Applications of Mathematics

27

Gerhard Winkler

Image Analysis, Random Fields and Dynamic Monte Carlo Methods

A Mathematical Introduction

图象分析、随机场和动态蒙特卡罗方法

Springer-Verlag

世界图书出版公司

Gerhard Winkler

Image Analysis, Random Fields and Dynamic Monte Carlo Methods

A Mathematical Introduction

With 59 Figures

Springer-Verlag
世界图书出版公司
北京·广州·上海·西安

Gerhard Winkler

Mathematical Institute, Ludwig-Maximilians Universität,
Theresienstraße 39, D-80333 München, Germany

Mathematics Subject Classification (1991):
68U10, 68U20, 65C05, 93Exx, 65K10, 65Y05, 60J20, 62M40

ISBN 3-540-57069-1 Springer-Verlag Berlin Heidelberg New York
ISBN 0-387-57069-1 Springer-Verlag New York Berlin Heidelberg

Library of Congress Cataloging-in-Publication Data.
Winkler, Gerhard, 1946-
Image analysis, random fields and dynamic Monte Carlo methods: a mathematical introduction
Gerhard Winkler. p. cm. (Applications of mathematics; 27)
Includes bibliographical references and index.
ISBN 3-540-57069-1 (Berlin: acid-free paper). – ISBN 0-387-57069-1 (New York: acid-free paper)
1. Image analysis–Statistical methods. 2. Markov random fields. 3. Monte Carlo method.
I. Title. II. Series. TA1637.W56 1995 621.367'015192–dc20 94-24251 CIP

Printed in Germany

This reprint has been authorized by Springer-Verlag (Berlin/Heidelberg/New York) for sale in the People's Republic of China only and not for export therefrom.
Reprinted in China by Beijing World Publishing Corporation, 1999

Applications of Mathematics

1 Fleming/Rishel, **Deterministic and Stochastic Optimal Control** (1975)
2 Marchuk, **Methods of Numerical Mathematics,** Second Edition (1982)
3 Balakrishnan, **Applied Functional Analysis,** Second Edition (1981)
4 Borovkov, **Stochastic Processes in Queueing Theory** (1976)
5 Liptser/Shiryayev, **Statistics of Random Processes I: General Theory** (1977)
6 Liptser/Shiryayev, **Statistics of Random Processes II: Applications** (1978)
7 Vorob'ev, **Game Theory: Lectures for Economists and Systems Scientists** (1977)
8 Shiryayev, **Optimal Stopping Rules** (1978)
9 *Ibragimov/Rozanov,* ***Gaussian Random Processes*** *(1978)*
10 Wonham, **Linear Multivariable Control: A Geometric Approach,** Third Edition (1985)
11 Hida, **Brownian Motion** (1980)
12 Hestenes, **Conjugate Direction Methods in Optimization** (1980)
13 Kallianpur, **Stochastic Filtering Theory** (1980)
14 Krylov, **Controlled Diffusion Processes** (1980)
15 Prabhu, **Stochastic Storage Processes: Queues, Insurance Risk, and Dams** (1980)
16 Ibragimov/Has'minskii, **Statistical Estimation: Asymptotic Theory** (1981)
17 Cesari, **Optimization: Theory and Applications** (1982)
18 Elliott, **Stochastic Calculus and Applications** (1982)
19 Marchuk/Shaidourov, **Difference Methods and Their Extrapolations** (1983)
20 Hijab, **Stabilization of Control Systems** (1986)
21 Protter, **Stochastic Integration and Differential Equations** (1990)
22 Benveniste/Métivier/Priouret, **Adaptive Algorithms and Stochastic Approximations** (1990)
23 Kloeden/Platen, **Numerical Solution of Stochastic Differential Equations** (1992)
24 Kushner/Dupuis, **Numerical Methods for Stochastic Control Problems in Continuous Time** (1992)
25 Fleming/Soner, **Controlled Markov Processes and Viscosity Solutions** (1993)
26 Baccelli/Brémaud, **Elements of Queueing Theory** (1994)
27 Winkler, **Image Analysis, Random Fields and Dynamic Monte Carlo Methods** (1995)

To my parents, Daniel and Micki

Preface

This text is concerned with a probabilistic approach to image analysis as initiated by U. GRENANDER, D. and S. GEMAN, B.R. HUNT and many others, and developed and popularized by D. and S. GEMAN in a paper from 1984. It formally adopts the Bayesian paradigm and therefore is referred to as 'Bayesian Image Analysis'.

There has been considerable and still growing interest in prior models and, in particular, in discrete Markov random field methods. Whereas image analysis is replete with ad hoc techniques, Bayesian image analysis provides a general framework encompassing various problems from imaging. Among those are such 'classical' applications like restoration, edge detection, texture discrimination, motion analysis and tomographic reconstruction. The subject is rapidly developing and in the near future is likely to deal with high-level applications like object recognition. Fascinating experiments by Y. CHOW, U. GRENANDER and D.M. KEENAN (1987), (1990) strongly support this belief.

Optimal estimators for solutions to such problems cannot in general be computed analytically, since the space of possible configurations is discrete and very large. Therefore, dynamic Monte Carlo methods currently receive much attention and stochastic relaxation algorithms, like simulated annealing and various dynamic samplers, have to be studied. This makes up a major section of this text. A cautionary remark is in order here: There is scepticism about annealing in the optimization community. We shall not advocate annealing as it stands as a universal remedy, but discuss its weak points and merits. Relaxation algorithms will serve as a flexible tool for inference and a useful substitute for exact or more reliable algorithms where such are not available.

Incorporating information gained by statistical inference on the data or '*training*' the models is a further important aspect. Conventional methods must be modified to become computationally feasible or new methods must be invented. This is a field of current research inspired for instance by the work of A. BENVENISTE, M. MÉTIVIER and P. PRIOURET (1990), L. YOUNES (1989) and R. AZENCOTT (1990)–(1992). There is a close connection to learning algorithms for Neural Networks which again underlines the importance of such studies.

The text is intended to serve as an introduction to the mathematical aspects rather than as a survey. The organization and choice of the topics are made from the author's personal (didactic) point of view rather than in a systematic way. Most of the study is restricted to finite spaces. Besides a series of simple examples, some more involved applications are discussed, mainly to restoration, texture segmentation and classification. Nevertheless, the emphasis is on general principles and theory rather than on the details of concrete applications. We roughly follow the classical mathematical scheme: motivation, definition, lemma, theorem, proof, example. The proofs are thorough and almost all are given in full detail. Some of the background from imaging is given, and the examples hopefully give the necessary intuition. But technical details of image processing definitely are not our concern here.

Given basic concepts from linear algebra and real analysis, the text is self-contained. No previous knowledge of image analysis is required. Knowledge of elementary probability theory and statistics is certainly beneficial, but not absolutely necessary. The text should be suitable for students and scientists from various fields including mathematics, physics, statistics and computer science. Readers are encouraged to carry out their own experiments and some of the examples can be run on a simple home computer. The appendix reviews the techniques necessary for the computer simulations. The text can also serve as a source of examples and exercises for more abstract lectures or seminars since the single parts are reasonably selfcontained.

The general model is introduced in Chapter 1. To give a realistic idea of the subject a specific model for restauration of noisy images is developed step by step in Chapter 2. Basic facts about Markov chains and their multi-dimensional analogue – the random fields – are collected in Chapters 3 and 4. A simple version of stochastic relaxation and simulated annealing, a generally applicable optimization algorithm based on the Gibbs sampler, is developed in Chapters 4 through 6. This is sufficient for readers to do their own experiments, perhaps following the guide line in the appendix. Chapter 7 deals with the law of large numbers and generalizations. Metropolis type algorithms are discussed in Chapter 8. It also indicates the connection with combinatorial optimization. So far the theory of dynamic Monte Carlo methods is based on DOBRUSHIN's contraction technique. Chapter 9 introduces to the method of 'second largest eigenvalues' and points to recent literature. Some remarks on parallel implementation can be found in Chapter 10. It is followed by a few examples of segmentation and classification of textures in Chapters 11 and 12. They mainly serve as a motivation for parameter estimation by the pseudo-likelihood method addressed in Chapters 13 and 14. Chapter 15 applies random field methods to simple neural networks. In particular, a popular learning rule is presented in the framework of maximum *likelihood* estimation. The final Chapter 16 contains a selected collection of other typical applications, hopefully opening prospects to higher level problems.

The text emerged from the notes of a series of lectures and seminars the author gave at the universities of Kaiserslautern, München, Heidelberg, Augsburg and Jena. In the late summer of 1990, D. Geman kindly gave us a copy of his survey article (1990): plainly, there is some overlap in the selection of topics. On the other hand, the introductory character of these notes is quite different.

The book was written while the author was lecturing at the universities named above and Erlangen-Nürnberg. He is indebted to H.G. Kellerer, H. Rost and K.H. Fichtner for giving him the opportunity to hold this series of lectures on image analysis. Finally, he would like to thank G.P. Douglas for proof-reading parts of the manuscript and, last but not least, D. Geman for his helpful comments on Part I.

Gerhard Winkler

Table of Contents

Introduction

In this first chapter, basic ideas behind the Bayesian approach to image analysis are introduced in an informal way. We freely use some notions from elementary probability theory and other fields with which the reader is perhaps not perfectly familiar. She or he should not worry about that – all concepts will be made thoroughly precise where they are needed.

This text is concerned with *digital image analysis*. It focuses on the extraction of information implicit in recorded digital image data by automatic devices aiming at an interpretation of the data, i.e. an explicit (partial) description of the real world. It may be considered as a special discipline in *image processing*. The latter encompasses fields like image digitization, enhancement and restoration, encoding, segmentation, representation and description (we refer the reader to standard texts like ANDREWS and HUNT (1977), PRATT (1978), HORN (1986), GONZALEZ and WINTZ (1987) or HARALICK and SHAPIRO (1992)).

Image analysis is sometimes referred to as 'inverse optics'. Inverse problems generally are underdetermined. Similarly, various interpretations may be more or less compatible with the data and the art of image analysis is to select those of interest. *Image synthesis*, i.e. the 'direct problem' of mapping a real scene to a digital image will not be dicussed in this text.

Here is a selection of typical problems :

- Image restoration: Recover a 'true' two-dimensional scene from noisy data.
- Boundary detection: Locate boundaries corresponding to sudden changes of physical properties of the true three-dimensional scene such as surface, shape, depth or texture.
- Tomographic reconstruction: Showers of atomic particles pass through the body in various directions (transmission tomography). Reconstruct the distribution of tissue in an internal organ from the 'shadows' cast by the particles onto an array of sensors. Similar problems arise in emission tomography.
- Shape from shading: Reconstruct a three-dimensional scene from the observed two-dimensional image.
- Motion analysis: Estimate the velocity of objects from a sequence of images.
- Analysis of biological shape: Recognize biological shapes or detect anomalies.

We shall comment on such applications in Chapter 2 and in Parts IV and VI. Concise introductions are GEMAN and GIDAS (1991), D. GEMAN (1990). For shape from shading and the related problem of shape from texture see GIDAS and TORREAO (1989). A collection of such (and many other) applications can be found in CHELLAPA and JAIN (1993). Similar problems arise in fields apparently not related to image analysis:

- Reconstruct the locations of archeological sites from measurements of the phosphate concentration over a study region (the phosphate content of soil is the result of decomposition of organic matter).
- Map the risk for a particular disease based on observed incidence rates.

Study of such problems in the Bayesian framework is quite recent, cf. BESAG, YORK and MOLLIÉ (1991).

The techniques mentioned will hopefully be helpful in high-level vision like object recognition and navigation in realistic environments.

Whereas image analysis is replete with ad hoc techniques one may believe that there is a need for theory as well. Analysis should be based on precisely formulated mathematical models which allow one to study the performance of algorithms analytically or even to design optimal methods. The probabilistic approach introduced in this text is a promising attempt to give such a basis. One characterization is to say it is Bayesian. As always in Bayesian inference, there are two types of information: prior knowledge and empirical data. Or, conversely, there are two sources of uncertainty or randomness since empirical data are distorted ideal data and prior knowledge usually is incomplete.

In the next paragraphs, these two concepts will be illustrated in the context of restoration, i.e. 'reconstruction' of a real scene from degraded observations. Given an observed image, one looks for a 'restored image' hopefully being a better represention of the true scene than was provided by the original records. The problem can be stated with a minimum of notation and therefore is chosen as the introductory example.

In general, one does not observe the ideal image but rather a distorted version. There may be a loss of information caused by some deterministic noninvertible transformation like blur or a masking deformation where only a portion of the image is recorded and the rest is hidden to the observer. Observations may also be subject to measurement errors or unpredictable influences arising from physical sources like sensor noise, film grain irregularities and atmospheric light fluctuations. Formally, the mechanism of distortion is a deterministic or random transformation $y = f(x)$ of the true scene x to the observed image y. 'Undoing' the degradations or 'restoring' the image ideally amounts to the inversion of f. This raises severe problems associated with invertibility and stability. Already in the simple linear model $y = Bx$, where the true and observed images are represented by vectors x and y, respectively, and the matrix B represents some linear 'blur operator', B is in general highly noninvertible and solutions x of the equation can be far apart. Other difficulties come in since y is determined by physical sampling and the elements of B are specified independently by system modeling. Thus the system of equations may be inconsistent in practice and have no solution at all. Therefore an error term enters the model, for example in the additive form $y = Bx + e(x)$.

Restoration is the object of many conventional methods. Among those one finds *ad hoc* methods like 'noise cleaning' via smoothing by weighted moving averages or – more generally – application of various linear filters to the image. Surprising results can be obtained by such methods and linear filtering is a highly developed discipline in engineering. On the other hand, linear filters only transform an image (possibly under loss of information), hopefully, to a better representation, but there is no possibility of *analysis*.

Another example is inverse filtering. A primitive example is least-square inverse filtering: For simplicity, suppose that the ideal and the distorted image are represented by rectangular arrays or real functions x and y on the plane giving the distribution of light intensity. Let $y = Bx + \eta$ for some linear operator B and a noise term η. An image $\hat{x}$ is a candidate for a 'restoration' of y if it minimizes the distance between y and Bx in the L^2-norm; i.e. the function $x \mapsto \|y - Bx\|_2^2$ (for an array $z = (z_s)_{s \in S}$, $\|z\|_2^2 = \sum_s z_s^2$). This amounts to the criterion to minimize the noise variance $\|\eta\|_2^2 = \|y - Bx\|_2^2$. A final solution is determined according to additional criteria. The method can be interpreted as minimization of the quadratic function $z \mapsto \|y - z\|_2^2$ under the 'rigid' constraint $z = Bx$ and the choice of some $\hat{x}$ satisfying $\hat{z} = B\hat{x}$ for the solution $\hat{z}$. The constraint $z = Bx$ mathematically expresses the prior information that x is transformed to Bx.

If the noise variance is known one can minimize $x \mapsto \|y - x\|_2^2$ under the constraint $\|y - Bx\|_2^2 = \sigma^2$ where σ^2 denotes noise variance. This is a simple example of constrained smoothing.

Bayesian methods differ from most of these methods in at least two respects: (i) they require full information about the (probabilistic) mechanism which degrades the original scene, (ii) rigid constraints are replaced by weak ones. These are more flexible: instead of classifying the objects in question into allowed and forbidden ones they are weighted by an 'acceptance function' quantifying the degree to which they are desired or not. Proper normalization yields a probability measure on the set of objects - called the 'prior distribution' or prior. The Bayesian paradigm allows one to consistently combine this 'weak constraint measure' with the data. This results in a modification of the prior called *posterior distribution* or posterior. Here the more or less rigid expectations compete with faithfulness to the data. By a suitable decision rule a solution to the inverse problem is selected, i.e. an image hopefully in proper balance between prior expectations and fidelity to the data.

To prevent fruitless discussions on the Bayesian philosophy, let us stress that though the model formally is Bayesian, the prior distribution can be just considered as a flexible substitute for rigid constraints and, from this point of view, it is at least in the present context an analytical rather than a probabilistic concept. Nevertheless, the name 'Bayesian image analysis' is common for this approach. Besides its formal merits the Bayesian framework has several substantial advantages. Methods from this mature field of statistics can be adopted or at least serve as a guideline for the development of more specific methods. In particular, this is helpful for the estimation of optimal solutions. Or, in texture classification, where the prior can only be specified up to a set of parameters, statistical inference can be adopted to adjust the parameters to a special texture.

All of this is a bit general. Though of no practical importance, the following simple example may give you a flavour of what is to come.

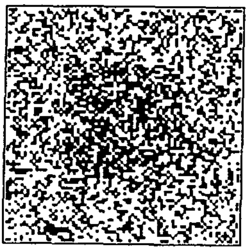

Fig. 0.1. A degraded image

Consider black and white pictures as displayed on a computer screen. They will be represented by arrays $(x_s)_{s\in S}$; S is a finite rectangular grid of 'pixels' s, $x_s = 1$ corresponds to a black spot in pixel s and $x_s = 0$ means that s is white. Somebody (nature ?) displays some image y (Fig. 1).

We are given two pieces of information about the generating algorithm: (i) it started from an image x composed of large connected patches of black and white, (ii) the colours in the pixels were independently flipped with probability p each. We accept a bet to construct a machine which roughly recovers the original image. There are 2^σ possible combinations of black and white spots, where σ is the number of pixels. In the figures we chose $\sigma = 80 \times 80$ and hence $2^\sigma \sim 10^{192}$; in the more realistic case $\sigma = 256 \times 256$ one has $2^\sigma \sim 10^{19,660}$. We want to restrict our search to a small subset using the information in (i). It is not obvious how to state (i) in precise mathematical terms. We may start selecting only the two extreme images which are either totally white or totally black (Fig. 2). Formally, this amounts to the choice of a feasible subset of the space $\mathbf{X} = \{0,1\}^S$ consisting of two elements. This is a poor formulation of (i) since it does not express the degrees to which for instance Fig. 3(a) and (b) are in accordance with the requirement: both are forbidden. Thus let us introduce the local constraints

- $x_s = x_t$ for all pixels s and t adjacent in the horizontal, vertical or diagonal directions.

In the example, we have $n = 80$ rows and columns, respectively, and hence $2n(n-1) = 12,640$ adjacent pairs s,t in the horizontal or vertical directions, and the same number of diagonally adjacent pairs. The feasible set is the same as before but weighting configurations x by the number $A(x)$ of valid constraints gives a measure of smoothness. Fig. 3(a) differs from the black

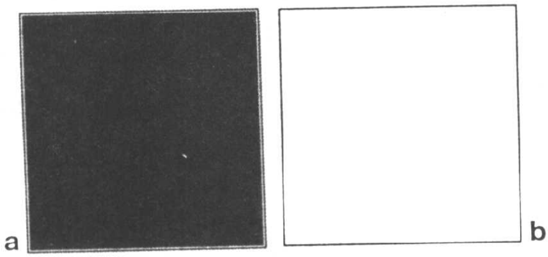

Fig. 0.2. Two very smooth images

image only by a single white dot and thus violates only 8 of the 25, 280 local constraints whereas (b) violates one half of the local constraints. By the rigid constraints both are forbidden whereas A differentiates between them. This way the rigid constraints are relaxed to 'weak constraints'. Hopefully, the reader will agree that the latter is a more adequate formulation of piecewise smoothness in (i) than the rigid ones.

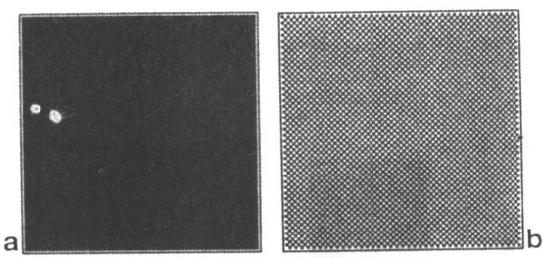

Fig. 0.3. (a) Violates few, (b) violates many local constraints

More generally, one may define local acceptor functions by

$$A_{st}(x_s, x_t) = \begin{cases} a_{st} & if \quad x_s = x_t \\ r_{st} & if \quad x_s \neq x_t \end{cases}$$

(a for 'attractive' and r for 'repulsive'). The numbers a_{st} and r_{st} control the degree to which the rigid local constraints are fulfilled. For the present, they are not completely specified. But if we agree that $A_{st}(x_s, x_t) > A_{st}(x'_s, x'_t)$