DISCRETIZATION AND BAYESIAN MODELING IN INVERSE PROBLEMS AND IMAGING

Tapio Helin

DISCRETIZATION AND BAYESIAN MODELING IN INVERSE PROBLEMS AND IMAGING

Tapio Helin

Dissertation for the degree of Doctor of Science in Technology to be presented, with due permission of the Faculty of Information and Natural Sciences, for public examination and debate in auditorium K at Aalto University, School of Science and Technology (Espoo, Finland) on the 12th of March 2010, at 12 o'clock noon.

Tapio Helin
Department of Mathematics and Systems Analysis
Aalto University
P.O. Box 11100,   FI-00076 Aalto,   Finland
E-mail: tapio.helin@tkk.fi

Abstract: In this thesis the Bayesian modeling and discretization are studied in inverse problems related to imaging. The treatise consists of four articles which focus on the phenomena that appear when more detailed data or a priori information become available. Novel Bayesian methods for solving ill-posed signal processing problems in edge-preserving manner are introduced and analysed. Furthermore, modeling photographs in image processing problems is studied and a novel model is presented.

Tiivistelmä: Väitöskirjassa tutkitaan bayesiläistä mallintamista ja diskretointia kuvantamiseen liittyvissä inversio-ongelmissa. Työ koostuu neljästä artikkelista, jotka tarkastelevat ilmiöitä liittyen tarkemman datan tai a priori tiedon hyödyntämiseen. Väitöskirjassa esitellään huonosti asetetuille signaalinkäsittelyongelmille uusia bayesiläisiä ratkaisumenetelmiä, jotka pyrkivät signaalin epäjatkuvuuksien tarkkaan rekonstruointiin. Lisäksi työssä tutkitaan valokuvien esittämistä resoluutiovapaasti ja esitellään uusi malli tähän tarkoitukseen.

Preface

This work was done at the Institute of Mathematics, Helsinki University of Technology in Finland, 2006–2009. I am thankful for Professor Olavi Nevanlinna, the director of the institute and my supervisor, for providing an encouraging research enviroment. Considerable amount of the thesis was written during my frequent visits to the Johann Radon Institute for Computational and Applied Mathematics (RICAM) in Linz, Austria. For having this opportunity I am greatly indebted to director of the institute, Professor Heinz W. Engl.

It has been a privilege to work under the guidance of an outstanding researcher such as Professor Matti Lassas. I am grateful for the time he spent listening to my troubles and the flexible attitude he had towards me. Not once did I leave his office empty-handed.

It was an honour to have Professor Thorsten Hohage and Professor Lassi Päivärinta as the pre-examiners of this dissertation. In addition, I sincerely thank Professor Päivärinta, Director of the Finnish Centre of Excellence in Inverse Problems, for the constant support and numerous enthusiastic discussions during this project. During the final hours of this journey, I am grateful to have Professor Martin Burger as my opponent.

I have had the pleasure to work with and discuss mathematics with many inspiring people. During the several brainstorming sessions with my co-author Professor Samuli Siltanen I learned how rewarding collaboration can be. At the end of a typical work week I had the chance to relax during the institute's floorball match. By far, the best passes I received during those games were delivered by Docent Nuutti Hyvönen. For these and maybe the more important good passes on the field of mathematics, I thank him sincerely. When faced with an abstract question about random variables, I found my solutions converging in the hands of PhD Petteri Piiroinen and Dr. Lasse Leskelä. During the entire process I made many good friends at TKK, University of Helsinki and RICAM. I thank them all for being such nice individuals.

Completion of this research would not have been possible without the financial support of the Emil Aaltonen Foundation, the Graduate School of Inverse Problems, the Vilho, Yrjö and Kalle Väisälä Foundation and Helsingin Sanomat Foundation, for which I am sincerely grateful.

Finally, there are three more individuals that receive my thanks beyond any written word. My parents, Markku and Sirpa, for the care and support since day one. And Hanna, for supporting my efforts in mathematics, but more importantly, for the unconditional love, understanding and patience.

Linz, February 19, 2010 Tapio Helin

The thesis consists of this overview and the following papers:

Publications

[I] T. HELIN, *On infinite-dimensional hierarchical probability models in statistical inverse problems*, Inverse probl. imaging **3**(4) (2009), pp. 567–597.

[II] T. HELIN AND M. LASSAS, *Bayesian signal restoration and Mumford–Shah functional*, Proc. Appl. Math. Mech. **7** (2007), pp. 2080013–2080014.

[III] T. HELIN AND M. LASSAS, *Hierarchical models in statistical inverse problems and the Mumford–Shah functional*, arXiv:0908.3396v2

[IV] T. HELIN, M. LASSAS AND S. SILTANEN, *Infinite photography: new mathematical model for high-resolution images*, J. Math. Imaging Vision **36**(2) (2010), pp. 140–168.

Own contribution to publications

[I] This paper is an independent research of the author.

[II] The major ideas and the writing are due to the author. The paper has appeared in a refereed conference proceedings and is based on the presentation of the author.

[III] The author has a major role in this paper. The numerical application is an independent work of the author. A major part of the writing is done by the author.

[IV] The author has made major contributions to the theoretical analysis of the model in Sections 2, 3 and 4. Most of the numerical applications and a major part of the writing are due to the author.

Contents

1 Introduction

This thesis studies Bayesian modeling and discretization in inverse problems related to imaging. Signal processing problems are the framework for study in papers [I-III] whereas paper [IV] focuses on two dimensional image processing. The common methodological goal is to study and give answers to some questions raised in previous work [44]. With these observations in mind we shed more light to the phenomena which appear when more data or more detailed a priori information become available. Furthermore, we introduce novel solutions in the Bayesian approach to signal processing problems and study the modeling of photographs in image processing. Let us begin by explaining some background to our work.

Defining an inverse problem simultaneously requires defining a forward problem. In imaging problems arising from physics this distiction is often intuitive. For the sake of argument, consider the physics behind the computed tomography (CT) imaging. This imaging modality is based on acquiring several X-ray projection images of an object from different directions. The images are then used to generate a three dimensional image of the attenuation of X-ray inside the object [54]. The forward problem is solving the X-ray images given the attenuation properties of the object. Such a problem is numerically stable and can be solved very reliably. However, the inverse problem, which is finding the attenuation given a set of X-ray images, is difficult and in mathematical terms called *ill-posed*. The vast majority of inverse problems are characterized by ill-posedness.

In mathematical terminology a problem is called *well-posed* in the sense of Hadamard [33] if it satisfies following three conditions:

(i) There exists a solution to the problem (existence).

(ii) There is at most one solution to the problem (uniqueness).

(iii) The solution depends continuously on the data (stability).

Accordingly, a problem is defined ill-posed if one or more of these conditions are not satisfied.

Violating anyone of the three properties produces very distinctive challenges in solving the problem. Clearly, the existence of a solution is guaranteed if the measurement belongs to the solution space of the forward problem. This may not be the case since any physical measurement can be contaminated by noise. In this case the notion of a solution may need adjusting. Similarly, uniqueness is of high importance in many situations and lack of uniqueness means that the data are not sufficient for determining the solution. For non-linear problems proving uniqueness in enough generality may turn out to be extremely challenging. For instance, consider the evolution of the inverse conductivity problem [5, 53, 65]. In practise, the non-uniqueness appears often due to the finite amount of data. In the example of CT imaging this reflects the inevitable practical problem that only a very limited number of X-ray images can be taken and hence the inverse problem becomes

underdetermined. In conclusion, a mathematical remedy to existence and uniqueness considerations can be the reformulation of the problem.

The most difficult property for numerical algorithms to handle is the instability of the problem. In other words, small measurement noise can cause an arbitrarily large error in the solution. This is the case for most inverse problems. In the classical *regularization* approach the goal is to produce a reasonable estimate of the solution. This is done by introducing a slight modification of the original problem in order to make it stable. Naturally, by doing so, one introduces some new error to the method and hence it is of interest to keep the modification as small as possible. Roughly, a regularization method balances between errors produced by the modification and the remaining instability. These methods have proved to be efficient and research remains active to date. Among a number of excellent textbooks on the topic we mention [24, 41, 66].

A fundamental assumption in regularization theory is that the magnitude of the noise, i.e. the noise level, in the measurement is known or can be accurately estimated. Standard regularization methods are known to converge with respect to the noise level. In a large number of applications, statistical modeling of the measurement error makes more sense [25]. In such cases, the problem becomes a question of statistical inference. Given the measurement and all other information available, what do we know about the solution?

The main paradigms of statistical inference are frequentist and Bayesian. The core difference between the two is in the interpretation of probability. Whereas frequentists see the probability of an event only as its relative frequency over time, the Bayesian school gives the notion of probability a subjective status [9]. The latter approach assumes probability measures the degree of belief in the outcome. Consider throwing of a dice (frequentist probability) or gambling odds (Bayesian probability) as an example of the difference in thinking.

The frequentist approach assumes that the value of the unknown is deterministic. Various statistical estimation techniques can be used, however, frequentists typically lean towards non-Bayesian statistical decision theory [10, 25]. This approach aims to build estimators of the unknown which minimize some specific *risk* function. A widely used non-Bayesian estimator is the *maximum likelihood* estimate which is the solution that is most likely to produce the measured data. It often fails to address the instability of inverse problems and requires deterministic regularization to also be applied.

In opposition to the frequentist approach, the Bayesian approach seeks to utilize prior information in the form of a probability distribution called the *prior* distribution. When the measurement is obtained the prior knowledge is updated into the *posterior* distribution using the measurement model and Bayes formula. Hence the fundamental difference between Bayesian and other approaches can be summarized by noting that the Bayesian solution to an inverse problem is a probability distribution rather than a single point estimate or a confidence interval. The posterior distribution yields a straightforward method to assess the reliability of any value of the solution.

In connection to classical methods the role of the prior information in the Bayesian paradigm is to regularize the problem. Many similarities exist with the frequentist approach when the goal is regularization of the ill-posedness. For example, operating with Gaussian distributions in Bayesian scheme and regularizing with Tikhonov-based methods in a frequentist scheme can be shown to produce same point estimates. From the point of view of statistical decision theory these methods should be seen to complement each other.

Bayesian methods have been widely applied to practical applications of inverse problems [7, 39, 42, 64], but there are still crucial open questions in the theory. The key question in this thesis is to determine if the Bayesian approach can be successful in infinite dimensional problems. This problem has been studied with Gaussian prior and noise distributions since the paper by Franklin [27], but the use of non-Gaussian information has turned out to be much more difficult. Such problems have been recently studied in [44, 45, 56]. The methodological goal in [I-III] is to show that so-called hierarchical models leading to non-Gaussian probability distributions can be used in infinite dimensional Bayesian inverse problems.

The second methodological question approached in [I-III] is related to discretization of inverse problems and performance of the hierarchical models when the discretization levels are increased. For example, in the signal processing problems this could be an increase in sampling frequency. Questions to consider are, if the noise model for each signal is Gaussian, does the solution stay stable? Do the solutions stay consistent with the prior? Papers [I-III] show that under hierarchical models the performance of different Bayesian point estimates depends on the asymptotics of the noise.

The novel aspect of solving ill-posed signal processing problems in [I-III] is in introducing Bayesian methods that are *edge-preserving*. This means that the method is specialized in preserving sharp edges. For example, Gaussian modeling often results to unwanted smoothing in the reconstructions. As a point of interest, computing CM estimate with the widely used total-variation priors has been shown to smoothen such edges asymptotically [44].

Let us finally discuss the results in paper [IV]. Among the different approaches to inverse problem the common step where all methods utilize prior information is when one chooses the underlying space for the unknown. In what space photographs should be modeled is an open question in image processing. There is a long tradition to consider them as functions of bounded variation and use methods based on variational calculus (see e.g. [60]) but controversy remains [30]. Recently, general Banach space models have been presented to tackle problems with rapid oscillation patterns [50]. In [IV] a novel image model is introduced which considers a simplified physical process of a CCD sensor. This problem falls outside the scope of typical inverse problems research, however, it is of fundamental value in image processing problems.

This text is organized as follows: Section 2 reviews the theoretical background of Bayesian inversion. The related methodological implications of the results in [I-III] are discussed in the light of the general theory. In Section

3 the edge-preserving Bayesian methods introduced in articles [I-III] and the related convergence results are explained in more detail. The findings of paper [IV] and the related research in image processing problems are reviewed in Section 4. Finally, Section 5 is devoted to a discussion about the results and their implications.

2 Bayesian inversion

2.1 The Bayes model

Consider a quantity u that can be observed indirectly through a relation

$$m = Au + e$$

where e is an additive noise. Throughout the thesis we assume that A is linear and bounded. Linearity does not play a central role in the results and is chosen to maintain a more readable presentation. Instead of linearity, suitable continuity and boundedness assumptions on A often suffice.

The Bayesian approach to find u given m is to reformulate the problem as a question of statistical inference incorporating all prior information available. The quantities are viewed as random variables where the subjective knowledge is encoded into the probability distributions [39]. Furthermore, the randomness depicts the lack of knowledge regarding the true values of the quantities. The Bayesian view is that the measurement is a sample of random variable

$$M = AU + \mathcal{E} \tag{1}$$

where U and $\mathcal{E}$ model the values of u and e, respectively. Consequently, the Bayesian paradigm asks

$$\text{given a realization } m = M(\omega_0), \text{ what we know about } U. \tag{2}$$

In real life situations the measurement and the computed solution are always finite dimensional. This thesis focuses on the asymptotical questions related to their dimensions and hence we distinguish a *computational model*

$$M_{kn} = A_k U_n + \mathcal{E}_k \tag{3}$$

from the *ideal model* (1). In equation (3) parameters $k, n \in \mathbb{N}$ are related to discretizing the measurement and the unknown, respectively.

The infinite dimensional linear model and its discretization have been studied extensively for Gaussian distributions in [26, 27, 43, 46, 47, 48]. Still, for non-Gaussian probability distributions the interplay between equations (1) and (3) remains one of the fundamental open questions in Bayesian inversion. We return to the convergence issues below but point out the following observation. Often in literature one implicitly assumes that the finite dimensional measurement is produced by random variable M_{kn}. This is not immediately clear since the nature itself does not discretize U and hence the

11

right model for the measurement should be $M_k = A_k U + \mathcal{E}_k$. An interesting perspective is to consider this discrepancy as *model reduction* [4, 40] such that the model error $A_k(U - U_n)$ is included in the random noise. This approach was first analysed in [40] with Gaussian distributions. Later it was shown in [45] that with Gaussian noise such assumption is not necessary.

Looking from a practical point of view, the process presented above can also be seen reversed. Namely, in a real life application one can have a range of possible choices for k and hence some idea how the statistics of $\mathcal{E}_k$ develops. In this case, the corresponding stability issues are obviously of interest. A Bayesian solution method that works coherently on all these scales is called *discretization invariant*. In the following sections we discuss and comment on this property in connection to converging posterior distributions as well as point estimates. Notice that the essence of discretization invariance is not only in convergence issues but also in discretization independent *qualitative* properties of the estimates. The qualitative incoherence of point estimates is discussed in Section 3. Finally, it should be recognized that a limiting model like (1) might not exist. We discuss this and the role of the limiting noise $\mathcal{E}$ more below.

2.2   The posterior distribution

The Bayesian answer to problem (2) is the conditional distribution of the unknown given the measurement. This is called the posterior distribution. When a density function presentation is available in the finite dimensional problems, the Bayes formula has a straightforward form.

Consider the following framework for the computational problem. Let $(H_1, \langle \cdot, \cdot \rangle_1)$ and $(H_2, \langle \cdot, \cdot \rangle_2)$ be two real Hilbert spaces such that $\dim H_1 = N$ and $\dim H_2 = K$. Suppose U_n obtains realizations in H_1 and M_{kn} is modeled in H_2. Furthermore, let $\mathcal{I} : H_1 \to \mathbb{R}^N$ and $\mathcal{J} : H_2 \to \mathbb{R}^K$ be two arbitrary isometries. Let us then map equation (3) to a matrix equation

$$\mathbf{M}_{kn} = \mathcal{J} M_{kn} = \mathbf{A}_{kn} \mathbf{U}_n + \mathbf{E}_k \tag{4}$$

where $\mathbf{A}_{kn} = \mathcal{J} A_k \mathcal{I}^{-1} \in \mathbb{R}^{K \times N}$, $A_k = P_k A$ and $\mathbf{U}_n = \mathcal{I} U_n : \Omega \to \mathbb{R}^N$. If all the probability distributions of the random variables above are absolutely continuous with respect to Lebesgue measure, by the Bayes formula the posterior density π_{kn} has a form

$$\pi_{kn}(\mathbf{u} \mid \mathbf{m}) = \frac{\Pi_n(\mathbf{u}) \Gamma_{kn}(\mathbf{m} \mid \mathbf{u})}{\Upsilon_{kn}(\mathbf{m})} \tag{5}$$

where $\mathbf{u} \in \mathbb{R}^N$ and $\mathbf{m} \in \mathbb{R}^K$. In equation (5) functions Π_n and Γ_{kn} represent the densities of the prior distribution of $\mathbf{U}_n$ and the *likelihood* distribution of $\mathbf{M}_{kn}$ given $\mathbf{U}_n = \mathbf{u}$, respectively, and Υ_{kn} is the density of $\mathbf{M}_{kn}$.

Notice that due to the connection between $\mathbf{m}$ and $\mathbf{u}$ via (3) the likelihood term depicts the distribution of the noise. We also point out that if $\Upsilon_{kn}(\mathbf{m}) = 0$ happens, the prior modeling has failed. Namely, the measurement can not be predicted by the model.

The density function presentation proves powerful in many respect. However, in infinite dimensional problems the Lebesgue measure is not available and hence this presentation is not meaningful. The first question about problem (1) is obviously whether the posterior distribution exists and can be defined. The answer turns out to be 'yes' even in the general framework of Polish spaces, i.e., separable and completely metrizable topological spaces. An existence result can be found in [23, Section 10.2].

Let us consider problem (1) and let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space. Furthermore, let $(G, \mathcal{G})$ and $(H, \mathcal{H})$ be measurable spaces such that $U : \Omega \to G$ and $M : \Omega \to H$ are random variables. Denote the probability distributions of U and M with μ_U and μ_M, respectively. The following theorem and a proof is given in [61, Theorem 1.31].

Theorem 1 *Let ν be a measure in $(H, \mathcal{H})$. Assume that there exists a conditional distribution of M given U and denote it by $\mu_{M|U}(\cdot \mid u)$. Furthermore, assume that $\mu_{M|U}(\cdot \mid u)$ is absolutely continuous with respect to ν for μ_U-almost all $u \in G$ with the Radon-Nikodym derivative*

$$\Gamma_{M|U}(m \mid u) := \frac{d\mu_{M|U}(\cdot \mid u)}{d\nu}(m).$$

In addition, suppose that the mapping $(u, m) \mapsto \Gamma_{M|U}(m \mid u)$ is measurable in $(G \times H, \mathcal{G} \times \mathcal{H})$. Then a conditional distribution of U given M exists, $\mu_{U|M}(\cdot \mid m)$ is absolutely continuous with respect to μ_U for μ_M-almost all $m \in H$, and

$$\mu_{U|M}(A \mid m) = \frac{\int_A \Gamma_{M|U}(m \mid u)\mu_U(du)}{\int_G \Gamma_{M|U}(m \mid u)\mu_U(du)}, \quad A \in \mathcal{G}, \tag{6}$$

for μ_M-almost all $m \in H$.

We notice that a 'reference measure' ν is still required to obtain a likelihood density. The key notion in the recent results in [45] is to use a choice of $\nu = \mu_M$ for linear inverse problems with additive Gaussian noise. This can be shown to satisfy the assumptions above due to the Cameron-Martin theorem [13]. Furthermore, in [45] the formula is used to obtain convergence of posterior distributions between problems (3) and (1). In paper [I] the convergence result is generalized by relaxing the assumption on the convergence of prior distributions. Finally, we point out that convergence of the posterior can be studied also from different perspectives [55, 56].

2.3 Point estimates

The role of Bayesian point estimates is, likewise to classical regularization methods, to describe the solution with a single value. However, the theory proves versatile since the estimate can be intuitively chosen to depict different properties of the posterior. The two most common point estimates used in Bayesian inversion and also studied in [I-III] are the conditional mean (CM)

and the maximum a posteriori (MAP) estimate. Let us first shed light to the first of these two.

The CM estimate is defined as the expectation value of the posterior distribution. Given the density function representation and the notations of Section 2.2 it is defined as the integral

$$\mathbf{u}_{kn}^{CM} = \int_{\mathbb{R}^N} \mathbf{u}\pi_{kn}(\mathbf{u} \mid \mathbf{m})d\mathbf{u}$$

and hence the CM estimate for problem (3) is

$$u_{kn}^{CM} = \mathcal{I}^{-1}\left(\mathbf{u}_{kn}^{CM}\right) \in H_1.$$

The numerical implementation requires computation of the integral. In practical applications the dimensionality of the integral increases rapidly and thus sophisticated numerical methods are required. Usually the value is approximated by Markov Chain Monte Carlo (MCMC) methods [34, 49]. In the numerical examples in [I] we have implemented a method called single component adaptive Metropolis [31, 32]. Unfortunately, even adaptive methods are not feasible without enormous computing power to problems like medical imaging where very high dimensional objects are involved. However, it is the opinion of the author that the evolution of computers together with efficient parallelization of the algorithm can develop it into a very competitive and stable choice.

The definition of a CM estimate can be extended to infinite dimensional setting using Bochner integrals (see [22]). Notice that the posterior measure obtained from Bayes formula (6) makes sense for almost all measurements. Hence, conditioning on a single measurement requires careful consideration. In [45] a concept of *reconstructor* is introduced to overcome this difficulty.

Definition 1 *Denote by $\mathcal{M} \subset \Sigma$ the σ-algebra generated by the random variable M. We say that any deterministic function*

$$\mathcal{R}_M(U \mid \cdot) : H \to G, \quad m \mapsto \mathcal{R}_M(U \mid m),$$

is a reconstructor of $U \in L^1(\Omega, \Sigma; G)$ with measurement M if

$$\mathcal{R}_M(U \mid M(\omega)) = \mathbb{E}(U \mid \mathcal{M})(\omega) \quad almost\ surely.$$

Replacing variables M and U by M_{kn} and U_n, respectively, defines the reconstructor for the computational model. This object coincides with the functional presentation of the conditional expectation (see [63]). However, a reconstructor is defined for every measurement $m \in H$ and hence the individual definition [45] is needed. The definition is non-unique and especially for a single measurement may be arbitrary. Thus one has to fix which reconstructor is used (see [I]). Furthermore, here the discrepancy of assuming measurement as realization of M_{kn} or M_k is visible. Namely, one has to make sure that the reconstructor of the computational model is defined for the data obtained from M_k.

Among inverse problems the most used point estimate is by far the MAP estimate. The standard definition requires

$$\mathbf{u}_{kn}^{MAP} \in \mathrm{argmax}_{\mathbf{u} \in \mathbb{R}^n} \pi_{kn}(\mathbf{u} \mid \mathbf{m})$$

where the set on the right-hand side consist of all points $\mathbf{u}$ maximizing $\pi_{kn}(\cdot \mid \mathbf{m})$. Hence the value of

$$u_{kn}^{MAP} = \mathcal{I}^{-1}\left(\mathbf{u}_{kn}^{MAP}\right) \in H_1$$

is commonly defined as the MAP estimate of the computational problem. One notices this value is obtained from an optimization problem and it is not necessarily unique given a multi-modal posterior distribution. In fact, the main critique towards MAP estimation is that it does not represent the posterior well. For the sake of argument, consider an example on real axis where a thin spike is located far away from most of the probability mass. However, in real life applications when dimensionality increases significantly, the computation of MAP estimates can be very efficient when numerical methods such as conjugate gradient or Newton's method can be applied. In the numerical work of [I] and [III] the difference between computation times is huge: for a similar 512 dimensional problem a MAP estimate was computed in a few seconds whereas the computation of a CM estimate with reasonable accuracy took over 50 hours. Clearly, this comparison is not completely fair since computing a CM estimate with MCMC methods produces also information regarding the whole posterior.

In some cases a MAP estimate can be defined for the infinite dimensional problems [35]. The definition is rather implicit and difficult to assess since it involves studying all neighborhoods around a candidate point. It can be shown that the definition coincides in finite dimensional cases. Giving any general convergence conditions remains an open problem.

Finally, we note that although the posterior density depends on the inner products $\langle \cdot, \cdot \rangle_1$ and $\langle \cdot, \cdot \rangle_2$ both point estimates are invariant with respect to such choices.

2.4 Modeling noise

Noise is always present in physical measurements and estimating the error in solutions requires some knowledge on how it contributes to the data. Since ill-posed problems are by definition sensitive to noise, it is a crucial part of the problem to model reliably how the noise is produced.

Most often in inverse problems the noise is modeled as additive and mutually independent of the unknown [39]. This is the simplest choice yielding accessible likelihood distribution. More complicated relationships have been used as well, e.g., multiplicative noise. Such model could appear, e.g., when signal goes through a noisy amplifier. In the convergence analysis in [45] and [I-III] additivity is a crucial ingredient of the result making it possible to use the Cameron-Martin theorem [13]. Furthermore, noise is assumed

Gaussian due to the same reason. Likewise, we recognize the variety of situations leading to different non-Gaussian models. Consider problems where the measurements are done according to a Poisson process [1, 37].

The most typical Gaussian noise model is the white noise. Roughly, if the measurement is written in an orthonormal basis, every coefficient in this presentation is contaminated by an independent and identically distributed Gaussian noise. Consider such a process on a torus $\mathbb{T}$. In the $L^2(\mathbb{T})$-orthonormal basis of trigonometrical functions $\{t \mapsto e^{i2\pi kt}, t \in \mathbb{T}\}_{k \in \mathbb{Z}}$ the white noise can be formally written as

$$W = \sum_{k \in \mathbb{Z}} X_k e^{i2\pi kt} \tag{7}$$

where $X_k \sim \mathcal{N}(0, 1)$ i.i.d. for all $k \in \mathbb{Z}$. The difficulty in this formulation lies in the knowledge that the sum in equation (7) does not converge in $L^2(\mathbb{T})$ almost surely. This phenomenon reflects a more general observation that the covariance operator of a Gaussian random variable in a Hilbert space is of the trace class. Consequently, the random variable W is not Gaussian in $L^2(\mathbb{T})$ although every finite truncation of the sum in (7) is. There are ways to approach this ambiguity by defining the random variable in more general setting [43, 28]. In [I-III] we have chosen to embed the L^2-type white noise into the Sobolev space $H^{-s}(\mathbb{T})$ for $s > 1/2$. Neither approach avoids all controversies.

The convergence considerations also raise a question about the interpretation of the limiting noise distribution. One could ask whether idealizing the measurement (by taking k to infinity) erases the measurement error. Some arguments can be given defending the existence of $\mathcal{E}$. First, a physical process always contains some *background noise*. Secondly, in the Bayesian approach the role of the noise is also to describe the inexact physical modeling. Notice that these two properties are not completely independent. In paper [III] the discrete noise $\mathcal{E}_k$ is defined via

$$\mathcal{E}_k = E_k + P_k \mathcal{E} \tag{8}$$

where E_k and $\mathcal{E}$ describe the *instrumentation* and the background noise, respectively. On the other hand, $\mathcal{E}$ might be neglectable from practical point of view. Hence both aspects can make sense and the right path is rather problem-dependent.

Let us finally describe the *scaled white noise* model used in [III]. The random variabe $\mathcal{E}_k : \Omega \to \text{Ran}(P_k) \subset L^2(\mathbb{T})$ is assumed to be a Gaussian random function on $\mathbb{T}$ with zero expectation and covariance

$$\mathbb{E}\left(\langle \mathcal{E}_k, \phi \rangle_{L^2} \langle \mathcal{E}_k, \psi \rangle_{L^2}\right) = K^{-\kappa} \langle \phi, \psi \rangle_{L^2} \tag{9}$$

for any $\phi, \psi \in \text{Ran}(P_k)$ where $K = \dim(\text{Ran}(P_k))$. The parameter κ describes the *scaling* of the distribution. Notice that equation (9) implies $\mathbb{E} \|\mathcal{E}_k\|_{L^2}^2 = K^{1-\kappa}$ and hence choice $\kappa \geq 1$ is reasonable if the L^2-norm of the noise is expected to vanish or stay stable.

3 Edge-preserving prior structures

In many imaging applications one seeks a solution that may contain discontinuities. The classical Tikhonov-type regularization methods often lead to some smoothing in the solutions. The same phenomenon appears in Bayesian models when Gaussian prior distributions are utilized. Hence it has become a challenge on its own to introduce efficient methods that focus on good reconstruction of the edges [15, 36, 52, 62].

In Bayesian paradigm this property has to be modeled in the prior distribution. Roughly speaking, the prior distribution should be constructed in such a way that the solutions which have high probability with respect to the posterior distribution are piecewise smooth and have rapidly changing values only in a set of small measure. Such reconstruction methods are often called *edge-preserving*.

In finite dimensional Bayesian inversion theory a number of methods have been introduced for obtaining edge-preserving reconstructions [14, 20, 29, 68]. Maybe the most widely used edge-preserving structure is the total variation prior [38, 42, 64]. In the seminal work [29] by Geman and Geman and also recently by Calvetti and Somersalo in [16, 17] the edge-preserving property is obtained using so-called *hierarchical* prior models. In a nutshell, if a parameter of the prior distribution is not known, it is a part of the inference problem [39]. In such cases the prior is referred to as hierarchical.

We point out that the majority of these methods concentrate in the MAP estimation and therefore the question of discretization invariance is usually clear or omitted. However, the asymptotic properties of the MAP estimate tell very little if nothing about the convergence of a non-Gaussian posterior distribution. In fact, it was proved in [44] regarding the total variation priors that the CM estimates lose their edge-preserving property asymptotically or even diverge with the typical choice of parameters. The methodological novelty of [I-III] lies in studying hierarchical prior models for which *the posterior distribution converges while the edge-preserving property is maintained*.

Many Bayesian prior models lift deterministic regularization ideas to the stochastic setting (cf. total variation priors [44]). Such lifting has also been the starting point of this thesis. Namely, the finite dimensional MAP estimation problem studied by Geman and Geman in [29] lead Mumford and Shah in [52] to introduce the celebrated continuous deterministic version of the method. The novel idea was to define a penalty term

$$F(u, K) = \int_{\mathbb{T} \backslash K} |Du|^2 dt + \sharp(K) \tag{10}$$

where $K \subset \mathbb{T}$ is a set and the notation $\sharp(K)$ stands for the number of points in K. Roughly, the minimization of

$$F(u, K) + \int_{\mathbb{T}} |u - m|^2 dt$$

with respect to u and K results to a piecewise smooth signal approximating m and a set describing the points where this approximation jumps. The original

framework for the functional was in two dimensional image segmentation in which case $\sharp(K)$ is replaced by the one dimensional Hausdorff measure of the edge set K. This setting has been extensively studied and a number of interesting theoretical questions remain open [21].

The functional $F(u, K)$ is known to be difficult to handle numerically and several approximations to the variational problem have been introduced. In [2, 3] it is shown that the Mumford–Shah functional can be approximated by elliptic functionals

$$F_\epsilon(u, v) = \int_{\mathbb{T}} \left((\epsilon^2 + v^2)|Du|^2 + \epsilon|Dv|^2 + \frac{1}{4\epsilon}(1 - v)^2 \right) dt$$

in the sense of Γ-convergence with respect to $\epsilon > 0$ [11, 12]. This is discussed in detail in [III]. For a minimizing pair (u_ϵ, v_ϵ) one sees that the auxiliary function v_ϵ describes the edges of u_ϵ.

The core idea in [I-III] is to create a hierarchical prior distribution so that samples with high probability produce small value for a functional similar to F_ϵ. Let us next formally define the discrete prior models studied in [I-III]. Set $N = 2^n$ and let points $t_j = j/N$, $j = 0, 1, ..., N$, and t_0 identified with t_N, denote an equispaced mesh on $\mathbb{T}$. We define $PL(n)$ to be the space of continuous functions $f \in C(\mathbb{T})$ such that f is linear on each interval $[t_j, t_{j+1}]$ for $0 \leq j < N$. Furthermore, let $PC(n)$ be the space of functions $f \in L^2(\mathbb{T})$ such that f is constant on each interval $[t_j, t_{j+1})$ for $0 \leq j < N$. Denote by $Q_n : L^2(\mathbb{T}) \to PC(n)$ the orthogonal projections with respect to $L^2(\mathbb{T})$ inner product and let $Q = Q_0$ be the projection to constant functions. Define the operator $D_q = D + \epsilon^q Q : H^1(\mathbb{T}) \to L^2(\mathbb{T})$ where $\epsilon > 0$, $q > 1$ and $D = \frac{d}{dt}$.

The hierarchical structure is defined in two steps. First, fix $\epsilon > 0$ and $\alpha \in \mathbb{R}$ and let $V_{n,\epsilon}$ be a Gaussian random variable in $PL(n)$ with density function

$$\Pi_{V_{n,\epsilon}}(v) = C \exp \left(-\frac{N^\alpha}{2} \int_{\mathbb{T}} \left(\epsilon|Dv|^2 + \frac{1}{4\epsilon}(1 - v)^2 \right) dt \right) \qquad (11)$$

where $v \in PL(n)$. Then choose $v_{n,\epsilon}$ to be a sample of $V_{n,\epsilon}$. The random variable $U_{n,\epsilon}$, conditioned on $v_{n,\epsilon}$, is then defined as a Gaussian random variable on $PL(n)$ with a density function

$$\Pi_{U_{n,\epsilon}|V_{n,\epsilon}}(u \mid v_{n,\epsilon}) = C'(v_{n,\epsilon}) \exp \left(-\frac{N^\alpha}{2} \int_{\mathbb{T}} (\epsilon^2 + |Q_n v_{n,\epsilon}|^2)|D_q u|^2 \, dt \right) \quad (12)$$

where $u \in PL(n)$. Following Section 2.2 we notice that the density function presentation is not unique without a definition of the inner product on $PL(n)$. The choice of the inner product is rather technical and for details see [I].

Roughly speaking, a sample $v_{n,\epsilon}$ has a high probability if it is continuous and the $L^2(\mathbb{T})$-distance from constant function $\mathbf{1}$ is small. When ϵ is decreased, the probable samples becomes less smooth. Consequently, a sample of $U_{n,\epsilon}$ has a high probability if it varies rapidly only near the points where $v_{n,\epsilon}$ is close to zero. Hence the role of ϵ is to control how sharp jumps $U_{n,\epsilon}$ can

have. We call parameter ϵ the *sharpness* of the prior. Further, the parameter α describes the *scaling* of the prior information. Increasing α results to a prior distribution that is more concentrated around the expectation values.

As a consequence of the construction above the probability density of the joint distribution of $(U_{n,\epsilon}, V_{n,\epsilon})$ has a form

$$\Pi_{(U_{n,\epsilon}, V_{n,\epsilon})}(u, v) = c \exp\left(-\frac{N^\alpha}{2} F_{\epsilon,n}^\alpha(u, v)\right)$$

where $(u, v) \in PL(n) \times PL(n)$ and

$$
\begin{aligned}
F_{\epsilon,n}^\alpha(u, v) \;=\; &\int_{\mathbb{T}} \Big(- N^{1-\alpha} \log(\epsilon^2 + |Q_n v|^2) + \\
&+ (\epsilon^2 + |Q_n v|^2)|D_q u|^2 + \epsilon|Dv|^2 + \frac{1}{4\epsilon}|1 - v|^2 \Big) dt. \quad (13)
\end{aligned}
$$

The logarithmic term in (13) appears due to the fact that the normalization constant C' in (12) depends on $v_{n,\epsilon}$.

Let us next describe the results in [I-III]. The papers study asymptotical properties of linear inverse problems (3) and the implementation of the hierarchical prior presented above. The forward operator A is assumed to be bounded in $L^2(\mathbb{T})$. For the sake of the presentation we simplify the setting by assuming that k and n are coupled although this is not necessary for the convergence results in [I]. Further, it is assumed that $\mathcal{E}_k$ is scaled white noise and $\kappa = \alpha$. This implies that the asymptotic behavior of the noise is well-known and the scaling of the prior distribution can be estimated accordingly. The case when $\kappa \neq \alpha$ is discussed in [III]. Due to these simplifications we drop the notation k and κ.

Under these assumptions the MAP estimate for $(U_{n,\epsilon}, V_{n,\epsilon})$ corresponding the measurement m_n sampled from M_n is a minimizer

$$\left(u_{n,\epsilon}^{MAP}, v_{n,\epsilon}^{MAP}\right) \in \operatorname{argmin}_{u,v \in PL(n)} \left(F_{\epsilon,n}^\alpha(u, v) + \|P_n A u - m_n\|_{L^2}^2\right). \quad (14)$$

Notice that the presence of an arbitrary bounded operator A produces some difficulties when analyzing the convergence of the MAP estimates with respect to ϵ. In fact, the Mumford–Shah penalization has been applied only recently to inverse problems [57, 58, 59]. In image segmentation the residual term provides a bound for the minimizers in $L^p(\mathbb{T})$ for some $p > 1$. The approach taken in [58, 59] is to assume a priori that the minimizers are in a bounded set of $L^\infty(\mathbb{T})$. In [III] we modify this approach by assuming instead that $A : H^{-s}(\mathbb{T}) \to L^2(\mathbb{T})$ for some $s < 1/2$ is invertible. We show that it is enough to prove a bound for the minimizers in $L^{1+\delta}(\mathbb{T})$ where $\delta > 0$ is a small number. Hence as a byproduct some novel results about the deterministic problem are proved in [III].

Consider the situation when $\alpha = 0$ and $\epsilon > 0$. With the additional assumption that $A : L^2(\mathbb{T}) \to H^1(\mathbb{T})$ is bounded the following convergence results are shown in [I] and [III] when $n \to \infty$:

(i) There exists a well-defined random variable $(U, V) : \Omega \to L^2(\mathbb{T}) \times L^2(\mathbb{T})$ to which $(U_{n,\epsilon}, V_{n,\epsilon})$ converges in distribution.

(ii) The posterior distributions converge weakly in $L^2(\mathbb{T}) \times L^2(\mathbb{T})$.

(iii) The CM estimates $(u_{n,\epsilon}^{CM}, v_{n,\epsilon}^{CM})$ converge in $L^2(\mathbb{T}) \times L^2(\mathbb{T})$ to the reconstructors of problem (1).

(iv) The MAP estimates $(u_{n,\epsilon}^{MAP}, v_{n,\epsilon}^{MAP})$ diverge.

The case when $\alpha \geq 1$ is studied in [III]. Recall that in this setting the expectation of the measurement error is bounded. In this context we assume that $A : L^p(\mathbb{T}) \to L^p(\mathbb{T})$ is bounded for $p \in \{1, 2\}$ and also that mapping $A : H^{-s}(\mathbb{T}) \to L^2(\mathbb{T})$ is bounded from below for some $s < \frac{1}{2}$. The following claims are proved in [III]:

(i') When $n \to \infty$ the MAP estimates $\left(u_{n,\epsilon}^{MAP}, v_{n,\epsilon}^{MAP}\right)$ converge weakly in $H^1(\mathbb{T}) \times H^1(\mathbb{T})$ to a minimizer, denoted $\left(u_{\epsilon}^{MAP}, v_{\epsilon}^{MAP}\right)$, of a perturbed Ambrosio–Tortorelli functional.

(ii') The functions $\left(u_{\epsilon}^{MAP}, v_{\epsilon}^{MAP}\right)$ are shown to converge with respect to $\epsilon > 0$ in $L^1(\mathbb{T}) \times L^1(\mathbb{T})$, up to a sequence, to a minimizer of a Mumford–Shah-type functional.

We discuss the implications of these results in Section 5.

4 Modeling the space of images

4.1 Current aspects

A digital image is a discretized quantization of the continuous world [6]. Most popular quantization is by far the one produced by common day digital cameras. Roughly, one can consider a regular grid with a number assigned to each square describing, e.g., the average brightness of the analogue image. Although the ideal undiscretized image is never available in practise, the notion is useful for resolution independent mathematical analysis.

The question of how to choose a mathematical space for modeling images is often problem-driven. In the PDE based methods both the uniqueness and the existence of the solution are very valuable because they help to characterize why and when the algorithms work. However, this way the mathematical basis of the problem describes what type of images can be processed and are produced.

For instance, there is a long tradition of using functions of bounded variations (BV) in problems like image restoration involving variational calculus. The main reason is that such functions can be discontinuous across hypersurfaces whereas the classical Sobolev functions do not have this property. For presenting the complexity of natural photographs this is mandatory. It

is typical to all image processing methods that one can describe situations when the model fails [30].

In some cases an image model aims at describing one particular property well. In image segmentation methods the goal is to find meaningful edges from the images. A famous method presented in earlier sections was introduced by Mumford and Shah in [52]. In the related research a subspace of BV functions called special functions of bounded variation is used in the analysis.

In the same spirit, consider the problem of characterizing strongly oscillating patterns in images. There is a line of research utilizing the Meyer image model [50]. The goal of this approach is to decompose images into a well-structured component with homogeneous regions and a component that contains oscillating patters such as texture and noise. First of these is modeled with BV functions and the latter in a space that roughly corresponds the dual of BV functions. In conclusion, the characterization of an image becomes far from intuitive. To this direction there are also studies involving image models outside typical function spaces. Namely, in a paper by Mumford and Gidas [51] it was shown that locally integrable functions do not provide a framework where the statistics of large sets of random natural images can be studied.

Finally, it appears to be difficult to say which approach is more natural than another. Hence it is a well-justified question to ask what are the mathematical objects that the physical process of imaging produces. Following this observation, in paper [IV] a novel image model is introduced by analyzing a simplified imaging process of a CCD sensor. Below we describe the model and the obtained results.

4.2   Infinite photography

The starting point of article [IV] is a simplified model of a monochromatic digital camera comprising a light-sensitive surface called *sensor*, an optical arrangement, and a shutter for preventing light from passing through the optics. Photographs are taken by opening the shutter for a suitable period of time, allowing a large number of photons to arrive at the sensor which is divided into square-shaped disjoint subsets called *pixels*. Each pixel then measures the wavelength-dependent energy delivered by the photons, and the gray level of the pixel is determined by the ratio between energy detected at that pixel and the total energy received by the whole sensor.

The sensor surface is modeled as a two dimensional torus $\mathbb{D} = \mathbb{T}^2$. Thus one can consider a unit square with edges identified. An *exposure* is an infinite ordered sequence $e = (e_1, e_2, e_3, \dots)$, where each $e_j = (z_j, \lambda_j)$ models a photon with wavelength $\lambda_j > 0$ arriving at the sensor at the location $z_j \in \mathbb{D}$. The set of exposures is denoted by E. We recognize that knowing the exact data is an unphysical assumption. Yet, the value of the model is in deeper understanding of the structure of digital photographs provided by mathematical asymptotic analysis, and in new image processing methods

that can be applied consistently at any given finite resolution.

When a digital photograph is taken, the first J photons in an exposure e reach the sensor. We describe the resolution level by parameter $n \in \mathbb{N}$ such that the pixels at level n are sets

$$
D_n(k, \ell) := \left\{ (x, y) \in \mathbb{D} \;\middle|\; \begin{array}{ccccc} (\ell - 1)2^{-n} & \leq & x & < & \ell \, 2^{-n}, \\ 1 - k \, 2^{-n} & \leq & y & < & 1 - (k - 1)2^{-n} \end{array} \right\}.
$$

Given a resolution parameter $n \geq 0$, the pixel image $\mathcal{I}_n(e, J)$ produced by the J first photons of e is the following $2^n \times 2^n$ matrix:

$$
\mathcal{I}_n(e, J) := \begin{bmatrix} \mathcal{I}_n(e, J, 1, 1) & \cdots & \mathcal{I}_n(e, J, 1, 2^n) \\ \vdots & \ddots & \vdots \\ \mathcal{I}_n(e, J, 2^n, 1) & \cdots & \mathcal{I}_n(e, J, 2^n, 2^n) \end{bmatrix}. \tag{15}
$$

Above the matrix elements, or pixel values, are defined by

$$
\mathcal{I}_n(e, J, k, \ell) := \frac{1}{J |D_n(k, \ell)|} \sum_{\substack{z_j \in D_n(k, \ell) \\ 1 \leq j \leq J}} g(\lambda_j) \tag{16}
$$

where $1 \leq k \leq 2^n$ is the row index and $1 \leq \ell \leq 2^n$ is the column index. Here $g : \mathbb{R} \to \mathbb{R}$ is the *spectral sensitivity function* of the sensor satisfying $0 \leq g(\lambda) \leq M$ with some sensor-dependent constant $M < \infty$.

The ideal images, called *snapshots*, are defined using the concept *Banach limit*. It has the crucial property that while being linear with respect to the sequence, the limit exists always. Hence when we have infinite amount of photons, i.e., $J \to \infty$, we define ideal pixel values as

$$
\mathcal{I}_n(e, k, \ell) := B\text{-}\lim_{J \to \infty} \mathcal{I}_n(e, J, k, \ell) \tag{17}
$$

and similarly $\mathcal{I}_n(e)$ denotes the corresponding $2^n \times 2^n$ matrix.

To discuss distance between two snapshots we introduce a metric on E. Let

$$
d(e, e') := \max_{n \geq 0} \frac{|D_n(k, \ell)|}{\sqrt{2}(n + 1)} \, \|\mathcal{I}_n(e) - \mathcal{I}_n(e')\|_2
$$

for every $e, e' \in E$. Define also an equivalence relation $\sim$ between exposures $e, e' \in E$ by

$$
e \sim e' \quad \text{if and only if} \quad d(e, e') = 0.
$$

This allows us to define a central component of the analysis in [IV].

Definition 2 *The set $P := E/\!\sim$ is called* the space of infinite photographs. *The metric d of P is called the* snapshot metric.

4.3 Connections and applications

The image model presented above has interesting connections to Borel measures. First, denote by $\mathcal{B}_M(\mathbb{D})$ the set of all positive Borel measures μ on $\mathbb{D}$ satisfying $\mu(\mathbb{D}) \leq M$. In the following we formally construct two mappings between P and $\mathcal{B}_M(\mathbb{D})$ and characterize in which sense they are inverses of each other.

Given a measure $\mu \in \mathcal{B}_M(\mathbb{D})$ one can construct an exposure e by sampling locations of the photons using the distribution μ and setting the wavelength as a suitable constant. Namely, let $Z^{(j)} : \Omega \to \mathbb{D}$ be independent and identically distributed random variables with probability distribution $\frac{1}{\mu(\mathbb{D})}\mu$. In [IV] it is shown that one can choose $\omega_\mu \in \Omega$ for every μ such that mapping $\mathcal{P} : \mathcal{B}_M(\mathbb{D}) \to P$, where

$$\mathcal{P}(\mu) := [\, ((Z^{(j)}(\omega_\mu), \lambda))_{j=1}^\infty \,]$$

and $\lambda > 0$ satisfies $g(\lambda) = \mu(\mathbb{D})$, is well-defined. We call $\mathcal{P}$ the *illumination map*.

Likewise, given an exposure e one can proceed to the opposite direction and define an outer measure μ^* on $\mathbb{D}$ such that

$$\mu^*(D) = \inf\left\{ \sum_{\nu=1}^\infty 2^{-2n_\nu} \mathcal{I}_{n_\nu}(e, k_\nu, \ell_\nu) \;:\; D \subset \bigcup_{\nu=1}^\infty D_{n_\nu}(k_\nu, \ell_\nu) \right\} \qquad (18)$$

for any subset $D \subset \mathbb{D}$. The *limit image map* $\mathcal{M} : P \to \mathcal{B}_M(\mathbb{D})$ is defined by

$$\mathcal{M}([e]) = \mu^e$$

where μ^e is a restriction of μ^* to Borel sets. It is rather straightforward to see that $\mathcal{M}$ is well-defined since the snapshot data of e and e' coincides if $e \sim e'$.

To obtain a connection between $\mathcal{P}$ and $\mathcal{M}$ let us give the following definition.

Definition 3 *An element $e \in E$ is a* regular exposure *if the equality*

$$\mu^e(D_n(k, \ell)) = \frac{\mathcal{I}_n(e, k, \ell)}{|D_n(k, \ell)|}$$

holds and the limit $\lim_{J\to\infty} \mathcal{I}_n(e; J, k, \ell)$ exists for any $n \geq 0$ and for all indices $1 \leq k \leq 2^n$ and $1 \leq \ell \leq 2^n$. The subset of regular exposures is denoted by $E_0 \subset E$.

Further, let P_0 be the subset of P such that every equivalence class $[e] \in P_0$ has a regular exposure representative. This P_0 is called the set of regular infinite photographs.

The core feature of modeling images with exposure data lies in the following properties shown in [IV] for mappings $\mathcal{M}$ and $\mathcal{P}$:

(i) $\mathcal{M} \circ \mathcal{P}$ is the identity map on $\mathcal{B}_M(\mathbb{D})$,

(ii) $\mathcal{P} \circ \mathcal{M} : P \to P$ is a projection onto P_0,

(iii) $\mathcal{P} \circ \mathcal{M}$ is the identity map on P_0.

In conclusion, the regular infinite photographs naturally correspond to Borel measures on $\mathbb{D}$.

Let us discuss what type of applications such observation has. The first question studied in [IV] is to analyze a part of image that cannot be presented in typical function space models. A method for finding the Lebesgue decomposition of an image is introduced based on the exposure data and some numerical examples are presented. Also, some practical algorithms are described based on infinite photography. Pixel images and more general Borel measures can be mapped to exposures using a Markov Chain Monte Carlo algorithm, leading to a scalable image representation. Furthermore, applying this method to a negative image and interpreting photons as spots of ink constitutes a novel and flexible stochastic halftoning method applicable in graphical printing technology. Moreover, the familiar techniques of anti-aliasing and blurring allow photon-based implementations.

5   Conclusions

This dissertation studies the effects of discretization in the modeling of signal and image processing problems. The major mathematical contributions of the thesis are in papers [I-III] where novel hierarchical Bayesian prior models are introduced. The approach leads to discretization invariant edge-preserving reconstructions via MAP and CM estimation. In the light of the convergence results presented in Section 3 these point estimation methods differ remarkably from each other.

From the Bayesian point of view the case with scaling parameter $\alpha = 0$ is interesting since the non-Gaussian posterior distribution converges simultaneously while the MAP estimate diverges. Careful consideration of this discrepancy leads to the conclusion that even if the limiting posterior distribution is well-defined, the MAP estimate can not be used to describe it. Technically, this phenomenon occurs due to the hierarchical modeling and it is a future challenge within this framework to better understand the role of the MAP estimate. The qualitative performance of the numerical simulations in [I] provide clear evidence that the CM estimate can be used to reconstruct sharp jumps. However, the scope of the simulations is limited and demonstrating the full potential of the method requires large scale implementation.

The important contribution of [I-III] is to characterize prior models that use the methods of Mumford–Shah segmentation in Bayesian inversion. In the case $\alpha \geq 1$ this connection can be shown rigorously and the MAP estimates are proved to approximate Mumford–Shah minimizers. It would be an interesting yet difficult, problem to show a similar connection for the CM

estimates when $\alpha = 0$. Similarly, the extension of the methods to higher dimensions is part of the future work.

Article [IV] is concerned about the nature behind digital pixel images. The model assumes that the idealized imaging situation measures an unknown infinite point process on a bounded domain. While recognizing such assumptions lead to unphysical modeling, value of the model is in resolution independent analysis, as explained in Section 4.3. The most crucial direction for future studies is to include noise in the model.

Finally, we note that paper [IV] omits some recent development in regularization theory. Namely, in [67] the tools of regularization are extended to the space of Borel measures and hence many image processing methods become theoretically available for the model. The future development of the image model presented in [IV] involves studying how such methods can be formulated and applied with the help of exposure data.

References

[1] A. ANTONIADIS AND J. BIGOT, *Poisson inverse problems*, Ann. Statist. **34** (2006), pp. 2132–2158.

[2] L. AMBROSIO AND V. M. TORTORELLI, *Approximation of functionals depending on jumps by elliptic functional via Γ-convergence*, Comm. Pure Appl. Math. **43** (1990), pp. 999–1036.

[3] L. AMBROSIO AND V. M. TORTORELLI, *On the approximation of free discontinuity problem*, Boll. Un. Mat. Ital. B **6** (1992), pp. 105–123.

[4] S. R. ARRIDGE, J. P. KAIPIO, V. KOLEHMAINEN, M. SCHWEIGER, E. SOMERSALO, T. TARVAINEN AND M. VAUHKONEN, *Approximation errors and model reduction in optical diffusion tomography*, Inverse Problems **22** (2006), pp. 175–195.

[5] K. ASTALA AND L. PÄIVÄRINTA, *Calderon's inverse conductivity problem in the plane*, Ann. of Math. **163** (2006), pp. 265–299.

[6] G. AUBERT AND P. KORNPROBST, *Mathematical Problems in Image Processing: Partial Differential Equations and the Calculus of Variations*, Springer, New York, (2006).

[7] G. E. BACKUS *Bayesian inference in geomagnetism*, Geophys. J. **92** (1988), pp. 125–142.

[8] M. BENNING AND M. BURGER, *Error estimation for variational models with non-Gaussian noise*, preprint.

[9] G. CASELLA AND R. L. BERGER, *Statistical Inference*, Wadsworth & Brooks Cole, Pacific Grove, CA, (1990).

[10] N. Bissantz, T. Hohage, and A. Munk, *Consistency and rates of Convergence of Nonlinear Tikhonov regularization with random noise*, Inverse Problems **20** (2004), pp. 1773–1791.

[11] A. Braides, *Approximation of free-discontinuity problems*, Springer, Berlin, (1998).

[12] A. Braides, *Γ-convergence for Beginners*, Oxford University Press, Oxford, (2002).

[13] V. Bogachev, *Gaussian measures*, AMS, Providence, RI, (1998).

[14] C. Bouman and K. Sauer, *A generalized Gaussian image model for edge-preserving MAP estimation*, IEEE Trans. Image Process. **2** (1996), pp. 296–310.

[15] M. Burger and S. Osher, *Convergence rates of convex variational regularization*, Inverse Problems **20** (2004), pp. 1411–1421.

[16] D. Calvetti and E. Somersalo, *Gaussian hypermodels and recovery of blocky objects*, Inverse Problems **23** (2007), pp. 733–754.

[17] D. Calvetti and E. Somersalo, *Hypermodels in the Bayesian imaging framework*, Inverse Problems **24** (2008), 034013.

[18] D. Calvetti and E. Somersalo, *An introduction to Bayesian scientific computing - Ten lectures on subjective computing*, Springer, New York, (2007).

[19] E. Candès, D. Donoho, *Recovering edges in ill-posed inverse problems: optimality of curvelet frames*, Ann. Statist. **30** (2002), pp. 784–842.

[20] G. K. Chantas, N. P. Galatsanos and A. C. Likas, *Bayesian Restoration Using a New Nonstationary Edge-Preserving Image Prior*, IEEE Trans. Image Process. **15** (2006), pp. 2987–2997.

[21] G. David, *Singular sets of minimizers for the Mumford-Shah functional*, Birkhäuser Verlag, Basel, (2005).

[22] J. Diestel and J. J. Uhl, Jr., *Vector measures*, AMS, Providence, RI, (1977).

[23] R. M. Dudley, *Real Analysis and Probability*, Cambridge University Press, New York, (2002).

[24] H. Engl, M. Hanke and A. Neubauer, *Regularization of inverse problem*, Kluwer, Dordrecht, (1996).

[25] S. N. Evans and P.B. Stark, *Inverse problems as statistics*, Inverse Problems **18** (2002), R55–R97.

[26] B. G. FITZPATRICK, *Bayesian analysis in inverse problems*, Inverse Problems **7** (1991), pp. 675–702.

[27] J. N. FRANKLIN, *Well-posed stochastic extensions of ill-posed linear problems*, J. Math. Anal. Appl. **31** (1970), pp. 682–716.

[28] I. M. GELFAND AND N. Y. VILENKIN, *Generalized Functions, Vol. 4: Applications of Harmonic Analysis*, Academic Press, New York-London (1964).

[29] S. GEMAN AND D. GEMAN, *Stochastic Relaxation, Gibbs Distribution and the Bayesian Restoration of Images*, IEEE Trans. Patt. Anal. Machine Intell. **6** (1992), pp. 721–741.

[30] Y. GOUSSEAU AND J. M. MOREL, *Are natural images of bounded variation?*, SIAM J. Math. Anal. **33** (2001), pp. 634–648.

[31] H. HAARIO, E. SAKSMAN AND J. TAMMINEN, *An adaptive Metropolis algorithm*, Bernoulli **7** (2001), pp. 223–242.

[32] H. HAARIO, E. SAKSMAN AND J. TAMMINEN, *Componentwise adaption for high dimensional MCMC*, Comput. Statist. **20** (2005), pp. 265–273.

[33] J. HADAMARD, *Sur les problèmes aux dèrivées partielles et leur signification physique*, Princeton Univ. Bull. **13** (1902), pp. 49–52.

[34] W. HASTINGS, *Monte Carlo sampling methods using Markov chains and their applications*, Biometrika **57** (1970), pp. 97–109.

[35] M. HEGLAND, *Approximate Maximum a Posteriori with Gaussian Process Priors*, Constr. Approx. **26** (2007), pp. 205–224.

[36] B. HOFMANN, B. KALTENBACHER, CH. PÖSCHL, AND O. SCHERZER, *Convergence Rates Result for Tikhonov regularization in Banach Spaces with Non-Smooth Operators*, Inverse Problems **23** (2007), pp. 987–1010.

[37] T. HOHAGE, *Variational regularization of inverse problems with Poisson data*, preprint.

[38] J. P. KAIPIO, V. KOLEHMAINEN, E. SOMERSALO, AND M. VAUHKONEN, *Statistical inversion and Monte Carlo sampling methods in electrical impedance tomography*, Inverse problems **16** (2000), pp. 1487–1522.

[39] J. P. KAIPIO AND E. SOMERSALO, *Statistical and computational inverse problems*, Springer, New York, (2005).

[40] J. P. KAIPIO AND E. SOMERSALO, *Statistical inverse problems: discretization, model reduction and inverse crimes*, J. Comp. Appl. Math. **198** (2006), pp. 493–504.

[41] A. Kirsch, *An Introduction to the Mathematical Theory of Inverse Problems*, Springer, Berlin, (1996).

[42] V. Kolehmainen, S. Siltanen, S. Järvenpää, J.P. Kaipio, P. Koistinen, M. Lassas, J. Pirttilä and E. Somersalo, *Statistical inversion for medical X-ray tomography with few radiographs II: Application to dental radiology*, Phys. Med. Biol. **48** (2003), pp. 1465–1490.

[43] S. Lasanen, *Discretizations of generalized random variables with applications to inverse problems*, Ann. Acad. Sci. Fenn. Math. Diss. **130**, (2002).

[44] M. Lassas and S. Siltanen, *Can one use total variation prior for edge preserving Bayesian inversion?,* Inverse Problems **20** (2004), pp. 1537–1564.

[45] M. Lassas, E. Saksman and S. Siltanen, *Discretization invariant Bayesian inversion and Besov space priors*, Inverse Probl. Imaging **3** (2009), pp. 87–122.

[46] M. Lehtinen, L. Päivarinta, and E. Somersalo, *Linear inverse problems for generalised random variables*, Inverse Problems **5** (1989), pp. 599–612.

[47] H. Luschgy, *Linear estimators and radonifying operators*, Theory Probab. Appl. **40** (1995), pp. 167–175.

[48] A. Mandelbaum, *Linear Estimators and Measurable Linear Transformations on a Hilbert Space*, Z. Wahrsch. Verw. Gebiete **65** (1984), pp. 385–397.

[49] N. Metropolis, A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller and E. Teller, *Equations of state calculations by fast computing machine*, J. Chem. Phys. **21** (1953), pp. 1087–1091.

[50] Y. Meyer, *Oscillating patterns in image processing and in some nonlinear evolution equations*, AMS, Providence, RI, (2001).

[51] D. Mumford and B. Gidas, *Stochastic models for generic images*, Quart. Appl. Math. **59** (2001), pp. 85–111.

[52] D. Mumford and J. Shah, *Optimal Approximations by Piecewise Smooth Functions and Associated Variational Problems*, Commun. Pure Appl. Math. **42** (1989), pp. 577–685.

[53] A. I. Nachman, *Global uniqueness for a two dimensional inverse boundary value problem*, Ann. of Math. **143** (1996), pp. 71–96.

[54] F. Natterer, *The mathematics of computerized tomography*, John Wiley and Sons, Ltd., Chichester, (1986).

[55] A. Neubauer and H. K. Pikkarainen, *Convergence results for the Bayesian inversion theory*, J. Inverse Ill-posed Probl. **16** (2008), pp. 601–613.

[56] P. Piiroinen, *Statistical Measurements, Experiments and Applications*, Ann. Acad. Sci. Fenn. Math. Diss. **143** (2005).

[57] R. Ramlau and W. Ring, *A Mumford-Shah level-set approach for the inversion and segmentation of X-ray tomography data*, J. Comp. Phys. **221** (2007), pp. 539–557.

[58] L. Rondi, *On the regularization of the inverse conductivity problem with discontinuous conductivities*, Inverse Probl. Imaging **2** (2008), pp. 397–409.

[59] L. Rondi and F. Santosa, *Enhanced Electrical Impedance Tomography via the Mumford-Shah Functional*, ESAIM Control Optim. Calc. Var. **6** (2001), pp. 517–538.

[60] L. Rudin, S. Osher and E. Fatemi, *Nonlinear Total Variation based noise removal algorithms*, Physica D **60** (1992), pp. 259–268.

[61] M. J. Schervish, *Theory of Statistics*, Springer-Verlag, New York, (1995).

[62] J. A. Sethian, *Level Set Methods and Fast Marching Methods*, Cambridge University Press, Cambridge, (1999).

[63] A. Shiryaev, *Probability*, Springer, New York, (1996).

[64] S. Siltanen, V. Kolehmainen, S. Järvenpää, J. P. Kaipio, P. Koistinen, M. Lassas, J. Pirttilä and E. Somersalo, *Statistical inversion for medical X-ray tomography with few radiographs I: General theory*, Phys. Med. Biol. **48** (2003), pp. 1437–1463.

[65] J. Sylvester and G. Uhlmann, *A global uniqueness theorem for an inverse boundary value problem*, Ann. of Math. **125** (1987), pp. 153–169.

[66] C. Vogel, *Computational Methods for Inverse Problems*, SIAM, Philadelphia, (2002).

[67] B. Walch and O. Scherzer, *Sparsity Regularization for Radon Measures*, SSVM Proc. (2009), pp. 452–463.

[68] J. Wang and N. Zabaras, *Hierarchical Bayesian models for inverse problems in heat conduction*, Inverse Problems **21** (2005), pp. 183–206.

(continued from the back cover)

A583   Wolfgang Desch, Stig-Olof Londen
        Semilinear stochastic integral equations in L_p
        December 2009

A582   Juho Könnö, Rolf Stenberg
        Analysis of H(div)-conforming finite elements for the Brinkman problem
        January 2010

A581   Wolfgang Desch, Stig-Olof Londen
        An L_p-theory for stochastic integral equations
        November 2009

A580   Juho Könnö, Dominik Schötzau, Rolf Stenberg
        Mixed finite element methods for problems with Robin boundary conditions
        November 2009

A579   Lasse Leskelä, Falk Unger
        Stability of a spatial polling system with greedy myopic service
        September 2009