmoid, then we cannot expect the residuals to be noise.

We want to choose  $\gamma_0$  on the basis of information contained in the map  $\gamma \mapsto \Phi_\gamma(v)$ . Therefore, we shall introduce equivariant data-driven parameter choices. This amounts to the construction of maps  $\Gamma : \mathbb{R}^n \rightarrow \mathbb{R}_+, v \mapsto \Gamma(v)$  (with  $\mathbb{R}_+ = [0, \infty)$ ) which assign to each data set  $v$  a hyperparameter  $\Gamma(v)$ . In combination with Potts filters, a sufficient condition for equivariance reads:

**Lemma 3.1** *If  $\Phi_{\Gamma(v)}(v) = \Phi_{\Gamma(av+c)/a^2}(v)$  then  $v \mapsto \Phi_{\Gamma(v)}(v)$  is  $\mathbb{A}$ -equivariant.*

This is a straightforward calculation using (16) and the assumption.

### 3.3 Interval Criteria

There is a host of model selection criteria like the classical ones from H. Akaike (1974) and G. Schwarz (1978). They provide rules to select (and reduce) the parameter dimension for a family of parametric models. It is easy to verify that the Akaike as well as the Schwarz information criterion provide equivariant Potts filters. Applied to the brain data, both methods more or less returned data, see Fig. 12, third frame.

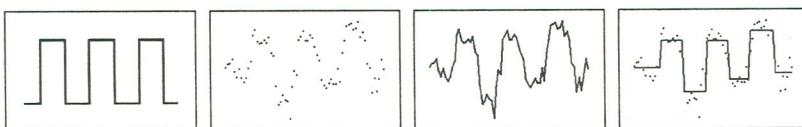


Figure 12: Brain data: stimulus, data,  $\Phi_\gamma(v)$  for hyperparameter from Akaike's and Schwarz' information criterion and longest interval from Section 3.3. The latter gives a decent estimate whereas the former basically return data.

All our suggestions or attempts to find equivariant data-driven choices of hyperparameters are based on Theorem 3.2. Since the minimizers of the Potts functional do not change over entire intervals of  $\gamma$ -values, we may decide on estimators depending only on these intervals and not on single values of  $\gamma$ . Such interval criteria use the complete information about the map  $\gamma \mapsto \Phi_\gamma(f)$ . Even morphological features of the outcomes may be incorporated.

A first naive idea is to aim at stability under changes of the hyperparameter  $\gamma$  and of data  $v$ . For this reason, and in view of the continuity results in Section 3.1, we tried  $\Phi_{\gamma^*(v)}(v)$  with (an arbitrary)  $\gamma^*(v)$  from the *longest interval of  $\gamma$ -values* between 0 and

$\gamma_{k(v)}(v)$  from Theorem 3.2. For the brain data from Example 1.1, this simple method worked. The outcome is contrasted to those of Akaike's and Schwarz' information criteria in Fig. 12. Another – and different – example are the gene expression data. There the crucial restriction is that the representation should be decreasing. With parsimony in mind, we chose – starting from the right – the last interval before the first nondecreasing estimate. Fig. 13 displays some of these estimates. The morphological information of monotony is not fully exploited by this method and further work on this topic has to be done.

The motivation for interval criteria is twofold: First of all, they yield equivariant filter maps as required for the construction of invariant jump maps. This will be shown below. To understand the second feature of interval criteria recall from Example 3.3 that it is easy to construct equivariant Potts filters by normalization of data. On the other hand, nothing is gained by normalization, since the hyperparameter still has to be determined. In case of the brain data, inspection of the results for the model selection criteria shows clearly that equivariance is only a minimal requirement. What we really want is a parameter choice, which relies only on the structure of the functionals and on data.

We are going to show now that the ‘longest interval filter’ is equivariant. Let  $G(v) = \{0, \gamma_1(v), \dots, \gamma_{k(v)}(v)\}$ . We choose a hyperparameter from the longest  $\gamma$ -interval; to be definite let  $\Gamma_l(v)$  be the middle point of the longest interval between 0 and  $\gamma_{k(v)}(v)$ , or, formally (with  $\gamma_0(v) = 0$ ):

$$\Gamma_l(v) = \begin{cases} \arg\max_{0 \leq \gamma < \gamma_{k(v)}(v)} 0 & \text{if } G(v) = \emptyset \\ \min_{0 \leq i \leq k(v)} |\gamma - \gamma_i(v)| & \text{otherwise} \end{cases}.$$

Proofs for the next results can be found in A. Kempe (2004).

**Proposition 3.1**  $\Phi_{\Gamma_l}$  is  $\mathbb{A}$ -equivariant. In particular, the associated jump map is  $\mathbb{A}$ -invariant.

In the next example – related to the gene expression data – we impose the additional restriction that  $\Phi_\gamma(v)$  is monotonically decreasing. Tracking  $\Phi_\gamma(v)$  from right to left, we find  $\gamma$ -intervals on which  $\Phi_\gamma(v)$  is monotonically decreasing and such where it is not. If we desire a monotone filter output, we take the (middle point of the) last interval in series where  $\Phi_\gamma(v)$  decreases. This defines a filter  $\Phi_{\Gamma_d(v)}(v)$ . Since the map  $v \mapsto av + c$  for  $a > 0$  preserves monotony the considerations from Example 3.2 apply and we can state

**Proposition 3.2**  $\Phi_{\Gamma_d}$  is  $\mathbb{A}$ -equivariant. In particular, the associated jump map is  $\mathbb{A}$ -invariant.

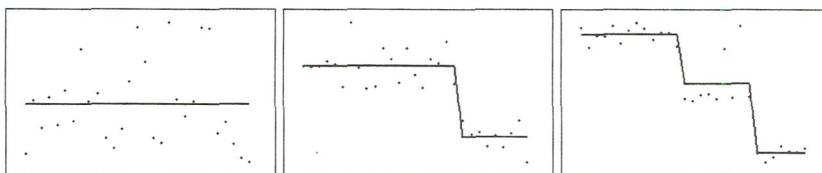


Figure 13: Estimates with leftmost  $\gamma$ -interval with decreasing  $\Phi_\gamma(v)$ : gene data.

## 4 Time-Continuous Signals and Revisited

For us, the Mumford-Shah functional is *the* typical example of variational approaches to edge preserving smoothing in the time-continuous setting. Therefore, it deserves a more detailed investigation. The general form as a combination of a data term, a penalty for the number of breaks, and a control term for smoothness clearly is the first, natural, choice, in particular in view of the time-discrete analogues, and therefore out of question. On the other hand, there are subtle questions about the notion of admissible signals, and in consequence, about the range of definition. Due to its simplicity, the one-dimensional setting is a perfect playground for such discussions.

### 4.1 Spaces of Signals

In this section, we discuss again suitable notions of signals, justify the segmentations introduced in Sections 1.3 and 2.2, and bridge the gap between segmentations and the function spaces used in most of the literature. We continue from Section 2.2.

Let us first append the basics of the theory underlying the discussion in Section 2.2. Let  $I$  be a compact interval on the real line. Recall that a locally integrable function  $h \in L^1_{\text{loc}}(I^\circ)$  is  *$k$  times differentiable in the generalized sense*, if there is a locally integrable function  $h^{(k)}$  such that

$$\int_{I^\circ} h(x) \frac{d^k \phi(x)}{dx} dx = (-1)^k \int_{I^\circ} h^{(k)}(x) \phi(x) dx$$

for every infinitely often differentiable test function  $\phi$  on  $I^\circ$  with compact support. Note that all  $h^{(l)}$ ,  $l \leq k$ , are square-integrable if  $h^{(k)}$  is square integrable. The generalized derivative is defined only almost surely.

Denote by  $\mathcal{W}^{k,2}(I^\circ)$  the *Hilbert-Sobolev space* of functions with  $k$  generalized derivatives and for which the  $k^{\text{th}}$  generalized derivative is square integrable. By the *Sobolev Embedding Theorem*, we have  $\mathcal{W}^{1,2}(I^\circ) \subset C_b(I)$  and even  $\mathcal{W}^{2,2}(I^\circ) \subset C^1(I)$  in the sense that in every equivalence class of functions in the respective Sobolev space there is a representative which extends continuously to the closure. In Section 2.2, the spaces  $W^{1,2}(I^\circ)$  were used as smoothness classes. This choice will be justified shortly from a different point of view.

Recall that  $U$  denotes the unit interval  $[0, 1]$ . At the first sight, the space  $BV(U)$  of *functions with finite total variation* seems to be the natural object. Plainly we have  $BV(U) \subset L^p(U)$  for every  $p \geq 0$ . Let us have a closer look at the functions  $f \in BV(U)$ . Each of them is *regular*, i.e. it has everywhere left and right limits  $f(b_+)$  and  $f(b_-)$ , cf. H. v. Weizsäcker and G. Winkler (1990), Proposition 5.1.8. It induces a finite signed measure  $m$  on the Borel- $\sigma$ -field on  $U$  by the relation

$$m((a, b]) = f(b_+) - f(a_-), \quad a < b \in \mathbb{R}.$$

A càdlàg cumulative distribution function

$$G_f(t) = m((-\infty, t]), t \in \mathbb{R}$$

is associated to  $f$ ; it is unique due to the normalization  $G_f(-\infty) = 0$ . The signed measures  $m$  have unique decompositions into three components, namely

$$m = m_d + m_r + m_s,$$

where  $m_r$  is the *regular part*, admitting a Lebesgue density, where  $m_d$  is the *singular-discrete part*, consisting of a discrete measure, and where  $m_s$  is the *singular-continuous part*, living on a Lebesgue nullset, but without discrete component. One reference is M. Reed and B. Simon (1980), Theorem I.14, page 22, and the preceding definition. Clearly, the support  $J_d(f)$  of the discrete measure  $m_d$  will play the rôle of the jump set of  $f$ . A major question in this context is whether the optimization problem for the Mumford-Shah functional is well-posed. For the version defined in (17) below the problem lies in the singular-continuous component, see A. Chambolle (2000), page 27. Hence we exclude it, and set

$$SBV(U) = \{f \in BV(U) : m_s = 0\}.$$

A member of this class is called *function of special bounded variation*. With this definition we meet the setting adopted in large parts of the literature. Note that a  $SBV(U)$ -function can be written as the sum of a continuous and a (not necessarily càdlàg) stepfunction. It follows easily from Proposition 4.1 below that it has a càdlàg version.

Let us now establish the connection to the Sobolev segmentations introduced in Section 2.2 before (10). To this end, we write the functions induced by Sobolev segmentations in closed form and call a function  $f$  a *piecewise  $(k,p)$ -Sobolev function* if there is a (possibly countable) càdlàg stepfunction  $t$  such that  $f - t \in W^{k,p}(I^\circ)$ . Let us further call a function  $f \in SBV(I^\circ)$  with  $f' \in L^p(I^\circ)$  of  $p$ -special bounded variation and denote the space of these functions by  $SBV_p(I^\circ)$ , cf. G. Dal Maso (1993).

**Proposition 4.1** *For each piecewise  $(1,2)$ -Sobolev function  $f$  there is  $\tilde{f} \in SBV_2(I^\circ)$  with  $f = \tilde{f}$  Lebesgue almost everywhere, and vice versa.*

This shows that the Definitions 10, based on segmentations, and (11) on  $L^2(U)$  from Section 1.3 are equivalent.

## 4.2 The Mumford-Shah Functional

As has been shown, the definition (11) of the Mumford-Shah functional makes sense. Let us formally state

**Definition 4.1** *The one dimensional Mumford-Shah functional with parameters  $\mu, \gamma > 0$  and  $g \in L^2(U)$  is given by*

$$(17) \quad E_{\mu,\gamma} : L^2(U) \longrightarrow \mathbb{R} \cup \{\infty\},$$

$$f \mapsto \begin{cases} \gamma|J(f)| + \int_U (f(t) - g(t))^2 + \frac{1}{\mu^2} |f'_I(t)|^2 dt & \text{if } f \in SBV_2(U) \\ \infty & \text{otherwise} \end{cases}.$$

One of the first and main problems is the mere existence of a minimizer. In two and higher dimension this is an intricate problem. Present solutions rely on special function spaces which are chosen with the explicit aim to guarantee a minimum. In one dimension the situation is much more pleasant.

**Theorem 4.1** *For all  $g \in L^2(U)$  the Mumford-Shah functional (17) has a minimizer. Each minimizer has a càdlàg representative.*

V. Liebscher et al. (2004) contains a proof. In general, minimizers are not unique.

**Remark 4.1** Let us briefly comment on the scales of hyperparameters. Consider the action of  $\mathbb{S} = \mathbb{R}_+ \times \mathbb{R}_+$  on  $L^2(U)$  given by

$$(a, a')f(x) = af(a'x).$$

Then

$$a'E_{\mu/a^2, \gamma/a', (a,a')g}((a, a')f) = E_{\mu, \gamma, g}(f).$$

This means that fixing parameters  $\mu$  and  $\gamma$  simply amounts to the choice of a special  $x$ - and  $y$ -scale for signals in  $L^2(U)$ .

## 5 The Space of Functionals

In Section 2.3 we gave an overview of various functionals in discrete and continuous time which were of similar form. To study their relation, and their behaviour in dependence of hyperparameters, as well as the asymptotics of time-discrete functionals as the sampling rate increases, suitable topological concepts are needed. A main requirement is that they should, under natural conditions, yield continuous dependence of minimizers on hyperparameters. For the numerical analysis, the notion of  $\Gamma$ -convergence is frequently adopted, which is a purely sequential approach. We prefer the topological concept of epiconvergence. It is based on the hit-and-miss topology on closed subsets of  $L^2(U) \times \bar{\mathbb{R}}$ ; restricted to the epigraphs of lower semicontinuous functionals it is called the *epitopology*. Continuous dependence in the epitopology of functionals and minimizers will be addressed in Section 5.2.

### 5.1 Epiconvergence, Semicontinuity, and $\Gamma$ -Convergence

A function  $f$  on  $X$  taking values in  $\bar{\mathbb{R}} = \mathbb{R} \cup \{-\infty, \infty\}$  is called a *numerical function*. A numerical function  $f$  is called *lower semicontinuous* (l.s.c.) if all sets  $\{f(x) > \lambda\}$ ,  $\lambda \in \bar{\mathbb{R}}$ , are open. The *epigraph* of  $f$  is the set

$$\Lambda(f) = \{(x, y) \in X \times \bar{\mathbb{R}} : y \geq f(x)\}.$$

Semi-continuity of a functional on  $X$  is equivalent to the closedness of its epigraph in  $X \times \bar{\mathbb{R}}$ , endowed with the product topology of  $\mathcal{G}$  and the usual topology on  $\bar{\mathbb{R}}$ .

**Proposition 5.1** *A numerical function is lower semi-continuous if and only if its epigraph is closed.*

For a proof see K.T. Rockafellar (1996) or G. Matheron (1975).

Recall the functionals  $F_{\varepsilon, \gamma, \mu}$  introduced in Section 2.3 (and in Section 4.2). To simplify notation let  $\mathbb{L} = \{1/n : n \in \mathbb{N}\} \cup \{0\}$  endowed with its natural topology.

**Theorem 5.1** *All functionals  $F_h : L^2(U) \rightarrow \bar{\mathbb{R}}$ ,  $h \in \mathbb{L} \times \mathbb{R}_+^2$ , are lower semicontinuous.*

A proof is given in V. Liebscher et al. (2004). Since lower semicontinuous functions are characterized by their (closed) epigraphs we need a suitable topology on the set of closed subsets of  $\tilde{X} = X \times \bar{\mathbb{R}}$ . The following definitions do not rely on the product structure of  $\tilde{X}$ . Let hence again  $(X, \mathcal{G})$  be a topological space with the topology  $\mathcal{G}$  of open subsets. It is convenient to start with two topologies on the set  $\mathcal{F}(X) = \mathcal{F}$  of closed subsets of  $X$ . Denote the set of all compact subsets of  $X$  by  $\mathcal{K}$ .

**Definition 5.1** *Let  $\mathfrak{T}_*$  and  $\mathfrak{T}^*$  denote the topologies on  $\mathcal{F}$  which are generated by the families*

$$\mathfrak{T}_*^h = \{\{A \in \mathcal{F} : A \cap G \neq \emptyset\} : G \in \mathcal{G}\}, \quad \mathfrak{T}_m^* = \{\{A \in \mathcal{F} : A \cap K = \emptyset\} : K \in \mathcal{K}\},$$

respectively. The **hit-and-miss topology**  $\mathfrak{T}$  is the topology generated by  $\mathfrak{T}_* \cup \mathfrak{T}^*$ .

The topologies  $\mathfrak{T}_*$  and  $\mathfrak{T}^*$  may be interpreted as the *hit- and miss-topologies* on  $\mathcal{F}$  with hit-sets in  $\mathfrak{T}_*^h$  and miss-sets in  $\mathfrak{T}_m^*$ , respectively. By definition, a sequence  $(A_k)_{k \geq 1}$  in  $\mathcal{F}$  converges to  $A \in \mathcal{F}$  with respect to  $\mathfrak{T}_*$ , if and only if (i)  $G \cap A \neq \emptyset$ ,  $G \in \mathcal{G}$ , implies  $G \cap A_k \neq \emptyset$  eventually. Similarly,  $(A_k)_{k \geq 1}$  converges to  $A$  in  $\mathfrak{T}^*$  if and only if (ii)  $K \cap A = \emptyset$ ,  $K \in \mathcal{K}$ , implies  $K \cap A_k = \emptyset$  eventually. We have convergence in  $\mathfrak{T}$  if and only if both (i) and (ii) hold. In view of Proposition 5.1, the following definition makes sense now:

**Definition 5.2** *Let us call the restriction of the  $\mathfrak{T}$ -topology on  $\tilde{\mathcal{F}}$  to the set*

$$\Lambda_X = \{\Lambda(f) \subset X \times \bar{\mathbb{R}} : f : X \rightarrow \bar{\mathbb{R}}, f \text{ l.s.c.}\}$$

of epigraphs the **epitopology**. A map  $h \mapsto F_h$  from a topological space into the space of l.s.c. functions on  $X$  will be called **epicontinuous** if the map  $h \mapsto \Lambda(F_h)$  is continuous for the epitopology.

A sequential version is known as  $\Gamma$ -convergence, cf. G. Dal Maso (1993). Let now  $\text{ArgMin } F$  denote the set of minimizers of a functional  $F$  on  $X$ .

**Lemma 5.1** *If  $F : X \rightarrow \bar{\mathbb{R}}$  is lower semicontinuous then  $\text{ArgMin } F$  is closed.*

The proof is an easy exercise. The following ‘semincontinuity’ assertion makes sense by Lemma 5.1 and holds without further hypotheses.

**Theorem 5.2** *The map*

$$\tilde{\mathcal{F}} \supset \Lambda_X \longrightarrow \mathcal{F}, \Lambda(F) \longmapsto \text{ArgMin } F$$

is continuous with respect to the  $\mathfrak{T}$ -topology on  $\Lambda_X$  and  $\mathfrak{T}^*$  on  $\mathcal{F}(X)$ .

Note that  $\mathfrak{T}^*$  in general does not enjoy the Hausdorff separation property.

## 5.2 Joint Continuity in Hyperparameters

The last topic in this paper is joint continuity of the functionals  $F_{\varepsilon, \gamma, \mu}$  in *all the parameters*  $\varepsilon$ ,  $\gamma$  and  $\mu$ , symbolically displayed in Fig. 7. Another question is how this carries over to their (sets of) *minimizers*. This fulfills the promise given in Section 2.3. Proofs of the main results are given in V. Liebscher et al. (2004).

**Theorem 5.3** *Let  $g \in L^2(U)$ . The map*

$$\mathbb{IL} \times \mathbb{IR}_+ \times \mathbb{IR}_+ \longrightarrow \mathbb{IR}, (\varepsilon, \gamma, \mu) \longmapsto F_{\varepsilon, \gamma, \mu}.$$

*is epicontinuous.*

The proof is divided into several parts according to the list in Section 2.3. It is given in V. Liebscher et al. (2004). For some specific examples in the purely time-continuous setting see G. Alberti (2000).

Concerning minimization, let us report the following results:

**Theorem 5.4** *Each functional  $F_{\varepsilon, \gamma, \mu}$  attains its minimum.*

Theorem 5.3 in combination with Theorem 5.2 implies:

**Theorem 5.5** *Let  $g \in L^2(U)$ . The map*

$$\mathbb{IL} \times \mathbb{IR}_+ \times \mathbb{IR}_+ \longrightarrow \mathcal{F}(L^2(U)), (\varepsilon, \gamma, \mu) \longmapsto \text{ArgMin } F_{\varepsilon, \gamma, \mu}.$$

*is continuous with respect to the miss topology  $\mathfrak{T}^*$  on  $\mathcal{F}(L^2(U))$ .*

A more concrete result in this spirit is the following one:

**Theorem 5.6** *Let  $h_k \in \mathbb{IL} \times \mathbb{IR}_+^2$ ,  $k \geq 1$ , and suppose that the sequence  $(h_k)_{k \geq 1}$  converges to  $h \in \mathbb{IL} \times \mathbb{IR}_+^2$ . Let further for each  $k \geq 1$  the function  $f_k^*$  be any minimizers of the functional  $F_{h_k}$ . Then the following hold:*

- (i) *The set  $\{f_k^* : k \geq 1\}$  is relatively compact in  $L^2(U)$ .*
- (ii) *The limit of each  $L^2(U)$ -convergent subsequence of  $(f_k^*)$  minimizes  $F_h$ .*

The situation is particularly pleasant if  $F_h$  has only one single minimizer.

**Corollary 5.1** *Suppose that  $F_h$  has one and only one minimizer  $f^*$ . Then any sequence  $(f_k^*)_{k \geq 1}$  with  $f_k \in \text{ArgMin } F_{h_k}$  converges to  $f^*$  in  $L^2(U)$ .*

With this statement we conclude the discussion of continuous dependence of minimizers on parameters and the sampling rate, and also this essay. The discussion is continued in the forthcoming papers L. Boysen et al. (2005) and V. Liebscher et al. (2004).

## Summary and Final Remarks

This essay had essentially three aims: Firstly, to give an integral view over a variety of variational approaches to signal analysis. Motivated by discrete-time models we argued that time-continuous models must be incorporated in order to describe the ‘closure’ of

the set of discrete-time models. Conversely, the transition from continuous to discrete time justifies the notion of smoothness in discrete time, captured by the penalty.

Secondly, we wanted to exemplify by way of the simplest of these models the first steps of a rigorous and in parts complete mathematical analysis.

The third aim was to establish the connection to relevant real-world data of a special kind. The main two properties of these data are that they consist of relatively short time-series, and that the relevant features are of qualitative, or more precisely, morphological nature.

Therefore the classical justifications like validity of model assumptions or asymptotics are idle in these cases, and we must resort to other, more basic criteria. A first one is equivariance, which for us was a partial justification of the interval estimators for the Potts functional. Plainly, this can only be the starting point for future discussions and investigations.

Let us finally mention, that closely related variational models appear in various disguises. Let us give an example from wavelet theory, cf. A. Chambolle et al. (1998).

**Example 5.1** Consider a function  $f \in L^2(\mathbb{R})$  with a wavelet decomposition

$$f = \sum_{j,k \in \mathbb{Z}} \langle f, \psi_{j,k} \rangle \psi_{j,k} =: \sum_{j,k \in \mathbb{Z}} c_{j,k} \psi_{j,k}, \quad \psi_{j,k}(t) = 2^{j/2} \psi(2^j t - k), \quad j, k \in \mathbb{Z},$$

where  $\psi \in L^2(\mathbb{R})$  is a ‘mother wavelet’ such that  $\{\psi_{j,k} : j, k \in \mathbb{Z}\}$  is an orthonormal basis of  $L^2(\mathbb{R})$ . *Soft Wavelet Thresholding* amounts to the operation

$$\phi_\lambda(c_{j,k}) = \begin{cases} c_{j,k} - \lambda & \text{if } c_{j,k} > \lambda \\ 0 & \text{if } |c_{j,k}| \leq \lambda \\ c_{j,k} + \lambda & \text{if } c_{j,k} < -\lambda \end{cases}, \quad \lambda > 0.$$

on the wavelet-coefficients and the subsequent transformation  $f \mapsto \tilde{f} = \sum_{j,k \in \mathbb{Z}} \phi_\lambda(c_{j,k}) \psi_{j,k}$ . Because of vanishing first moments, large coefficients correspond to steep parts of  $f$ , this can be used for edge preserving smoothing. The shrinkage  $c_{j,k} \mapsto c_{j,k} \pm \lambda$  removes further noise related to the surviving coefficients. Then  $\tilde{f}$  solves the variational problem

$$(18) \quad \tilde{f} = \underset{g \in L^2(\mathbb{R})}{\operatorname{argmin}} C \|g\|_1^\psi + \|f - g\|_2^2,$$

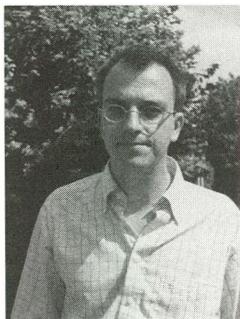
where  $\|g\|_1^\psi$  is the wavelet dependent  $\ell^1$ -norm on the coefficients ( $\langle g, \psi_{j,k} \rangle$ ). Under standard conditions on  $\psi$  this norm is equivalent to the usual Besov norm. It generalizes the  $W^{1,2}$ -notion of smoothness.

**Acknowledgement.** K. Hahn and D. Auer provided the brain data, J. Beckers and A.L. Drobyshev the gene expression data. M. Wild pointed out Example 5.1; K. Gröchenig discussed wavelets with us. Simulations were performed with *AntsInFields* by F. Friedrich; a CD-ROM is attached to G. Winkler (2003), free download under F. Friedrich (2003).

## References

- [1] H. Akaike. A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, 19(6):716–723, 1974.
- [2] G. Alberti. Variational models for phase transitions, an approach via gamma-convergence. In G. Buttazzo et al., editor, *Differential Equations and Calculus of Variations*. Springer Verlag, 2000.
- [3] A. Blake. The least disturbance principle and weak constraints. *Pattern Recognition Lett.*, 1:393–399, 1983.
- [4] A. Blake and A. Zisserman. *Visual Reconstruction*. The MIT Press Series in Artificial Intelligence. MIT Press, Massachusetts, USA, 1987.
- [5] L. Boysen, A. Munk, V. Liebscher, and O. Wittich. Jump-penalized least squares: Consistencies and rates of convergence. Schriftenreihe des IBB, 05-3, January 2005. 38 pages.
- [6] A. Chambolle. Inverse problems in Image processing and Image segmentation: some mathematical and numerical aspects. Lecture given at the School on Mathematical Problems in Image Processing, Ceremade, Université de Paris-Dauphine, September 2000.
- [7] A. Chambolle, R.A. DeVore, N-Y Lee, and B.J. Lucier. Nonlinear wavelet image processing: Variational problems, compression, and noise removal through wavelet shrinkage. *IEEE Trans. Im. Proc.*, 7(3):1–22, 1998.
- [8] G. Dal Maso. *An introduction to  $\Gamma$ -convergence*., volume 8 of *Progress in Nonlinear Differential Equations and their Applications*. Birkhäuser, Basel, 1993.
- [9] P.L. Davies. Data features. *J. of the Netherlands Society for Statistics and Operations Research*, 49(2):185–245, July 1995.
- [10] P.L. Davies and A. Kovac. Local extremes, runs, strings and multiresolution. *Ann. Stat.*, 29(1):1–65, 2001.
- [11] A. L. Drobyshev, Chr. Machka, M. Horsch, M. Seltmann, V. Liebscher, M. Hrabé de Angelis, and J. Beckers. Specificity assessment from fractionation experiments, (SAFE): a novel method to evaluate microarray probe specificity based on hybridization stringencies. *Nucleic Acids Res.*, 31(2):1–10, 2003.
- [12] F. Friedrich. AntsInFields: Stochastic simulation and Bayesian inference for Gibbs fields, 2003. URL: [www.AntsInFields.de](http://www.AntsInFields.de).
- [13] S. Geman and D. Geman. Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images. *IEEE Trans. PAMI*, 6:721–741, 1984.
- [14] F.R. Hampel, E.M. Ronchetti, P.J. Rousseeuw, and W.A. Stahel. *Robust Statistics*. Wiley Series in Probability and Mathematical Statistics. Wiley & Sons, New York, 1986.
- [15] P.J. Huber. *Robust statistics*. Wiley Series in Probability and Mathematical Statistics. John Wiley & Sons, New York etc., 1981.
- [16] E. Ising. Beitrag zur Theorie des Ferromagnetismus. *Z. Physik*, 31:253, 1925.
- [17] A. Kempe. *Statistical analysis of discontinuous phenomena with Potts functionals*. PhD thesis, Institute of Biomathematics and Biometry, National Research Center for Environment and Health, Munich, Germany, 2004.
- [18] V. Liebscher, O. Wittich, A. Kempe, and G. Winkler. Segmentation of time series: A case study. Preprint, 45 pages, June 2004.
- [19] E. Mammen and S. van de Geer. Locally adaptive regression splines. *The Annals of Statistics*, 25(1):387–413, 1997.
- [20] D. Marr. *Vision*. W.H. Freeman and Company, New York, 1982.
- [21] G. Matheron. *Random sets and integral geometry*. Wiley Series in Probability and Mathematical Statistics. John Wiley & Sons, New York etc., 1975.
- [22] D. Mumford and J. Shah. Boundary detection by minimizing functionals I. In *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, pages 22–26, Washington, June 1985. CVPR '85, San Francisco, IEEE Computer Society Press.
- [23] D. Mumford and J. Shah. Optimal approximation by piecewise smooth functions and associated variational problems. *Comm. Pure Appl. Math.*, 42:577–685, 1989.

- [24] R.B. Potts. Some generalized order-disorder transitions. *Proc. Camb. Phil. Soc.*, 48:106–109, 1952.
- [25] M. Reed and B. Simon. *Methods of Modern Mathematical Physics*, volume I: Functional Analysis. Academic Press, San Diego, revised and enlarged edition, 1980.
- [26] K.T. Rockafellar. *Convex analysis*. Princeton University Press, Princeton, New Jersey, 1996.
- [27] B. Russell. On the notion of cause, with applications to the free-will problem. In H. Feigl and M. Brodbeck, editors, *Readings in the philosophy of science*, pages 387–407. Appleton-Century-Crofts, New York, 1953.
- [28] G. Schwarz. Estimating the dimension of a model. *The Annals of Statistics*, 6(2):461–464, 1978.
- [29] J. Serra. *Image analysis and mathematical morphology*. Acad. Press, London, 1982.
- [30] J. Serra. *Image analysis and mathematical morphology: Theoretical Advances*. Academic Press, London etc., 1988.
- [31] H.v. Weizsäcker and G. Winkler. *Stochastic Integration*. Vieweg Advanced Lectures in Mathematics. Friedrich Vieweg & Sohn, Braunschweig, Wiesbaden, 1990.
- [32] G. Winkler. *Image Analysis, Random Fields and Markov Chain Monte Carlo Methods. A Mathematical Introduction*, volume 27 of *Applications of Mathematics*. Springer Verlag, Berlin, Heidelberg, New York, second edition, 2003. Completely rewritten and revised, Year of publication: 2002.
- [33] G. Winkler, V. Aurich, K. Hahn, A. Martin, and K. Rodenacker. Noise reduction in images: Some recent edge-preserving methods. *Pattern Recognition and Image Analysis. Advances in Mathematical Theory and Applications*, 9(4):749–766, 1999.
- [34] G. Winkler and V. Liebscher. Smoothers for discontinuous signals. *J. Nonpar. Statist.*, 14(1–2):203–222, 2002.
- [35] O. Wittich, A. Kempe, G. Winkler, and V. Liebscher. Complexity penalized sums of squares for time series': Rigorous analytical results. Schriftenreihe des IBB, 05-1, January 2005. 18 pages.



# Der Multiplikationssatz der Mengenlehre

Oliver Deiser

## Abstract

- Mathematics Subject Classification: 01-02, 03-03, 03 E 10, 03 E 20, 03 E 25
- Keywords and Phrases: multiplication theorem, cardinal arithmetic, history of set theory

Wir diskutieren die Geschichte des Multiplikationssatzes der Mengenlehre, der die Frage nach der Mächtigkeit des kartesischen Produkts einer unendlichen Menge mit sich selbst beantwortet.

Eingegangen: 23.01.2004, in überarbeiteter Fassung am 03.03.2005

Oliver Deiser, Mathematisches Institut der LMU München,  
[deiser@mathematik.uni-muenchen.de](mailto:deiser@mathematik.uni-muenchen.de)

**DMV**  
JAHRESBERICHT  
DER DMV  
© B. G. Teubner 2005

## 1 Einführung

Der Multiplikationssatz der Mengenlehre lautet:

*Ist  $M$  eine unendliche Menge, so existiert eine Bijektion zwischen  $M \times M$  und  $M$ .*

Dieser Satz bedarf aufgrund seiner überzeugend schlichten Antwort auf ein sehr natürliches und zunächst schwieriges Problem keinerlei Werbemaßnahmen, jedoch ist das Interesse, das dem Satz zukommt, vielgestaltig: Von der mengertheoretischen Sternwarte aus ist der Satz Teil des Ergebnisses, daß alle kanonischen arithmetischen Operationen mit Mächtigkeiten letztendlich trivial sind mit Ausnahme der Exponentiation, deren Natur dafür um so schwerer zu ergründen ist. Geschichtlich trug der Spezialfall des Satzes für die reellen Zahlen – die Existenz einer Bijektion zwischen der Ebene und der Geraden –, viel zur Fundierung der Analysis bei, die sich gegen Ende des 19. und zu Anfang des 20. Jahrhunderts vollzog. Weiter spiegelt die sich über drei Jahrzehnte erstreckende Geschichte des allgemeinen mengentheoretischen Satzes den gesellschaftlichen Wachstum einer Theorie wieder, die aus einer schöpferischen Einzelleistung hervorging, und dann etwas fröhreif als blue chip der mathematischen Börse dastand und ihren Crash auslöste. Was für die gesamte Mengenlehre gilt, gilt hier speziell für ein einzelnes Resultat: Georg Cantor geht den Weg zunächst ganz alleine, und läßt dabei genügend Raum für seine Epigonen übrig, seine Ideen nicht nur weiterzuentwickeln, sondern für sich neu zu ordnen und zu interpretieren. Seine schwindende mathematische Kraft gegen Ende des 19. Jahrhunderts markiert eine Unstetigkeitsstelle in der Entwicklung des Gebiets, die Tradierung seiner Intuition und seines Wissens gelingt nur bruchstückhaft. Seine Einsichten etwa über die Paradoxien der „absolut unendlichen Vielheiten“, der „großen“ echten Klassen im Gegensatz zu den „kleinen“ Mengen sind heute nur brieflichen Spuren folgend zu rekonstruieren. Daß Cantor einen Beweis des Multiplikationsatzes gesehen hat, ist nicht einmal handschriftlich dokumentiert, sondern nur einer Bemerkung von Felix Bernstein in seiner Dissertation von 1901 zu entnehmen, die dann einige Jahre lang in nebelhafter Weise verwendet wird.

Im Jahre 1905 veröffentlichte ein bis heute obskurer Herr A. E. Harward, angeregt durch nichts als Russells „Principles of Mathematics“ und zwei nicht gerade glänzende Artikel von Philip Jourdain einen ersten vollständigen Beweis, der von der Mathematikergilde nicht wahrgenommen wurde. Harward war Verwaltungsangestellter in Indien, und gibt der ganzen Geschichte eine kuriose und exotische Note. Er skizziert in einem Anhang seiner Arbeit zudem noch einen zweiten Beweis, den Hausdorff erst 1914 wiederentdecken wird, und der im wesentlichen den heute üblichen darstellt. Innerhalb der professionellen Mathematik gelingt Gerhard Hessenberg unabhängig von Harward ungefähr zeitgleich ein weiterer Beweis, der 1906 in seinem Lehrbuch „Grundbegriffe der Mengenlehre“ erscheint, und seither als erster Beweis des Satzes zitiert wird.

Ein wichtiger Satz der Mengenlehre, der zudem mit dem heute zentralen Begriff der Regularität von Kardinalzahlen in enger Beziehung steht, blieb durch einen mathematischen Generationenwechsel, gepaart mit der Nichtbeachtung eines Außenseiters, an der Jourdain nicht unschuldig ist, schätzungsweise ein ganzes Jahrzehnt verdeckt. Das Unglück der frühen Mengenlehre ist die mangelnde Zentrumsbildung um Cantor in den 80er und 90er Jahren, ihr Glück die Wiederaufnahme seiner Ideen in zwei völlig ver-

schiedenen Richtungen durch Ernst Zermelo und Felix Hausdorff nach dem Jahrhundertwechsel, die in eine Zeit des allgemeinen Interesses an mathematischen Grundlagenfragen fiel.

Wir diskutieren im folgenden ausführlich die Geschichte des Satzes, können die fünf darin eine Rolle spielenden Beweise aber aufgrund physikalischer Limitationen nur andeuten. Den an vollständigen mathematischen Beweisen interessierten Leser müssen wir auf die modernen Kommunikationsmittel verweisen: Auf der Homepage des Autors findet sich eine längere Version der Arbeit, die vollständige Beweise enthält. Es wurde versucht, sie so zu präsentieren, daß sie nach Möglichkeit jedem Mathematiker zugänglich sind.

Weiter kann auch auf die Geschichte des Dimensionsbegriffs hier nicht genauer eingegangen werden; siehe hierzu [Katetov / Simon 1997] für eine eingehende Darstellung und weitere Literatur.

## 2 Notationen und zentrale mengentheoretische Sätze

Wir referieren kurz die wichtigsten Bausteine der Mächtigkeitstheorie und der Theorie der Wohlordnungen. Das Bühnenbild ist das der klassischen Mathematik. Technisch gesprochen heißt das: Wir arbeiten in der üblichen Mengenlehre mit Auswahlaxiom. Die Verwendung des Auswahlaxioms ist für den behandelten Gegenstand von Bedeutung und wird daher jeweils notiert.

Sind  $M$  und  $N$  Mengen, so schreiben wir  $|M| = |N|$ , falls eine Bijektion von  $M$  auf  $N$  existiert;  $|M| \leq |N|$ , falls eine Injektion von  $M$  nach  $N$  existiert;  $|M| < |N|$ , falls  $|M| \leq |N|$ , aber  $|M| \neq |N|$  gilt. Ist  $|M| = |N|$ , so nennen wir  $M$  und  $N$  gleichmächtig  $f$ ; ist  $|M| < |N|$ , so sagen wir, daß die Mächtigkeit von  $M$  kleiner ist als die Mächtigkeit von  $N$ . Dies ist die relationale Definition der Mächtigkeiten – wir haben nicht definiert, was „die Mächtigkeit von  $M$ “ selbst ist, und brauchen dies auch nicht tun.

Eine Menge  $M$  heißt (Dedekind-) unendlich, falls es eine echte Teilmenge  $N$  von  $M$  gibt mit  $|N| = |M|$ . Äquivalent hierzu ist: Es gilt  $|N| \leq |M|$ . (Diese Äquivalenz verwendet das Auswahlaxiom nicht. Es wird aber verwendet, um zu zeigen:  $M$  ist genau dann endlich, wenn gilt: es gibt ein  $n \in \mathbb{N}$  mit  $|M| = |\{0, \dots, n-1\}|$ .)

Wir notieren nun offiziell das Thema dieses Artikels:

### Multiplikationssatz (Harward 1905, Hessenberg 1906)

Ist  $M$  eine unendliche Menge, so ist  $|M \times M| = |M|$ .

Hessenberg beweist zugleich auch den verwandten Additionssatz: Die Vereinigung zweier disjunkter Kopien einer unendlichen Menge ist gleichmächtig zur Ausgangsmenge. Hierzu definieren wir für Mengen  $M$  und  $N$ :  $M + N = M \times \{0\} \cup N \times \{1\}$ . Die Mengen  $M' = M \times \{0\}$  und  $N' = N \times \{1\}$  sind gleichmächtige „Kopien“ von  $M$  bzw.  $N$ , und es gilt  $M' \cap N' = \emptyset$ . Wir schreiben dann den Additionssatz wie folgt:

### Additionssatz

Ist  $M$  eine unendliche Menge, so ist  $|M + M| = |M|$ .

Für viele Beweise des Multiplikationssatzes ist der Additionssatz ein fahrplanmäßiger Halt auf der Beweisstrecke. Zeigt man den Multiplikationssatz direkt, so ergibt sich umgekehrt der Additionssatz als Korollar mit Hilfe des folgenden Satzes, der pausenlos im Einsatz ist:

**Satz** (*Satz von Cantor-Bernstein; Cantor 1883, Dedekind 1887, Bernstein 1897*)  
Seien  $M, N$  Mengen mit  $|M| \leq |N|$  und  $|N| \leq |M|$ .  
Dann gilt  $|M| = |N|$ .

Cantor hat diesen Sachverhalt 1883 formuliert und einen Beweis angekündigt. Aber erst Felix Bernstein konnte den Satz 1897 in einem von Cantor veranstalteten Seminar in Halle vollständig beweisen. Dedekind hatte bereits 1887 einen Beweis entdeckt, der sich heute in seinem Nachlaß findet.

Für den Beweis des Satzes muß das Auswahlaxiom nicht verwendet werden.

Offenbar gilt  $|M| \leq |M + M| \leq |M \times M|$  für Mengen  $M$  mit mindestens zwei Elementen, und daher folgt der Additionssatz aus dem Multiplikationssatz mit Hilfe von Cantor-Bernstein.

Max Zorn hat 1944 einen Beweis des Multiplikationssatzes gegeben, der keine Ordinalzahlen verwendet. Das wesentliche Hilfsmittel ist hier:

**Satz** (*Zornsches Lemma; Zorn 1935*)

Ist  $P$  eine partiell geordnete Menge, in der jede total geordnete Teilmenge eine obere Schranke besitzt, so existiert ein maximales Element der Ordnung, d. h. ein  $x \in P$  mit: Ist  $y \in P$  und  $x \leq y$ , so ist  $x = y$ .

Das Zornsche Lemma ist geeignet, den recht filigranen Begriff der Wohlordnung aus bestimmten Argumenten zu vertreiben, und ist auch zu dieser zuweilen etwas grobschlächtigen Anwendung ins Leben gerufen worden: Zorn gab 1935 das Lemma als Prinzip ohne Beweis an. Das Zornsche Lemma ist ein einfaches Korollar des Hausdorffschen Maximalitätsprinzips [Hausdorff 1914]. Aufgrund seiner einfachen Handhabung wurde es zum Exportschlag.

Das Zornsche Lemma ist, auf der Basis der anderen Axiome der Mengenlehre, äquivalent zum Auswahlaxiom. Gleichermaßen gilt für den folgenden Satz:

**Satz** (*Vergleichbarkeitssatz; Cantor 1878, Zermelo 1904*)

Seien  $M, N$  Mengen. Dann gilt  $|M| \leq |N|$  oder  $|N| \leq |M|$ .

### Beweis

O. E. sind  $M, N \neq \emptyset$ . Sei  $P$  die Menge aller Injektionen  $f: M' \rightarrow N$  mit  $M' \subseteq M$ , geordnet durch Inklusion, d. h.  $f \leq g$ , falls  $f \subseteq g$  (d. h.  $g$  setzt  $f$  fort). Das Zornsche Lemma findet Anwendung. Sei also  $f \in P$  maximal.

Dann ist  $f: M \rightarrow N$  injektiv oder  $f^{-1}: N \rightarrow M$  injektiv.

Der Vergleichbarkeitssatz wurde von Cantor zunächst als „offensichtlich“ angesehen [vgl. den ersten Absatz von Cantor 1878], später hat er den Satz dann als Problem formuliert [Cantor 1895, Ende § 2] und bewiesen, aus heutiger Sicht jedoch nicht in vollständiger Strenge. Erst der Zermelosche Wohlordnungssatz von 1904 lieferte einen lückenlosen Beweis. Die Äquivalenz zum Auswahlaxiom (und damit zum Zornschen Lemma) hat Friedrich Hartogs gezeigt [Hartogs 1915].

Eine totale Ordnung  $\langle W, < \rangle$  ist eine *Wohlordnung*, falls jede nichtleere Teilmenge von  $W$  ein  $<$ -kleinstes Element besitzt.  $X \subseteq W$  ist ein (*echtes*) *Anfangsstück* von  $W$ , falls ein  $b \in W$  existiert mit  $X = \{a \in W \mid a < b\}$ . Wir schreiben im folgenden oftmals nur  $W$  für eine Wohlordnung  $\langle W, < \rangle$ . Eine die Menge  $W$  wohlordnende Relation  $< \subseteq W \times W$  ist dann stillschweigend mit dabei.

Zwei Wohlordnungen  $W_1$  und  $W_2$  sind *ordnungsisomorph* oder *gleichlang*, falls eine Bijektion  $f: W_1 \rightarrow W_2$  existiert, sodaß für alle  $a, b \in W_1$  gilt:  $a < b \Leftrightarrow f(a) < f(b)$  [ $\Leftrightarrow$ ]

steht wie in der Logik üblich für „genau dann, wenn“].  $W_1$  ist (*strikt*) kürzer als  $W_2$ , falls  $W_1$  isomorph zu einem Anfangsstück von  $W_2$  ist.

Ein Hauptmerkmal von Wohlordnungen ist die Vergleichbarkeit ihrer Längen:

**Satz** (*Vergleichbarkeitssatz für Wohlordnungen; Cantor 1897*)

Seien  $W_1$  und  $W_2$  zwei Wohlordnungen. Dann tritt genau einer der drei folgenden Fälle ein:

- (i)  $W_1$  und  $W_2$  sind gleichlang,
- (ii)  $W_1$  ist kürzer als  $W_2$ .
- (iii)  $W_2$  ist kürzer als  $W_1$ .

Insbesondere sind Wohlordnungen niemals gleichlang zu ihren eigenen Anfangsstücken.

Der Vergleichbarkeitssatz für Wohlordnungen ist ohne Auswahlaxiom beweisbar. Allgemein wird das Auswahlaxiom beim Jonglieren mit Wohlordnungen nie gebraucht, da man immer auf einen *kleinsten Zeugen* innerhalb irgend etwas Nichtverschwindendem zugreifen kann, anstatt nur *auf einen Zeugen*, was der Job einer Auswahlfunktion wäre. Lediglich um die nackte Existenz von Wohlordnungen zu sichern ist das Auswahlaxiom bitter nötig:

**Satz** (*Wohlordnungssatz; Zermelo 1904, zweiter Beweis 1908*)

Jede Menge lässt sich wohlordnen:

Ist  $M$  eine Menge, so existiert eine Wohlordnung  $<$  auf  $M$ .

Der Zermelosche Wohlordnungssatz ist äquivalent zum Auswahlaxiom.

Noch ein paar Worte zu Ordinalzahlen und Kardinalzahlen. In der Mengenlehre definiert man *Ordnungstypen* oder *Ordinalzahlen* nach Cantor informal als das allen Wohlordnungen gleicher Länge Gemeinsame, oder formal nach von Neumann und Zermelo als bestimmte natürliche und uniform definierbare Repräsentanten für Wohlordnungen – je ein Repräsentant pro Länge. Die *transfiniten Zahlen* sind dann in beiden Versionen einfach diejenigen Ordinalzahlen, die den unendlichen Wohlordnungen zugeordnet sind.

Man kann mit Wohlordnungen (und folglich mit Ordinalzahlen) arithmetisch operieren: Hintereinanderhängen zweier Ordnungen führt zur Summe, lexikographische Ordnung des kartesischen Produkts (oder iterierte Summation) führt zur Multiplikation, iterierte Multiplikation zur Exponentiation.

Ganz ähnlich kann man *Kardinalzahlen* informal als das allen Mengen gleicher Mächtigkeit Gemeinsame definieren. Eine formale Definition ist möglich, auch in einer Mengenlehre ohne Auswahlaxiom (mit einer nichttrivialen Konstruktion). *Alephs* sind nun diejenigen Kardinalzahlen, die zu den unendlichen und wohlordenbaren Mengen gehören. Mit Hilfe des Wohlordnungssatzes sind alle Mengen gleichmächtig zu einer Wohlordnung, und die Alephs fallen dann mit den unendlichen Kardinalzahlen zusammen. In einer Mengenlehre ohne Auswahlaxiom bilden die Alephs eine Teilkasse der unendlichen Kardinalzahlen, und bzgl. der Vergleichbarkeit von Kardinalzahlen kann dann nur noch für die Alephs Garantie übernommen werden.

Mit Kardinalzahlen kann man ebenfalls rechnen, die Operationen sind hier über die Mächtigkeiten von Summe, kartesischem Produkt und, für die Exponentiation, der

Menge aller Funktionen von einer Menge der Mächtigkeit des Exponenten in eine Menge der Mächtigkeit der Basis definiert.

Diese knappen Bemerkungen genügen hoffentlich, weite Strecken des folgenden historischen Teils für jeden Leser zugänglich zu machen.

### 3 Abriß der Geschichte des Multiplikationssatzes

Georg Cantor hat das Konzept der Mächtigkeit von unendlichen Mengen in den siebziger Jahren des 19. Jahrhunderts entwickelt und untersucht. Zu dieser Zeit bewies er auch den Multiplikationssatz für abzählbare Mengen und für Mengen der Mächtigkeit der reellen Zahlen. Der erste Beweis des allgemeinen Satzes erschien 1905 in einer furiosen, aber kaum bekannten Arbeit von A. E. Harward, „Indian Civil Servantin Calcutta“. Unabhängig hiervon erschien 1906 Hessenbergs erster Beweis. Ein Jahr später veröffentlichte Hessenberg dann noch einen zweiten Beweis des Satzes, seiner Natur nach wie der erste arithmetisch, abervon ihm doch wesentlich verschieden. Ein dritter Beweis wurde 1908 von Philip Jourdain geführt, und eine Vereinfachung dieses Beweises, die sich ebenfalls schon bei Harward 1905 findet, fand dann Eingang in das Buch von Felix Hausdorff von 1914, und dadurch weite Verbreitung. Auch der heute übliche Beweis folgt der Harward-Hausdorffschen „zweidimensionalen Argumentationslinie“, die als direkteste Verallgemeinerung der Cantorschen Diagonalaufzählung von  $\mathbb{N}^2$  angesehen werden kann. Hessenbergs Beweise dagegen sind in Vergessenheit geraten – sicher zu unrecht, zumal besonders Hessenbergs erster Beweis direkt auf Cantors Arbeiten aufbaut, und dadurch noch den unverwechselbaren Glanz der erwachenden Mengenlehre an sich trägt. Schließlich gab Max Zorn 1944 noch einmal einen ganz anderen Beweis mit Hilfe des Zornschen Lemmas, der Wohlordnungen ganz vermeidet. Auch dieser Beweis ist heute großflächig vergessen.

### 4 Cantors Paarungsfunktion und die Mächtigkeit der Ebene (1878)

In einem Brief an Richard Dedekind vom 29.11.1873 stellt Cantor die Frage, ob sich die natürlichen Zahlen bijektiv auf die reellen Zahlen abbilden lassen. Er bemerkt, daß sich eine Bijektion zwischen den natürlichen Zahlen und den (positiven) rationalen Zahlen leicht angeben läßt. Die Gleichung  $|\mathbb{N} \times \mathbb{N}| = |\mathbb{N}|$  war ihm zu diesem Zeitpunkt bereits völlig klar. Wenige Tage später kann Cantor die von ihm gestellte Frage selbst beantworten. In der 1874 erschienenen Arbeit „Über eine Eigenschaft des Inbegriffes aller reellen algebraischen Zahlen“ zeigt Cantor dann der mathematischen Welt die Abzählbarkeit der algebraischen Zahlen und die Überabzählbarkeit der reellen Zahlen. Die Arbeit ist von Cantor mit „Berlin, den 23. Dezember 1873“ unterzeichnet, doch was hier als hübsch verpacktes Weihnachtsgeschenk für die Kollegen daherkommt entpuppt sich schnell als Startschuß für eine aufregende Epoche des fundamental Neuen und der neuen Fundamente. Insbesondere fließt aus Cantors Ergebnis so ganz nebenbei die Existenz transzendornter Zahlen, ein großes Ergebnis von Liouville aus dem Jahre 1847.

In „Ein Beitrag zur Mannigfaltigkeitslehre“ von 1878 gibt Cantor, eher beiläufig, die bekannte Diagonalaufzählung von  $\mathbb{N}^2$  konkret an: Das einfache Polynom zweiten Grades

$$f(n, m) = m + (m + n)(m + n + 1)/2$$

bildet die Menge aller Paare von natürlichen Zahlen bijektiv auf natürlichen Zahlen ab – und ist als Polynom zweiten Grades in dieser Hinsicht einzigartig. Das Hauptresultat der Arbeit von 1878 ist jedoch, daß  $|\mathbb{R}^n| = |\mathbb{R}|$  gilt für alle natürlichen Zahlen  $n \geq 1$ . Das Problem hatte Cantor bereits Jahre früher in einem Brief an Dedekind vom 5.1.1874 aufgeworfen. In dem Brief heißt es:

„Läßt sich eine Fläche (etwa ein Quadrat mit Einschluß der Begrenzung) eindeutig auf eine Linie (etwa eine gerade Strecke mit Einschluß der Endpunkte) eindeutig beziehen, so dass jedem Puncte der Fläche ein Punct der Linie und umgekehrt zu jedem Puncte der Linie ein Punct der Fläche gehört?“

Mir will es Augenblick noch scheinen, dass die Beantwortung dieser Fragen, – obgleich man auch hier zum *Nein* sich so gedrängt sieht, daß man den Beweis dazu fast für überflüssig halten möchte, – grosse Schwierigkeiten hat. – “

(Diese und die weiteren zitierten Briefstellen finden sich in [Cantor 1991].)

Die Lösung ließ diesmal nicht Tage, sondern Jahre auf sich warten. Cantor teilte einen Beweis seines kontraintuitiven Resultates Dedekind brieflich am 20.6.1877 mit – in der Tat ist die Frage mit einem *Ja* zu beantworten. Er verwendet das „Reißverschlußverfahren“, um zwei reelle Zahlen des Einheitsintervalls, die in unendlicher Dezimaldarstellung vorliegen, zu einer neuen reellen Zahl zu verschmelzen: Die Nachkommastellen der beiden Zahlen werden abwechselnd aneinandergereiht, aus 0,1223... und 0,9267... wird etwa 0,19222637... Dedekind antwortet ihm, daß die entstehende Abbildung nicht surjektiv ist – Cantors Beweis zeigt nur  $|\mathbb{R}^2| \leq |\mathbb{R}|$ . Die Zahl 0,11101010... etwa liegt nicht im Bild der Funktion. Cantor bemerkt, daß dieser Einwand nicht das Herz der Sache trafe: Sein Beweis liefert eine Bijektion zwischen  $\mathbb{R}^2$  und einer Teilmenge von  $\mathbb{R}$ , und das scheint ja irgendwie noch mehr zu sein, als er wollte. Dennoch sucht er nach einem Beweis von  $|\mathbb{R}^2| = |\mathbb{R}|$ . Der Satz von Cantor-Bernstein stand ihm damals noch nicht zur Verfügung, und Cantor verwickelt sich in unangenehme technische Probleme. Die gewünschte Bijektion wird sehr aufwendig konstruiert, an die Stelle der Dezimalbruchentwicklungen treten Kettenbrüche, und Cantor argumentiert umständlich, daß die irrationalen Zahlen gleichmächtig zu den reellen Zahlen sind. Und obwohl er schließlich ein einfaches Argument hierfür sieht und auch angibt [Cantor 1878, § 6], will er in der Veröffentlichung auf seine kompliziertere Konstruktion nicht verzichten. Die Arbeit ist, ganz abgesehen von dem überraschenden Resultat, in vielerlei Hinsicht interessant: Sie ist Photographie des Nebels, der über neuen Begriffen in ihren Morgenstunden liegt, und zugleich ein Dokument der Psyche von Wissenschaftlern, denen allzu einfache Lösungen oft gar nicht so gelegen kommen, da diese vorangehende Kraftakte überflüssig machen.

Nachdem Cantor die Gleichung  $|\mathbb{R}^2| = |\mathbb{R}|$  bewiesen und sein Vorgehen diskutiert hat, beendet er die Arbeit mit der Kontinuumshypothese: Jede überabzählbare Menge von reellen Zahlen ist gleichmächtig mit  $\mathbb{R}$ . Möglicherweise war der entdeckte Zusammenfall der beiden vermeintlich verschiedenen Mächtigkeiten  $|\mathbb{R}|$  und  $|\mathbb{R}^2|$  der Auslöser für die Hypothese, die ja auch den Zusammenfall vieler Mächtigkeiten vermutet.

Die Gleichmächtigkeit der Ebene  $\mathbb{R}^2$  mir der Geraden  $\mathbb{R}$  war für die damalige Zeit eine große Überraschung, wobei schon die Fragestellung ungewöhnlich genug erschien. Cantor schreibt in einem Brief an Dedekind vom 25.6.1877:

„Die meisten, welchen ich diese Frage [nach der Gleichmächtigkeit von  $\mathbb{R}^n$  mit  $\mathbb{R}$ ] vorgelegt, wunderten sich sehr darüber, dass ich sie habe stellen können, da es sich ja von selbst verstände, dass zur Bestimmung eines Punctes in einer Ausgedehntheit von  $p$  Dimensionen immer  $p$  unabhängige Coordinaten gebraucht werden. Wer jedoch in den Sinn der Frage eindrang musste bekennen, dass es mindestens eines Beweises bedürfe, warum sie mit dem ‚selbstverständlichen‘ nein zu beantworten sei. Wie gesagt gehörte ich selbst zu denen, welche es für das Wahrscheinlichste hielten, dass jene Frage mit einem Nein zu beantworten sei, – bis ich vor ganz kurzer Zeit durch ziemlich verwinkelte Gedankenreihen zu der Überzeugung gelangte, dass jene Frage ohne all Einschränkung zu bejahen ist. Bald darauffand ich den Beweis, welchen Sie heute vor sich sehen.“

Gemeint ist der komplizierte Beweis von der Arbeit von 1878. Es gibt einen einfachen Beweis von  $|\mathbb{R}^2| = |\mathbb{R}|$  mit Hilfe einer modifizierten Reißverschlußtechnik, der den Satz von Cantor-Bernstein nicht benötigt, und als Trick von Julius König bekannt ist. Cantor hat dieses einfache Argument übersehen. Die Idee ist, das Reißverschlußverfahren mit sog. Ziffernblöcken anstelle von einzelnen Ziffern durchzuführen; ist  $a \neq 0$  eine Ziffer einer reellen Zahl in unendlicher Dezimaldarstellung, so bildet  $a$  zusammen mit allen vorangehenden Nullen einen Block der Zahl, der dann also die Form  $00..0a$  hat. Die ersten Blöcke von  $0,102002304\dots$  sind etwa  $1, 02, 002, 3, 04$ . Werden nun je zwei reelle Zahlen des offenen Einheitsintervalls I durch Verzahnung von Blöcken anstatt von Einzelziffern amalgamiert, so entsteht eine bijektive Abbildung von I  $\times$  I nach I.

Die erste dem Autor bekannte Referenz auf die Beobachtung von König ist die im Buch von Schoenflies [1900, p. 23]. Dort heißt es: „... und zweitens denke man sich die eventuellen Nullen mit der ersten auf sie folgenden Ziffer [ungleich 0] zu je einer Gruppe verbunden, und dehne das Abbildungsgesetz [das Reißverschlußverfahren] auf diese Zahlengruppen aus“<sup>1)</sup>. Die Anmerkung 1 hierzu ist dann: „1) Dieser Gedanke röhrt von J. Königher.“ Auch in [Fraenkel 1928] – vielfach ein historisches Miniaturenmuseum – findet sich lediglich die Bemerkung: „Durch diesen von J. König stammenden Kunstgriff wird also der vorstehende Beweis lückenlos“[eb., p. 100].

In seinen Antwortschreiben wies Dedekind auf die Unstetigkeit der konstruierten Bijektionen hin, undwarf damit neue Fragen auf. Giuseppe Peano gab dann 1890 eine stetige Surjektion von der Geraden auf die Ebene an. Die Frage, ob die Injektivität einer solchen Abbildung notwendig verletzt sein müsse, blieb offen. Erst 1911 gelang Brouwer der vollständige Beweis des Satzes, daß es keine stetigen Bijektionen zwischen verschiedenendimensionalen Kontinua geben kann.

Aus heutiger Sicht – oder genauer seit Hausdorffs Einführung der allgemeinen Topologie 1914 [vgl. speziell hierzu Hausdorff 1914, p. 377f] – ist das Resultat für  $\mathbb{R}$  und  $\mathbb{R}^2$  relativ einfach zu beweisen: Die Ebene bleibt nach Entfernung eines Punktes zusammenhängend, die Gerade wird dagegen durch Entfernung eines Punktes unzusammenhängend. Es folgt, daß es keine stetige Bijektion zwischen  $\mathbb{R}^2$  und  $\mathbb{R}$  geben kann. Wegen der in diesem Fall automatischen Stetigkeit der Umkehrabbildung kann es dann weiter auch keine stetige Bijektion zwischen  $\mathbb{R}$  und  $\mathbb{R}^2$  geben.

## 5 Cantors Kalkül der Kardinalzahlarithmetik (1895)

Den nächsten größeren Fortschritt in der Geschichte des Problems bildet die Entdeckung des algebraischen Kalküls der Kardinalzahlarithmetik durch Cantor. Ein Beweis von  $|\mathbb{R}|^2 = |\mathbb{R}|$  liest sich darin dann einfach so:

$$|\mathbb{R}|^2 = (2^{\aleph_0})^2 = 2^{2\cdot\aleph_0} = 2^{\aleph_0} = |\mathbb{R}|.$$

Cantor führt den Kalkül im ersten Teil seiner „Beiträge zur Begründung der transfiniten Mengenlehre“ von 1895 ein, und notiert begeistert, daß sich die Beweise der Arbeit von 1878 nun auf „wenige Striche“ verkürzen. Der Kalkül selbst wurzelt in Cantors Entdeckung des Diagonalverfahrens für Abbildungen einer Menge  $M$  in die Menge  $\{0, 1\}$ . (Cantor trug das Diagonalverfahren auf der ersten Jahrestagung der DMV 1891 vor.) Solche 0-1-Belegungen suggerieren eine Potenzierung für Kardinalzahlen, etwa  $2^{|M|}$ , und die vertrauten und leicht zu beweisenden Rechengesetze, wie etwa  $(2^{|M|})^{|N|} = 2^{|M| \times |N|}$  liefern nützliche, zuvor nur durch mühsame Manipulation von Bijektionen zu gewinnende Ergebnisse. Entscheidend ist zudem die Gleichung  $|\mathbb{R}| = |\{f \mid f: \mathbb{N} \rightarrow \{0, 1\}\}|$ , die die Brücke zwischen  $|\mathbb{R}|$  und  $2^{\aleph_0}$  schlägt.

## 6 Zermelo über die Addition von Kardinalzahlen (1901)

Zermelo widmet seine erste mengentheoretische Arbeit der Untersuchung der Addition von unendlichen Kardinalzahlen. Er zeigt (hier in kardinalzahlfreier Notation wiedergegeben): Gilt  $|M| = |M + N_n|$  für Mengen  $M, N_n, n \in \mathbb{N}$ , so gilt  $|M| = |M + \bigcup_{n \in \mathbb{N}} N_n|$ . Insbesondere folgt hieraus: Gilt  $|M| = |M + N|$  für zwei Mengen  $M$  und  $N$ , so gilt auch die Gleichung  $|M| = |M + (\mathbb{N} \times N)|$ . Die Argumentation erweist sich interessanterweise als eine Verallgemeinerung des Bernsteinschen Beweises des Satzes von Cantor-Bernstein, und dieser Punkt scheint ein Hauptanliegen von Zermelo gewesen zu sein. Eine allgemeine Additions- oder Multiplikationshypothese stellt Zermelo in seinem Artikel nicht auf, und auch ein Additions- oder Multiplikationssatz für wohlordnbare Mengen wird nicht diskutiert.

Zermelo spielt in der Geschichte des Satzes noch in zweierlei Hinsicht eine Rolle: Zum einen beweist er 1904 den Wohlordnungssatz, der den Multiplikationssatz auf das Problem reduziert,  $|M^2| = |M|$  für wohlgeordnete Mengen  $M$  zu beweisen. Alle Beweise mit Ausnahme des Beweises von Zorn folgen dieser Reduktionsmöglichkeit. Für den Beweis von  $|\mathbb{N}^2| = |\mathbb{N}|$  durch diagonale Aufzählung ist die Ordnung von  $\mathbb{N}$  wesentlich, und ebenso hilft eine einer unendlichen Menge  $M$  zugrundeliegende Wohlordnung für einen Beweis von  $|M^2| = |M|$ . Zum anderen hat Zermelo die Lückenhaftigkeit der Argumentation einer frühen Arbeit von Jourdain betont [Jourdain 1904b], in der Jourdain zum mindesten den Spezialfall des Multiplikationssatzes für Mengen der kleinsten überabzählbaren Mächtigkeit beweisen wollte, also die Gleichung  $\aleph_1 \cdot \aleph_1 = \aleph_1$ . In seiner Arbeit von 1908 weist Jourdain auf Zermelos berechtigte frühere Kritik hin, und dankt ihm weiter für kritische Kommentare zur vorliegenden Arbeit: „I must here refer gratefully to the trouble which Prof. Zermelo has taken in repeatedly critising weak points in my proofs and suggesting improvements.“ [Jourdain 1908, p. 512]. Zermelo brachte also

dem Problem kontinuierliches lebhaftes Interesse entgegen, und er unterstützte in seiner bekannten kritischen Art Jourdain auf seinem langen und holprigen Weg zu einem korrekten Beweis.

Ein dritter Punkt betreffend Zermelos Anteil und Teilnahme am Multiplikationsproblem ist die Ankündigung, daß Zermelo einen andersartigen Beweis des Satzes gefunden habe und ihn demnächst veröffentlichen werde. Die Ankündigung findet sich im Buch von Hessenberg 1906 und wird in der Arbeit von Jourdain 1908 referiert, aber nicht weiter erläutert. Obwohl Zermelo also Jourdains Beweise zum Problem prüfte und kritisierte, scheint er ihm seinen eigenen Beweis nicht mitgeteilt zu haben. Wir kommen unten auf die Bemerkung bei Hessenberg noch kurz zurück.

## 7 Die Hypothek der Dissertation von Bernstein (1901)

Zur Multiplikation noch größerer Mengen als der Menge der reellen Zahlen hat Cantor nichts veröffentlicht und auch in seinen Briefen äußert er sich hierzu nicht. Aus seinem algebraischen Kalkül folgen aber unmittelbar Gleichungen wie  $|P(\mathbb{R})^2| = |P(\mathbb{R})|$ , und damit gilt der Multiplikationssatz auch für die Menge aller reellen Funktionen, also für  $F = \{f \mid f : \mathbb{R} \rightarrow \mathbb{R}\}$ .

Nach Aussagen von Felix Bernstein hat Cantor aber einen Beweis des Multiplikationssatzes für wohlordenbare Mengen gesehen, und ihm diesen mündlich mitgeteilt. In seiner Dissertation von 1901 verwendet Bernstein den Multiplikationssatz für wohlordenbare Mengen als Hilfsatz, und schreibt zum Beweis lakonisch:

„Den Beweis des Satzes, den ich aus mündlicher Mitteilung von G. Cantor kenne, führt man analog wie im einfachsten Falle ... [für  $\mathbb{N}$ ] durch Verwandlung einer Doppelreihe in eine einfache Reihe [d. h. man verwandelt eine doppelt indizierte Folge in eine gewöhnliche Folge] ...“ [Bernstein 1905, § 12].

Die Dissertation von Bernstein wurde 1905 in den Mathematischen Annalen veröffentlicht, als eine „bis auf einige Verbesserungen und Bemerkungen ... unveränderte Wiedergabe [der Dissertation]“. Zitiert wird hier nach dieser leichter zugänglichen Arbeit.

Bernstein beläßt es also in seiner Doktorarbeit bei einem argumentum ad au-toritatem, anstatt den Cantorschen Beweis durchzuführen. Es ist gut vorstellbar, daß der junge Bernstein das von Cantor wahrscheinlich nur skizzierte Argument in schweigender Bewunderung zur Kenntnis genommen hatte und es später nicht genau rekonstruieren konnte: Die Details einer Aufzählung von  $M \times M$  bei „wohlgeordneten Achsen“ sind nicht ganz trivial, „diagonal“ im einfachen Sinne wie für  $\mathbb{N} \times \mathbb{N}$  kann man sie nicht so ohne weiteres durchführen. Bemerkenswerterweise hat nun ein strenger Beweis des Multiplikationssatzes für Wohlordnungen durch eine Art Diagonalaufzählung bis zur Arbeit von Jourdain 1908 auf sich warten lassen, obwohl für einen solchen Beweis der Zermelosche Wohlordnungssatz nicht gebraucht wird, und obwohl die Beweisstrategie durch Bernsteins Bemerkung vorgezeichnet war. Hessenbergs Beweise von 1906 und 1907 kann man nicht als eine direkte Verallgemeinerung der Diagonalaufzählung von  $\mathbb{N}^2$  ansehen. Und auch Jourdains Beweis verschleiert die Dinge noch unnötig, erst die Konstruktion von Hausdorff 1914 machte den Weg frei für die aus heutiger Sicht auf der Straße liegende Verallgemeinerung der Cantorschen Paarungsfunktion auf kartesische Produkte mit Achsen beliebiger Länge.

Arthur Schoenflies schreibt im zweiten Teil seines Berichts über den Stand der Mengenlehre 1908:

„Es liegt zunächst nahe den Beweis [des Multiplikationssatzes für Wohlordnungen] in ähnlicher Wiese zu führen wie den [Cantorschen Beweis von  $|\mathbb{N} \times \mathbb{N}| = |\mathbb{N}|$ ] durch Diagonalaufzählung... Doch bedarf der Begriff der diagonalen Anordnung [für allgemeine Wohlordnungen] ... einer präzisen Erläuterung<sup>1)</sup> ... [Fußnote 1]: In obiger Form erscheint der Beweis z. B. in Bernsteins Dissertation ... Eine ausführliche Darstellung des Beweises liegt nicht vor.“ [Schoenflies 1908, p. 13]

Der Bericht von Schoenflies ist mit „Königsberg i. Pr., im Oktober 1907“ unterzeichnet.

Der Multiplikationssatz für beliebige Mengen spielt weiter in der Bernsteinschen Dissertation eine wichtige Rolle. Er erscheint dort als eine Voraussetzung, aus der sich Vergleichbarkeitsresultate gewinnen lassen. Bernstein zeigt: Gilt für ein unter Addition abgeschlossenes System S von Mengen der Multiplikationssatz, d. h.  $|M^2| = |M|$  für alle  $M \in S$ , so sind die Elemente des Systems bzgl. ihrer Mächtigkeit vergleichbar, d. h. es gilt  $|M| \leq |N|$  oder  $|N| \leq |M|$  für alle  $M, N \in S$  [Bernstein 1905, § 4]. Bernstein verwendet im Beweis implizit das Auswahlaxiom. Dennoch bildet sein trickreiches Argument dann das Herzstück im Beweis des Satzes von Bernstein-Tarski [Tarski 1924], daß nämlich der Multiplikationssatz zusammen mit den restlichen Axiomen der Mengenlehre das Auswahlaxiom impliziert.

## 8 Die Hausdorff-Formel und eine Regularitätsbehauptung (1904)

Im Zusammenhang mit Bernsteins Rückgriff auf einen mündlichen Beweis des Multiplikationssatzes für Wohlordnungen ist weiter die erste mengentheoretische Arbeit von Felix Hausdorff „Der Potenzbegriff in der Mengenlehre“ aus dem Jahr 1904 interessant. In dieser skizzenhaften Arbeit mit dem Untertitel „aus dem Sprechsaal“ behauptet Hausdorff die sog. Regularität von Nachfolger-Alephs. Diese Behauptung ist, wie schnell zu zeigen ist, mit dem Multiplikationssatz für Wohlordnungen gleichwertig. (Zum Beweis der Regularität muß zusätzlich zum Multiplikationssatz für Wohlordnungen auch das Auswahlaxiom verwendet werden.) Hausdorff braucht die Regularität zum Beweis seiner „Hausdorff-Formel“ der Kardinalzahlarithmetik. Sie stellt eine falsche Behauptung aus der Dissertation von Bernstein richtig, die ihrerseits zu dem falschen Beweis der Nichtwohlordenbarkeit des Kontinuums geführt hat, den Julius König 1904 auf dem Heidelberger Mathematikerkongress zur großen Aufregung seiner Hörer – unter ihnen Cantor – vortrug. Hausdorff scheint hier, wie Bernstein in seiner Dissertation, den Multiplikationssatz für Wohlordnungen aus dem Cantorschen Mund zu übernehmen.

Hausdorff reicht keinen vollständigen Beweis der Hausdorff-Formel, d. h. kein Argument für die nur behauptete Regularität von Nachfolger-Alephs nach. In [Hausdorff 1908] gibt er ohne Beweis und ohne Referenz „die Alefgleichung  $\aleph_\alpha^2 = \aleph_\alpha$ “ als Begründung für die Regularität von  $\aleph_{\alpha+1}$  an. In seinem Buch von 1914 wird die Regularitätsbehauptung dann unmittelbar nach einem ausführlichen Beweis der Alefgleichung – gleichwertig: des Multiplikationssatzes für Wohlordnungen – bewiesen [Hausdorff 1914, p. 129]. Dort wird zudem in einer Anmerkung ausdrücklich Hessenberg als derjenige genannt, der den Multiplikationssatz 1906 zuerst bewiesen hatte. Hausdorffs eigene Arbeit von 1904 bleibt unerwähnt.

Im Jahr 1904 zeigt Zermelo den Wohlordnungssatz, und damit wird der Multiplikationssatz für Wohlordnungen gleichwertig zum vollen Multiplikationssatz für alle Mengen. Die nackt im Raum stehende Verallgemeinerung von Cantors Gleichung  $|\mathbb{N}|^2 = |\mathbb{N}|$  auf alle wohlgeordneten unendlichen Mengen – und mit Zermelos starkem Resultat damit dann sogar auf alle unendlichen Mengen – sehnte sich nun in zunehmendem Maße nach einem richtigen mathematischen Gewand, oder etwas weniger bildhaft: einem lupenreinen Argument.

Bemerkenswert am Duo [Zermelo 1901] und [Hausdorff 1904] ist, daß elementare Fragen der Kardinalzahlarithmetik in den mengentheoretischen Erstlingen der beiden wichtigsten Erforscher der Mengenlehre nach Cantor eine zentrale Position einnehmen.

## 9 Jourdains Versuch (1904) und Harwards Beweis (1905)

Jourdain hat in einer im „Philosophical Magazine“ 1904 veröffentlichten Arbeit versucht, den Multiplikationssatz für Wohlordnungen zu beweisen [Jourdain 1904b, p. 298–300]. Er beschreibt aber letztendlich das Problem nur, und beweist gar nichts (vgl. auch Abschnitt 7). A. E. Harward war es dann, ein mathematischer Außenseiter mit bislang unaufgelösten Initialen, der, von Indien aus agierend und auf der vergleichsweise recht dünnen Wissensgrundlage von Russells „Principles of Mathematics“ von 1903 und Jourdains Artikelpaar im „Philosophical Magazine“ von 1904 stehend, einen ersten vollständigen Beweis des Multiplikationssatzes für Wohlordnungen geben konnte. Harward verwendet den Wohlordnungssatz ähnlich wie Cantor als „Quasi-axiom“, und erhält so den vollen Multiplikationssatz.

Harwards Arbeit „On the transfinite numbers“ von 1905 ist eine klargestriebene Einführung in eine informal-axiomatische Mengenlehre. Harward formuliert ein Ersetzungssaxiom, und begründet wie Cantor in seinen späten Briefen den Wohlordnungssatz durch Abzählen einer Menge entlang der Ordinalzahlen: Dieses muß irgendwann enden, da sonst die Menge gleichmächtig zur echten Klasse der Ordinalzahlen wäre. Weiter zeichnet sich die Arbeit durch eine große Sicherheit und Originalität in der Ordinalzahlarithmetik aus, und der Beweis des Multiplikationssatzes für Wohlordnungen ruht auf feinen arithmetischen Beobachtungen. Das Argument verwendet allerdings versteckt das Auswahlaxiom. Diese Schwäche lässt Harwards Beweis des Multiplikationsatzes für Wohlordnungen aus heutiger Sicht zweitrangig erscheinen. Doch:

In einem Anhang, innerhalb einer Kritik des Arguments von [Jourdain 1904b], skizziert Harward noch einen zweiten Beweis, und schlägt hier genau die Aufzählung der Hälfte eines diagonal halbierten Quadrats über einer wohlgeordneten Achse vor, die Hausdorff 1914 als eine Vereinfachung der Jourdain-schen Konstruktion von 1908 wiederentdecken wird. Harward schreibt über seine beiden Beweise im Anhang seiner Arbeit:

„In order to complete the proof on the lines indicated by Mr. Jourdain, it is necessary that some rule or formula [analog zur Cantorschen Paarungsfunktion für  $\mathbb{N}^2$ ] should be given by which the required correlation can be established once for all. As I could not succeed in constructing such a formula, I adopted a different method of proof [gemeint ist der im Hauptteil der Arbeit dargestellte Beweis].

I have recently discovered that by a slight modification of Mr. Jourdain's method a simple an-digorous proof can be obtained...“ [Harward 1905, p. 458, in „Note A“ zu seinem Artikel].

Wir diskutieren die Konstruktion von Harward-Hausdorff in Abschnitt 12. Harward unterscheidet allerdings zeittypisch nicht genau zwischen dem ungeordneten Paar  $\{\alpha, \beta\}$  und dem geordneten Paar  $(\alpha, \beta)$ , was aber der Sache und der klaren Skizze der Konstruktion und Beweisidee keinen Abbruch tut.

Es ergibt sich ein bemerkenswerter Zirkel: Jourdain, der Harwards Arbeit mit dem Autor diskutiert hat [vgl. Harward 1905, p. 439], kennt Harwards zweiten Beweis, stellt aber 1908 seinen eigenen endlich richtigen Beweis ins Rampenlicht, den dann Hausdorff 1914, ohne Kenntnis des Artikels von Harward, zu dem zweiten Harwardschen Beweis „resimplified“! Jourdain hat in seinem Artikel von 1908 sehr unsauber auf Harwards Leistung hingewiesen, er diskutiert den Artikel zusammen mit Details seiner eigenen früheren Arbeiten, anstelle klar anzugeben, daß und wie Harward den Beweis vor ihm geführt hat.

Gregory Moore hat 1976 in einer Notiz auf die alleine schon aus rein axiomatischer Hinsicht bemerkenswerte Arbeit von Harward hingewiesen, und ihre Vernachlässigung als ein „establishment“-Phänomen interpretiert: „Harward [1905] contained the core of what could have been a worthwhile axiomatization of set theory. Nevertheless, his paper provoked no public response, even from Jourdain who had suggested changes in it prior to publication. This silence was partly due to the fact that Harward was a self-confessed amateur vis-à-vis set theory...“ [Moore 1976].

Auch nach Moores (nicht gerade an mathematisch auffälliger Stelle veröffentlichtem) Hinweis blieb Harwards Beitrag zur Fundierung der Mathematik oder zum Multiplikationsproblem in (fast?) allen alten wie neueren mathematischen wie historischen Texten zur Mengenlehre vollkommen unbeachtet ([Deiser 2004] leider eingeschlossen, in guter Gesellschaft mit z. B. [Hausdorff 2002, p. 33 und p. 597f]). Die folgende Geschichte des Satzes liest sich dann auch, als hätte es Harwards Artikel von 1905 gar nicht gegeben.

## 10 Hessenbergs erster Beweis des Satzes (1906)

Unabhängig von Harward gelang Hessenberg ein Beweis des allgemeinen Additions- und Multiplikationssatzes durch ein einfaches, hübsches und heute fast vergessenes Argument, das die sogenannte Cantorsche Normalform [Cantor 1897, § 17] für Ordinalzahlen verwendet. Der Beweis erstreckt sich über die §§ 75–77 der „Grundbegriffe“ [Hessenberg 1906].

Die Cantorsche Normalform ist, cum grano salis, die  $\mathbb{N}$ -adische Darstellung transfiniter Zahlen. Ganz so, wie sich natürliche Zahlen eindeutig in der Form

$$10^{n_1} \cdot a_1 + 10^{n_2} \cdot a_2 + \dots + 10^{n_k} \cdot a_k$$

schreiben lassen mit natürlichen Zahlen  $n_1 > n_2 > \dots > n_k > 0$ ,  $1 \leq a_i \leq 9$ , lassen sich Ordinalzahlen eindeutig in der Form

$$\mathbb{N}^{a_1} \cdot a_1 + \mathbb{N}^{a_2} \cdot a_2 + \dots + \mathbb{N}^{a_k} \cdot a_k$$

schreiben mit (endlich vielen!) Ordinalzahlen  $\alpha_1 > \alpha_2 > \dots > \alpha_k \geq 0$ , und Koeffizienten  $a_i \in \mathbb{N} - \{0\}$ . Hierbei ist  $\mathbb{N}^\alpha = \omega^\alpha$  die von Cantor definierte Ordinalzahlexponentiation zur Basis  $\mathbb{N}$ .

Hessenberg bringt zwei Ordinalzahlen in ihre Cantorsche Normalform, und definiert dann die von ihm sogetaufte *natürliche Summe* der beiden Zahlen wie folgt. Er addiert zuerst paarweise die in den Normaldarstellungen auftretenden Glieder mit gleichen Exponenten. Diese Summen haben also die Form  $\mathbb{N}^{\alpha_i} \cdot a_i$  oder  $\mathbb{N}^{\alpha_i} \cdot (a_i + b_i)$  oder  $\mathbb{N}^{\alpha_i} \cdot (a_i + b_i)$  oder  $\mathbb{N}^{\beta_i} \cdot b_i$ . Anschließend summiert er alle diese Einzeladditionen zu einer neuen Normalform auf (d. h. die größeren Glieder kommen zuerst). So verschmelzen zwei Ordinalzahlen zu einer neuen. Betrachtet man die Operation, so sieht man schnell, daß immer nur endlich viele Paare dasselbe Ergebnis dieser Verschmelzung hinterlassen. Der Prozeß ist also „fast“ injektiv. Zudem liegt die natürliche Summe recht nahe am Maximum der beiden Summanden. Hessenberg gewinnt aus diesen Beobachtungen dann relativ leicht den Additions- und Multiplikationssatz für Wohlordnungen, und mit Hilfe des Zermeloschen Wohlordnungssatzes folgen dann die uneingeschränkten Versionen.

Hessenberg hat Bernsteins Andeutungen über einen Cantorschen Beweis des Satzes nicht unkommentiert gelassen. Sowohl im Vorwort seines Buches als auch am Ende des Beweises des Multiplikationssatzes spricht er von einer „Mitteilung“ von Felix Bernstein, derzu folge das Resultat bereits von Cantor bewiesen worden war [Hessenberg 1906, p. IV und Ende § 77]. Hierbei ist nicht klar, ob sich Hessenberg lediglich auf die Bernsteinsche Dissertation bezieht oder ob es eine zusätzliche Korrespondenz mit Bernstein hierüber gab. (Ersteres erscheint wahrscheinlicher.) Im Vorwort schreibt er über diese Mitteilung von Felix Bernstein:

„Ob der in dieser Mitteilung flüchtig skizzierte Beweis derselbe ist, den ich hier [in §§ 75–77 dieser Abhandlung] darstelle, vermag ich nicht zu beurteilen.“

Daß die Beweisidee dieselbe ist, erscheint nicht unmöglich, da alle Zutaten des Hessenbergschen Beweises Cantorsche Eigengewächse sind, allen voran die Normaldarstellung transfiniter Zahlen. Wahrscheinlicher ist aber, daß Cantor einen Beweis gesehen hatte, der enger mit der Argumentation von Harward und Jourdain verwandt ist.

Weiter spricht Hessenberg von einem „wesentlich verschiedenen“ Beweis des Satzes, der ihm „in jüngster Zeit“ von Ernst Zermelo mitgeteilt worden sei, und demnächst veröffentlicht werde [Hessenberg 1906, Ende § 77]. Wie Cantors Beweis hat auch Zermelos Beweis nie das Licht der Welt gesehen. Im Hinblick auf die Arbeiten und Argumente von Zermelo aus dieser Zeit ist der Zermelosche Beweis vermutlich dem Zornschen Beweis des Satzes ähnlicher als den Beweisen von Hessenberg und Jourdain.

## 11 Hessenbergs zweiter Beweis des Satzes

Hessenberg hat 1907 noch einen zweiten arithmetischen Beweis des Multiplikationssatzes gegeben, der die Cantorsche Normalform nicht verwendet, dafür aber von der feinen Dynamik der Ordinalzahlexponentiation in anderer Weise gebraucht macht. Die Exponentiation wurde von Cantor 1897 [Cantor 1897, § 18] eingeführt, und hat Ruhm erlangt als erstes Beispiel einer Definition durch transfinite Rekursion. Hessenbergs Beweis benutzt jedoch eine äquivalente rekursionsfreie Definition der Exponentiation, die auf Hausdorff zurückgeht. Modulo dieser Definition ist der zweite Hessenbergsche Beweis dann sehr einfach zu führen.

Ziel ist es zu zeigen, daß unendliche Kardinalzahlen abgeschlossen unter Ordinalzahlmultiplikation sind, daß also  $\alpha \cdot \beta < \kappa$  gilt für alle  $\alpha, \beta < \kappa$ , wobei  $\kappa$  eine unendliche Kardinalzahl ist. Hierfür verwendet Hessenberg die Abschätzung  $\alpha \cdot \beta \leq 2^\alpha \cdot 2^\beta = 2^{\alpha+\beta} \leq 2^\kappa$ . (Daß  $\alpha + \beta < \kappa$  gilt, ist einfach zu sehen.) Die Behauptung folgt nun einfach daraus, daß für unendliche Kardinalzahlen  $2^\kappa = \kappa$  gilt. ( $2^\kappa$  ist hier die Ordinalzahlexponentiation, nicht die identisch notierte Kardinalzahlexponentiation, für die immer  $2^\kappa > \kappa$  gilt.) Für den Beweis dieser Gleichung verwendet Hessenberg die rekursionsfreie Hausdorff-Hessenberg-Darstellung der Ordinalzahl-Exponentiation: Er braucht, daß die Mächtigkeit des Ergebnisses der Exponentiation  $2^\alpha$  nur von der Mächtigkeit von  $\alpha$  abhängt, was aus der rekursiven Definition von  $2^\alpha$  im Gegensatz zur Hausdorff-Hessenberg-Darstellung nicht unmittelbar hervorgeht.

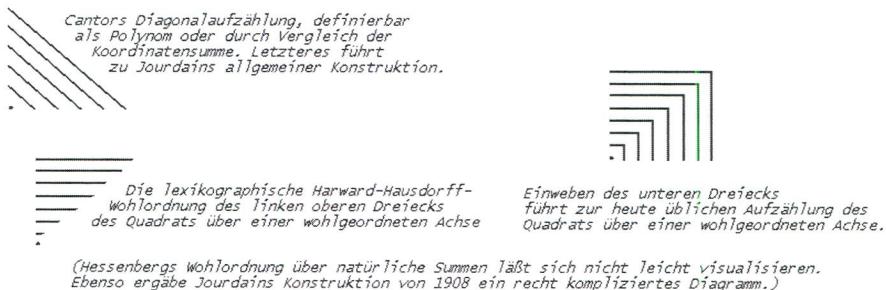
Schoenflies skizziert die beiden Hessenbergschen Beweise von 1906 und 1907 im zweiten Teil seines „Berichts“ [Schoenflies 1908, p. 13 f.]. In seiner Neufassung des ersten Teils von 1913 findet sich dann nur noch der erste Beweis [Schoenflies 1913, p. 131 ff.]. Das schöne Argument der Arbeit von 1907 gerät in Vergessenheit.

## 12 Jourdain (1908), Hausdorff (1914), und der heute übliche Beweis

Zwei Jahre nach Hessenberg veröffentlicht Philip Jourdain einen weiteren Beweis des Multiplikationssatzes. Er bildet die Grundlage für Hausdorffs Beweis in den „Grundzügen der Mengenlehre“ von 1914, der, wie erwähnt, mit Harwards zweitem Beweis von 1905 zusammenfällt. Hausdorff bezeichnet Hessenbergs Beweis von 1906 in einer Anmerkung als „umständlich“ [Hausdorff 1914, p. 456], was auf die Darstellung zutrifft, aber nicht auf den Beweis selber. Der heute übliche Beweis des Satzes folgt der Harvard-Hausdorffschen Argumentation, und unterscheidet sich von ihr nur durch etwas bessere Feinmechanik und läuft dadurch etwas glatter; er ist Allgemeingut und mit keinem weiteren Namen verbunden (außer daß die im Beweis implizit konstruierte Funktion manchmal als Gödelsche Paarungsfunktion bezeichnet wird). Alle Beweise verwenden wie Hessenberg Ordinalzahlen, wobei die benutzten arithmetischen Operationen im Lauf der Zeit immer einfacher werden – Jourdain verwendet noch die Addition von Ordinalzahlen, der heutige Beweis kommt mit der trivialen Operation der Maximum-Bestimmung zweier Elemente einer Wohlordnung aus. Die Kernbeobachtung ist, daß aus einer Wohlordnung einer Menge  $M$  sehr einfach eine Wohlordnung von  $M \times M$  konstruiert werden kann, und daß dies zudem in einer fortsetzbaren Weise geschehen kann: Wird die Wohlordnung von  $M$  verlängert zu einer Wohlordnung von  $M'$ , so ist die konstruierte Wohlordnung auf  $M \times M$  ein Anfangsstück der Wohlordnung auf  $M' \times M'$ . Mit Hilfe dieser Fortsetzungseigenschaft zeigt man dann, daß in vielen wichtigen Spezialfällen die Wohlordnung auf  $M \times M$  ordnungsisomorph zur Wohlordnung auf  $M$  selbst ist. Dies zeigt den Multiplikationssatz für Wohlordnungen, und wie bei Hessenberg folgt das allgemeine Ergebnis unter Heranziehung des Zermeloschen Wohlordnungssatzes. Und erst für diesen letzten Schritt wird das Auswahlaxiom verwendet.

Diese Schlüsselidee läßt sich ohne Arithmetik umsetzen, und der Beweis läßt sich daher rein mit Hilfe des Wohlordnungsbegriffs bequem führen; es ist nicht notwendig mit Ordinalzahlen zu arbeiten, nicht einmal aus Notationsgründen. Andererseits wird die zugrundeliegende Arithmetik nicht wirklich eliminiert, sondern lediglich verborgen:

Die konstruierten Wohlordnungen sind immer einfach definierbare wohlgeordnete Summen von Wohlordnungen, und damit letztendlich arithmetischer Natur.



Allen Konstruktionen der einfachen Wohlordnung auf  $M \times M$  liegt eine Variation der Cantorschen Diagonalaufzählung von  $\mathbb{N} \times \mathbb{N}$  zugrunde. Jourdain schreibt in der Einleitung seines Artikels [Jourdain 1908]: „In order to prove that  $\aleph_\gamma \cdot \aleph_\gamma = \aleph_\gamma$ , where  $\gamma$  is any (finite or transfinite) ordinal number, we shall generalise Cantor's [...] method of proving that  $\aleph_0 \cdot \aleph_0 = \aleph_0$ .“ Jourdain gibt in einer Fußnote die Cantorsche Bijektion  $f: \mathbb{N}^2 \rightarrow \mathbb{N}$  als Polynom an, jedoch kommt es ihm hier nur auf die induzierte Ordnung der Elemente von  $\mathbb{N}^2$  untereinander an: Will man  $(a, b)$  und  $(c, d) \in \mathbb{N}^2$  bzgl. der Diagonalaufzählung miteinander vergleichen, so ist es nicht nötig,  $f(a, b)$  und  $f(c, d)$  auszurechnen und diese Werte miteinander zu vergleichen. Denn es gilt:  $(a, b) < (c, d) \text{ gdw } a + b < c + d$  oder  $a + b = c + d$  und zudem  $a < c$ . Die Wohlordnung der Diagonalaufzählung kann geradezu in dieser Weise definiert werden, und diese Definition liefert dann, wie Jourdain erkannt hat, eine Wohlordnung von  $\gamma \times \gamma$  für beliebige Ordinalzahlen  $\gamma$ : Seien  $(\alpha, \beta), (\alpha', \beta') \in \gamma \times \gamma$ . Dann setzen wir:  $(\alpha, \beta) < (\alpha', \beta')$ , falls  $\alpha + \beta < \alpha' + \beta'$  oder falls  $\alpha + \beta = \alpha' + \beta'$  und  $\alpha < \alpha'$ . Jourdain zeigt nun, daß diese Wohlordnung von  $\gamma \times \gamma$  ordnungsisomorph zu  $\gamma$  ist, falls  $\gamma$  eine unendliche Kardinalzahl ist. Damit ist  $|\gamma \times \gamma| = |\gamma|$  für unendliche Kardinalzahlen  $\gamma$ . Der allgemeine Multiplikationssatz folgt dann wie bei Hessenberg mit Hilfe des Wohlordnungssatzes von Ernst Zermelo.

Jourdains Beweis beruht also auf einer direkten Verallgemeinerung eines die Cantorsche Diagonalaufzählung von  $\mathbb{N}^2$  definierenden Vergleichskriteriums. Jourdains Konstruktion besitzt allerdings die oben erwähnte Fortsetzungseigenschaft nicht in voller Allgemeinität.

Dies ist anders bei der Harward-Hausdorff Konstruktion, welche – bei Harward und Hausdorff mit Ordinalzahlen, hier in der Sprache der Wohlordnungen formuliert – aus einer Wohlordnung auf  $M$  eine Wohlordnung nicht auf  $M^2$ , sondern auf der Menge  $\{(a, b) \mid a, b \in M, a < b\}$  erzeugt, also auf „dem linken oberen Dreieck“ von  $M \times M$ : Die Elemente dieser Menge werden lexikographisch geordnet, zunächst nach ihrer zweiten Komponente, und bei gleicher zweiter Komponente nach der ersten. Das gleichmäßige Einweben der analogen Ordnung auf dem anderen Dreieck von  $M \times M$  (unter Einschluß der Diagonalen) führt zur heute üblichen „kanonischen“ Wohlordnung auf  $M \times M$ , und verkürzt den Beweis um einen oder zwei Hilfssätze; Hausdorff muß z. B. zusätzlich zeigen, daß das „obere linke Dreieck“ von  $M \times M$  gleichmächtig zur ganzen Menge  $M \times M$  ist für unendliche Mengen  $M$ .

## 13 Problembeußtsein und Reaktionen

Es scheint, daß der allgemeine Multiplikationssatz nie „offiziell“ als offenes Problem formuliert worden ist. Cantor selbst konzentrierte sich auf Mengen der Mächtigkeit von  $\mathbb{N}$  oder  $\mathbb{R}$ , und trug mit dem Kontinuumsproblem – ob nämlich jede unendliche

Teilmenge von  $\mathbb{R}$  die Mächtigkeit von  $\mathbb{N}$  oder die von  $\mathbb{R}$  selbst habe – schwer genug auf den Schultern. Aus der Bemerkung in Bernsteins Dissertation geht aber hervor, daß sich Cantor mit dem Multiplikationssatz für Wohlordnungen zumindest gedanklich auseinandersetzt hat. Auch zur Jahrhundertwende standen die Mächtigkeiten von  $\mathbb{N}$  und  $\mathbb{R}$  noch immer im Rampenlicht des Interesses. Als eine Menge noch größerer Mächtigkeit als  $\mathbb{R}$  wurde die Menge F der reellwertigen Funktionen untersucht, und allgemein zeigte Cantors Diagonalargument von 1891, daß die Menge aller Funktion einer Menge M in sich selbst immer größere Mächtigkeit hat als M selbst [siehe etwa Schoenflies 1900, p. 26]. Aber selbst eine elementar zu beweisende Gleichung wie  $|F^2| = |F|$  findet sich z. B. im Bericht von Schoenflies von 1900 nicht. Man rang noch mit relativ kleinen, erdnahen Mächtigkeiten, für die Beschäftigung mit astronomischen Größen war es noch etwas zu früh. Als große in voller Allgemeinheit noch zu bewältigende Aufgaben diskutierte man das Wohlordnungsproblem und das Problem der Vergleichbarkeit von Mengen, daneben machten aber auch die kleinen wiederum erdnahen Dinge größere Schwierigkeiten, etwa die Frage, ob eine nach der Mächtigkeit von  $\mathbb{N}$  nächstgrößere Mächtigkeit existiere: „Der einzige Fortschritt [im Kontinuumsproblem] ist der, daß man inzwischen wenigstens eine bestimmte Menge als zweite [unendliche] Mächtigkeit zu definieren gelernt hat.“ So schreibt Schoenflies zur Jahrhundertwende [eb., p. 27].

Selbst von einer „zweiten Mächtigkeit“ wie im Zitat von Schoenflies konnte man streng genommen nur für die wohlordenbaren Mengen sprechen: Die „bestimmte Menge“, sagen wir A, war wohlordenbar, aber man wußte nicht in voller mathematischer Klarheit, daß  $|A| \leq |M|$  gilt für alle Mengen M mit  $|\mathbb{N}| < |M|$ . Man sah die Gültigkeit dieser Aussage für wohlordenbare M. Der allgemeine Fall benötigt eine weit über die Mengenlehre ohne Auswahlaxiom hinausgehende abgeschwächte Form des Wohlordnungssatzes, und erst Zermelos Arbeiten im ersten Jahrzehnt des 20. Jahrhunderts schalteten in diesem Gewölbe voller verwirrender Fragen das Licht an. In einer Bemerkung im zweiten Teil des Berichts geht Schoenflies auf dieses Problem der zweiten Mächtigkeit ein: „Erst dieser Satz [der Vergleichbarkeit von Mengen] würde uns das Recht geben,  $\aleph_1$  [oben A genannt] als die zweite Mächtigkeit zu bezeichnen... Freilich hat der Sprachgebrauch sich längst an diese Bezeichnung gewöhnt. Er ist aber ohne den obigen Satz nicht gestattet.“ [Schoenflies 1908, p. 32].

Zwischen Cantors letzter bedeutender mengentheoretischer Arbeit von 1897 und bis zum wirkungsvollen Auftreten von Zermelo und Hausdorff 1904 fehlte schlichtweg ein kreativer und jugendlich-kraftvoller mathematischer Kopf ersten Ranges, dervon Cantor den stilsicheren Umgang mit den neuen transfiniten Zahlen geerbt hätte. Aus mathematischer Sicht machte erst Hausdorffs allgemeine Untersuchung linearer Ordnungen ab 1906 die Definition von komplizierteren Ordnungen durch verschiedenste Vergleichskriterien zur handwerklichen Selbstverständlichkeit, und Hausdorff war es dann auch, der 1914 den ersten vollständig ausgeführten Beweis zu Papier brachte, der noch heute in Darstellung und Inhalt gleichermaßen überzeugt.

Interessant ist auch, die Reaktionen auf die Lösungen des Multiplikationsproblems zu verfolgen. In seinem Bericht von 1900 erwähnt Schoenflies ein allgemeines Multiplikationsproblem nicht. Im zweiten Teil des Berichts [Schoenflies 1908], beklagt er dann aber das Fehlen eines ausführlichen Beweises dessen, was Bernstein in seiner Dissertation nur andeutet (siehe das Zitat in Abschnitt 3). Anschließend referiert er die beiden Beweise von Hessenberg von 1906 und 1907. Harwards Beweis dringt, wie erwähnt, in die

mathematische Fachwelt gar nicht ein. In der umgearbeiteten Fassung des ersten Teils aus dem Jahre 1913 schreibt Schoenflies dann zum Hessenbergschen Resultat und zum Beweis von Jourdain:

„Es liegt sehr nahe, diese Methode [der Diagonalaufzählung von  $\mathbb{N}^2$ ] auf beliebige Alephs zu übertragen. Tatsächlich hat die Formel [der Multiplikationssatz für wohlgeordnete Mengen] deshalb auch stets als richtig gegolten, ehe man einen präzisen Beweis für sie besessen hat. Ph. Jourdain war der erste, der den Cantorschen Grundgedanken in ausführlicher Darstellung auf den Fall beliebiger Alephs ausgedehnt hat... Einen rein arithmetischen Beweis hat Hessenberg geliefert...“ [Schoenflies 1913, p. 132]

Das „stets“ kann sich hier doch wohl nur auf die Zeit nach der Bernsteinschen Dissertation beziehen, bezeichnet also lediglich einen Zeitraum von fünf Jahren bis zur endgültigen Lösung des Problems 1906: Eine wichtige Formel, die als richtig gilt aber noch nicht vollständig bewiesen ist, sollte in einem Bericht wie dem Schoenfliesschen von 1900 irgendwo notiert sein. Gleichungen wie  $|\mathbb{N}^2| = |\mathbb{N}|$ ,  $|\mathbb{R} \times \mathbb{N}| = |\mathbb{R}|$ ,  $|\mathbb{R}^n| = |\mathbb{R}|$  werden dort ausführlich diskutiert; für die Menge  $F$  aller Funktionen von  $\mathbb{R}$  nach  $\mathbb{R}$  findet sich die Relation  $|\mathbb{R}| < |F|$ , aber die Frage  $|F^2| = |F|$  wird ebensowenig erwähnt wie eine allgemeine Multiplikationshypothese für Mengen oder auch nur für wohlordnbare Mengen.

Das Multiplikationsproblem scheint also in den beiden letzten Jahrzehnten des 19. Jahrhunderts nur halbwissentlich gewesen zu sein; eine größere Rolle spielt es dann erst seit der Dissertation von Felix Bernstein 1901. Die Tatsache, daß Hessenberg sein Resultat im knappen Vorwort seines Buches von 1906 ausführlich diskutiert, und die Besorgtheit von Jourdain um einen kleinen Platz in der Geschichte, die aus seinen Arbeiten herauszu hören ist, weisen dann darauf hin, daß man sich über die Bedeutung des nun völlig bewußten Problems und seiner Lösung sofort im klaren war. Eine ungewöhnliche Zwischenstufe in der Geschichte bilden die Bemerkungen von Bernstein über Cantors Einsichten in das Problem.

Die Nachwelt griff dann ab [Hausdorff 1914] den Beweis von Harward-Hausdorff mit Referenzen an Jourdain auf, unwissend um Harwards Priorität. Wie erwähnt, gilt Hausdorff die Hessenbergsche Argumentation als umständlicher. Es mag sein, daß sich der Beweis nach Harward-Hausdorff in den „normalen“ auf Kompaktheit ausgerichteten Aufbau eines Textes oder einer Vorlesung zwangloser einfügt; die Autorität von Hausdorffs „Grundzügen der Mengenlehre“ spielte aber sicher eine Rolle im Selektionsprozeß „survival of the fittest proof“. Man darf vermuten: Hätte Cantor selber den ersten Hessenbergschen Beweis gesehen und ihn als einfaches Korollar zu seiner Normalform notiert [etwa in Cantor 1997, § 19], wäre das Argument heute ein bekannter mengentheoretischer Klassiker.

Ein bemerkenswertes historisches Detail in der Rezeptionsgeschichte des Multiplikationssatzes ist, daß in der „Einleitung in die Mengenlehre“ von Abraham Fraenkel [Fraenkel 1928] das Multiplikationsproblem und seine Lösung nicht behandelt werden. Das Buch von Fraenkel richtet sich zwar an einen weiteren Leserkreis als die Überblicksartikel von Schoenflies oder das Buch von Hausdorff, jedoch ließe die Fülle des diskutierten Materials auf eine Erwähnung des Multiplikationsproblems schließen. Fraenkels Text glänzt mit dem vollständigsten und sorgfältigsten Literaturverzeichnis der Zeit, die Arbeiten [Jourdain 1908] und [Harward 1905] allerdings fehlen, obwohl Fraenkel fünf Arbeiten von Jourdain zwischen 1905 und 1922 auflistet. Das Fehlen des Multiplikationsproblems bedeutet hier sicher keine mangelnde Wertschätzung, sondern ist eher ein Beleg dafür, daß sich das Problem aus einer einführenden Darstellung gut ausklammern läßt – hat man

doch schon genug Mühe, den Lesern Gleichungen wie  $|IR| = |IR|$  in ihrer historischen Dimension vor Augen zu führen. Als Übungsaufgaben für den Leser notiert Fraenkel aber immerhin die Gleichungen  $|F + F| = |F|$  und  $|F \times F| = |F|$  für  $F = \{f \mid f: IR \rightarrow IR\}$  [Fraenkel 1928, p. 102 und p. 120].

## 14 Multiplikationssatz und Auswahlaxiom (1924)

Alle diskutierten Beweise des Multiplikationssatzes verwendeten mehr oder weniger direkt das Auswahlaxiom. Das Auswahlaxiom wird nicht gebraucht, um den Multiplikationssatz für wohlordenbare Mengen zu zeigen. Für das allgemeine Resultat wird aber dann der Zermelosche Wohlordnungssatz herangezogen, der zum Auswahlaxiom äquivalent ist. Mit der immer sensibler werdenden Wahrnehmung seiner Verwendung stellt sich auch die Frage, ob ein Rückgriff auf das Auswahlaxiom für einen Beweis des Multiplikationssatzes unumgänglich ist. Dies ist in der Tat der Fall: Nimmt man zu den um das Auswahlaxiom reduzierten Axiomen der Mengenlehre die Aussage des Multiplikationssatzes – quasi als Axiom -hinzu, so läßt sich in dieser Theorie das Auswahlaxiom beweisen. Auswahlaxiom und Multiplikationssatz sind also äquivalent (auf der Basis der übrigen Axiome). Dieses Resultat wird gewöhnlich einer Arbeit von Tarski aus dem Jahre 1924 zugeordnet [Tarski 1924], jedoch reichen die Vorarbeiten bis in das Jahr 1901 zurück. Felix Bernstein bewies in seiner Dissertation den folgenden Satz:

- (+) Seien  $M$  und  $N$  Mengen, und es gelte  $|M \times N| = |M + N|$ . Dann sind die Mächtigkeiten von  $M$  und  $N$  vergleichbar, d. h. es gilt  $|M| \leq |N|$  oder  $|N| \leq |M|$  [Bernstein 1905, p. 131ff].

Der allgemeine Vergleichbarkeitssatz stand damals noch nicht zur Verfügung, und die Suche nach hinreichenden Bedingungen war eine natürliche Problemstellung der Zeit. Bernsteins Argument ist hübsch und kurz, aber nur vermeintlich logisch-elementar: Er verwendet versteckt das Auswahlaxiom. (Vgl. hierzu auch [Schoenflies 1913, p. 47f.], wo das Auswahlaxiom immer noch versteckt eingeht.) Dessen ungeachtet ist der Satz von Interesse, insbesondere aufgrund der binomischen Gleichung

$$(++) |(M + N)^2| = |M^2 + \{0, 1\} \times M \times N + N^2|.$$

Setzt man nämlich den Multiplikationssatz voraus, so kann man das Quadrat links in (++) weglassen, und die Gleichung zeigt dann insbesondere  $|M \times N| \leq |M + N|$ . Die Ungleichung  $|M + N| \leq |M \times N|$  ist für Mengen mit mehr als einem Element trivial, und mit Cantor-Bernstein haben wir also  $|M \times N| = |M + N|$  aus dem Multiplikationssatz abgeleitet. Aus (+) folgt dann die Vergleichbarkeit von  $M$  und  $N$ . Soweit findet sich alles bereits in der Dissertation von Bernstein 1901. Leider wird aber das Auswahlaxiom im Beweis von (+) verwendet. Es ist aber aus dem Beweis von (+) abzulesen, daß das Auswahlaxiom nicht gebraucht wird, wenn man eine der beiden Mengen  $M$  und  $N$  in (+) als wohlordenbar voraussetzt. Gegeben den allgemeinen Multiplikationssatz und eine beliebige Menge  $M$ , wählt Tarski nun als  $N$  eine Wohlordnung, die sich nicht in  $M$  injektiv einbetten läßt. Daß eine derart lange Wohlordnung immer existiert, hat Friedrich Hartogs 1915 innerhalb der Zermeloschen Mengenlehre ohne Auswahlaxiom (und insbesondere ohne Ersetzungssatz) gezeigt [Hartogs 1915]. Nun läuft Bernsteins Ar-

gument ohne Auswahlaxiom durch, und wir erhalten  $|M| \leq |N|$  oder  $|N| \leq |M|$ . Letzteres ist nach Wahl von  $N$  ausgeschlossen, also gilt  $|M| \leq |N|$ . Somit ist  $M$  wohlordenbar, da sich  $M$  in eine wohlordenbare Menge einbetten lässt. Dieses Argument liefert also den Wohlordnungssatz aus dem Multiplikationssatz ohne irgendeine Verwendung des Auswahlaxioms!

Tarski stellte dann die sich aufdrängende Frage, ob der Additionssatz ebenfalls zum Auswahlaxiom äquivalent ist. Dies ist jedoch nicht der Fall: Der Additionssatz ist echt schwächer als das Auswahlaxiom (aber dennoch echt stärker als eine Mengenlehre ganz ohne Auswahlaxiom). Beweise dieses Resultats wurden unabhängig voneinander und mit verschiedenen Methoden von Gershon Sageev sowie Dan Halpern und Paul Howard gegeben [Segeev 1975, Halpern / Howard 1976]. Es gibt eine Fülle verwandter Resultate, siehe hierzu z. B. [Jech 1973] oder [Halbeisen / Shelah 2001].

## 15 Max Zorn: Ein wohlordnungsfreier Beweis (1944)

Max Zorn hat 1935 sein heute nach ihm benanntes Lemma veröffentlicht – fortan fester Bestandteil im Werkzeugkoffer des Algebraikers –, und es 1944 für einen Beweis des Additions- und Multiplikationssatzes verwendet [Zorn 1944].

Der Begriff der Wohlordnung kann so ganz vermieden werden, auf Kosten der anschaulichkeit und feinen Struktur des Arguments. Der Beweis selbst beruht auf einer doppelten Anwendung des Zornschen Lemmas. Zunächst wird der Additionssatz gezeigt (oder bequemer eine verschärzte Form wie etwa  $|N \times M| = |M|$  für unendliche Mengen  $M$ , aus der der Additionssatz sofort folgt). Anschließend gewinnt man den Multiplikationssatz, unter Verwendung der Zerlegung von  $(X \cup Y)^2$  in die vier Teile  $X \times X$ ,  $Y \times Y$ ,  $X \times Y$  und  $Y \times X$ . Beide Schritte benutzen das Zornsche Lemma. Die zugrundeliegenden partiellen Ordnungen sind jeweils kanonisch. Sparsamkeit ist hier nicht recht am Ort, und so spielt es keine Rolle, daß neben dem Satz von Cantor-Bernstein auch noch der Vergleichbarkeitssatz verwendet wird.

Bemerkenswert ist schließlich, daß die Möglichkeit eines wohlordnungsfreien Beweises in die Standardliteratur der Mengenlehre allenfalls marginal eingegangen ist [z. B. in Levy 1979, p. 163]. Die relative Unzugänglichkeit der Arbeit [Zorn 1944] mag eine Rolle gespielt haben. Hauptursache scheint aber die gut begründbare Ansicht zu sein, daß der Beweis von Zorn zu grob ist, um den Feinheiten der Gleichung  $|M^2| = |M|$  gerecht zu werden.

## 16 Literatur

- Bernstein, Felix**, 1905, Untersuchungen aus der Mengenlehre; (*Zeitschriften-Veröffentlichung der Dissertation von Felix Bernstein, Göttingen 1901*); *Mathematische Annalen* 61, 111–155.
- Brouwer, Luitzen**, 1911, Beweis der Invarianz der Dimensionszahl; *Mathematische Annalen* 70, 161–165.
- Cantor, Georg** 1874, Über eine Eigenschaft des Inbegriffs aller reellen algebraischen Zahlen; *Journal für die reine und angewandte Mathematik* 77, 258–262.
- 1878, Ein Beitrag zur Mannigfaltigkeitslehre; *Journal für die reine und angewandte Mathematik* 84, 242–258.

- 1890/1891, Über eine elementare Frage der Mannigfaltigkeitslehre; *Jahresbericht der Deutschen Mathematiker Vereinigung 1 (1890/1891)*, 75–78.
  - 1895, Beiträge zur Begründung der transfiniten Mengenlehre; *Mathematische Annalen 46 (1895)*, 481–512.
  - 1897, Beiträge zur Begründung der transfiniten Mengenlehre (zweiter Artikel); *Mathematische Annalen 49 (1897)*, 207–246.
  - 1932, Gesammelte Abhandlungen mathematischen und philosophischen Inhalts; Herausgegeben von Ernst Zermelo. Springer, Berlin. Repro graphisch er Nachdruck Georg Olms Verlagsbuchhandlung, Hildesheim 1962. Bibliographisch ergänzte Neuauflagen 1980, 1990 bei Springer, Berlin.
  - 1984, Über unendliche, lineare Punktmannigfaltigkeiten. Arbeiten zur Mengenlehre aus den Jahren 1872–1884; Herausgegeben von Günter Asser, B. G. Teubner, Leipzig.
  - 1991, Briefe; Herausgegeben von Herbert Meschkowski und Winfried Nilson. Springer, Berlin.
- Deiser, Oliver** 2004, Einführung in die Mengenlehre; zweite erweiterte und korrigierte Auflage. Springer, Berlin.
- Fraenkel, Abraham** 1928, Einleitung in die Mengenlehre; 3. Auflage, Springer, Berlin.
- Gödel, Kurt** 1940, The consistency of the axiom of choice and of the generalized continuum hypothesis with the axioms of set theory; *Annals of mathematical studies 3 (1940)*. Princeton University Press.
- Halbeisen, Lorenz / Shelah, Saharon**, 2001, Relations between some cardinals in the absence of the axiom of choice; *The Bulletin of Symbolic Logic 7/2 (2001)*, 231–261.
- Halpern, Dan / Howard, Paul**, 1976, The law of infinite cardinal addition is weaker than the axiom of choice; *Transactions of the American Mathematical Society 220 (1976)*, 195–204.
- Hartogs, Friedrich**, 1915, Über das Problem der Wohlordnung; *Mathematische Annalen 76*, 438–443.
- Harward, A.E.**, 1905, On the transfinite numbers; *Philosophical Magazine 6/10 (1905)*, 439–460.
- Hausdorff, Felix**, 1904, Der Potenzbegriff in der Mengenlehre; *Jahresbericht der Deutschen Mathematiker-Vereinigung 13 (1904)*, 569–571.
- 1906, Untersuchungen über Ordnungstypen; *Berichte über die Verhandlungen der Königlich Sächsischen Gesellschaft der Wissenschaften zu Leipzig, mathematisch-physikalische Klasse 58 (1906)*, 106–169.
  - 1908, Grundzüge einer Theorie der geordneten Mengen; *Mathematische Annalen 65 (1908)*, 435–508.
  - 1914, Grundzüge der Mengenlehre; Veit & Comp., Leipzig.
  - 2002, Gesammelte Werke in 8 Bänden, Band II: Grundzüge der Mengenlehre. Hg. E. Brieskorn et al.; Springer, Berlin.
- Hessenberg, Gerhard**, 1906, Grundbegriffe der Mengenlehre; Vandenhoeck & Ruprecht, Göttingen.
- 1907, Potenzen transfiniter Ordnungszahlen; *Jahresbericht der Deutschen-Mathematiker-Vereinigung 16 (1907)*, 130–137.
- Jech, Thomas**, 1973, The Axiom of Choice; North-Holland, Amsterdam/London.
- 2002, Set Theory; Springer, Berlin.
- Jourdain, Philip**, 1904a, On the transfinite cardinal numbers of well-ordered aggregates; *Philosophical Magazin 6/7 (1904)*, 61–75.
- 1904b, On the transfinite cardinal numbers of number-classes in general; *Philosophical Magazin 6/7 (1904)*, 294–303.
  - 1908, On the multiplication of alephs; *Mathematische Annalen 65 (1908)*, 506–512.
- Kanamori, Akihiro**, 1996, The mathematical development of set theory from Cantor to Cohen; *The Bulletin of Symbolic Logic, Volume 2, Number 1 (1996)*, S. 1–71.
- Katetov, Miroslav / Simon, Petr**, 1997, Origins of dimension theory; in: C.E.Aull/R. Lowen (Hrsg.): *Handbook of the History of General Topology, Volume 1*. Kluwer Academic Publishers, Dordrecht.
- Levy, Azriel**, 1979, Basic Set Theory; Springer, Berlin.
- Meschkowski, Herbert**, 1967, Probleme des Unendlichen. Werk und Leben Georg Cantors; Vieweg, Braunschweig.

- Moore, Gregory H.**, 1976, Ernst Zermelo, A. E. Harward, and the axiomatization of set theory; *Historia Mathematica 3* (1976), 206–209.
- Sageev, Gershon**, 1975, An independence result concerning the axiom of choice; *Annals of Mathematical Logic 8* (1975), 1–184.
- Schoenflies, Arthur**, 1900, Die Entwicklung der Lehre von den Punktmanigfaltigkeiten. 1. Teil; *Jahresbericht der Deutschen Mathematiker-Vereinigung 8*, Heft 2, 1–250, B. G. Teubner, Leipzig.
- 1908, Die Entwicklung der Lehre von den Punktmanigfaltigkeiten. 2 .Teil; *Jahresbericht der Deutschen Mathematiker- Vereinigung, 2. Ergänzungsband*; B. G. Teubner, Leipzig.
  - 1913, Entwicklung der Mengenlehre und ihrer Anwendungen. Umarbeitung des im VIII. Bande der Jahresberichte der Deutschen Mathematiker-Vereinigung erstatteten Berichts; B. G. Teubner, Leipzig.
- Tarski, Alfred**, 1924, Sur quelques théorèmes qui équivalent à l'axiom du choix; *Fundamenta Mathematicae 5* (1924), 147–154.
- Zermelo, Ernst**, 1901, Über die Addition transfiniter Kardinalzahlen; *Nachrichten von der Königlichen Gesellschaft der Wissenschaften zu Göttingen, mathematisch physikalische Klasse* (1901), 34–38.
- 1904, Beweis, daß jede Menge wohlgeordnet werden kann; *Mathematische Annalen 59* (1904), 514–516.
- Zorn, Max**, 1935, A remark on method in transfinite algebra; *Bulletin of the American Mathematical Society 41* (1935), 667–670.
- 1944, Idempotency of infinite cardinals; *University of California Publications 2*, 9–12.



# Elmar Thoma zum Gedächtnis 1926 – 2002

Eberhard Kaniuth und  
Günter Schlichting



## Abstract

- Mathematics Subject Classification: 01 A 70
- Keywords and Phrases: Elmar Thoma, Lebenslauf, wissenschaftliches Werk, ausgewählte Schriften, Doktoranden

This article describes the life of Elmar Thoma, parts of his mathematical work and subsequent developments. The names of his Ph. D. students and selected publications are listed.

Eingegangen: 2. 12. 2004

Eberhard Kaniuth, Institut für Mathematik, Universität Paderborn  
D-33095 Paderborn, kaniuth@math.uni-paderborn.de

Günter Schlichting, Fakultät für Mathematik, Technische Universität  
München, D-80290 München, schlicht@mathematik.tu-muenchen.de

**DMV**  
**JAHRESBERICHT**  
**DER DMV**  
© B. G. Teubner 2005



## Der Lebenslauf

Am 23. Juli 2002 verstarb Prof. Dr. Elmar Thoma, emeritierter Ordinarius für Mathematik an der Technischen Universität München im Alter von 75 Jahren.

Elmar Thoma wurde am 10. September 1926 in Baden-Baden geboren, wuchs jedoch in Neumarkt in der Oberpfalz auf und besuchte das Gymnasium in Nürnberg. Nach dem Abitur studierte er von 1946 an Mathematik und Physik an den Universitäten Regensburg und Erlangen und legte 1951 in Erlangen die erste Staatsprüfung für das Lehramt an Gymnasien ab. Bereits im Jahre 1952 erfolgte die Promotion bei Otto Haupt in Erlangen. Nach einer Industrietätigkeit bei der Firma Siemens in Erlangen war er ab 1954 wissenschaftlicher Assistent bei Georg Aumann an der Ludwig-Maximilians-Universität

München, wo er sich 1957 für das Fach Mathematik habilitierte. Daran schloß sich ein zweijähriger Aufenthalt als Gastprofessor in den USA an, und zwar an der University of Washington in Seattle.

Nach drei Jahren an der Universität Heidelberg als wissenschaftlicher Rat und außerplanmäßiger Professor folgte Thoma 1964 einem Ruf auf einen Lehrstuhl für Mathematik an der Westfälischen Wilhelms-Universität Münster. In unruhigen und schwierigen Zeiten war er dort mit 42 Jahren Dekan der großen Mathematisch-Naturwissenschaftlichen Fakultät und wesentlich beteiligt am beginnenden Ausbau der Mathematik. Obwohl er sich in Münster ausnehmend wohl fühlte, war doch die Verbundenheit mit der bayerischen Heimat zu stark, um einen ergangenen Ruf an die Technische Universität München abzulehnen.

Dort war Elmar Thoma als Nachfolger von Robert Sauer von 1970 bis zu seiner Emeritierung im Jahre 1994 Inhaber des Lehrstuhls für Höhere Mathematik und Analytische Mechanik. Dazu kamen Ämter in der akademischen Selbstverwaltung. Er war 1973/74 Dekan der damals noch großen Fakultät für Allgemeine Wissenschaften, später Geschäftsführender Direktor des Mathematischen Instituts und viele Jahre Vorsitzender des Prüfungsausschusses für Mathematiker. Bis kurz vor seiner Emeritierung war er Beauftragter für die Fachbereichsbibliothek.

Über zehn Jahre lang bis 1990 fungierte Elmar Thoma als einer der Herausgeber der Mathematischen Annalen, seit 1969 war er gemeinsam mit Horst Leptin Organisator der im zweijährigen Turnus stattfindenden internationalen Tagung ‚Harmonische Analyse und Darstellungstheorie topologischer Gruppen‘ im Mathematischen Forschungsinstitut Oberwolfach. Er war Mitveranstalter der gemeinsamen Arbeitstagung über Gruppen und topologische Gruppen der Universitäten Erlangen, Freiburg, Trient, Technische Universität München und Würzburg. Aus diesen und anderen Aktivitäten entstanden zahlreiche Kontakte zu Wissenschaftlern in vielen verschiedenen Ländern.

Elmar Thoma hat sich nie als spezialisierter Vertreter seines Faches Mathematik verstanden, seine Interessen waren vielseitig, von der Physik und Chemie bis hin zur Geschichte, insbesondere der bayerischen Geschichte.

Thoma verstand es, seine Begeisterung für Mathematik auf die Hörer seiner Vorlesungen zu übertragen. Seine Studenten schätzten seinen lebendigen Vortragsstil. In Münster etwa versammelte sich alsbald eine ansehnliche Zahl von Schülern um ihn. Für diese war er jederzeit ansprechbar, und auf Grund seiner ausgeprägten Diskutierfreudigkeit und seiner Ideen ein anregender und hilfreicher Gesprächspartner. Seinen Mitarbeitern gewährte er bei ihrer Tätigkeit ein Maximum an Freiheit, deren fachliche Entwicklung stand für Thoma im Vordergrund. Der Erfolg des akademischen Lehrers Elmar Thoma wird insbesondere auch dadurch eindrucksvoll dokumentiert, daß immerhin neun seiner Doktoranden im Universitätsbereich verblieben sind.

## Zum wissenschaftlichen Werk

Das Forschungsgebiet von Elmar Thoma war die Darstellungstheorie und Harmonische Analysis lokalkompakter, insbesondere diskreter, Gruppen. Während dieses Gebiet, in das Gruppentheorie, Topologie, Funktionalanalysis, Maß- und Integrations-theorie u.a. einfließen, Ende der 50er Jahre etwa auf Grund der fundamentalen Arbeiten von Mackey in den USA bereits entscheidende Fortschritte gemacht hatte, wartete es auf seine Erschließung und Wiederbelebung im Nachkriegsdeutschland.

Gewiß hat der Aufenthalt in Seattle und das durch Edwin Hewitt geprägte mathematische Umfeld Thoma darin bestätigt, Harmonische Analysis zu betreiben. Er hat jedoch wiederholt erwähnt, daß während seiner Tätigkeit bei Siemens Fragen theoretischer Physiker zur Darstellungstheorie klassischer Gruppen ihn zur Beschäftigung mit Darstellungstheorie lokalkompakter Gruppen angeregt hätten.

Ein Kernstück, auf das hier näher eingegangen werden soll, des wissenschaftlichen Werkes von Thoma war die Charakterisierung der diskreten Gruppen vom Typ I [1, 6], die in der internationalen Fachwelt allergrößtes Aufsehen erregte. Grob gesprochen sind die lokalkompakten Gruppen vom Typ I diejenigen, deren irreduzible Darstellungen man (zumindest prinzipiell) bestimmen kann. Thomas Ergebnis, das eine Vermutung von Kaplansky [Ka] bestätigt, besagt, daß eine diskrete Gruppe genau dann vom Typ I ist, wenn sie einen abelschen Normalteiler von endlichem Index besitzt. Insbesondere sind also dann nicht nur alle irreduziblen Darstellungen endlich-dimensional, sondern die Darstellungsgrade sind sogar beschränkt durch diesen Index. Dieses Resultat stellt eines der tiefsten in der Darstellungstheorie diskreter Gruppen überhaupt dar und war in der Folge Anlaß für viele weitere Untersuchungen anderer Mathematiker.

Ein Beispiel hierfür ist die Mooresche Beschreibung aller derjenigen lokalkompakten Gruppen, deren irreduzible Darstellungen sämtlich endlich dimensional sind [Mo]. Es war naheliegend, daß im Anschluß an Thomas Satz die Frage gestellt wurde, ob eine diskrete Gruppe bereits dann einen abelschen Normalteiler von endlichem Index besitzen muß, falls ihre reguläre Darstellung vom Typ I ist. Dies wurde in [K1] bejaht. Später schlossen sich weitere Verallgemeinerungen an [K2, Sc, T1, T2, KS], wie etwa: Explizite Beschreibung der maximalen Typ I-Projektion in der von Neumann-Algebra der regulären Darstellung, Erweiterung der Resultate auf eine größere Klasse als die der diskreten Gruppen und auf Darstellungen, die im Unendlichen verschwinden.

Um das Problem der Charakterisierung der diskreten Gruppen vom Typ I angehen zu können, entwickelte Thoma eine Theorie der sogenannten Charaktere [1,2]. Für eine (diskrete) Gruppe  $G$  bezeichne  $K(G)$  die Menge aller positiv-definiten Funktionen  $\varphi$  auf  $G$  mit  $\varphi(e) = 1$  und  $\varphi(y^{-1}xy) = \varphi(x)$  für alle  $x, y \in G$ . Die Menge  $K(G)$  ist konvex und, versehen mit der Topologie der punktweisen Konvergenz, kompakt. Die Elemente der Menge  $E(G)$  aller Extremalpunkte von  $K(G)$  heißen Charaktere. Im abelschen Fall sind dies also gerade die Elemente der dualen Gruppe von  $G$ , im Falle einer endlichen Gruppe  $G$  die normierten Spuren irreduzibler Darstellungen von  $G$ . Vermöge der Gel'fand-Naimark-Segal-Konstruktion definiert jedes  $\varphi \in E(G)$  eine sogenannte Faktordarstellung  $\pi_\varphi$  von  $G$  von endlichem Typ, und eine solche ist nur dann vom Typ I, wenn sie endlich-dimensional ist. Die Essenz des Beweises des Thomaschen Satzes ist also zu zeigen, daß die Gruppe  $G$  einen abelschen Untergruppe von endlichem Index besitzen muß, wenn alle diese  $\pi_\varphi, \varphi \in E(G)$ , endlich-dimensional sind.

Es war offenkundig, daß zumindest für diskrete Gruppen  $G$  der Raum  $E(G)$  ein nützliches duales Objekt darstellt. Für klassenfinite Gruppen etwa gibt die Zuordnung  $\varphi \rightarrow \pi_\varphi$  Anlaß zu einem Homöomorphismus zwischen  $E(G)$  und dem Raum der primitiven Ideale der  $C^*$ -Gruppenalgebra von  $G$ . Gleicher gilt für endlich-erzeugte nilpotente Gruppen, wenn man  $E(G)$  mit der Seitentopologie versieht.

Thoma hat, wohl auch seiner Neigung zu konkreten Berechnungen folgend, für die Gruppe  $S_\infty$  aller endlichen Permutationen der Menge der natürlichen Zahlen  $E(S_\infty)$  bestimmt [3]. Der Beweis war tief sinnig und benutzte Aussagen über ganze Funktionen. Die Arbeit [3] initiierte ein immenses Interesse an der Gruppe  $S_\infty$  und anderer diskreter Gruppen wie  $GL(\infty, F)$  für endliche Körper  $F$  [Sk]. Unter verschiedenen Gesichtspunkten wurden die Charaktere der  $S_\infty$  von Kerov und Vershik studiert [KV, VK]. Insbesondere haben diese einen Beweis gefunden, der die Approximation der Charaktere der  $S_\infty$  durch die Charaktere der endlichen symmetrischen Gruppen benutzt. Einen weiteren Beweis hat Okounkov, Ideen von Olshanskii fortentwickelt, gegeben [Ok]. Große Serien irreduzibler Darstellungen der  $S_\infty$  wurden von mehreren Autoren, etwa Hirai [Hi], konstruiert.

## Doktoranden

Eberhard Kaniuth (1967), Wolfgang Krieger (1968), Rolf Wim Henrichs (1970), Günter Schlichting (1970), Gerhard Zumbusch (1971), Hans-Ludwig Skudlarek (1972), Rainer Felix (1973), Norbert Riedel (1977), Ulrich Stegmeir (1978), Winfried Geißler (1978), Wolfram Ziegler (1981), Kurt Allinger (1982), Stefan Wolfenstetter (1984), Michael Voit (1988), Martin Leitner (1989), Josef Obermaier (1990), Margit Rösler (1992).

## Ausgewählte Schriften

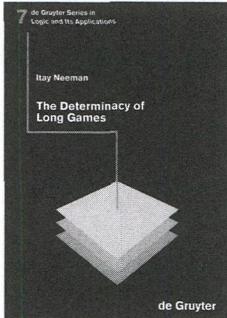
1. Über unitäre Darstellungen abzählbarer, diskreter Gruppen, *Math. Ann.* **153** (1964), 111–138.
2. Über positiv-definite Klassenfunktionen abzählbarer Gruppen, *Math. Z.* **84** (1964), 389–402.
3. Die unzerlegbaren, positiv-definiten Klassenfunktionen der abzählbar unendlichen symmetrischen Gruppe, *Math. Z.* **85** (1964), 40–61.

4. Invariante positiv-definite Klassenfunktionen und ergodische Maße, *Math. Ann.* **162** (1965), 172–189.
5. Zur harmonischen Analyse klassenfiniter Gruppen, *Invent. Math.* **3** (1967), 20–42.
6. Eine Charakterisierung diskreter Gruppen vom Typ I, *Invent. Math.* **6** (1968), 190–196.
7. Die Einschränkung der Charaktere von  $\mathrm{GL}(n, q)$  auf  $\mathrm{GL}(n-1, q)$ , *Math. Z.* **119** (1971), 321–338.

## Weitere Referenzen

- [Hi] T. Hirai: Construction of irreducible unitary representations of the infinite symmetric group  $S_\infty$ , *J. Math. Kyoto* **31** (1991), 495–541.
- [K1] E. Kaniuth: Der Typ der regulären Darstellung diskreter Gruppen, *Math. Ann.* **182** (1969), 334–339.
- [K2] E. Kaniuth: Die Struktur der regulären Darstellung lokalkompakter Gruppen mit invarianter Umgebungsbasis der Eins, *Math. Ann.* **194** (1971), 225–248.
- [KS] E. Kaniuth and G. Schlichting: The type structure of multiplier representations which vanish at infinity, *Arch. Math.* **39** (1982), 443–454.
- [Ka] I. Kaplansky: Functional analysis. 1958, in: *Some aspects of analysis and probability*, pp. 1–34; *Surveys in Applied Math.* Vol. 4, Wiley, 1958.
- [KV] S.V. Kerov and A.M. Vershik: The characters of the infinite symmetric group and probability properties of the Robinson-Schensted-Knuth algorithm, *SIAM J. Alg. Disc. Math.* **7** (1986), 116–124.
- [Mo] C.C. Moore: Groups with finite dimensional irreducible representations, *Trans. Amer. Math. Soc.* **166** (1972), 401–410.
- [Ok] A. Okounkov: Thoma's theorem and representations of an infinite symmetric group (Russian), *Translation in: Func. Anal. Appl.* **28** (1994), 100–107.
- [Sc] G. Schlichting: Eine Charakterisierung gewisser diskreter Gruppen durch ihre reguläre Darstellung, *Manuscripta Math.* **9** (1973), 389–409.
- [Sk] H.L. Skudlarek: Die unzerlegbaren Charaktere einiger diskreter Gruppen, *Math. Ann.* **223** (1976), 213–231.
- [T1] K.F. Taylor: The type structure of the regular representation of a locally compact group, *Math. Ann.* **222**, (1976), 211–224.
- [T2] K.F. Taylor: Group representations which vanish at infinity, *Math. Ann.* **252** (1980), 185–190.
- [VK] A.M. Vershik and S.V. Kerov: Characters and factor representations of the infinite symmetric group, *Soviet Math. Dokl.* **23** (1981), 389–392.

## Just released!



Itay Neeman

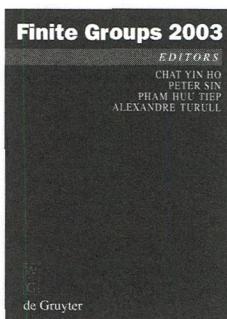
### ■ The Determinacy of Long Games

2004. XI, 317 pages. Cloth.

€ [D] 128.00 / sFr 205.00 / for USA, Canada, Mexico US\$ 128.95.

ISBN 3-11-018341-2

In this volume the author develops and applies methods for proving, from large cardinals, the determinacy of definable games of countable length on natural numbers. The determinacy is ultimately derived from iteration strategies, connecting games on natural numbers with the specific iteration games that come up in the study of large cardinals.



### ■ Finite Groups 2003

#### Proceedings of the Gainesville Conference on Finite Groups, March 6–12, 2003

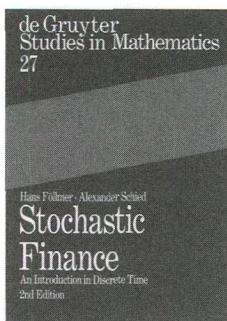
Ed. by Chat Yin Ho / Peter Sin / Pham Huu Tiep / Alexandre Turull

2004. XV, 417 pages. 1 Frontispiece. Cloth.

€ [D] 168.00 / sFr 269.00 / for USA, Canada, Mexico US\$ 168.95.

ISBN 3-11-017447-2

This is a volume of research articles related to finite groups. Topics covered include the classification of finite simple groups, the theory of  $p$ -groups, cohomology of groups, representation theory and the theory of buildings and geometries.



Hans Föllmer / Alexander Schied

### ■ Stochastic Finance

#### An Introduction in Discrete Time 2nd rev. and extend. ed.

2004. XI, 459 pages. Cloth.

€ [D] 58.00 / sFr 93.00 / for USA, Canada, Mexico US\$ 59.95.

ISBN 3-11-018346-3

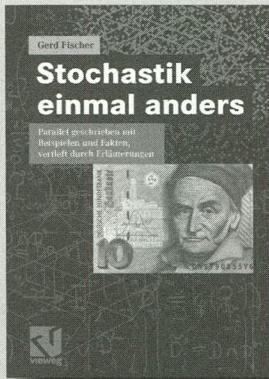
This book is an introduction to financial mathematics. In addition to many corrections and improvements, this second edition contains several new sections, including a systematic discussion of law-invariant risk measures and of the connections between American options, superhedging, and dynamic risk measures.



de Gruyter  
Berlin · New York

*Prices are subject to change.*

# So versteht man die Stochastik leicht



Gerd Fischer

## Stochastik einmal anders

Parallel geschrieben mit Beispielen und Fakten, vertieft durch Erläuterungen

2005. VIII, 327 S. Br. EUR 24,90

ISBN 3-528-03967-1

## INHALT

Beschreibende Statistik - Wahrscheinlichkeitsrechnung - Schätzen - Testen von Hypothesen - Anhang: Ergänzungen und Beweise

## DAS BUCH

Eine Einführung in die Fragestellungen und Methoden der Wahrscheinlichkeitsrechnung und Statistik (kurz Stochastik) sowohl für Studierende, die solche Techniken in ihrem Fach benötigen, als auch für Lehrer, die sich für den Unterricht mit den nötigen fachlichen Grundlagen vertraut machen wollen.

Der Text hat einen besonderen Aufbau - als Trilogie ist er in Beispiele, Fakten und Erläuterungen aufgeteilt.

Was überall in der Mathematik gilt, ist noch ausgeprägter in der Stochastik: Es geht nichts über markante Beispiele, die geeignet sind, die Anstrengungen in der Theorie zu rechtfertigen. Um dem Leser dabei möglichst viele Freiheiten zu geben, ist der Text durchgehend parallel geführt: links die Beispiele, rechts die Fakten.



Abraham-Lincoln-Straße 46  
D-65189 Wiesbaden  
Fax 0611.78 78-420

Änderungen vorbehalten.  
Erhältlich beim Buchhandel oder beim Verlag.

**G. Aubert, P. Kornprobst**

**Mathematical Problems in Image Processing, Partial Differential Equations and the Calculus of Variations**

G. Aubert, P. Kornprobst  
**Mathematical Problems in Image Processing, Partial Differential Equations and the Calculus of Variations**

Appl. Math.  
Sciences 147

Berlin, u. a., Springer, 2002, 286 S., € 74,95

Dies ist ein wunderbares Buch. Der Weg zu dieser Erkenntnis war allerdings etwas länger.

Beim Lesen des Titels dieses Buches wurden meine Erwartungen sehr hoch geschaubt:

Endlich, das längst überfällige Buch, das die Grundlagen der mathematischen Bildverarbeitung knapp und zugleich umfassend darstellt und die faszinierende Entwicklung der letzten Jahre dieses vergleichsweise neuen Wissenschaftsbereichs würdigt. Ein derartiges Buch würde eine seit langem beobachtete Lücke füllen, die jeder zu spüren bekommt, der sich bei der Vorbereitung auf eine Vorlesung zu diesem Thema schon einmal über die unzureichende Literaturlage geärgert hat.

Auch das sehr anregend geschriebene Vorwort von Olivier Faugeras erhöht noch die Erwartungen an Inhalt und Lesbarkeit des Textes: er empfiehlt dieses Buch auch der Computer Vision Gemeinde, die eher in der Informatik zu Hause ist. Diese Empfehlung ist allerdings mit der Warnung versehen, dass diese Leserschaft vielleicht nicht alle Details der Beweise verstehen aber zumindest doch auch schon beim ersten Lesen die wesentlichen Ideen erfassen könnte.

Dieser Satz ist wohl dem verständlichen aber in diesem Fall irreführenden Bestreben des Verlags, eine möglichst breite Leserschaft anzusprechen, geschuldet. Bereits der Untertitel des Buches ‚partial differential equations and the calculus of variations‘ hätte den auf-

merksamen Betrachter mit etwas Vorsicht auf das Kommende ausstatten können.

Dieses Buch konzentriert sich ausschließlich auf den Themenbereich des Untertitels, der in den vergangenen Jahren einige der bedeutendsten neuen Beiträge zur mathematischen Theorie der Bildverarbeitung geleistet hat. Ein vollständiges Buch zur Mathematischen Bildverarbeitung ist es allerdings nicht, zentrale Themen aus Anwendersicht (Image Fusion, Pattern recognition, Klassifikation, etc.) oder einige der wichtigsten mathematischen Methoden (Abtasttheorie und die Theorie diskreter Filter, Wavelet Analysis) fehlen ebenso wie alle elementaren Basismethoden (Morphologische Verfahren, Histogramm-Methoden, u.v.a.).

Und noch eine Einschränkung sollte hinzugefügt werden: selbst der Untertitel ist noch zu weit gefasst. Die Autoren beschreiben im wesentlichen die Ergebnisse der französischen Schule. Den Ergebnissen von Weickert werden ganze 4 Seiten gewidmet, der Rudin-Osher-Ansatz und seine weitreichenden Weiterentwicklungen werden auf 2 Seiten abgehandelt, die fruchtbaren Verbindungen zu Operatorgleichungen und inversen Problemen werden in Kapitel 3.2 zwar angeprochen, die schönen Ergebnisse von Scherzer fehlen allerdings völlig ebenso wie die sehr gut strukturierten und anschaulichen Arbeiten von Keeling. Auch die mathematischen Probleme der numerischen Umsetzung werden im Anhang kurz und im Sinne einer groben Programmieranleitung gut verständlich beschrieben, die mathematischen Aspekte dieser Fragestellungen werden kaum berührt.

Nachdem diese leichte Verärgerung über den irreführenden Titel verflogen ist, erkennt man allerdings schnell, dass dies ein sehr schönes und äußerst sorgfältig geschriebenes Buch ist. Es gibt meines Erachtens kein vergleichbares Buch, das die Ergebnisse der französischen Schule zu diesem Thema für ein mathematisches Publikum derart kompetent und umfassend darstellt.

Die Voraussetzungen, die der Leser oder ein Student höheren Semesters, der eine sich

an diesem Buch orientierende Vorlesung besucht, mitbringen sollte, sind allerdings nicht gering. Grundkenntnisse der nichtlinearen Funktionalanalysis und ein solides Grundwissen über die Existenztheorie elliptischer und parabolischer Probleme sowie solide Kenntnisse der Theorie von Funktionenräumen werden vorausgesetzt.

Nun aber endlich zum Inhalt des Buchs: Die Autoren sind im Grunde genommen strenge Analytiker, dementsprechend beginnt das Buch in Kapitel 2 mit einem knapp 40-seitigen Kapitel Mathematical Preliminaries. Hier werden u. a. einige Ergebnisse der nichtlinearen Funktionalanalysis und Variationsrechnung inklusive der Gamma-Konvergenz, Eigenschaften des BV-Raumes sowie einige Ergebnisse der Lösungstheorie von PDE'S (Eikonal-Gleichung, Viskositätslösungen) zusammengefasst.

Dieses Kapitel wäre ohne die vorangestellte 20-seitige, hervorragende Einleitung mit einem detaillierten ‚Plan‘ des Buchs sehr abschreckend. Dieser in Kapitel 1.4 vorgestellte Plan hat auf mich wie ein Ruhepol beim Lesen des Buchs gewirkt. Man kann immer wieder zu diesem Kapitel zurückkehren, um z. B. die Einordnung des gerade Gelesenen in dem Gesamtzusammenhang zu erkennen oder um einen gewagten Sprung über einige Kapitel des Buchs vorzubereiten, oder auch nur, um die wesentlichen Aspekte dieses Buches nochmals in komprimierter Form zu überfliegen.

Die nachfolgenden Kapitel heißen *Image restoration, segmentation problem, und other challenging applications*. D. h. sie sind nach den Anwendungsaspekten geordnet, im Kern sind sie aber mathematisch aufgebaut: image restoration steht für Energie-Funktionale und nicht-lineare Diffusion, Segmentation für mean curvature flow, Mumford-Shaw-Funktionale und , Level-Set Methoden.

Der abschließende Anhang zu Finiten Differenzen wird einen Numeriker nicht zufrieden stellen. Er ist aber gut geeignet, um Studenten einer Vorlesung im Hauptstudium eine gut verständliche Anleitung für eigene,

einfache Simulationsrechnungen zu geben, ohne vorher eine komplette Veranstaltung Numerik partieller Differentialgleichungen besuchen zu müssen.

Es folgen noch drei kurze Bemerkungen zum Inhalt des Buches.

1. Als einen typischen Testfall für die Lesbarkeit des Buche habe ich versucht, auf kürzestem Weg den berühmten Beweis von Catte, Lions et al. zur Existenz einer Lösung des regularisierten Perona-Malik-Modells nachzuvollziehen (S. 111 ff). Zwar wäre hierfür ein Einstieg über das vorhergehende Kapitel (nonlinear diffusion) leichter, aber auch so lässt sich dieses zentrale Ergebnis sehr gut verstehen.

2. Diese Buch ist außergewöhnlich sorgfältig aufgebaut und äußerst gut geschrieben. Besonders der Aufbau des Buches ist positiv hervorzuheben. Der Leser wird über mehrere Stufen (Detailed Plan Kapitel 1.4, mathematical preliminaries Kapitel 2 und die nachfolgenden ausführlichen Kapitel zu den einzelnen Spezialthemen) angenehm an die Thematik heranführt. Außerdem muss die sehr sorgfältige Gestaltung des Buches gelobt werden, es ist eine wirkliche Freude dieses Buch in die Hand zu nehmen!

3. Der Wille der Autoren zur Abstraktion ist stark ausgeprägt. Gelegentlich hätte die Beschränkung auf konkrete Funktionenräume die Lesbarkeit erhöht. Von dem Leser wird ein selbstverständlicher Umgang mit z. B. der BV- $w^*$ -Topologie erwartet, den z. B. der Referent dieses Buches nicht vorzuweisen hat.

Wer wird dieses Buch lesen? Dieses Buch ist meines Erachtens ein Muss für jeden Mathematiker, der sich mit Bildverarbeitung beschäftigt. Das Maß an notwendiger abstrakter Analysis ist für diesen Leserkreis sehr gut gewählt und notwendig, wenn man in knapper Form die zentralen Ergebnisse dieses Forschungsgebietes vorstellen will.

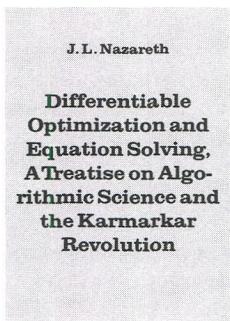
Ebenso hervorragend geeignet ist das Buch für einen Analytiker, der sich diesem Gebiet von der theoretischen Seite nähern möchte. Er wird mit der analytischen Präzision dieses Buches mehr als zufrieden sein

und – hoffentlich – mit einer inneren Befriedigung und einem Erstaunen erkennen, welche Beiträge die angewandte Analysis zu der Entwicklung dieses Anwendungsgebiets geleistet hat.

Zu guter Letzt sollte noch betont werden, dass die zahlreichen, perfekt illustrierten und dabei anspruchsvollen Beispiele dieses Buches ein optischer Genuss sind und dieses Buch für Mathematiker aller Ausrichtungen und Spielarten interessant macht.

Bremen

P. Maaß



Berlin u. a., Springer, 2003, 256 S., € 79,95

Die vorliegende Monographie beschäftigt sich mit einigen ausgewählten Verfahren zur Lösung von differenzierbaren Optimierungsproblemen und nichtlinearen Gleichungssystemen. Das Buch umfasst 15 Kapitel, die sich hauptsächlich mit der unrestringierten Optimierung und den linearen Programmen auseinandersetzen. Es enthält so gut wie keine Theoreme und auch keine Übungsaufgaben, ist allerdings auch nicht als Lehrbuch konzipiert. Vielmehr erwartet der Autor bereits entsprechende Grundkenntnisse aus dem Bereich der algorithmischen Optimierung.

Die Auswahl des Stoffes in den 15 Kapiteln hängt sehr stark von den persönlichen Forschungsinteressen des Autors ab. Im Wesentlichen werden hier die eigenen Forschungsleistungen des Autors zusammen-

hängend dargestellt. Aus diesem Grund fehlen wohl auch eine Reihe von wichtigen Teilgebieten der differenzierbaren Optimierung. Beispielsweise wird die nichtlineare restriktive Optimierung sträflich vernachlässigt, obwohl auch hier durch die sonst ausführlich vertretenen Inneren-Punkte-Methoden wesentliche Neuerungen stattgefunden haben.

Die einzelnen Kapitel beschäftigen sich meistens mit einem bestimmten Verfahren oder einer Klasse von Verfahren. Dabei handelt es sich überwiegend um bekannte Methoden aus dem Bereich der Optimierung. Diese Verfahren werden entweder nur angegeben oder, in einigen Fällen, auch hergeleitet. Danach werden mögliche Varianten besprochen, wie man sie in der Originalliteratur noch nicht findet. Somit enthält das Buch eine Reihe von Ideen, die allerdings nicht weiter ausgeführt werden.

Ein Beispiel mag die Vorgehensweise besser illustrieren. Wir betrachten dazu das Kapitel 5, welches sich mit den CG-Verfahren beschäftigt. Hier wird zunächst an das lineare CG-Verfahren zur Lösung eines linearen Gleichungssystems mit positiv definiter Koeffizientenmatrix erinnert. Anschließend folgt eine Wiederholung der wichtigsten CG-Verfahren zur Lösung des nichtlinearen Optimierungsproblems

$$\text{minimiere } f(x)$$

mit einer glatten Funktion  $f : \mathbb{R}^n \rightarrow \mathbb{R}$ . Diese sind meist von der Gestalt

$$x_{k+1} := x_k + \alpha_k d_k, \quad k = 0, 1, \dots$$

für eine Schrittweite  $\alpha_k > 0$  und eine Suchrichtung  $d_k$ , die bei den nichtlinearen CG-Verfahren ebenfalls einer Rekursion der Gestalt

$$d_{k+1} := -\nabla f(x_{k+1}) + \beta_k d_k$$

genügt, wobei unterschiedliche Vorschriften für  $\beta_k \in R$  verschiedene Verfahren liefern. Zwei bekannte Vertreter sind

$$\beta_k := \frac{(g_{k+1})^T g_{k+1}}{(g_k)^T g_k} \quad (\text{Fletcher-Reeves})$$

und

$$\beta_k := \frac{(g_{k+1})^T y_k}{(g_k)^T y_k} \quad (\text{Hestenes-Stiefel})$$

mit  $g_k := -\nabla f(x_k)$  und  $y_k := g_{k+1} - g_k$ . Der Autor schlägt nun vor, die neue Formel

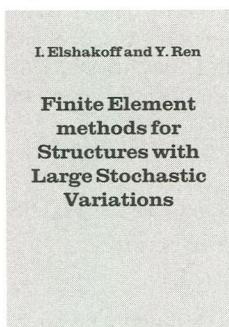
$$\beta_k := \frac{\lambda_k(g_{k+1}^T g_{k+1}) + (1 - \lambda_k)(g_{k+1}^T y_k)}{\mu_k(g_k^T g_k) + (1 - \mu_k)(d_k^T y_k)}$$

für zwei Parameter  $\lambda_k, \mu_k \in [0, 1]$  zu untersuchen. Konkrete Vorschläge für eine möglichst gute Wahl dieser Parameter finden sich nicht, und die zugehörigen numerischen Resultate sind ebenfalls nicht sehr aufschlussreich.

Diese Vorgehensweise ist typisch für die meisten Kapitel. Wer dies mag, dem wird das Buch eine Reihe von ähnlichen Ideen geben. Für alle anderen ist die vorliegende Monographie eher weniger geeignet.

Würzburg

C. Kanzow



Oxford University Press, 2003, 260 S., £ 45,-

This monograph is written by engineers for engineers, though applied mathematicians may find it an interesting source of information on the types of problems that arise in engineering and how engineers tackle them. There is an extensive list of over 300 references and the authors provide many useful historical comments about the development of the subject, which, loosely speaking, is

about stochastic finite element methods, although, strictly speaking, the finite elements considered are deterministic and the parameters in the problems being treated are random or stochastic. The authors thus use the expression finite element methods for stochastic problems, which they abbreviate FEMSP. They concentrate on the mean and covariance analysis of displacements in structures, restricting attention to linear and static problems. Their goal is to introduce non-perturbative methods, which allow large stochastic variations to be handled.

There are seven chapters, an epilogue, eleven short appendices as well as an extensive bibliography, author index and subject index. Chapter 1 briefly reviews the finite element formulation for beam bending and planar stress/strain analysis. The finite element methods for stochastic structures (FEMSS) is reviewed in Chapter 2, where traditional perturbation methods involving series expansions as well as homogeneous chaos are discussed and improvements suggested, which are then illustrated and compared in a very simple but instructive example. Chapter 3 treats FEMSS when an exact inverse of the stiffness matrix is available and Chapter 4 introduces exact solutions of stochastic shear and Bernoulli-Euler beams as benchmark problems. Variational principle-based FEM for stochastic beams, including stochastic versions of Bubunov-Galerkin and Rayleigh-Ritz methods, are discussed in Chapter 5 and element-level flexibility-based FEM for stochastic structures in Chapter 6. Finally, Chapter 7 provides a brief comparison of stochastic and interval matrix methods for the problems under consideration.

The monograph is very well written, the material is clearly explained and is illuminated through many examples. The numerous quotations from the cited literature further enhance its readability.

Frankfurt

P. Kloeden

# Alles ist Zahl!



Rudolf Taschner

## Der Zahlen gigantische Schatten

Mathematik im Zeichen der Zeit

2., verb. Aufl. 2005. 200 S. Geb.

EUR 39,90

ISBN 3-528-13211-6

## INHALT

Pythagoras: Zahl und Symbol - Bach: Zahl und Musik - Hofmannsthal: Zahl und Zeit - Descartes: Zahl und Raum - Leibniz: Zahl und Logik - Laplace: Zahl und Politik - Bohr: Zahl und Materie - Pascal: Zahl und Geist

## DAS BUCH

Wie sehr Zahlen die vielfältigen Aspekte des Daseins durchdringen, ist wenig bekannt, und kaum jemand scheint bisher ermessen zu haben, wie unfassbar weit der Zahlen lange Schatten reichen. Nicht was die Zahlen sind, wird hier erzählt, sondern was sie bedeuten.

Dass ein halbes Jahr nach Erscheinen der ersten Auflage bereits der Druck einer zweiten Auflage erfolgt, belegt die These, dass viele Menschen Mathematik vor allem als wesentlichen Bestandteil unserer Kultur empfinden und darüber mehr erfahren wollen. In der zweiten Auflage wurden einige Druckfehler korrigiert.

Die Anregung, das Buch durch einen Index zu ergänzen, hat der Verlag aufgegriffen; dadurch hat das Buch eine wertvolle Abrundung gewonnen.



Abraham-Lincoln-Straße 46  
D-65189 Wiesbaden  
Fax 0611.78 78-420

Änderungen vorbehalten.  
Erhältlich beim Buchhandel oder beim Verlag.

