ISNM

International Series of Numerical Mathematics

Volume 150

Managing Editors:

K.-H. Hoffmann, Bonn

D. Mittelmann, Tempe

Associate Editors:

R. E. Bank, La Jolla

H. Kawarada, Chiba

R. J. LeVeque, Seattle

C. Verdi, Milano

Honorary Editor:

J. Todd, Pasadena

Nonlinear Smoothing and Multiresolution Analysis

Carl Rohwer

Birkhäuser Verlag
Basel · Boston · Berlin

Author:

Carl Rohwer
Department of Mathematics
University of Stellenbosch
7602 Stellenbosch
South Africa

2000 Mathematics Subject Classification: Primary 00A69, 41A46, 42C40; Secondary 06, 47, 62, 65, 94

A CIP catalogue record for this book is available from the Library of Congress, Washington D.C., USA

Bibliographic information published by Die Deutsche Bibliothek

Die Deutsche Bibliothek lists this publication in the Deutsche Nationalbibliografie; detailed bibliographic data is available in the Internet at <http://dnb.ddb.de>.

ISBN 3-7643-7229-X Birkhäuser Verlag, Basel – Boston – Berlin

This work is subject to copyright. All rights are reserved, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, re-use of illustrations, broadcasting, reproduction on microfilms or in other ways, and storage in data banks. For any kind of use whatsoever, permission from the copyright owner must be obtained.

© 2005 Birkhäuser Verlag, P.O. Box 133, CH-4010 Basel, Switzerland

Part of Springer Science+Business Media

Printed on acid-free paper produced of chlorine-free pulp. TCF ∞

Printed in Germany

ISBN-10: 3-7643-7229-X

ISBN-13: 978-3-7643-7229-3

9 8 7 6 5 4 3 2 1

www.birkhauser.ch

Contents

Foreword	vii
Introduction	ix
1. Operators on Sequences	1
2. Basic Rank Selectors, Pulses and Impulses	9
3. <i>LULU</i> -Smoother, Signals and Ambiguity	21
4. <i>LULU</i> -Intervals and Similar Smoother	31
5. Smoothing and Approximation with Signals	43
6. Variation Reduction and Shape Preservation	51
7. Multiresolution Analysis of Sequences	71
8. The Discrete Pulse Transform	91
9. Fair Comparison with Linear Smoother	109
10. Interpretation and Future	127
References	131
Index	135

Foreword

This monograph is intended as a simple introduction to the so-called *LULU*-theory and the practical use of *LULU*-smoothers leading up to a full Multiresolution Analysis of any finite sequence. The attempt has been to present the subject in a way that is retrospectively ordered to some extent, but preserves some of the twisted paths that intuition initially suggested. It is based on the few publications in the subject, some already submitted and lectures given at conferences and on invitation at public and private institutions.

The ideas originated more than fifteen years ago when the author was involved in practical problems at the Institute for Maritime Technology in Simonstoun and advanced while the author was at the BMI in Stellenbosch. The author hopes that publication of this book will stimulate the interest of other researchers and takes responsibility for any remaining typographical or technical errors.

A growing love for this particular subject was one of the prime reasons for a return to academic life, and a perceived opportunity to consolidate various perspectives and results. It is however only the encouragement, guidance and help of several respected people that have resulted in the effort presented here. There is a deep gratitude for; Siegfried Göldner for teaching Mathematics, Gunther Meinardus for patience, good advice and courage, Johan de Villiers for researching with me in Spline Function, Vilmos Toti for pointing us towards Wavelets, Charles Chui for teaching these and Doron Lubinsky for trust, encouragement and walking many an academic mile. I also thank many colleagues and friends who have contributed by discussion or collaboration to the subject: Willie Conradie, Johannes Cronje, Tertius de Wet, Hans Eggers, Sven Ehrich, Frank Hampel, Trevor Hastie, Rolf Jeltsch, Henri Johnson, Gerhard Joubert, Odei Kao, Sergue Lapine, Dirk Laurie, Eberhard Malkowsky, Colin Mallows, Dana Murray, Lothar Reichel, Birgit Rohwer, Lourens Toerien, Johan van der Merwe, Marcel Wild, and the students that have attended my courses on the subject, especially J.P. Harper for his help with the sketches.

Finally, most of all, I thank Jan Malan for encouragement and support during the birth of the ideas, Lauretta Adams for her courage, patience and excellent word processing, Marguerite and our children for sharing me with *LULU*, and my parents to whom I dedicate this effort.

Introduction

The order that I create, my father said, is the order of life. To create this order I shape a face permitting adoration. Order is a sign of life and not it's cause. Introducing life creates order, and introducing order co-opts death. Order for the sake of order is a parody of life.

Antoine de Saint-Exupéry. *

Unlikely events do occur, but science, measurement and information are primarily concerned with tendencies, which we refer to in this book as “trends”. Perhaps the most basic trend is monotonicity, as exhibited in the arrow of time. Tides and waves exhibit a form of local trend, if the minor local trends and large global trends are overlooked (smoothed out). Anomalies, characterized by an apparent lack of trend do occur, and are often inherently brief in space and time. When there is a lack of obvious trend, there is still possibly a statistical trend in the shape of a distribution.

Numbers and mathematics have evolved from measurement and economical information storage and transfer. If mathematics is considered the language of science, and the functions of a language are examined, it is reasonable to deduce an order in the functions (uses) as for example in the first four functions of a language as given by Popper (18). Corresponding to each function, an associated value comparison is apparent.

	Function	Value
1.	Expressive	Revealing/not revealing
2.	Signalling	Efficiency/inefficiency
3.	Descriptive	Truth/falsity
4.	Argumentative	Validity/invalidity

It is natural to consider higher functions (uses), like the poetic function, since even a strict formal language like mathematics can convey wisdom by using metaphors and analogies for modelling events, and leads to the power of interpolation and extrapolation as evident in the impressive accomplishments of modern science. Science develops from simple models of relatively constant things, including recurring

*Imprompto translated by the author from a German translation.

events of constant period or frequency, and builds up to more complicated trends. Increasing speed, accuracy and data transfer rates have often led to an overabundance of data from which a desired “signal” or trend is to be extracted in a stable and computationally economical way. An early example is the effort produced by Gauss to determine an elliptic orbit (“signal”) from a sequence of measurements corrupted by undetermined and partly undeterminable perturbations produced by ray bending, instrument resolution, computational approximation, truncation and human error. A recent example is the transmission of information from an instrument on a distant planet. The relatively high data rate possible is an advantage but needs to be handled economically for good extraction of “signal”, given perturbations essentially of similar type as those handled by Gauss. Two distinct changes are visible; The amount of work involved to extract the wanted signal and the type of noise regularly introduced by digital processing. Celestial orbits and the determination of their parameters, particularly with the limits of technology of the times, have been replaced by more complicated models for which parameters have to be estimated. Specifically, the advent of digital measuring, calculation and transmission of an ever increasing amount of data has led to problems that are not only different in detail but in philosophical basis.

Whereas extrapolation requires the use of well motivated non-local models, interpolation can use simpler local models for predicting values between measurements under reasonably mild assumptions on the behavior of a function. Spline functions have been a natural development linking global continuity requirements with local support in a computationally convenient way, permitting a well-known idea, the Raleigh–Ritz–Galerkin approximation for solutions of differential equations, to expand into a powerful tool called the Finite Element Method. Using computationally convenient interpolants by functions with local support, and comparing with the best global approximant with the Lebesgue Inequality, justifies the procedure in that it shows the error to be no greater than a small factor of the best possible in a given norm. Thus local procedures have been made respectable, and the art of numerical analysis has been saved from the suspicion of being a collection of clever ad-hoc techniques to calculate parameters in a complex model for the purpose of explaining or predicting events. What spline theory has done to link global and local analysis, wavelet theory has done to periodic events; localizing some frequency based phenomena coherently (8).

Intelligent ad-hoc local procedures for the removal of “impulsive” noise from data have received considerable attention in the last two decades. Wisdom and experience has resulted in the widespread use of the (running-) median smoother and compositions of these. A lack of clear and consistent behavior has been widely lamented, but experience indicates that these are effective procedures. Examples of problems from practice suffice to indicate the extent of the problem.

Suppose a sequence of numbers is transmitted with a reasonable trend in that all are measurements of light intensity at a pixel in a camera. If any one of the pixels now and then registers a “glint”, or if one of the numbers now and then has a spurious influence in a significant bit, like a sign bit or exponent bit

of the number, there is the annoying presence of “impulsive” noise. This noise is characterized by briefness and unreasonable size relative to neighboring data. Even though this type of noise was also present in previous times, human intervention and wisdom easily identified such an “error” and interpolated from neighboring data to leave a relatively undamaged signal. For high speed data processing the problem is to balance complexity with speed and automation.

Two different concrete examples will suffice to illustrate the problem of impulsive noise in measurement. In the first a high sample rate representation of a missile trajectory, as measured by radar, is to be analyzed for various purposes. Various sections are seriously contaminated by large impulses. An elaborate analysis can reconstruct the general trajectory shape, and even features like control oscillations and other possible vibrations can be extracted by elaborate techniques. If this is to be done in real time, it is a different problem altogether.

Figure 1. Trajectory data corrupted by impulsive noise.

The second example is data from a speedometer of a small fast boat on a rough sea. The downward impulses, clearly due to the boat leaping out of the water, are significantly larger than the upward impulses. This bias is dependent on sea state. How is the bias to be removed to give a good speed estimate over a large range of sea states?

Statistics has provided an impressive intellectual framework for containing the uncertainty introduced by unexplained or random perturbation (“noise”) as exhibited by deviation from the expected “signals”. A philosophically consistent way of treating impulsive noise, in the presence of the usual random noise, is to consider the distribution of the random noise as “heavy tailed”. This leads to Robust Statistics, as developed by Huber, Hampel and others.

When economy of effort or speed is required a different approach was advocated by Tukey and others. The idea is to view a sequence x of data as the sum of signals, well-behaved noise from a reasonable distribution and added impulsive noise (arbitrarily large amplitude impulses occurring occasionally). A simple selector, like the running median, is used to remove impulses of prescribed width. This is followed by linear smoothing and/or linear signal extraction. The underlying idea is that a linear process “smears” the energy of impulses over neighboring sequence elements, and makes it more difficult to extract afterwards. It is therefore advan-

Figure 2. Data from a boat speedometer with impulsive noise.

tageous to remove and replace it, by comparison to immediate neighbors as early as possible, to lessen the damage by the subsequent linear smoothing processes. This is easy to understand and implement and relatively fast and economical, and leads to the subsequent established theory for signal extraction.

These simple nonlinear smoothers (selectors and combinations of them) have over the last two decades received considerable attention, particularly in the engineering literature. Comparisons and analysis have been based mainly on the transfer of distributions if the noise is identically independently distributed random noise. A characterization of “roots”, or sequences that remain unchanged by smoothing, did slowly emerge in terms of “edges”, “constant regions” and “oscillations”, but the “enigmatic behavior” and lack of theory has been almost universally lamented (31). The design and choice of such smoothers has been based mainly on wisdom and experience. Tukey and others popularized smoothers like the three point median (M_1), the three point median applied twice and followed by the five point running median ($M_2M_1^2$), the repeated three point median ($R = M_1^\infty$) and the recursive version (M_1^*) of the three point median. Other rank order selectors are often mentioned, particularly the use of min/max operators in image processing, but the emphasis was on running medians and compositions (“concatenations”) of them.

The idea of the median smoother links naturally with that of linear smoothers (filters) in the following way. Instead of projecting a given sequence x onto a chosen subspace (“signals”), a process that may be non-local, it is practical to consider sections of the sequence separately, by passing a “window” over it. Considering an element x_i of x and the neighbors in the window $\{x_{i-n}, \dots, x_i, \dots, x_{i+n}\}$, they are used in a projection onto a subspace (constant sequences, linear sequences or sampled higher order polynomials or trigonometric functions). The norm usually used

is the $p = 2$ norm, leading to a linear mapping, for instance onto the best constant approximation to $\{x_{i-n}, \dots, x_i, \dots, x_{i+n}\}$, from which a smoothed value, (the average of the elements in the window), is substituted for. This is repeated for each i .

Another motivation for this particular norm is usually loosely given as the “Central Limit Theorem”. Many decades ago Whitaker and Robinson, in their book “Calculus of Observations”, already pointed out the often inappropriate addiction to, or misinterpretation of, the central limit theorem by quoting a witty remark of Poincaré;

“Everybody believes in the exponential law of errors; the experimenters, because they think it can be proved by mathematicians; and the mathematicians, because they believe it has been established by observation.”

The Gaussian (normal-) distribution for “noise” on measurement can be motivated by the observation that if a large amount n of numbers, with typical roundoff error from a uniform distribution, are added or subtracted, the result is a spline distribution of order n results. For $n = 6$ this already is almost indistinguishable on a typical graph from a Gaussian. When other operations like division are however also employed, they may result in vastly different error distributions. This problem has been amplified by the vast increase in the number of calculations performed in a modern system. Different methods of adjusting have to be explored.

The convenience of a linear projection associated with the least squares norm is however a good motivation, leading to a simple computational and conceptual procedure. The value x_i in a sequence is replaced by the weighted average of it and its neighbors. (Or the output sequence is the discrete convolute with a fixed local sequence.) Although the linear projection in the window is idempotent, the global smoothing of the sequence, considered as a mapping between input sequence and output sequence, is not.

In the presence of impulsive noise it is natural to replace the least squares projection with an approximation in a more robust norm, the ℓ_1 -norm. This leads to the replacing of x_i by median $\{x_{i-n}, \dots, x_{i+n}\}$, and thus the (“running-”) median smoother M_n . Since a median is a good estimator of the average, the median and running average smoothers are similar in effect when the noise is “reasonable”, but when the noise is from a “heavy-tailed” distribution the median becomes superior.

It is natural to compare a median and similar smoothers by considering its “linear part” (14), providing a method of comparison. A general perception did persist that other methods of comparison and analysis are required to augment this comparison, but the concept of eigenanalysis of nonlinear operators, which is natural for linear operators, reeks of folly. There may however be considerable merit in it. Even a clear definition of signal, in the context of median smoothing, has generally been avoided and only emerged slowly. Once a clear characterization of “signals”, or the significant “roots” of median smoothers, emerged, it became natural to consider the problem of “best approximation” onto signals (4). Although this leads to a computationally unappetizing process, there is merit in the insight obtained and a surprising twist in interpretation.

A consistent framework for analysis and comparison of smoothers has been lacking. For linear filters the vectorspace framework has been useful, leading naturally to eigenanalysis and the associated analysis in the “frequency domain”.

A natural alternative framework seems to be the semi-group generated by the unsymmetric minimum and maximum selectors, and the natural order on this semi-group. Not only does this lead to computationally convenient smoothers, almost naturally vectorisable, but it leads to a coherent conceptual framework for analysis and comparison of other rank order smoothers and their composition. The so-called *LULU*-structure has emerged as worth aiming for, also in dimensions higher than 1. The *LULU*-operators are actually particular cases of Morphological Filters, as developed by Serra (29) and others. This was discovered by the author in the mid-1990s, and independently by Maragos and Schafer (15), who informed me shortly afterwards. There is however a strong belief that the specific case has particular value. It is an interesting quirk of logic that that slight relaxation of one of the smoother axioms of Mallows leads to the possibility of adding additional useful and natural axioms, that can be satisfied by practical and consistent smoothers.

The development of the so called *LULU*-theory in the following chapters is based on initial practical work on real problems, for which the author had not been educationally prepared, and subsequent occasional publications on the subject (19)–(25). It introduces the ideas and methods of proof developed previously, but deviates where hindsight suggests a development that is simpler, clearer and more suggestive in indicating search directions in the many facets not yet investigated. It attempts to introduce scientists and mathematicians with a general background into a complementary method of analysis of measurements. As Collatz (10) pointed out long ago, the concepts of order, inequalities and lattices are perhaps the more fundamental.

The final result presented is a novel Multiresolution Analysis (MRA), complementary to that of the discrete wavelet (MRA), which decomposes a sequence into subsequences of linearly decreasing resolution. It does so in a very consistent fashion and explains the accepted good performance of Median Transforms and provides many pleasant and unexpected surprises.

To prove a strong consistency, very important attributes of an operator, like Full Trend Preservation, Neighbor Trend Preservation and Co-idempotence are introduced, and demonstrated to be useful and natural. A similarly important measure of smoothness, Local Monotonicity is introduced, motivated and used extensively, and shown to link with the excepted natural measure of smoothness; Total Variation.

Beauty, simplicity and elegance of poetic metaphors have always been useful when exploring in science and engineering, as in chess. The author is convinced, from his own experience, that the *LULU*-theory is the natural self-contained route into the wonderful field of Mathematical Morphology for all scientists and engineers educated in the usual type of undergraduate curriculum. The hope is that it becomes a well used and useful scenic highway.