Hierarchical models in statistical inverse problems and the Mumford–Shah functional

T Helin and M Lassas

1 Johann Radon Institute for Computational and Applied Mathematics (RICAM), Austrian Academy of Sciences, Altenbergerstrasse 69, A-4040 Linz, Austria
2 Department of Mathematics and Statistics, University of Helsinki, PO Box 68 (Gustaf Hallstromin katu 2b) FI-00014, Finland

E-mail: tapio.helin@ricam.oeaw.ac.at and matti.lassas@helsinki.fi

Received 21 June 2010, in final form 16 September 2010
Published 16 December 2010
Online at stacks.iop.org/IP/27/015008

Abstract

The Bayesian methods for linear inverse problems are studied using hierarchical Gaussian models. The problems are considered with different discretizations, and we analyse the phenomena which appear when the discretization becomes finer. A hierarchical solution method for signal restoration problems is introduced and studied with arbitrarily fine discretization. We show that the maximum a posteriori estimate converges to a minimizer of the Mumford–Shah functional, up to a subsequence. A new result regarding the existence of a minimizer of the Mumford–Shah functional is proved. Moreover, we study the inverse problem under different assumptions on the asymptotic behaviour of the noise as discretization becomes finer. We show that the maximum a posteriori and conditional mean estimates converge under different conditions. This paper concentrates on the results regarding the maximum a posteriori estimate. The convergence results of the conditional mean estimate are proven in Helin (2009 Inverse Problems Imaging 3 4).

(Some figures in this article are in colour only in the electronic version)

1. Introduction

We study hierarchical Bayesian methods for linear inverse problems. In particular, we consider inverse problems with different discretizations and the phenomena which appear when the discretization is refined. The effect of fine discretization has recently been studied for Gaussian inverse problems in [36, 37, 44], and motivated by this development we consider hierarchical Gaussian models. More precisely, we introduce a hierarchical solution method and analyse its properties with arbitrarily fine discretization.
The inverse problem we consider is the linear signal restoration problem where the measurement \( m(t) \) relates indirectly to the unknown signal \( u(t) \) via

\[
    m(t) = Au(t) + e(t), \quad t \in \mathbb{T}.
\]

Here, \( \mathbb{T} \) is the unit circle which we frequently consider as the interval \([0, 1]\) with the end points identified. Furthermore, \( A \) is a smoothing linear integral operator and \( e(t) \) is the random noise. The signals are considered on the unit circle \( \mathbb{T} \) to avoid the complicated boundary effects that fall outside the scope of this paper. We discuss the prospects of extending the domain to unperiodic functions and higher dimensional domains at the end of the paper in remark 3.

In the Bayesian approach, \( u(t) \) and \( e(t) \) are modelled as random functions. Let us denote by \( U(t, \omega) \) and \( E(t, \omega) \) random functions where \( \omega \in \Omega \) is an element of a complete probability space \((\Omega, \Sigma, \mathbb{P})\) and \( t \in \mathbb{T} \). The distribution of \( U(t, \omega) \) and \( E(t, \omega) \) models our a priori knowledge on the unknown signal \( u(t) \) and error \( e(t) \), respectively, before the measurement is obtained. Below, the variable \( \omega \) is often omitted. The ideal measurement is considered to be a realization of the random function \( M(t) = AU(t) + E(t) \) on \( t \in \mathbb{T} \). In Bayesian inversion the aim is to make statistical inference on \( U \) given a realization \( m \) of the random function \( M \), and the Bayesian solution to the inverse problem means finding the conditional probability distribution of \( U \), called the posterior distribution, or some estimates for this distribution. Typically studied point estimates are the expectation of the posterior distribution called the conditional mean (CM) estimate and the maximum a posteriori (MAP) estimate.

In Bayesian inversion a reconstruction method is said to be edge-preserving if the functions \( u \) which have high probability with respect to the posterior distribution are roughly speaking piecewise smooth and have rapidly changing values only in a set of small measure. In the finite-dimensional Bayesian inversion theory, a number of methods have been introduced for obtaining edge-preserving reconstructions \([10, 21, 27, 48]\). In this paper the prior distribution of the random function \( U \) has a Gaussian distribution such that its covariance depends on an auxiliary random function \( V \). Moreover, the random function \( V \) has a Gaussian distribution. Such a model is called the hierarchical Gaussian model. With a fixed discretization, similar models have been studied in inverse problems in \([57]\). Furthermore, in the work by Calvetti and Somersalo \([15, 16]\) hierarchical methods have been used for image processing problems to obtain edge-preserving and numerically efficient reconstruction algorithms. We also mention that the edge-preserving reconstruction methods have been extensively studied in the deterministic problem setting, see e.g. \([13, 22, 43, 46, 53, 55, 56]\). Our main result in this paper connects computing the MAP estimate of a hierarchical Gaussian model to the minimization of the Mumford–Shah functional \([43]\) used in image processing. As a byproduct we also present new results concerning the existence of a minimizer of the Mumford–Shah functional.

Let us next discuss the discretization of Bayesian inverse problems. Above we have considered \( U(t) \) and \( M(t) \) as random functions defined on the unit circle. For any practical computations, such models have to be discretized, i.e. to be approximated by random variables taking values in a finite-dimensional space. Roughly speaking, a Bayesian model is said to be discretization invariant, if for fixed model parameters it works coherently at any level of discretization. For an extended discussion on the discretization invariance and the relation of the practical measurement models and the computational models considered below, see \([39]\).

In the ideal model the noise \( E \) can be considered as a background noise. In this paper we will further assume that the practical measurement setting produces an additional instrumentation noise. More precisely, we assume that the practical measurement can be modelled as the realizations of a random variable \( M_k = P_k M + E_k, \) where
the operator $P_k$ is a finite-dimensional projection. The random variable $E_k$ describes the instrumentation noise and it takes values in the range of $P_k$. Increasing the number $k$ corresponds here to the case when we make more or finer observations of the ideal measurement signal $M(t)$. Moreover, in practical computations also $U$ needs to be approximated by a finite-dimensional random variable $U_n$ which leads us to consider the computational model

$$M_{kn} = A_kU_n + \mathcal{E}_k$$

where $k, n \in \mathbb{N}$ are parameters related to discretizing the measurement and the unknown, respectively. In equation (2) we have $A_k = P_kA$ and $\mathcal{E}_k$ is a random variable in the range of $P_k$ satisfying

$$\mathcal{E}_k = P_kE + E_k.$$ (3)

In developing new Bayesian algorithms, it is important to study if the posterior distribution given by problem (2) or some preferred estimate converges when $k$ or $n$ increases. This question is often non-trivial. For example, for the total variation prior it is proven in [38] that the MAP and CM estimates converge under different conditions as discretization is refined. Moreover, if the free parameters of the discrete total variation priors are chosen so that the posterior distribution converges, then the limit is a Gaussian distribution. Hence, the key property of the total variation prior is lost in very fine discretizations. This example illustrates the difficulty involved in discretizing non-Gaussian distributions. Also in this paper we will observe that the convergence of the MAP and CM estimates occurs in different cases.

Let us next formally define the discrete models we study. Set $N = 2^n$ and let points $t_j = j/N$, $j = 0, 1, \ldots, N$, 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 $Q_n : L^2(\mathbb{T}) \rightarrow PC(n)$ 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 $D_\epsilon = D + \epsilon^q Q : H^1(\mathbb{T}) \rightarrow L^2(\mathbb{T})$ where $\epsilon > 0$, $q > 1$ and $D$ is the derivative with respect to $t \in \mathbb{T}$.

The hierarchical structure is defined in two steps. First, let $V_{n,\epsilon}$ be a Gaussian random variable in $PL(n)$ with the density function

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

where $v \in PL(n)$, $\alpha \in \mathbb{R}$ and $N$ is the number of mesh points. Here and below $c$ is a generic constant whose value may vary. 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 the density function

$$\Pi_{U_{n,\epsilon}|V_{n,\epsilon}}(u|v_{n,\epsilon}) = c' \exp \left( -\frac{N^2}{2} \int_{\mathbb{T}} (\epsilon^2 + |Q_n v_{n,\epsilon}(t)|^2) |D_u(t)|^2 dt \right)$$ (5)

where $u \in PL(n)$. Note that the constant $c'$ depends on $v_{n,\epsilon}$. Since the density function presentation in finite-dimensional Hilbert spaces is non-standard, we give in section 2 a definition of the random variables $U_{n,\epsilon}$ and $V_{n,\epsilon}$ based on the coordinate representation.

Roughly speaking, the sample $v_{n,\epsilon}$ has a high probability if it varies from 1 only little and this variation becomes less smooth if $\epsilon$ is decreased. 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 the parameter $\epsilon > 0$ is to control how sharp in jumps $U_{n,\epsilon}$ can be and consequently, we call it
the sharpness of the prior. Furthermore, the parameter \( \alpha \) describes the scaling of the prior information. The bigger the value of \( \alpha \) the more concentrated the prior distribution.

In consequence of the construction above, the probability density of the joint distribution of \((U_n, \epsilon, V_n, \epsilon)\) has a form

\[
\Pi_{(U_n, V_n)}(u, v) = c \exp \left( -\frac{N^\alpha}{2} F_{\epsilon,n}^\alpha(u, v) \right)
\]

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

\[
F_{\epsilon,n}^\alpha(u, v) = \int_T \left( -N^{1-\alpha} \log(\epsilon^2 + |Q_n v|^2) + (\epsilon^2 + |Q_n v|^2)|D_q u|^2 \\
+ \epsilon |Dq v|^2 \right) dt.
\]

(6)

The logarithmic term in equation (6) appears due to the fact that the normalization constant \(c'\) in equation (5) depends on \(v_n, \epsilon\). It turns out that the functional \(F_{\epsilon,n}^\alpha\) is closely connected to the widely studied segmentation method in deterministic image and signal processing, namely the Mumford–Shah functional

\[
F(u, K) = \int_{T \setminus K} |Du|^2 dt + \sharp(K) + \int_T |Au - m|^2 dt
\]

(7)

with respect to the function \(u\) and the set \(K\) of the points where \(u\) jumps [43]. The notation \(\sharp(K)\) stands for the number of points in \(K\). This functional is known to be difficult to handle numerically, and a number of approximations to the variational problem of minimizing (7) have been introduced. In [6, 7] it is shown that the Mumford–Shah functional can be approximated by elliptic functionals in the sense of \(\Gamma\)-convergence. These Ambrosio–Tortorelli functionals are the key element in our presentation.

Let us describe our main results. We study the behaviour of the MAP estimate in the case when the discretization parameters \(k\) and \(n\) are coupled. For the sake of presentation, we assume \(k = n\) and drop \(k\) from the notations. Furthermore, we assume that \(E_n\) is the white noise with variance depending on \(n\) and the scaling parameter \(\kappa\). More precisely, \(E_n\) is a Gaussian random function on \(T\) taking values in \(\text{Ran}(P_n)\) with zero expectation and covariance

\[
\mathbb{E} \left( \langle E_n, \phi \rangle_{L^2} \langle E_n, \psi \rangle_{L^2} \right) = N^{-\kappa} \langle \phi, \psi \rangle_{L^2}
\]

for any \(\phi, \psi \in \text{Ran}(P_n)\). Note that consequently \(\mathbb{E} \|E_n\|_{L^2}^2 = N^{1-\kappa}\) and the choice of \(\kappa\) describes how the norm of the noise is expected to behave asymptotically. We emphasize that the case when \(\kappa > 1\) corresponds to an assumption that more measurements produce better accuracy expectation, whereas with \(\kappa = 1\) one assumes that the accuracy in the norm of \(L^2(T)\) is expected to stay stable. An example of the case \(\kappa \geq 1\) is when the background noise \(E\) is negligible and the instrumentation noise follows asymptotics (8). The case \(\kappa = 0\) corresponds to the discretization of the Gaussian white noise, see [39]. To be able to prove positive results for the convergence of the MAP estimates, we will assume

\[
\kappa = \alpha.
\]

(9)

This implies that the scaling parameter of the prior distribution is determined by the variance of the noise in discretized measurements. The case when (9) is not valid is discussed in remark 1. Due to equality (9) we drop the notation \(\kappa\) and use \(\alpha\) as the scaling parameter of the noise distribution.

Under these assumptions, the MAP estimate for \((U_{n,e}, V_{n,e})\) corresponding to the measurement \(m_{dn}\)

\[
\lim_{n \to \infty} m_n = m \text{ in } L^2(T),
\]

is a minimizer:

\[
(u_{n,e}^{\text{MAP}}, v_{n,e}^{\text{MAP}}) \in \arg\min_{(u, v) \in PL(n) \times PL(n)} \left( F_{\epsilon,n}^\alpha(u, v) + \|A_n u - m_n\|_{L^2}^2 \right).
\]

(10)
In theorems 3 and 4, we prove for the MAP estimates:

(a) For $\alpha = 0$, the minimization problems (10) diverge as $n \to \infty$.

(b) For $\alpha \geq 1$, the MAP estimates $(u_{MAP}^n, v_{MAP}^n)$ converge to the minimizer, denoted $(u_{MAP}, v_{MAP})$, of a perturbed Ambrosio–Tortorelli functional as $n \to \infty$. Moreover, the functions $(u_{MAP}^n, v_{MAP}^n)$ are shown to converge up to a subsequence to the minimizer of the Mumford–Shah functional (7) as $\epsilon \to 0$.

In [31] and remark 2 the following is shown, with slightly different assumptions on the operator $A$, for the convergence of prior distributions and the CM estimate:

(a’) For $\alpha = 0$, the random variables $(U_{n,\epsilon}, V_{n,\epsilon})$ converge in distribution on $L^2(T) \times L^2(T)$ and the CM estimates $(U_{CM}^n, V_{CM}^n)$ converge in $L^2(T) \times L^2(T)$ as $n \to \infty$;

(b’) For $\alpha \geq 1$, the random variables $(U_{n,\epsilon}, V_{n,\epsilon})$ converge to zero as $n \to \infty$.

The type of convergence in (b’) is discussed in remark 2. Consequently, the results (a), (b) and (a’), (b’) illustrate how the convergence properties of the MAP and CM estimates are different for hierarchical Gaussian models.

Let us recall that the CM and MAP estimates coincide for finite-dimensional Gaussian inverse problems [35]. Typically the MAP estimates are computationally faster to obtain than the CM estimates and thus in inverse problems close to Gaussian ones the MAP estimate is used as an approximation of the CM estimate. The above results show that this is not the case for the hierarchical Gaussian models in general.

Finally, let us consider the current perspectives to Bayesian modelling and how this paper connects to earlier work. Bayesian inversion in infinite-dimensional function spaces was first studied by Franklin in [26]. This research has been continued and generalized in [25, 40, 42]. The convergence of the posterior distribution is studied in [31, 37, 39, 47]. In relation to result (b) the convergence of posterior distribution is studied in [32, 33, 44] when objective information becomes more accurate with Gaussian prior and noise distributions. For a general resource on the Bayesian inverse problems theory and computation, see [17, 35]. The Mumford–Shah functional has been applied to inverse problems for instance in [49–51], and for related work in image processing problems, see [8, 14, 18, 20]. Finally, we mention that variational approximation with $\Gamma$-convergence is used earlier in the context of inverse problems in e.g. [28, 38, 50, 51].

This paper is organized as follows. In section 2 we introduce the stochastic model and necessary tools to tackle the convergence problems related to MAP estimates. In section 3 we cover the main results and postpone the proofs to sections 5 and 6. In section 4 we study the existence of MAP estimates and in section 5 we discuss the cases when desired convergence does not take place. In section 6 we give the proofs related to $\Gamma$-convergence and equi-coerciveness of the functionals. Finally, in section 7 we illustrate the method in practice by numerical examples.

2. Definitions

In this section we cover the stochastic model introduced in [31] and furthermore give the main tools and theoretical results concerning the variational problem of the MAP estimate. Let us first introduce some notation. Most function spaces in our presentation have the structure of a real separable Hilbert space. We often use the $L^2$-based Sobolev spaces $H^s(T)$ for any $s \in \mathbb{R}$ equipped with Hilbert space inner product

$$\langle \phi, \psi \rangle_{H^s} := \int_T ((I - \Delta)^{s/2} \phi)(t)((I - \Delta)^{s/2} \psi)(t) \, dt$$
for any \( \phi, \psi \in H^s(\mathbb{T}) \) where \( \Delta = \frac{d^2}{dx^2} \). However, we also study the Banach structure of \( H^s(\mathbb{T}) \) with the dual space \( H^{-s}(\mathbb{T}) \). In this setting the Banach dual pairing is denoted by \( \langle \cdot, \cdot \rangle_{H^{-s} \times H^s} \). We also denote \( H^s(\mathbb{T}; [a, b]), s \geq 0 \), for the functions \( f \in H^s(\mathbb{T}) \) such that \( a \leq f \leq b \) a.e. for \( a, b \in \mathbb{R} \). Furthermore, we discuss the spaces \( H^s(a, b) = \{ f | f = g|_{[a, b]}, g \in H^s(\mathbb{R}) \} \) for \( a, b, s \in \mathbb{R} \). We say that a sequence \( \{ x_j \}_{j=1}^\infty \) converges to \( x \) strongly in the Banach space \( X \), if \( \lim_{j \to \infty} \| x - x_j \|_X = 0 \) as \( j \to \infty \).

Recall from Hilbert space valued stochastics \cite{9} that a covariance operator \( C_X \) of a Gaussian random variable \( X : \Omega \to H \) is defined by the equality

\[
\mathbb{E}((X - \mathbb{E}X, \phi)_H (X - \mathbb{E}X, \psi)_H) = (C_X \phi, \psi)_H
\]

for all \( \phi, \psi \in H \). We call a Gaussian random variable centred if \( \mathbb{E}X = 0 \).

We use a perturbed derivative \( D_q = D + \epsilon^q Q \), where \( D = \frac{d}{dx} \) and

\[
(Q f)(t) = \left( \int_\mathbb{T} f(t') \, dt' \right) \mathbf{1}(t)
\]

for \( \mathbf{1}(t) \equiv 1 \) and any \( f \in L^1(\mathbb{T}) \). This construction guarantees that \( D_q : H^1(\mathbb{T}) \to L^2(\mathbb{T}) \) and \( D_q|_{PL(n)} : PL(n) \to PC(n) \) are invertible mappings.

### 2.1. Bayes modelling

Let us now shortly describe how we define the Bayesian maximum a posteriori estimate for the computational model given in equation (2). Let \( (H_1, \langle \cdot, \cdot \rangle_1) \) and \( (H_2, \langle \cdot, \cdot \rangle_2) \) be two real Hilbert spaces such that \( \dim H_1 = J \) and \( \dim H_2 = K \). Assume that \( U_n \) obtains realizations in an \( H_1 \) and the range of the measurement projection \( P_k \) is \( H_2 \). Furthermore, let \( \mathcal{I} : H_1 \to \mathbb{R}^J \) and \( \mathcal{J} : H_2 \to \mathbb{R}^K \) be two arbitrary isometries. Let us now map equation (2) to a matrix equation

\[
M_{kn} = \mathcal{J} M_{kn} = A_{kn} U_n + E_k
\]

where \( A_{kn} = \mathcal{J} A_k \mathcal{I}^{-1} \in \mathbb{R}^{K \times J} \), \( A_k = P_k A \), \( E_k = \mathcal{J} E_k \) and \( U_n = \mathcal{I} U_n : \Omega \to \mathbb{R}^J \). If the a priori and likelihood distributions above are absolutely continuous with respect to Lebesgue measure, the posterior distribution can be obtained by the Bayes formula: the posterior density \( \pi_{kn} \) then has the form

\[
\pi_{kn}(u|m) = \frac{\Pi_n(u) \Gamma_{kn}(m|u)}{\Upsilon_{kn}(m)}
\]

where \( u \in \mathbb{R}^J \) and \( m \in \mathbb{R}^K \). In equation (12), the functions \( \Pi_n \) and \( \Gamma_{kn} \) are the prior and likelihood densities, respectively, and \( \Upsilon_{kn} \) is the density of \( M_{kn} \) \cite{35}. The standard definition of the maximum a posteriori estimate is then

\[
\mathbf{u}^{MAP}_{kn} = \arg\max_{u \in \mathbb{R}^J} \pi_{kn}(u|m)
\]

where the set on the right-hand side consists of all points \( u \) maximizing \( \pi_{kn}(\cdot|m) \). The value of \( \mathbf{u}_{kn}^{MAP} = \mathcal{I}^{-1} (\mathbf{u}_{kn}^{MAP}) \in H_1 \) is commonly defined as the MAP estimate of problem (2). Another point estimate in Bayesian inversion is the CM estimate which is defined as the integral

\[
\mathbf{u}_{kn}^{CM} = \int_{\mathbb{R}^J} \mathbf{u} \pi_{kn}(\mathbf{u}|m) \, d\mathbf{u}
\]

and \( \mathbf{u}_{kn}^{CM} = \mathcal{I}^{-1} (\mathbf{u}_{kn}^{CM}) \in H_1 \).

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. For more information about the point estimates, see \cite{39} for CM and \cite{29} for MAP estimation in Hilbert spaces.
2.2. The prior model

In this subsection we introduce the prior model discussed in the introduction and explain the meaning of the density function representation in equations (4) and (5). Let us first review the infinite-dimensional prior model introduced in [31]. Consider a centred Gaussian distribution $\lambda^v$ on $L^2(\mathbb{T})$ with covariance operator $C_U(v) = L \Lambda(v)L^* : L^2(\mathbb{T}) \to L^2(\mathbb{T})$ with $L = D_q^{-1} : L^2(\mathbb{T}) \to L^2(\mathbb{T})$ and multiplication operator $\Lambda(v) : L^2(\mathbb{T}) \to L^2(\mathbb{T})$ defined as

$$(\Lambda(v)f)(t) = \frac{1}{\epsilon^2 + v(t)^2} \cdot f(t), \quad t \in \mathbb{T},$$

for any $v \in L^2(\mathbb{T})$. Let us now formally discuss the qualitative behaviour of $\lambda^v$. Such a distribution has the property that in a set of large samples from the distribution $\lambda^v$ are likely to be smooth. Vice versa, in sets where $v(t)^2$ is small, the distribution allows more rapid changes.

Next we set the prior distribution of a random variable $U$ to be $\lambda^v$. However, the crucial step in hierarchical modelling is to model the values of $v$ with a random variable $V$. Thus, instead of knowing the exact locations of the jumps, we model how they are distributed. In [31] the random variable $V$ is Gaussian with the expectation $E V = I$ and the covariance operator $C_V = (\frac{1}{\alpha}I - \epsilon \Delta)^{-1}$ on $L^2(\mathbb{T})$. Denote the distribution of $V$ on $L^2(\mathbb{T})$ by $v$. The joint distribution of the random variable $(U, V) : \Omega \to L^2(\mathbb{T}) \times L^2(\mathbb{T})$ is then defined to satisfy

$$\lambda(E \times F) = \int_E \lambda_v(E) \, \text{d}v(v)$$

for any Borel measurable sets $E, F \subset L^2(\mathbb{T})$. This construction is shown in [31] to be well defined.

In the following we define the finite-dimensional prior structure studied in this paper with all scalings $\alpha \in \mathbb{R}$. In [31], these random variables are shown to converge to $U$ and $V$ in distribution on $L^2(\mathbb{T}) \times L^2(\mathbb{T})$ when $\alpha = 0$. First define two inner products on $H^1(\mathbb{T})$, namely

$$(f, g)_1 := (D_q f, D_q g)_1$$

and

$$(f, g)_2 := (C_V^{-\frac{1}{2}} f, C_V^{-\frac{1}{2}} g)_2$$

for any $f, g \in H^1(\mathbb{T})$. Next construct two orthonormal basis $\{f_j\}_{j=1}^\infty, \{g_j\}_{j=1}^\infty \subset H^1(\mathbb{T})$ with respect to the inner products $(\cdot, \cdot)_1$ and $(\cdot, \cdot)_2$, respectively, in the following way: for any $n \in \mathbb{N}$, we have $\{f_j\}_{j=1}^N, \{g_j\}_{j=1}^N \subset PL(n)$ where $N = 2^n, n \in \mathbb{Z}_+$. Such a construction can be obtained, e.g., using the Gram–Schmidt orthonormalization procedure. To simplify our notation, we assume that the probability space $(\Omega, \Sigma, P)$ has the additional structure $\Omega = \Omega_1 \times \Omega_2, \Sigma = \Sigma_1 \otimes \Sigma_2$ and $P = P_1 \otimes P_2$. We denote $\omega = (\omega_1, \omega_2) \in \Omega_1 \times \Omega_2$.

**Definition 1.** Define $V_{n,\alpha} : \Omega_2 \to PL(n) \subset L^2(\mathbb{T})$ as

$$V_{n,\alpha}(\omega_2) = \sum_{j=1}^N V_{j,\alpha}^{(\alpha_2)} g_j + 1$$

where the random vector $V_{n,\alpha} = \{V_{j,\alpha}\}_{j=1}^N : \Omega_2 \to \mathbb{R}^N$ is a centred Gaussian random variable with covariance matrix $C_{V_{n,\alpha}} = N^{-\alpha} I \in \mathbb{R}^{N \times N}$.
Definition 2. Let $U_{\alpha,n,\epsilon}^\alpha : \Omega \to L^2(\mathbb{T})$ be the random variable

$$U_{\alpha,n,\epsilon}^\alpha(\omega_1, \omega_2) = \sum_{j=1}^N U_{j,n,\epsilon}(\omega_1, \omega_2) f_j$$

where the random vector $U_{\alpha,n,\epsilon}^\alpha(\omega) = (U_{j,n,\epsilon}^\alpha)_{j=1}^N \in \mathbb{R}^N$ is given the following structure. Denote by $\omega_2 \mapsto C(\omega_2) \in \mathbb{R}^{N \times N}$ a random matrix such that

$$C_{jk}(\omega_2) = N^{-\alpha} \langle A_{\alpha}(V_{\alpha,n,\epsilon}^\alpha(\omega_2)), Dq_j, Dq_k \rangle_{L^2}.$$

Due to the positive definiteness of $C$ we can define

$$U_{\alpha,n,\epsilon}^\alpha(\omega) = C(\omega_2)^{1/2} W_N(\omega_1)$$

where $W_N : \Omega_1 \to \mathbb{R}^N$ is a centred Gaussian random variable with identity covariance matrix.

Following the procedure shown in section 2.1 choose $I_1, I_2 : P L(n) \to \mathbb{R}^N$ to be two isometries with respect to the inner products $\langle \cdot, \cdot \rangle_1$ and $\langle \cdot, \cdot \rangle_2$, respectively, and with the usual inner product of $\mathbb{R}^N$. Clearly, it follows that the vector presentation is then

$$U_{\alpha,n,\epsilon}^\alpha = I_1 U_{\alpha,n,\epsilon}^\alpha \quad \text{and} \quad V_{\alpha,n,\epsilon}^\alpha = I_2 (V_{\alpha,n,\epsilon}^\alpha - 1).$$

In [31] it was shown that if $u, v \in P L(n)$ are arbitrary and $u = I_1 u, v = I_2 v \in \mathbb{R}^N$, then it holds that the probability density function of $V_{\alpha,n,\epsilon}^\alpha$ in $\mathbb{R}^N$ is

$$\Pi_{V_{\alpha,n,\epsilon}^\alpha}(v) = c \exp \left( -\frac{N^\alpha}{2} \left( \epsilon \|Dv\|_{L^2}^2 + \frac{1}{4\epsilon} \|v - 1\|_{L^2}^2 \right) \right)$$

and the conditional probability density function of $U_{\alpha,n,\epsilon}^\alpha$ in $\mathbb{R}^N$ is

$$\Pi_{U_{\alpha,n,\epsilon}^\alpha|V_{\alpha,n,\epsilon}^\alpha}(u|v) = c \exp \left( -\frac{N^\alpha}{2} \int_T \left( -N^{1-\alpha} \log(\epsilon^2 + (Q_n v)^2) + (\epsilon^2 + (Q_n v)^2) |Dq_u|^2 \right) \right).$$

With the same assumptions, the joint prior density then takes the form

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

where the functional $F_{\epsilon,n}^\alpha$ is given in the following definition.

Definition 3. For any $\epsilon > 0$, $n \in \mathbb{N}$, and $\alpha \in \mathbb{R}$ let $F_{\epsilon,n}^\alpha : H^1(\mathbb{T}) \times H^1(\mathbb{T}) \to \mathbb{R} \cup \{\infty\}$ be functional such that

$$F_{\epsilon,n}^\alpha(u, v) = \int_T \left( -N^{1-\alpha} \log(\epsilon^2 + (Q_n v)^2) 

+ (\epsilon^2 + (Q_n v)^2) |Dq_u|^2 + \epsilon |Du|^2 + \frac{1}{4\epsilon} (1 - v)^2 \right) dx$$

when $(u, v) \in P L(n) \times P L(n)$ and $F_{\epsilon,n}^\alpha(u, v) = \infty$ when $(u, v) \in (H^1(\mathbb{T}) \times H^1(\mathbb{T})) \setminus (P L(n) \times P L(n))$. 8
2.3. Variational approximation and the functions of bounded variation

In this section we recall the definition and some important properties of \( \Gamma \)-convergence and the functions of bounded variation. The concept of \( \Gamma \)-convergence was first introduced by De Giorgi in the 1970s. For a comprehensive presentation on the topic, see [24]. Let \( X \) be a separable Banach space endowed with a topology \( \tau \) and let \( G, G_j : X \to [-\infty, \infty] \) for all \( j \in \mathbb{N} \).

Definition 4. We say that \( G_j \) \( \Gamma \)-converges to \( G \) for the topology \( \tau \) and denote \( G = \Gamma \lim_{j \to \infty} G_j \) if

(i) for every \( x \in X \) and for every sequence \( x_j \) \( \tau \)-converging to \( x \) in \( X \), we have \( G(x) \leq \inf_{j \to \infty} G_j(x_j) \);

(ii) for every \( x \in X \) there exists a sequence \( x_j \) \( \tau \)-converging to \( x \) in \( X \) such that \( G(x) \geq \limsup_{j \to \infty} G_j(x_j) \).

Note that an equivalent definition is obtained by replacing condition (ii) with

(ii') For every \( x \in X \) there exists a sequence \( x_j \) \( \tau \)-converging to \( x \) in \( X \) such that \( G(x) = \lim_{j \to \infty} G_j(x_j) \).

Definition 5. A functional \( G : X \to [-\infty, \infty] \) is said to be coercive if the condition

\[ \lim_{j \to \infty} \|x_j\|_X = \infty \implies \lim_{j \to \infty} G(x_j) = \infty. \]

We call a sequence of functionals \( G_j : X \to [-\infty, \infty], j \in \mathbb{N} \), equi-coercive in the topology \( \tau \) if for every \( t \geq 0 \) there exists a compact set \( K_t \subset X \) such that \( \{x \in X \mid G_j(x) \leq t\} \subset K_t \) for all \( j \in \mathbb{N} \).

The following theorem summarizes some of the known results regarding \( \Gamma \)-convergence. For proofs, see [24].

Theorem 1. Let \( G, G_j : X \to [-\infty, \infty], j \in \mathbb{N} \), be a sequence of equi-coercive functionals in the topology \( \tau \) and \( G = \Gamma \lim_{j \to \infty} G_j \). Then the following three properties hold:

(i) If the \( \Gamma \)-limit of \( G_j \) exists, it is unique and lower semicontinuous.

(ii) For any continuous \( H : X \to \mathbb{R} \), we have \( G + H = \Gamma \lim_{j \to \infty} (G_j + H) \).

(iii) Let \( x_j \in X \) be such that \( \{G_j(x_j) - \inf_{\|x\|_X} G_j(x)\} \leq \delta_j \) where \( \delta_j \to 0 \). Then any accumulation point \( y \) of \( \{x_j\}_{j=1}^\infty \subset X \) is a minimizer of \( G \) and moreover \( \lim_{j \to \infty} G_j(x_j) = G(y) \).

Note the immediate corollary to (iii): suppose that the assumptions in theorem 1 hold and \( x_j \) is a minimizer of \( G_j \) for \( j \in \mathbb{N} \). Then any converging subsequence of \( \{x_j\}_{j=1}^\infty \subset X \) converges to a minimizer of \( G \).

Let us now turn to the related function spaces. Let \( u : T \to \mathbb{R} \) be a measurable function and fix \( t \in T \). We say that \( z \in \mathbb{R} \cup \{\infty\} \) is the approximate limit of \( u \) at \( t \) and write \( z = \text{aplim}_{s \to t} u(s) \) if for every neighbourhood \( \mathcal{N} \) of \( z \) in \( \mathbb{R} \cup \{\infty\} \) it holds that

\[ \lim_{\rho \to 0} \frac{1}{\rho} \mathbb{1}_{\{|s - t| < \rho, u(s) \notin \mathcal{N}\}} = 0. \]

We use the notation \( \bar{u}(t) = \text{aplim}_{s \to t} u(s) \) when the limit exists. Denote the set of points \( t \in \mathbb{T} \) where the approximate limit does not exist by \( S_u \). When \( u \in L^1(\mathbb{T}) \) it follows that \( |S_u| = 0 \), see [5].

Denote by \( BV(\mathbb{T}) \) the Banach space of functions of bounded variation. A function \( u \) belongs to \( BV(\mathbb{T}) \) if and only if \( u \in L^1(\mathbb{T}) \) and its distributional derivative \( Du \) is a bounded signed measure. We endow \( BV(\mathbb{T}) \) with the usual norm \( \|u\|_{BV} = \|u\|_{L^1} + |Du|(\mathbb{T}) \) where \( |Du| \) is the total variation of the distributional derivative.
Recall that due to the Lebesgue decomposition of measures, the distributional derivative \( Du \) can be written as a unique sum \( Du = D^f u + D^g u \) where \( D^f u \) is absolutely continuous and \( D^g u \) is singular with respect to the Lebesgue measure \(|·|\). Denote the density of \( D^g u \) with respect to the Lebesgue measure by \( \nabla u \). We call the function \( \nabla u \) the approximate gradient of \( u \). Moreover, denote \( D^f u = D^f u|_{S_0} \) and \( D^g u = D^g u|_{\Gamma \setminus S_0} \) where we have used the notation \( \mu|_X(Y) = \mu(X \cap Y) \) for the measurable sets \( X, Y \subset \mathbb{T} \). These restrictions are called the jump part and the cantor part, respectively. We say that \( u \in SBV(\mathbb{T}) \) or \( u \) is a special function of bounded variation if \( u \in BV(\mathbb{T}) \) and \( D^g u \equiv 0 \). Furthermore, denote by \( GSBV(\mathbb{T}) \) the Borel functions \( u : \mathbb{T} \to \mathbb{R} \) that satisfy \( \min(k, \max(u, -k)) \in SBV(\mathbb{T}) \) for all \( k \in \mathbb{N} \). The space \( GSBV \) is called the space of generalized special functions of bounded variation.

It turns out that the generalized special functions of bounded variation inherit most important features of \( SBV \) functions. First of all the set \( S_0 \) is well defined and enumerable for \( u \in GSBV(\mathbb{T}) \), and the approximate gradient \( \nabla u \) exists almost at every point in \( \mathbb{T} \). We refer to [5, 11] for a detailed presentation on these properties.

### 2.4. Mumford–Shah and Ambrosio–Tortorelli functionals

The idea of the weak formulation of the Mumford–Shah functional is to use the function space \( GSBV \) as a framework for the minimization problem and identify the set of jumps \( K \) in (7) with the set \( S_0 \) defined above. Let us drop the residual term from functional (7) for the moment and denote

\[
MS(u, v) = \begin{cases} 
\int_{\mathbb{T}} |\nabla u|^2 \, dx + \mathcal{E}(S_0) & \text{if } u \in GSBV(\mathbb{T}) \text{ and } v = 1 \text{ a.e.,} \\
\infty & \text{otherwise.}
\end{cases}
\]

The role of the auxiliary function \( v \) becomes clear later. The regularization term \( MS \) has been widely used in problems related to image segmentation problems. The application to ill-posed problems has been less extensive since in general with the non-invertible forward operator \( A \), the compactness of any minimizing sequence is not known. For the inverse conductivity problem in [50, 51], the compactness is obtained by posing an \textit{a priori} assumption that the minimizers are bounded in \( L^\infty \). In section 4 we prove a compactness result without such an assumption for mildly ill-posed problems.

Next we define the Ambrosio–Tortorelli functionals [6, 7]. First denote \( X = H^1(\mathbb{T}) \times H^1(\mathbb{T}; [0, 1]) \) and the regularizing term

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

for \((u, v) \in H^1(\mathbb{T}) \times H^1(\mathbb{T})\). A comprehensive proof for the next theorem can be found in [11] when \( p = 1 \) and [19] for the case \( p = 2 \).

**Theorem 2.** (Ambrosio–Tortorelli). The following statement holds for \( p = 1 \) and \( p = 2 \). Define the functional \( F^{AT}_\epsilon : L^p(\mathbb{T}) \times L^p(\mathbb{T}) \to (-\infty, \infty] \) so that

\[
F^{AT}_\epsilon(u, v) = \begin{cases} 
AT_\epsilon(u, v) & \text{when } (u, v) \in X, \\
\infty & \text{otherwise.}
\end{cases}
\]

Then we have that \( \Gamma-\lim_{\epsilon \to 0} F^{AT}_\epsilon = MS \) in the strong topology of \( L^p(\mathbb{T}) \times L^p(\mathbb{T}) \).

### 3. Main results

Let us now return to the computational model (2) and the prior distributions introduced in section 2.2. For the results shown in [30, 31] no dependence of the discretization parameters...
For all $k$ and $n$ is assumed. However, in this paper we need to require that $k$ and $n$ are coupled, i.e. the discretization can be characterized with only one parameter ($k = k(n)$ and $\lim_{n \to \infty} k(n) = \infty$). For the sake of clarity in the following, we assume $k = n$ and hence we drop the notation $k$.

Furthermore, the computational model (2) simply becomes

$$M_n = A_n U_n + E_n.$$  \hspace{1cm} (18)

Before stating the assumptions concerning the Bayesian inverse problems (18) for $n \in \mathbb{N}$, let us first introduce a definition similar to the one used in [31, 39].

**Definition 6.** The finite-dimensional measurement projections $P_n$, $n \in \mathbb{N}$, are called proper measurement projections if they satisfy the following conditions:

(i) We have $\text{Ran}(P_n) \subset H^1(\mathbb{T})$ and it holds that $\|P_n\|_{L(H^1)} \leq C$ and $\|P_n\|_{L(L^2)} \leq C$ for some constant $C$ with all $n \in \mathbb{N}$.

(ii) For $t \in \{-1, 0, 1\}$ we have $\lim_{n \to \infty} \|P_n f - f\|_{H^t} = 0$ for all $f \in H^t(\mathbb{T})$.

(iii) For all $\phi, \psi \in L^2(\mathbb{T})$ it holds that $\langle P_n \phi, \psi \rangle_{L^2} = \langle \phi, P_n \psi \rangle_{L^2}$.

For a discussion about the assumptions regarding the measurement, see [39]. In section 7 we provide an example of the projections $P_n$ that satisfy definition 6.

**Assumption 1.** For the problems in equation (18), there exist proper measurement projections $P_n$, $n \in \mathbb{N}$, and fixed parameters $\alpha \in \mathbb{R}$, $\epsilon > 0$ and $s > 0$ such that

(i) there exists a bounded linear operator $A : L^2(\mathbb{T}) \rightarrow L^2(\mathbb{T})$ which satisfies

$$\|u\|_{H^{-s}} \leq C \|Au\|_{L^2}$$  \hspace{1cm} (19)

for any $u \in L^2(\mathbb{T})$ with some constant $C > 0$ and $A_n = P_n A$ for all $n \in \mathbb{N}$.

(ii) The additive noise $E_n$ is a Gaussian random variable in $PL(n)$ such that $\mathbb{E}E_n = 0$ and for any $\phi, \psi \in L^2(\mathbb{T})$, the covariance satisfies

$$\mathbb{E}\langle E_n, \phi \rangle_{L^2} \langle E_n, \psi \rangle_{L^2} = \mathbb{N}^{-\alpha} \langle P_n \phi, P_n \psi \rangle_{L^2}.$$

(iii) The prior structure is modelled with random variables $(U_{n, \epsilon}^\alpha, V_{n, \epsilon}^\alpha)$ introduced in section 2.2.

(iv) The measurements $m, m_n \in L^2(\mathbb{T})$, $n \in \mathbb{N}$, satisfy $\lim_{n \to \infty} m_n = m$ in $L^2(\mathbb{T})$.

Note that condition (ii) means that $E_n$ has white noise statistics and in the case $\alpha = 0$, the random variables $E_n$ converge to white noise in the sense of generalized random variables as $n \to \infty$ [39, 52]. Now with assumption 1 the variational problem of finding the MAP estimates for equation (18) becomes

$$\min_{(u, v) \in PL(n) \times PL(n)} \left( \mathbb{F}_{\epsilon, n}(u, v) + \|P_n Au - m_n\|_{L^2}^2 \right).$$  \hspace{1cm} (20)

Below we study the behaviour of the MAP estimates with respect to the parameters $n, \alpha$ and $\epsilon$ using the variational approximation methods presented in section 2.3. In order to describe the $\Gamma$-limits of the functionals in equation (20) we have to introduce some new notation.

Let us first denote an auxiliary domain

$$X_\epsilon = H^1(\mathbb{T}) \times H^1(\mathbb{T}; [0, 1 + 30\epsilon])$$  \hspace{1cm} (21)

for sufficiently small $\epsilon$. Details about this choice of domain are given in appendix A.1 and we discuss it in more detail below. For now, it suffices to point out that the domain $X_\epsilon$ formally approaches $X$ when $\epsilon$ decreases. Denote the auxiliary operators

$$L_\epsilon(v) = \int_{\mathbb{T}} -\log(\epsilon^2 + v^2) \, dt$$
and

\[ S_{\epsilon}^q(u, v) = \int_{T} (\epsilon^2 + v^2)(2\epsilon^2 (Qu)Du + \epsilon^2 (Qu)^2) \, dt \]

for \((u, v) \in H^1(T) \times H^1(T)\). We motivate these notations after the next definition. Recall that \(X\) denotes the domain \(H^1(T) \times H^1(T; [0, 1])\).

**Definition 7.** Let us define the functionals \(F_\alpha^\epsilon : L^1(T) \times L^1(T) \to (-\infty, \infty], \epsilon > 0\) and \(\alpha \geq 1\), so that for \(\alpha = 1\)

\[ F_1^\epsilon (u, v) = \begin{cases} L_\epsilon(u, v) + AT_\epsilon(u, v) + S_{\epsilon}^q(u, v) & \text{when } (u, v) \in X_\epsilon, \\ \infty & \text{otherwise}, \end{cases} \]

and for \(\alpha > 1\)

\[ F_\alpha^\epsilon (u, v) = \begin{cases} AT_\epsilon(u, v) + S_{\epsilon}^q(u, v) & \text{when } (u, v) \in X, \\ \infty & \text{otherwise}. \end{cases} \]

Let us now discuss definition 7. The reason for the particular choice of \(X_\epsilon\) is two-fold. First, it turns out that the minimizer of the functional \(L_\epsilon + AT_\epsilon + S_{\epsilon}^q\) in \(H^1(T) \times H^1(T)\) may be located outside \(X\). Secondly, a pointwise bound for \(v\) provides easier proofs concerning the \(\Gamma\)-convergence results of the functionals \(F_\alpha^\epsilon\) in section 6.

Furthermore, it is straightforward to see that

\[ AT_\epsilon(u, v) + S_{\epsilon}^q(u, v) = \int_{T} \left( \epsilon^2 + v^2 |Dqu|^2 + \epsilon |Dv|^2 + \frac{1}{4\epsilon} (1 - v)^2 \right) \, dx \tag{22} \]

everywhere in \(H^1(T) \times H^1(T)\). Hence the role of \(S_{\epsilon}^q\) can be understood as a small perturbation that yields a lower bound for \(|Qu|\) and thus coercivity for \(F_\alpha^\epsilon\). On the other hand, compared to the Ambrosio–Tortorelli approach, a new term \(L_\epsilon\) appears due to the Bayesian hierarchical modelling.

In addition to problem (20), we will consider three different minimization problems throughout the paper. Two of them are the modified Ambrosio–Tortorelli minimization problem

\[ \min_{(u, v) \in H^1(T) \times H^1(T)} F_\alpha^\epsilon(u, v) + \|Au - m\|_{L^2}^2 \tag{23} \]

and the Mumford–Shah problem

\[ \min_{(u, v) \in L^1(T) \times L^1(T)} MS(u, v) + R(u). \tag{24} \]

In (24) we assume that \(A : L^p(T) \to L^p(T)\) is continuous for \(p \in [1, 2]\) and the residual \(R(u)\) is defined by

\[ R(u) = \begin{cases} \|Au - m\|_{L^2}^2 & \text{when } Au \in L^2(T), \\ \infty & \text{otherwise}. \end{cases} \]

In the following we often use the notation \(\|Au - m\|_{L^2}^2\) for \(R(u)\) when convenient. To describe cases when the edge-preserving property of MAP estimates is lost asymptotically, we consider the Tikhonov-type minimization problem

\[ H(u) = \begin{cases} \int_{T} |Du|^2 \, dt + \|Au - m\|_{L^2}^2 & \text{when } u \in H^1(T), \\ \infty & \text{otherwise}. \tag{25} \end{cases} \]

Note that the solution to problem (25) is obtained by \(u_{\text{min}} = (-\Delta + A^*A)^{-1}A^*m\). With the definitions given above we are ready to state the main results. Denote the conditional mean estimates introduced in section 2.1 of problem (18) by \((u_{CM}^{n, \epsilon}, v_{CM}^{n, \epsilon})\).
Theorem 3. Let the computational model (18) satisfy assumption 1 with prior parameters $s > 0$, $\alpha = 0$ and $\epsilon > 0$, and let the operator $A : L^2(\mathbb{T}) \rightarrow H^1(\mathbb{T})$ be bounded. Then the following statements hold.

(i) The CM estimates $(u_{CM}^{n,\epsilon}, v_{CM}^{n,\epsilon})$ converge in $L^2(\mathbb{T}) \times L^2(\mathbb{T})$ as $n \to \infty$.

(ii) The MAP estimates $(u_{MAP}^{n,\epsilon}, v_{MAP}^{n,\epsilon})$ diverge as $n \to \infty$.

In addition, the following holds for coupled parameters

(iii) If $\epsilon = \epsilon(n) \to 0$ as $n \to \infty$, then it follows that either the minimum values in formula (20) diverge to $-\infty$ or the MAP estimates $(u_{MAP}^{n,\epsilon(n)}, v_{MAP}^{n,\epsilon(n)})$ converge towards a minimizer of functional (25).

The statement (i) of theorem 3 is proved in [31] and statements (ii) and (iii) are proven in section 5. Note that even the coupling of $\epsilon$ and $n$ in statement (iii) does not yield convergence to Mumford–Shah minimizers. Namely the diverging minimum values immediately contradict with condition (i) in definition 4 since the functional MS is positive. Furthermore, the convergence to a minimizer of functional (25) implies that the edge-preserving property of the MAP estimates is lost. We point out that statement (iii) does not imply that this property is lost with all couplings.

Our main positive result regarding the convergence of the MAP estimates is the following.

Theorem 4. Let the computational model (18) satisfy assumption 1 with prior parameters $s < \frac{1}{2}$, $\alpha \geq 1$ and $\epsilon > 0$, and let the operator $A : L^p(\mathbb{T}) \rightarrow L^p(\mathbb{T})$ be bounded for $p = 1, 2$.

(i) The MAP estimates $(u_{MAP}^{n,\epsilon}, v_{MAP}^{n,\epsilon})$ have a subsequence converging to a minimizer $(u_\epsilon, v_\epsilon)$ of problem (23) in the weak topology of $H^1(\mathbb{T}) \times H^1(\mathbb{T})$ as $n \to \infty$.

(ii) The minimizers $(u_\epsilon, v_\epsilon) \in H^1(\mathbb{T}) \times H^1(\mathbb{T})$, $\epsilon > 0$, of problem (23) have a subsequence converging to a minimizer of the Mumford–Shah problem (24) in $L^1(\mathbb{T}) \times L^1(\mathbb{T})$ as $\epsilon \to 0$.

The result (ii) in theorem 4 can also be considered as a new interpretation of the Mumford–Shah functional; the minimizer of the Mumford–Shah functional can be approximated by the MAP estimates of Bayesian inverse problems. The proof for theorem 4 is given in section 6.

4. Well-posedness of the minimization problems

In this section, we study the properties of the individual problems (20), (23) and (24) with fixed parameters $\epsilon, \alpha$ and $n$. Our aim is to show three results. First, the existence of a minimizer of problem (24) is proved in theorem 5. Second, we show in lemma 4 that with the choice of the domain $X_\epsilon$ we do not exclude any pairs $(u, v) \in H^1(\mathbb{T}) \times H^1(\mathbb{T})$ which give a smaller value in problem (23). Finally, we show that the functionals $F_{e,n}^\alpha$ and $F_{e,n}^\alpha$ are coercive in $H^1(\mathbb{T}) \times H^1(\mathbb{T})$ which yields the existence of minimizers in problems (20) and (23).

Let us now study the existence of the solution to problem (24). The following compactness and semicontinuity theorem in GSBV is well known.

Lemma 1. Suppose a sequence $\{u_j\}_{j=1}^\infty \subset GSBV(\mathbb{T})$ satisfies

$$\|u_j\|_{L^p} + \sharp(S_{u_j}) + \int_\mathbb{T} |\nabla u_j|^2 \, dt \leq C$$

(26)

for some $1 < p \leq 2$. Then there exist $u \in GSBV(\mathbb{T}) \cap L^p(\mathbb{T})$ and a subsequence $\{u_k\}_{k=1}^\infty$ such that
(i) $u_{jk} \to u$ strongly in $L^1(\mathbb{T})$.
(ii) $\nabla u_{jk} \rightharpoonup \nabla u$ weakly in $L^2(\mathbb{T})$ and
(iii) $\sharp(S_u) \leq \liminf_{k \to \infty} \sharp(S_{u_{jk}})$.

**Proof.** If $\{u_j\}_{j=1}^\infty \subset GSBV(\mathbb{T})$ satisfies (26), then also
\[
\|u_j\|_{L^p} + \sharp(S_{u_j}) + \int_T |\nabla u_j|^p \, dt \leq C.
\] (27)
By [1–3] (see [23, theorem 2.1] for a short exposition) it holds that there exists a subsequence $\{u_{jk}\}_{k=1}^\infty$ and $u \in GSBV(\mathbb{T}) \cap L^p(\mathbb{T})$ such that conditions (i) and (iii) are satisfied. Furthermore, since $\nabla u_j$ is bounded in $L^2(\mathbb{T})$ due to the Banach–Alaoglu theorem we can extract a subsequence such that condition (ii) holds. 

The next lemma shows that the assumption in lemma 1 is in a sense self-improving and one can extend it for the purpose of mildly ill-posed problems.

**Lemma 2.** If a function $u \in GSBV(\mathbb{T}) \cap L^1(\mathbb{T})$ satisfies
\[
\|u\|_{H^{-s}} + \sharp(S_u) + \int_T |\nabla u|^2 \, dt \leq C'
\] (28)
for some $0 \leq s < \frac{1}{2}$, then for $p > 1$ such that $\frac{1}{p} = s + \frac{1}{2}$ it also satisfies inequality (26) for some constant $C$ depending on $s$ and $C'$.

**Proof.** Let us denote by $t_j$ the points in $S_u$, such that $S_u = \{t_1, t_2, \ldots, t_L\}$, where $t_1 < t_2 < \cdots < t_L$ and $L = \sharp(S_u)$ is bounded. Furthermore, denote by $I_j = (t_{j-1}, t_j) \subset \mathbb{T}$, $1 \leq j \leq L$, the interval between neighbouring points. Here $t_0$ and $t_L$ were identified. We can estimate the average of $u$ over the interval $I_j$ by
\[
\frac{1}{|I_j|} \left| \int_{I_j} u \, dt \right| \leq C|I_j|^{-\frac{1}{2}-s} \|u\|_{H^{-s}(I_j)},
\]
where we have used lemma 10. Now the Poincaré inequality states that
\[
\left\| u - \frac{1}{|I_j|} \int_{I_j} u \, dt \right\|_{L^p(I_j)} \leq C|I_j| \|\nabla u\|_{L^p(I_j)}
\]
and we obtain $\|u\|_{L^p(I_j)} \leq C(|I_j| + 1)$. By using the knowledge $\sum_{j=1}^L |I_j| = 1$, we deduce that $\|u\|_{L^p(\mathbb{T})} \leq C''$ where the constant $C''$ depends only on $s$ and $C'$. This proves the claim.

Clearly any sequence $u_j$ satisfying inequality (26) belongs to $L^\infty(\mathbb{T})$ and thus also $SBV(\mathbb{T})$. However, the bound in (26) does not control this norm and hence without any additional bound in $L^\infty$ the limit does not necessarily belong to $SBV(\mathbb{T})$. As the existence of a Mumford–Shah minimizer has interest for inverse problems in general we have formulated an independent proof to the following theorem.

**Theorem 5.** Let $A$ be a bounded linear operator in $L^p(\mathbb{T})$ for $p = 1, 2$ such that it satisfies inequality (19) for some $s < \frac{1}{2}$. Then the minimization problem
\[
\inf_{u \in L^1(\mathbb{T})} (MS(u, 1) + R(u))
\] (29)
has a solution $u \in GSBV(\mathbb{T}) \cap L^1(\mathbb{T})$. 

The first claim is obvious since
\[\inf_{u \in L^1(\mathbb{T})} (MS(u, 1) + R(u)) = \lim_{j \to \infty} (MS(u_j, 1) + R(u_j)).\]

Then the sequence \( u_j \) satisfies inequality (28) which in turn yields conditions in lemma 1. Consequently, we may extract a subsequence \( u_{j_k} \) converging in \( L^1(\mathbb{T}) \) to \( u \in GSBV(\mathbb{T}) \cap L^1(\mathbb{T}) \). Note that the residual term \( R(u) \) is lower semicontinuous in the \( L^1(\mathbb{T}) \) topology. Denoting the infimum in (29) by \( \mathcal{I} \) we obtain using lemma 1 that
\[\mathcal{I} \leq MS(u, 1) + R(u) \leq \liminf_{k \to \infty} MS(u_{j_k}, 1) + \liminf_{k \to \infty} R(u_{j_k}) \leq \liminf_{k \to \infty} (MS(u_{j_k}, 1) + R(u_{j_k})) \leq \mathcal{I}.\]
The claim follows from \( (u, 1) \) being a minimizer.

Next we discuss the choice of the domains \( X \) and \( X_\epsilon \) in definition 7. Denote by \( \psi_r, \psi_r : \mathbb{R} \to [0, r], r > 0 \), the functions \( \psi_r(t) = (r - |t|)\chi_{[0, r]}(t) \) and
\[\psi_r(t) = \sum_{j \in \mathbb{Z}} \psi_r(t - 2jr). \tag{30}\]
We note that for any function \( f \) and \( r > 0 \), the mapping \( \psi_r \circ f \) satisfies \( 0 \leq (\psi_r \circ f)(t) \leq r \) for all \( t \in \mathbb{T} \). Due to such property we call this operation folding. We list the following three properties of \( \psi_r \) as a lemma.

**Lemma 3.** For any \( f \in H^1(\mathbb{T}) \), it holds that

(i) \( |(\psi_r \circ f)(t)| \leq |f(t)| \) for any \( t \in \mathbb{T} \),
(ii) \( r - (\psi_r \circ f)(t) \leq |r - f(t)| \) and
(iii) \( |D(\psi_r \circ f)| = |Df| \) almost everywhere on \( \mathbb{T} \).

**Proof.** The first claim is obvious since \( \psi_r(f(t)) = \text{sgn}(f(t))|f(t)| \) when \( f(t) \in [-r, r] \) and also \( 0 \leq \psi_r(f(t)) \leq r \) for any \( t \in \mathbb{T} \). Claim (ii) also follows from the definition of the function \( \psi_r \). For claim (iii), note that since \( f \in H^1(\mathbb{T}) \), by the Sobolev embedding theorem \( f \) must be bounded, i.e. \( \sup_{t \in \mathbb{T}} |f(t)| < C \). In consequence, \( \psi_r \circ f \) can be written as a finite sum over the functions \( \psi_r(f(t) - 2jr) \). Now the result follows from a generalization of the chain rule (see e.g. [4]).

In particular, lemma 3 yields that \( \psi_r \circ f \in H^1(\mathbb{T}) \) whenever \( f \in H^1(\mathbb{T}) \). In the proof of the next lemma, we use the idea that in some cases \( \psi_r \circ v \) with suitable choices of \( r > 0 \) produces a lower value than \( v \) for the considered functionals. Consequently, we obtain information that the minimizers must lie in \( X \) or \( X_\epsilon \).

**Lemma 4.** Let \( \alpha \geq 1 \). For every \( v \in H^1(\mathbb{T}) \) there exists \( w \in H^1(\mathbb{T}; [0, 1 + 30\epsilon]) \) such that
\[F^w_\epsilon(u, w) \leq \delta_{u,1} L_\epsilon(v) + AT_\epsilon(u, v) + S_\epsilon^0(u, v), \tag{31}\]
for all \( u \in H^1(\mathbb{T}) \) where \( \delta_{u,1} = 1 \) when \( \alpha = 1 \) and is otherwise zero.

**Proof.** Consider a function \( \Psi_1 \) defined by equation (30). Due to lemma 3 and equation (22) we have \( F^w_\epsilon(u, \Psi_1 \circ v) \leq F^w_\epsilon(u, v) \) for any \( (u, v) \in H^1(\mathbb{T}) \times H^1(\mathbb{T}) \) and \( \alpha > 1 \). This immediately yields inequality (31) with \( w = \Psi_1 \circ v \) for \( \alpha > 1 \).
Let us then consider the case $\alpha = 1$ and let $(u, v) \in H^1(\mathbb{T}) \times H^1(\mathbb{T})$. To apply folding denote $E_- = \{v(t) < 0\}$, $E_0 = \{0 \leq v(t) \leq 1 + 3\epsilon\}$ and $E_+ = \{v(t) > 1 + 3\epsilon\}$ and by $1_E$ the indicator function of $E$. We write

$$v = v \cdot 1_{E_0} + (v - v_0) \cdot 1_{E_0}.$$ 

We construct $w$ by applying the folding operation to each restriction separately. First, recall identity (22) and denote

$$G(x, E) = \int_E \left( -\log(\epsilon^2 + v^2) + \frac{1}{4\epsilon} (1 - v^2) \right) dt,$$

with any measurable $E \subset \mathbb{T}$. Denote then

$$w_- = \Psi_1 \circ v_-, \quad w_+ = \Psi_{3\epsilon} \circ (v_+ - 1) + 1 \quad \text{and} \quad w = w_- + v_0 + w_+.$$ 

Clearly $0 \leq w \leq 1 + 3\epsilon$ and $(u, w) \in X_\alpha$.

First, we see due to claim (i) in lemma 3 that

$$|w(t)| = |w_-(t)| + |v_0(t)| + |w_+(t)| \leq |v(t)|$$

for all $t \in \mathbb{T}$. Furthermore, claim (iii) in lemma 3 implies $|Dw(t)| = |Dv(t)|$ almost everywhere on $\mathbb{T}$. These yield

$$\int_\mathbb{T} (|\epsilon^2 + w^2| |Du|^2 + \epsilon |Dv|^2) dt \leq \int_\mathbb{T} (|\epsilon^2 + v^2| |Du|^2 + \epsilon |Dv|^2) dt.$$  \hfill (33)

Let us next consider the integrand $g_\epsilon(t) = -\log(\epsilon^2 + t^2) + \frac{1}{4\epsilon} (1 - t^2)$ in equation (32) and apply the results in lemma 7 to the functions $w_-, w_0$ and $w_+$. Due to claims (i) and (iii) in lemma 7, it is straightforward to see $G(w_-, E_-) \leq G(v_-, E_-)$. Furthermore, claim (i) in lemma 7 implies $G(w_+, E_+) \leq G(v_+, E_+).$ From this we conclude

$$G(w, \mathbb{T}) = G(w_+, E_+) + G(v_0, E_0) + G(w_-, E_-)$$

$$\leq G(v_+, E_+) + G(v_0, E_0) + G(v_-, E_-) = G(v, \mathbb{T}).$$ \hfill (34)

Now inequalities (33) and (34) together with identity (22) yield the result. \hfill \Box

**Theorem 6.** The functionals $F^{\alpha}_\epsilon$ for $\alpha \geq 1$ and $F^{\alpha}_\epsilon_n$ for $\alpha \in \mathbb{R}$ are coercive in $H^1(\mathbb{T}) \times H^1(\mathbb{T})$ for any fixed $\alpha$, $n \in \mathbb{N}$ and $\epsilon > 0$.

**Proof.** Recall that a functional $G : X \to \mathbb{R}$ is coercive if we have a lower bound $G(x) \geq C \|x\|_X$ for $x \in X$ such that $\|x\|_X$ is large enough. By lemma 8 in the appendix we know that the functionals are bounded from below. One can deduce that

$$\int_\mathbb{T} \epsilon^2 |Du|^2 dt = \int_\mathbb{T} \epsilon^2 |Du|^2 dt + \epsilon^{2\gamma + 1} (Qu)^2 dt \geq C(\epsilon) \|u\|_{H^1}^2.$$ 

The lower bound for $\|u\|_{H^1}^2$ can be obtained from the term $\int_\mathbb{T} \left( \epsilon |Dv|^2 + \frac{1}{4\epsilon} (1 - v^2) \right) dt$. Hence it follows that both $F^{\alpha}_\epsilon(u, v)$ and $F^{\alpha}_\epsilon(u, v)$ go to infinity when $\|u\|_{H^1}$ or $\|v\|_{H^1}$ goes to infinity. \hfill \Box

5. Non-edge-preserving scaling

In this section we study the case when $s > 0$ and $\alpha = 0$ and prove theorem 3. Recall that claim (i) is shown in [31].

**Proof of claim (ii) in theorem 3.** Consider the value of $F^{\alpha}_\epsilon(u, v)$ at the function $(u(t), v(t)) \equiv (0, s)$ where $s > 1$, namely

$$F^{\alpha}_\epsilon(u, v) = g_N(s) = -N \log(\epsilon^2 + s^2) + \frac{1}{4\epsilon} (1 - s)^2$$
where \( N = 2^n \). With fixed \( s > 1 \) we have then \( \lim_{n \to \infty} g_N(s) = -\infty \). Also, it is easy to see that the minimizing values \( s = s(N) \) of \( g_N(s) \) go to infinity if \( n \to \infty \). Suppose now that the pair \((u_n, v_n) \in PL(n) \times PL(n)\) is a minimizer for problem (20) and that \((u_n, v_n)\) converges in \( H^1(\mathbb{T}) \times H^1(\mathbb{T})\). Clearly then
\[
F_{0,a}^0(u_n, v_n) + \| A_n u_n - m_n \|_{L^2}^2 \leq g_N(s(N)) + \| m_n \|_{L^2}^2
\]
for all \( n \in \mathbb{N} \). Since the terms of \( F_{0,a}^0 \) are all positive except for the logarithm term and the measurements \( m_n \) are bounded in \( L^2(\mathbb{T}) \), we have
\[
\int_{\mathbb{T}} -N \log(\epsilon^2 + (Q_n v_n)^2) \, dt \leq g_N(s(N)) + C
\]
for some constant \( C > 0 \). The assumption that \( v_n \) converge in \( H^1(\mathbb{T}) \) yields that \( \| v_n \|_\infty < C' \) for all \( n \in \mathbb{N} \) with some \( C' > 0 \). Thus, inequality (35) implies \(-NC'' \leq g_N(s(N)) + C\) for all \( n \in \mathbb{N} \) where \( C'' = \log(\epsilon^2 + (C')^2) \). This leads to a contradiction since \( \lim_{n \to \infty} \frac{N}{g_N(s(N))} = 0 \) and proves the claim.

We point out that the above proof also applies for the case \( 0 < \alpha < 1 \). The immediate question after result (ii) in theorem 3 is whether an appropriate coupling of \( \epsilon \) and \( n \) guarantees the convergence of MAP estimates. In the following we give some negative results about this.

Consider first how the discretization scheme affects the convergence in \( \| \cdot \|_\infty \)-norm.

**Theorem 7.** Assume that \( u_n, v_n \in PL(n) \) for \( n \in \mathbb{N} \), \( N = 2^n \) and \( F_{0,a}^0(u_n, v_n) \leq C \) for some constant \( C > 0 \). Then there exists a constant \( C' > 0 \) such that
\[
\| 1 - v_n \|_\infty \leq C\sqrt{\epsilon N + \epsilon^2 N^3}.
\]

**Proof.** The boundedness of \( F_{0,a}^0(u_n, v_n) \) and lemma 8 yield
\[
-C_1 \epsilon N^2 + \int_{\mathbb{T}} \frac{1}{\epsilon} (1 - v_n)^2 \, dt \leq C_2
\]
for some constants \( C_1, C_2 > 0 \). This immediately results in
\[
\| 1 - v_n \|_{L^2}^2 \leq C_2 \epsilon + C_1 \epsilon^2 N^2.
\]

First denote \( t_j = k/N \) for all \( 0 \leq k < N \). Suppose that \( f \in PL(n) \) achieves its maximum at \( t_j \) and denote by \( \phi_j \in PL(n) \) a function that satisfies \( \phi_j(t_k) = \delta_{jk} \) for all \( 0 \leq k < N \). Then by using the simple fact that \( \int_j^{j+1} f(t) \delta(t) \, dt \leq \int_j^{j+1} |f(t)| \, dt \), we have
\[
\| 1 - v_n \|_\infty = N \int_{j-1}^{j+1} \| 1 - v_n \|_\infty \delta_j(t) \, dt \leq \sqrt{N} \int_{\mathbb{T}} (1 - v_n(t))^2 \, dt
\]
where in the last inequality we have used the Cauchy–Schwarz inequality. Now inequality (36) proves the claim. \( \square \)

**Corollary 1.** Let \( \epsilon = \epsilon(n) \) such that \( \lim_{n \to \infty} \epsilon(n) = 0 \) and let \( (u_n, v_n) \in PL(n) \times PL(n) \) be a minimizer of \( F_{0,a}^0 \). Then the following statements hold:

(i) If \( \lim_{n \to \infty} \sqrt{\epsilon(n)^2 n} < \infty \), then the function \( v_n \) converges uniformly to \( 1 \) with respect to \( n \).

(ii) If \( \lim_{n \to \infty} \sqrt{\epsilon(n)^2 n} = \infty \), then the minimum values \( F_{0,a}^0 \) diverge, i.e.
\[
\lim_{n \to \infty} F_{0,a}^0(u_n, v_n) = -\infty.
\]
Proof. Let us first note that
\[
F_{e,n}^0(u_n, v_n) \leq \inf_{v \in P_L(n)} F_{e,n}^0(0, v) = \inf_{v \in P_L(n)} \int_{\mathbb{T}} \left(-2^n \log(\epsilon^2 + (Q_n v)^2) + \frac{1}{4\epsilon} (1 - v)^2\right) \, dt.
\]
Statement (ii) follows the upper bound given in lemma 8.

Assume now that \( \lim_{\epsilon \to 0} \sqrt{\epsilon(n)} 2^n \epsilon \to \infty \). By using inequality \( \log(1 + x) \leq x \) it follows that also \( \lim_{\epsilon \to 0} (2^n \log(1 + O(\sqrt{\epsilon}))) \to \infty \). By simple computations one can show that \( \lim_{\epsilon \to 0} \log(1 + O(\sqrt{\epsilon})) = 0 \) for any \( p > \frac{1}{2} \) and hence the quantity
\[
2^{3n} \epsilon^2 = (2^n \log(1 + O(\sqrt{\epsilon}))) \left(\frac{\epsilon^2}{\log(1 + O(\sqrt{\epsilon}))}\right)^3
\]
converges to zero. The convergence of \( 2^{2n} \epsilon^2 \) to zero follows by the same argument. Consequently, result (i) follows from theorem 7.

For \( n \in \mathbb{N} \) define the functionals
\[
H_n(u) = \int_{\mathbb{T}} h_n^2 |D_qu|^2 \, dt + \|P_nAu - m_n\|^2_{L^2}.
\]
(37)
for \( u \in P_L(n) \) where \( h_n \in H^1(\mathbb{T}; \mathbb{R}) \) converges to \( I(t) \equiv 1 \) uniformly.

Theorem 8. Let \( \epsilon = \epsilon(n) \) such that \( \lim_{n \to \infty} \epsilon(n) = 0 \). We have that
(a) \( H = \Gamma - \lim_{n \to \infty} H_n \) in the weak topology of \( H^1(\mathbb{T}) \)
(b) the functionals \( \{H_n\}_{n \in \mathbb{N}} \) are equi-coercive in the weak topology of \( H^1(\mathbb{T}) \).

Proof. Let \( u_n \rightharpoonup u \) weakly in \( H^1(\mathbb{T}) \) as \( n \to \infty \) where \( u_n \in P_L(n) \). By lower semicontinuity of the norm we have
\[
\int_{\mathbb{T}} |D_u|^2 \, dt \leq \liminf_{n \to \infty} \int_{\mathbb{T}} |D_u_n|^2 \, dt = \liminf_{n \to \infty} \int_{\mathbb{T}} h_n^2 |D_u u_n|^2 \, dt.
\]
Furthermore, by the Sobolev embedding theorem we see that \( u_n \rightharpoonup u \) in \( L^2(\mathbb{T}) \) and hence \( \lim_{n \to \infty} P_nAu_n - m_n = Au - m \) in \( L^2(\mathbb{T}) \). Together these imply \( H(u) \leq \liminf_{n \to \infty} H_n(u_n) \). This proves condition (i) in definition 4.

To prove condition (ii) in definition 4 it is sufficient to consider any sequence \( u_n \in P_L(n) \) such that \( u_n \rightharpoonup u \) in the \( H^1(\mathbb{T}) \)-norm as \( n \to \infty \). This proves claim (a) here. Let us then study claim (b) and assume that \( u_n \in P_L(n) \) for every \( n \in \mathbb{N} \) and
\[
H_n(u_n) = \int_{\mathbb{T}} h_n^2 |D_u u_n|^2 \, dt + \|P_n Au_n - m_n\|^2_{L^2} \leq C
\]
(38)
for some constant \( C > 0 \). In particular, we have \( \|P_nAu_n\|_{L^2} \leq C \) for all \( n \in \mathbb{N} \).

Next we show that also the sequence \( \|Au_n\|_{L^2} \) is uniformly bounded. Assume for the moment that this is not the case and \( \lim_{n \to \infty} \|Au_n\|_{L^2} = \infty \). Recall that the operator \( Q \) was defined as \( Qf = (\int_{\mathbb{T}} f(t) \, dt)1 \) for any \( f \in L^2(\mathbb{T}) \). Due to inequality (38) we have that \( \|Du_n\|_{L^2} \leq C \) and, in consequence, \( \|(I - Q)u_n\|_{L^2} \leq C \). Moreover, this yields
\[
\lim_{n \to \infty} \|QAu_n\|_{L^2} = \lim_{n \to \infty} (\|Au_n\|_{L^2} - \|A(I - Q)u_n\|_{L^2}) = \infty.
\]
(39)
By setting \( c_n = \int_{\mathbb{T}} u_n(t) \, dt \), equation (39) implies that
\[
\lim_{n \to \infty} |c_n| \|A1\|_{L^2} = \infty.
\]
(40)
The boundedness of \( \|A(I - Q)u_n\|_{L^2} \) yields together with \( \|P_nAu_n\|_{L^2} \leq C \) that
\[
|c_n| \|P_n A1\|_{L^2} = \|P_n AQ u_n\|_{L^2} \leq \|P_n (I - Q) u_n\|_{L^2} + \|P_n Au_n\|_{L^2} \leq C.
\]
(41)
By condition (40) we have \( \lim_{n \to \infty} |c_n| = \infty \) and by condition (41) it follows that \( \lim_{n \to \infty} \| P_n A_1 \|_{L^2} = 0 \). Due to condition (ii) in definition 6 this yields \( A_1 = 0 \). However, this contradicts equation (40). Consequently, we have proven that \( \lim_{n \to \infty} \| A u_n \|_{L^2} \leq C \) for some constant \( C > 0 \).

By the assumption on \( \Delta \) with \( s > 0 \), we have \( \| u_n \|_{H^s} \leq C \| A u_n \|_{L^2} \). As \( 1 \in (H^{-s}(\mathbb{T}))' = H^s(\mathbb{T}) \), we have \( |Q u_n| \leq C \| u_n \|_{H^{-s}} \) and hence we obtain by the Poincaré inequality that \( \| u_n \|_{H^1} \) is bounded. By the Banach–Alaoglu theorem there exists a converging subsequence which completes the proof. \( \square \)

Finally, we conclude this section by completing the proof of theorem 3.

**Proof of claim (iii) in theorem 3.** Suppose that \( \lim_{n \to \infty} \epsilon(n) = 0 \) and that \( (u_n, v_n) \in PL(n) \times PL(n) \) minimizes \( F_{\epsilon(n), \alpha}^0 \). By corollary 1 either the minimum values of \( F_{\epsilon(n), \alpha}^0 \) diverge to \( -\infty \) or \( v_n \to \) 1 uniformly.

Consider the latter case. Then the functions \( u_n \) clearly solve minimization problems \( \min_{\epsilon \in PL(\epsilon)} H_{\epsilon}(u) \) where \( H_{\epsilon}(u) \) is defined in equation (37) with \( h_n = \epsilon^2 + v_n^2 \). By theorems 1 and 8 it follows that \( u_n \) converge to a minimizer of \( H \). \( \square \)

### 6. Convergence proofs

#### 6.1. Convergence with respect to \( n \)

In this section we prove \( \Gamma \)-convergence of \( F_{\epsilon(n), \alpha}^0 \) with respect to \( n \) for all scalings \( \alpha \geq 1 \). Throughout the section, \( 0 < s < \frac{1}{2} \).

**Theorem 9.** For \( \alpha \geq 1 \) we have \( F_{\epsilon}^0 = \Gamma-\lim_{n \to \infty} F_{\epsilon(n), \alpha}^0 \) in the weak topology of \( H^1(\mathbb{T}) \times H^1(\mathbb{T}) \).

**Proof.** Let us assume that \( (u_n, v_n) \) converges weakly to \( (u, v) \in H^1(\mathbb{T}) \times H^1(\mathbb{T}) \). By the Sobolev embedding theorem, \( H^1(\mathbb{T}) \) embeds compactly to the space of Hölder continuous functions with exponent less than 1/2 and thus we have \( v_n \to v \) strongly in \( C^{0,1}(\mathbb{T}) \) for any \( \tau < 1/2 \). Furthermore, it follows that

\[
\sup_{t \in \mathbb{T}} |Q_n v_n(t) - v_n(t)| \leq \sup_{t \in \mathbb{T}} \int_{j(t)}^{j(t)+1} |v_n(t) - v_n(t')| dt' \leq N^{-\tau} \| v_n \|_{C^{0,\tau}},
\]

where \( j(t) \) is such that \( t \in \mathbb{J}(j(t)) = \left[ \frac{j(t)}{N}, \frac{j(t)+1}{N} \right) \). Now we see

\[
\| Q_n v_n - v \|_{L^2} \leq \| Q_n v_n - v_n \|_{L^\infty} + \| v_n - v \|_{L^2} \to 0
\]

as \( n \to \infty \). The immediate consequence is that

\[
\lim_{n \to \infty} \int_\mathbb{T} -N^{1-\alpha} \log(\epsilon^2 + (Q_n v_n)^2) dt = -\delta_{\alpha,0} \int_\mathbb{T} \log(\epsilon^2 + v^2) dt
\]

when \( \alpha \geq 1 \). Moreover, it also holds that

\[
\lim_{n \to \infty} \int_\mathbb{T} \frac{1}{4\epsilon} (1 - v_n)^2 dt = \int_\mathbb{T} \frac{1}{4\epsilon} (1 - v)^2 dt.
\]

Let us now consider condition (i) in definition 4 of \( \Gamma \)-convergence. Assume that \( (u_n, v_n) \to (u, v) \) in weak topology of \( H^1(\mathbb{T}) \times H^1(\mathbb{T}) \) as \( n \to \infty \). Since \( Dv_n \to Dv \) weakly in \( L^2(\mathbb{T}) \) and since a norm is lower semicontinuous, we deduce that

\[
\int_\mathbb{T} \epsilon |Dv|^2 dt \leq \liminf_{n \to \infty} \int_\mathbb{T} \epsilon |Dv_n|^2 dt.
\]
Without losing any generality we may also assume \( F_{\epsilon,n}(u_n, v_n) \leq C < \infty \) since otherwise there is nothing to prove. Hence in particular \( \int_{\Omega} |D_q u_n|^2 \, dt \leq C/\epsilon^2 \) and
\[
\lim_{n \to \infty} \int_{\Omega} \left( \frac{\epsilon^2}{2} + v_n^2 \right) |D_q u_n|^2 \, dt \leq \lim_{n \to \infty} \left\| v_n^2 \right\| \infty \cdot \frac{C}{\epsilon^2} = 0.
\]
Consequently, we obtain
\[
\int_{\Omega} \left( \epsilon^2 + v^2 \right) |D_q u|^2 \, dt \leq \liminf_{n \to \infty} \int_{\Omega} \left( \epsilon^2 + v_n^2 \right) |D_q u_n|^2 \, dt + \lim_{n \to \infty} \int_{\Omega} \left( v_n^2 - v^2 \right) |D_q u_n|^2 \, dt
\]
\[
= \liminf_{n \to \infty} \int_{\Omega} \left( \epsilon^2 + v_n^2 \right) |D_q u_n|^2 \, dt,
\]
due to lower semicontinuity of the norm \( \| \cdot \| = \| \sqrt{\epsilon^2 + v^2} \|_{L^2} \) and the weak convergence of \( D_q u_n \). By combining all inequalities above, it follows that \( F_{\epsilon}^q(u, v) \leq \liminf_{n \to \infty} F_{\epsilon,n}^q(u_n, v_n) \). This proves condition (i) in definition 4.

For condition (ii') in definition 4 we note that for an arbitrary \((u, v) \in L^1(\Omega) \times L^1(\Omega)\) there exists a sequence \((u_n, v_n) \in PL(n) \times PL(n)\) converging to \((u, v)\) in the strong topology. It is easy to see that one can then change \( \liminf \) into \( \lim \) and inequalities to equalities in (43) and (44). This yields the claim.

6.2. Convergence with respect to \( \epsilon \)

Let us prove the \( \Gamma \)-convergence for a modified functional. Define \( \Xi_{\epsilon} : L^1(\Omega) \times L^1(\Omega) \to (-\infty, \infty] \) as
\[
\Xi_{\epsilon}(u, v) = \begin{cases} 
F_{\epsilon}^1(u, v) & \text{when } (u, v) \in X, \\
\infty & \text{otherwise.}
\end{cases}
\]

**Theorem 10.** It holds that \( MS = \Gamma- \lim_{\epsilon \to 0} \Xi_{\epsilon} \) in the strong topology of \( L^1(\Omega) \times L^1(\Omega) \).

**Proof.** First we show that condition (i) of definition 4 holds. Suppose that \( \lim_{\epsilon \to 0}(u_\epsilon, v_\epsilon) = (u, v) \) in \( L^1(\Omega) \times L^1(\Omega) \). As in the previous \( \Gamma \)-convergence proofs, we may assume without losing any generality that
\[
\liminf_{\epsilon \to 0} \Xi_{\epsilon}(u_\epsilon, v_\epsilon) \leq C < \infty
\]
and \((u_\epsilon, v_\epsilon) \in X\). By using the same technique as in lemma 8 we can show a lower bound
\[
\Xi_{\epsilon}(u_\epsilon, v_\epsilon) \geq L_\epsilon(v_\epsilon) + \int_{\Omega} \frac{1}{8\epsilon} (1 - v_\epsilon)^2 \, dt \geq -C' \epsilon
\]
for some constant \( C' > 0 \). Moreover, inequalities (45) and (46) yield
\[
\int_{\Omega} \frac{1}{8\epsilon} (1 - v_\epsilon)^2 \, dt \leq C + C' \epsilon
\]
and hence in particular \( v_\epsilon \to 1 \) in \( L^2(\Omega) \) as \( \epsilon \to 0 \) and \( v = 1 \) a.e. Since \( 0 \leq v_\epsilon \leq 1 \), we have by lemma 9 that \( \lim_{\epsilon \to 0} L_\epsilon(v_\epsilon) = 0 \).

Let us next show that \( \lim_{\epsilon \to 0} S_{\epsilon}^q(u_\epsilon, v_\epsilon) = 0 \). Again since \( v_\epsilon \leq 1 \), we have
\[
|S_{\epsilon}^q(u_\epsilon, v_\epsilon)| \leq Ce^q b_\epsilon \int_{\Omega} |D_q u_\epsilon| \, dt \leq Ce^q b_\epsilon \sqrt{\int_{\Omega} |D_q u_\epsilon|^2 \, dt}
\]
where \( b_\epsilon = \int_{\Omega} u_\epsilon \, dt \). Since \( u_\epsilon \to u \) in \( L^1(\Omega) \), also \( b_\epsilon \to \int_{\Omega} u \, dt \) \( < \infty \). By assumption it holds that \( \epsilon^2 \int_{\Omega} |D_q u|^2 \, dt \leq C \) so that since \( q > 1 \), we obtain that the right-hand side of inequality (48) converges to zero as \( \epsilon \to 0 \).
Now theorem 2 implies

\[
MS(u, v) \leq \liminf_{\epsilon \to 0} F^A_T(u_\epsilon, v_\epsilon) + \lim_{\epsilon \to 0} L_\epsilon(v_\epsilon) + \lim_{\epsilon \to 0} S^\alpha_\epsilon(u_\epsilon, v_\epsilon) = \liminf_{\epsilon \to 0} \Xi_\epsilon(u_\epsilon, v_\epsilon).
\]

This yields condition (i).

Next let us consider condition (ii). By theorem 2 for any \((u, v) \in L^1(\mathbb{T}) \times L^1(\mathbb{T})\) there exists a sequence \(\{(u_\epsilon, v_\epsilon)\} \subset X\) such that \(\limsup_{\epsilon \to 0} F^A_T(u_\epsilon, v_\epsilon) \leq MS(u, v)\). By assuming that \(MS(u, v)\) is bounded, we obtain inequality (47) for \(v_\epsilon\) and lemma 9 yields

\[
\limsup_{\epsilon \to 0} \Xi_\epsilon(u_\epsilon, v_\epsilon) = \limsup_{\epsilon \to 0} F^A_T(u_\epsilon, v_\epsilon) + \lim_{\epsilon \to 0} L_\epsilon(v_\epsilon) + \lim_{\epsilon \to 0} S^\alpha_\epsilon(u_\epsilon, v_\epsilon) \leq MS(u, v).
\]

This proves condition (ii) in definition 4 and hence the claim follows.

\[\square\]

**Theorem 11.** We have that \(MS = \Gamma - \lim_{\epsilon \to 0} F^A_T\) in \(L^1(\mathbb{T}) \times L^1(\mathbb{T})\) for any \(\alpha \geq 1\).

**Proof.** We prove only the case when \(\alpha = 1\). For \(\alpha > 1\) the proof is obtained by leaving out the considerations regarding the term \(L_\epsilon\).

Note that the functionals \(F^1_\epsilon\) and \(\Xi_\epsilon\) differ only in the set \(X_\epsilon \setminus X\). Obviously since \(X \subset X_\epsilon\), condition (ii) in definition 4 follows immediately from inequality \(F^1_\epsilon \leq \Xi_\epsilon\) and theorem 10.

Let us then consider condition (i) and let \((u_\epsilon, v_\epsilon) \in X_\epsilon\) be a sequence converging to \((u, v) \in L^1(\mathbb{T}) \times L^1(\mathbb{T})\) as \(\epsilon \to 0\). By assuming that the sequence \(F^1_\epsilon(u_\epsilon, v_\epsilon)\) is bounded we note as in the proof of theorem 10 that \(v = 1\) almost everywhere. Consider the folding operation \(\Psi_1\) defined in equation (30). One can easily show that since \(0 \leq v_\epsilon \leq 1 + 30\epsilon\), we must have \(v_\epsilon = \Psi_1 \circ v_\epsilon \to v\) in \(L^1(\mathbb{T})\). Furthermore, due to lemma 3 we have

\[
\Xi_\epsilon(u_\epsilon, \Psi_1 \circ v_\epsilon) \leq F^1_\epsilon(u_\epsilon, v_\epsilon) + L_\epsilon(\Psi_1 \circ v_\epsilon) - L_\epsilon(v_\epsilon).
\]

Clearly we have \(L_\epsilon(\Psi_1 \circ v_\epsilon) - L_\epsilon(v_\epsilon) \to 0\) as \(\epsilon \to 0\) and thus

\[
MS(u, v) \leq \liminf_{\epsilon \to 0} \Xi(u_\epsilon, v_\epsilon) \leq \liminf_{\epsilon \to 0} F^1_\epsilon(u_\epsilon, v_\epsilon)
\]

which yields the result.

\[\square\]

**Lemma 5.** Let \(\alpha \geq 1\) and \(u \in SBV(\mathbb{T})\). For any sequence \(\{\epsilon_j\}_{j=1}^{\infty}, \epsilon_j \to 0\), there exist the functions \(\{(u_{\epsilon_j}, v_{\epsilon_j})\}_{j=1}^{\infty} \subset X\) converging to \((u, 1)\) in \(L^2(\mathbb{T}) \times L^2(\mathbb{T})\) such that

\[
MS(u, 1) = \lim_{j \to \infty} F'^A_T(u_{\epsilon_j}, v_{\epsilon_j}).
\]

**Proof.** Since \(u \in SBV(\mathbb{T}) \subset L^\infty(\mathbb{T})\), we may apply theorem 2 for \(p = 2\) to see that there are \(\{(u_{\epsilon_j}, v_{\epsilon_j})\}_{j=1}^{\infty} \subset X\) converging to \((u, 1)\) in \(L^2(\mathbb{T}) \times L^2(\mathbb{T})\) such that \(MS(u, 1) = \lim_{j \to \infty} F^A_T(u_{\epsilon_j}, v_{\epsilon_j})\). Following the proof of theorem 10, we can show

\[
\lim_{j \to \infty} (L_{\epsilon_j}(v_{\epsilon_j}) + S^\alpha_{\epsilon_j}(u_{\epsilon_j}, v_{\epsilon_j})) = 0
\]

and the claim follows.

\[\square\]
6.3. Convergence of minima

Let us show the equi-coercivity for the sequences of functionals studied above.

**Lemma 6.** Assume that \( v \in H^1(a, b) \), \( a < b \), \( a, b \in \mathbb{R} \), and \( \max_{t \in [a, b]} v(t) - \min_{t \in [a, b]} v(t) \geq T \). Then it holds that \( \int_a^b |Dv(t)|^2 \, dt \geq \frac{T^2}{b-a} \).

**Proof.** By the Sobolev embedding theorem, \( v \) can be extended to a continuous function on \([a, b]\). Denote by \( t_+ \) and \( t_- \) points in \([a, b]\) where

\[
v(t_+) = \max_{t \in [a, b]} v(t) \quad \text{and} \quad v(t_-) = \min_{t \in [a, b]} v(t).
\]

Without losing the generality we may assume that \( t_+ > t_- \). Then using the fundamental theorem of calculus, we see that

\[
T \leq v(t_+) - v(t_-) \leq \int_{t_-}^{t_+} |Dv(t)| \, dt \leq \|Dv\|_{L^2([a,b])} \sqrt{b-a}.
\]

This proves the statement. \( \square \)

Let us next prove the equi-coerciveness of the functionals \( F_\alpha \).

**Theorem 12.** Let \( \alpha \geq 1 \), \( C > 0 \), and \( (u_\epsilon, v_\epsilon) \in H^1(\mathbb{T}) \times H^1(\mathbb{T}) \), \( \epsilon > 0 \), be a sequence such that

\[
F_\alpha(u_\epsilon, v_\epsilon) + \|Au_\epsilon - m\|_{L^2}^2 \leq C.
\] (50)

Then there exists a subsequence \( (u_{\epsilon_j}, v_{\epsilon_j})_{j=1}^\infty \) which converges in \( L^1(\mathbb{T}) \times L^1(\mathbb{T}) \).

**Proof.** The proof is principally the same for both cases \( \alpha = 1 \) and \( \alpha > 1 \). First note that assumption (50) yields \( \|v_\epsilon - 1\|_{L^2}^2 \leq C \epsilon \) and hence the convergence of \( v_\epsilon \) to \( 1 \) in \( L^1(\mathbb{T}) \) is clear. The case for \( u_\epsilon \) follows by considering carefully how the convergence of \( v_\epsilon \) takes place. Let us first fix some \( \epsilon > 0 \) and divide the domain \( \mathbb{T} \) into \( K = \lfloor \frac{1}{\epsilon} \rfloor + 1 \) half-open intervals

\[
I^k = [\frac{k-1}{K}, \frac{k}{K})\]

where \( \lfloor \cdot \rfloor \) denotes the largest integer less than or equal to \( 1/\epsilon \). Moreover, denote

\[
\mathcal{I}_\epsilon = \left\{ 0 \leq k < K : \max_{t \in I^k} v_\epsilon(t) - \min_{t \in I^k} v_\epsilon(t) \geq \frac{1}{4} \right\}.
\]

From lemma 6 and inequality (50), we deduce that

\[
\sharp(\mathcal{I}_\epsilon) \cdot \frac{K}{16} \leq \int_\mathbb{T} |Dv|^2 \, dt \leq \frac{C}{\epsilon},
\]

and hence \( \sharp(\mathcal{I}_\epsilon) \leq 16C' \). Furthermore, denote

\[
\mathcal{J}_\epsilon = \left\{ k \in \{0, \ldots, N-1\} \setminus \mathcal{I}_\epsilon : \min_{t \in I^k} v_\epsilon(t) < \frac{1}{4} \right\}.
\]

If \( j \in \mathcal{J}_\epsilon \), we observe that the minimum of \( v_\epsilon \) on the interval \( I^k \) is less than \( 1/4 \) but also the oscillation is less than \( 1/4 \). In consequence, we have that \( I^k \subset \{ t \in \mathbb{T} | v_\epsilon(t) \leq \frac{1}{2} \} \) and the boundedness in inequality (50) yields

\[
\sharp(\mathcal{J}_\epsilon) \cdot \frac{1}{K} \leq 4 \int_\mathbb{T} (1 - v_\epsilon)^2 \, dt \leq 16\epsilon C
\]

and thus \( \sharp(\mathcal{J}_\epsilon) \leq 16C'' \). Consider the union of all intervals in the complement of \( \mathcal{I}_\epsilon \cup \mathcal{J}_\epsilon \). Since the number of indices in \( \mathcal{I}_\epsilon \cup \mathcal{J}_\epsilon \) is less than \( L = \lfloor 16(C + C'') \rfloor + 1 \), it follows that its
complement \((\mathcal{I}_s \cup \mathcal{J}_s)^c\) can be presented as a union of at most \(L\) half-open connected intervals \(K_j\) so that

\[
J(\epsilon) := \bigcup_{0 \leq k < K, k \in \mathcal{I}_s \cup \mathcal{J}_s} I_k^K = \bigcup_{j=1}^L K_j.
\]

Next we obtain \(L^q\)-boundedness for \(u_\epsilon\) with some \(q > 1\) by applying the Poincaré inequality on each interval \(I_k^K\), \(j \in \mathcal{I}_s \cup \mathcal{J}_s\), and to every \(K_j\). Let \(I \subset \mathbb{T}\) be an open connected interval and \(p > 1\) such that \(\frac{1}{p} = \frac{1}{2} + s\). By the Poincaré inequality, we have

\[
\|u_\epsilon - b_\epsilon\|_{L^p(I)} \leq c |I| \|Du_\epsilon\|_{L^p(I)},
\]

where \(b_\epsilon = \frac{1}{|I|} \int_I u_\epsilon(t) \, dt\) is the average of \(u_\epsilon\) on \(I\) which by lemma 10 satisfies \(|b_\epsilon| \leq C |I|^{-p} \). Using the triangle inequality to estimate the left-hand side of (51) we obtain

\[
\|u_\epsilon\|_{L^q(I)} \leq c |I| \|Du_\epsilon\|_{L^q(I)} + |I|^{\frac{1}{q'}} |b_\epsilon| \leq C |I| \|Du_\epsilon\|_{L^p(I)} + 1.
\]

Next, by inequality (55) we obtain

\[
\|u_\epsilon\|_{L^q(I)} \leq C(|I| + 1) \|Du_\epsilon\|_{L^p(I)} \leq C.
\]

Applying inequalities (52), (54) and (55) to (53) yields

\[
\|u_\epsilon\|_{L^q(I)} \leq C \left( \sum_{k \in \mathcal{I}_s \cup \mathcal{J}_s} \right) \left( e^{\frac{s}{2}} + 1 \right) \left( \sum_{j=1}^L (|K_j| + 1) \right) \leq C
\]

since \(L\) and \(\mathfrak{z}(\mathcal{I}_s \cup \mathcal{J}_s)\) were bounded. Note that this bound is uniform with respect to \(\epsilon\). By the Banach–Alaoglu theorem we can extract a subsequence, denoted also by \(u_\epsilon\), which converges to some \(u \in L^p(\mathbb{T})\) weakly.

Next, by inequality (55) we obtain \(\|u_\epsilon\|_{W^{1,q}(J(\epsilon_k))} < C\) for some constant \(C > 0\) independent of \(\epsilon\). Let us then extract a subsequence \(\{u_{\epsilon_j}\}_{j=1}^\infty\) in the following way: denote \(Z_\epsilon = \{k \in K : k \in \mathcal{I}_s \cup \mathcal{J}_s\} \subset \mathbb{T}\). Let \(\epsilon_j\) be such that the set \(Z_{\epsilon_j}\) converges in the Hausdorff topology to some discrete set \(Z\) such that \(\mathfrak{z}(Z) \leq L\). Note that \(J(\epsilon_j)^c\) is included in an \(\epsilon_j\)-neighbourhood of the set \(Z_{\epsilon_j}\). Then it follows that for every \(\ell \in Z_{\epsilon_j}\) and \((1/\ell)\)-neighbourhood \(U_\ell\) of \(Z\) we have \(\lim_{j \to \infty} \|u_{\epsilon_j}\|_{W^{1,q}(U_\ell)} < C\). Furthermore, by the Banach–Alaoglu theorem we can extract another subsequence, denoted also by \(u_{\epsilon_j}\), which converges weakly in \(W^{1,q}(U_\ell^c)\) for all \(\ell\). The Sobolev embedding theorem then yields that this subsequence also converges strongly in \(L^1(U_\ell^c)\). We conclude that

\[
\lim_{j \to \infty} \|u - u_{\epsilon_j}\|_{L^1(U_\ell^c)} \leq \lim_{j \to \infty} \left( \|u - u_{\epsilon_j}\|_{L^1(U_\ell^c)} + \|u - u_{\epsilon_j}\|_{L^1(U_\ell)} \right)
\]

\[
\leq \lim_{j \to \infty} \left( \frac{2L}{\ell} \right)^{1/p'} \leq C \left( \frac{2L}{\ell} \right)^{1/p'}
\]

for any \(\ell \in Z\), where \(1/p + 1/p' = 1\). Finally, the result follows since \(\ell\) was arbitrary. 

\[\square\]
Theorem 13. Consider fixed $\epsilon > 0$. Let $(u_n, v_n) \in PL(n) \times PL(n)$ be a sequence such that
\[
F_{\epsilon,a}(u_n, v_n) + \|A_n u_n - m\|_{L^2}^2 \leq C
\]
for some constant $C < \infty$. Then there exists a subsequence $(u_{n_j}, v_{n_j})$ which converges weakly in $H^1(\mathbb{T}) \times H^1(\mathbb{T})$.

Proof. We see that from theorem 12 we have a subsequence for some constant $c < 1$ and let $(u_{n_j}, v_{n_j}) \in X_\epsilon$ be a minimizer of problem (23). By the equi-coercivity theorem 12 we have a subsequence $\{(\epsilon_j, \alpha_j)_j\}$ such that $v_{\alpha_j}$ converges to $v$ in $L^1(\mathbb{T})$ and $\alpha_j$ to some $\alpha_\infty \in L^1(\mathbb{T})$. Let us then prove that $\alpha_\infty$ is a minimizer of problem (24). Let $u \in L^1(\mathbb{T})$ be such that $MS(u, 1) + R(u) < \infty$. Then clearly $u \in SBV(\mathbb{T})$ and by lemma 5 there exists a sequence $(u_j, v_j)_j \subset X_\epsilon$ such that $\lim_{j \to \infty} \alpha_j = \alpha_\infty$. Since the residual $R$ is continuous in $L^2(\mathbb{T})$ and lower semicontinuous in $L^1(\mathbb{T})$ we obtain using theorem 11 that
\[
MS(u_\infty, 1) + R(u_\infty) \leq \inf_{j \to \infty} \lim \inf F_{\epsilon_j}(u_{\alpha_j}, v_{\alpha_j}) + \lim \inf R(u_{\alpha_j})
\]
\[
\leq \lim \inf_{j \to \infty} \left( F_{\epsilon_j}(u_{\alpha_j}, v_{\alpha_j}) + R(u_{\alpha_j}) \right)
\]
\[
= MS(u, 1) + R(u).
\]
This proves that $(u_\infty, 1)$ is a minimizer of $MS + R$ since $u \in L^1(\mathbb{T})$ was an arbitrary function for which (57) is finite. The case $\alpha > 1$ follows identically.

Finally, before the numerical results, we present two remarks on issues discussed in the introduction.

Remark 1. Let us consider the case when $\alpha \neq \kappa$ in equation (8). Then the minimization problem of finding MAP estimates can be written as
\[
\min_{u, v \in PL(n)} \int_\mathbb{T} \left( -N^{1-a} \log(e^2 + (Q_n v)^2) + (e^2 + (Q_n v)^2)|D^t u|^2 
\right.
\]
\[
+ \epsilon |Dv|^2 + \frac{1}{4\epsilon} (1 - v^2 + N^{\kappa-a} |P_n Au - m|^2) \right) dt.
\]
In consequence, the residual term becomes over- or under-weighted in the limit regardless of the particular choices of $\alpha$ or $\kappa$.

For practical purposes, consider $N = cK$ with a positive constant $c \neq 1$, i.e. the discretizations are coupled but not equal. In this case it is straightforward to show that the constant $c$ appears as an additional weight in the MAP estimation problem (20) in the following way:
\[
\min_{(u, v) \in PL(n) \times PL(n)} \left( F_{\epsilon,a}(u, v) + c^\alpha \|P_n Au - m\|_{L^2}^2 \right).
\]
Results 4 and 3 remain valid when the change of the likelihood term is taken into account.

If $\alpha = \kappa \geq 1$, then our results are summarized in theorem 4 and the MAP estimates converge. Note that in this case the effect of the logarithm term vanishes asymptotically and the theorems concerning the limiting functionals coincide with the famous results by Ambrosio and Tortorelli [6].
Remark 2. Let us consider statement (b’) in the introduction. When $\alpha \geq 1$,

$$
\mathbb{E}\left((V_{n,\epsilon}^\alpha - 1, \phi)^2_{L^2}\right) = N^{-\alpha} \left\| \frac{1}{4\epsilon} I - \epsilon \Delta \right\|_{L^2}^{-1/2} \phi^2_{L^2}
$$

and

$$
\mathbb{E}\left((U_{n,\epsilon}^\alpha, \phi)^2_{L^2}\right) \leq \epsilon^{-2} N^{-\alpha} \left\| D_q^{-1} \phi \right\|_{L^2}^2,
$$

where $\phi \in L^2(\mathbb{T})$. Applying this for the orthonormal basis $\{e_j\}_{j=-\infty}^\infty \subset L^2(\mathbb{T})$, $e_j(t) = \exp(2\pi i j t)$, yields easily with a fixed $\epsilon > 0$ that

$$
\lim_{n \to \infty} \mathbb{E}\left\| V_{n,\epsilon}^\alpha - 1 \right\|_{L^2}^2 = 0 \quad \text{and} \quad \lim_{n \to \infty} \mathbb{E}\left\| U_{n,\epsilon}^\alpha \right\|_{L^2}^2 = 0.
$$

In particular, this implies that the random variable $(U_{n,\epsilon}^\alpha, V_{n,\epsilon}^\alpha)$ in $L^2(\mathbb{T}) \times L^2(\mathbb{T})$ converges to $(0, 1)$ in probability as $n \to \infty$ [34].

Remark 3. In this paper we have assumed the unknown to be a periodic one-dimensional function. It is an interesting question for future studies to apply the methodology used here to higher dimensional domains and nonperiodic functions.

For the higher dimensions, it is pointed out in [31] that the realizations of the random variable $V$ are not Hölder continuous. In fact, they belong to $L^2$ with probability zero in domains with dimension higher than 1. This produces additional difficulties for the analysis of the infinite-dimensional prior structure and the convergence of the CM estimates. Likewise, nonperiodic functions as priors require additional technical considerations for the stochastic modelling done in [31]. For example, paper [38] takes the boundary effects into account in the stochastic analysis for total variation priors.

The key issue for extending the results in this paper is the proof of equi-coerciveness in theorem 12. Straightforward remedy can be provided by making additional $L^\infty$-boundedness assumptions on the priors as discussed in section 2.4. However, this would also require reconsideration of the stochastic models.

7. Numerical considerations

In this section we study the qualitative behaviour of MAP estimates by giving a numerical example with the scaling $\alpha = 1$. Our purpose is to demonstrate that the MAP estimates do behave numerically in a similar manner in all discretizations, i.e. for different choices of parameter $n$. This can be expected given the results in theorem 4.

The numerical simulations for convergence of the CM estimates are demonstrated in [31] in the case $\alpha = 0$.

7.1. The model problem

We consider a Bayesian deblurring problem with linear operator $A = (I - \Delta)^{-s/2} : L^2(\mathbb{T}) \to H^s(\mathbb{T})$ for a given $0 < s < 1/2$. Note that this operator satisfies the condition $\|u\|_{H^{-s}} \leq \|Au\|_{L^2}$ for any $u \in H^{-s}(\mathbb{T})$. We assume the measurements to be obtained via the projections $P_n f = \sum_{j=1}^N \langle f, e_j \rangle_{H^{-s} \times H^s} e_j$ for any $f \in H^s(\mathbb{T})$ where the $L^2$-orthonormal basis functions $\{e_j\}_{j=-N}^N$ are $e_j(t) = \exp(-2\pi i j t)$ for $t \in [0, 1)$. It is straightforward to show that the projections $P_n$ are proper measurement projections in the sense of definition 6.

Let us now introduce some notation. For any $n \in \mathbb{N}$ we denote $\phi^n_j \in PL(n)$ a function such that $\phi^n_j(k/n) = \delta_{kj}$. The basis $\{\phi^n_j\}_{j=1}^n \subset PL(n)$ is called the roof-top functions. Let then $B_n \in \mathbb{R}^{N \times N}$ be a matrix such that $(B_n)_{jk} = \langle \phi^n_j, \phi^n_k \rangle_{L^2}$ for $1 \leq j, k \leq N$. 

25
In the following we use bold symbols to denote the coefficients of any function \( f \in PL(n) \) presented in the roof-top basis, i.e. if \( f = \sum_{j=1}^{N} f_j \phi_j \), then \( f = (f_1, \ldots, f_N) \in \mathbb{R}^N \). Furthermore, denote by \( D_{\phi} \) and \( Q_{\phi} \) the matrix presentations of operators \( D_{\phi} : PL(n) \to PC(n) \), respectively, from the roof-top basis to the basis of piecewise constant functions \( \{ \chi_{IN} \}_{j=1}^{N} \subset PC(n) \) where \( I_j^{N} = [j/N, (j+1)/N) \). Furthermore, denote the matrix \( \Lambda_n(v) = \text{diag}(\epsilon^2 + (Q_n v)^2) \in \mathbb{R}^{N \times N} \). With these notations the functional \( F_{\epsilon,n}^L \) written in terms of the coefficients in the roof-top basis functions has the form

\[
F_{\epsilon,n}^L(u, v) = -N^{-\alpha} \log(\det \Lambda(v)) + \frac{1}{N} (D_{\lambda} u)^T \Lambda_n(v) (D_{\lambda} u)
+ \frac{\epsilon}{N} \|D_n v\|_2^2 + \frac{1}{4\epsilon} (1 - v)^T B_n (1 - v).
\] (58)

Furthermore, assuming a white noise with \( \sigma^2 I \) covariance, the likelihood term has the form \( \frac{1}{\sigma^2} \|A_n u - m\|_2^2 \), where \( A_n \in \mathbb{R}^{(2N+1) \times N} \) maps \( u \) to the coefficients of \( P_n A u \) in the basis \( \{e_j\}_{j=-N}^{N} \). The components of matrix \( A_n \) satisfy \( (A_n)_{jk} = \langle e_j, (I - \Delta)^{-s/2} \phi_n \rangle_{L^2} \) for \(-N \leq j \leq N \) and \( 1 \leq k \leq N \).

### 7.2. Computational methods

Because of the non-quadratic terms in \( F_{\epsilon,n}^L \) we have chosen to implement an alternate minimization scheme (see e.g. [12]). The convergence of such a method is studied in [45] in a setting without the logarithm term \( L \). Producing a convergence result in our case lies outside the focus of this section.

Let us now write in pseudo-code how the minimizers are achieved.

1. Initialize \( u^0, v^0 \in \mathbb{R}^N \) and set \( j := 1 \).
2. Solve the minimization problem
   \[
   \min_{u \in \mathbb{R}^N} \frac{1}{N} (D_{\lambda} u)^T \Lambda_n(v) (D_{\lambda} u) + \frac{1}{\sigma^2} \|A_n u - m\|_2^2 ,
   \]
   which is equivalent to solving the equation \( \frac{1}{N} D_{\lambda}^T \Lambda_n(v^{-1}) D_{\lambda} + \frac{1}{\sigma^2} A_n^T A_n \) \( u = \frac{1}{\sigma^2} A_n^T m \) and set \( u^j = u \).
3. Solve the minimization problem
   \[
   \min_{v \in \mathbb{R}^N} \left(-N^{-\alpha} \log(\det \Lambda_n(v)) + \frac{1}{N} (D_{\lambda} u^j)^T \Lambda_n(v) (D_{\lambda} u^j)
   \right.
   \]
   \[+ \frac{\epsilon}{N} \|D_n v\|_2^2 + \frac{1}{4\epsilon} (1 - v)^T B_n (1 - v) \]
   \[\left. \right) \]
   and set \( v^j = v \).
4. If \( (u^j, v^j) \) satisfies \( F_{\epsilon,n}^L(u^j, v^j) \leq F_{\epsilon,n}^L(u^{j-1}, v^{j-1}) - \delta \), go to step (2); else stop.

### 7.3. Results

We implemented the problem with operator \( A \) having the parameter \( s = 0.35 \) and measurement noise with variance \( \sigma = 5 \times 10^{-3} \), i.e. replace \( N^{-s} \) in equation (8) with \( \sigma N^{-s} \). Furthermore, the scaling of the prior is assumed to be \( \alpha = 1 \). We used six different sets of data with two true values of \( u \) and three discretization sizes \( N = 64, N = 512 \) and \( N = 2048 \). The MAP estimates were computed with sharpness parameters \( \epsilon = 2 \times 10^{-2}, 1 \times 10^{-2}, 6 \times 10^{-3} \). The reconstructions are shown in figures A1 and A2.
In figure A1 the true value of $u$ is a simple step function. We have weighted the residual with constant $c = 14$. In figure A2 the true value is piecewise smooth with $\sharp(S_u) = 4$ and the residual was weighted with $c = 10$. The initial values in all the computations were vectors $u^0 = 0$ and $v^0 = 1$. Step (2) in the algorithm was implemented by using Matlab’s backslash function and in step (3) we used a gradient-descent method by choosing step-sizes with a line search algorithm. The minimization in step (3) was stopped when either no satisfying step-size was found or the values of the functional did not change by high accuracy. All computations were stopped at 50 iterations.

To achieve faster convergence we applied adaptation to the choice of $\sigma$. We chose $\delta = 10^{-4}$ for the first 20 iterations and $\delta = 10^{-6}$ for the rest.

We perform all the computations with Matlab 7.6 running in a desktop PC computer with Dual Intel Xeon processor running at 2.80 GHz and 4 GB of RAM. Computations took less than 10 s for $N = 512$ and less than 80 s for $N = 2048$.

7.4. Discussion

A visible feature of figures A1 and A2 is that the reconstructions do not change qualitatively by increasing the discretization parameter $n$. This result is in line with theorem 4, i.e. if one fixes $\epsilon > 0$ and takes $n$ to infinity, then the minimizers converge to Ambrosio–Tortorelli minimizers.
Figure A2. A piecewise smooth function. Left: the true value of $u$ and the noisy measurement $m_n = M_n(ω_0)$. Middle: $u_{\text{MAP}}$ estimates. Right: $v_{\text{MAP}}$ estimates. The thick, dashed and thin lines represent reconstructions with sharpness $\epsilon = 2 \times 10^{-2}, 1 \times 10^{-2}$ and $6 \times 10^{-3}$, respectively. Axis limits are the same in each plot.

It is also evident that the parameter $\epsilon$ controls how sharp reconstructions one can obtain. In figure A2 this is visible with the second peak. Namely, with the value $\epsilon = 0.02$ this peak is smoothed for $N = 512$ and $N = 2048$, whereas with the other values the reconstruction becomes sharp.

The convergence of the algorithm was satisfactory especially for $u$. In most of the runs the value of $u$ was achieved very accurately with less than ten iteration steps. However, the function $v$ still evolved slowly after this and a satisfactory estimate was obtained with 50 iteration steps where also each run was stopped. The authors expect that this slowness can be overcome by a more sophisticated minimization algorithm in step (3).

Acknowledgments

The first author was supported by Emil Aaltonen foundation, Graduate School of Inverse Problems (Academy of Finland) and Finnish Centre of Excellence in Inverse Problems Research (Academy of Finland). The authors are grateful to Luca Rondi and Samuli Siltanen for helpful comments.

Appendix. Technical lemmata

Properties of the domain $X_\epsilon$

In the definition of the domain $X_\epsilon$ in equation (21), one restricts the values of the function $v$ in the pair $(u, v) \in H^1(\mathbb{T}) \times H^1(\mathbb{T})$ to the interval $[0, 1 + 30\epsilon]$. Let us now discuss some
properties related to this choice. Define the function \( g_\epsilon(t) = -\log(\epsilon^2 + t^2) + \frac{1}{4\epsilon}(1-t)^2 \) for \( t \in \mathbb{R} \) where \( \epsilon > 0 \) is fixed.

**Lemma 7.** Assume that \( 0 < \epsilon < \frac{1}{4} \). The function \( g_\epsilon \) has a unique minimizer \( t_\epsilon \), which satisfies \( 1 \leq t_\epsilon \leq 1 + 30\epsilon \). Furthermore, the inequality \( g_\epsilon(t) \leq g_\epsilon(s) \) holds when \( s \) and \( t \) satisfy one of the following conditions:

(i) \( 1 \leq t \leq 1 + 30\epsilon \) and \( s > 1 + 30\epsilon \),
(ii) \( t \in [0, 1] \) and \( s \leq -1 \) or \( t \in [0, 1] \) and \( s = -t \).

**Proof.** Clearly, one has \( g_\epsilon(t) \leq g_\epsilon(-t) \) for any \( t \in \mathbb{R} \). This proves claim (iii) and since \( \lim_{\epsilon \to \infty} g_\epsilon(t) = \infty \), this also shows that the global minimizer has to be located in \( \mathbb{R}_+ \).

The derivative \( Dg_\epsilon \) has the form \( Dg_\epsilon(t) = -\frac{2t}{\epsilon^2} + \frac{1}{2}(t-1) \) for \( t \in \mathbb{R} \). The first term is negative everywhere in \( \mathbb{R}_+ \). Since the second term increases linearly and is positive for \( t > 1 \), the zeros of \( Dg_\epsilon \) on \( \mathbb{R}_+ \) have to be greater than 1. Also since \( \lim_{\epsilon \to \infty} Dg_\epsilon(t) = \infty \) and the first term is strictly decreasing for \( t > 1 \), the function \( D_\epsilon \) has a unique zero \( t_\epsilon \) for \( t > 1 \). This yields the existence of a unique minimizer for \( g_\epsilon \). Furthermore, claim (ii) can be easily deduced since \( Dg_\epsilon(t) < 0 \) for \( t < 1 \) when \( \epsilon < 1/8 \).

Let us now show an upper bound to \( t_\epsilon \). Apply inequality \( \frac{t}{\epsilon^2} \leq \frac{1}{4} \) for \( t > 1 \) to obtain a lower bound to the function \( Dg_\epsilon \). By solving the equation \( Dg_\epsilon(t) = -\frac{1}{t} + \frac{1}{2}(t_\epsilon - 1) = 0 \) for \( t_\epsilon > 1 \), one obtains a bound \( t_\epsilon \leq t_\epsilon \). A short computation yields \( t_\epsilon = 1 + \frac{1}{2}\sqrt{1 + 8\epsilon - 1} \leq 1 + 2\epsilon \).

Finally, let us study claim (i). From the above it is evident that there exists a unique point \( s_\epsilon \) such that \( g_\epsilon(s_\epsilon) = g_\epsilon(1) \). In the following we show that \( s_\epsilon < 1 + 30\epsilon \). Denote \( h_\epsilon(t) = g_\epsilon(t) - g_\epsilon(1) \). Then we have that

\[
 h_\epsilon(t) = -\log \frac{\epsilon^2 + t^2}{\epsilon^2 + 1} + \frac{1}{4\epsilon}(1-t)^2 \geq 1 - \frac{\epsilon^2 + t^2}{\epsilon^2 + 1} + \frac{1}{4\epsilon}(1-t)^2
\]

for any \( t \geq 1 \) where we have used the inequalities \( -\log x \geq -x + 1 \) for \( x \geq 0 \). The quadratic function on the right-hand side has a zero in \( t_\epsilon = 1 \) and with a detailed calculation, one can show that the second zero satisfies \( t_2 < 1 + 30\epsilon \) for \( \epsilon < \frac{1}{8} \). \( \square \)

**Auxiliary bounds**

Here we show auxiliary technical lemmata. Define

\[
 G_{\epsilon,n}(v, b) = \int \left( -N \log(\epsilon^2 + (Q_nv)^2) + \frac{b}{4\epsilon}(1 - v_n)^2 \right) \, dt
\]

where \( v_n \in PL(n), b, \epsilon > 0 \), \( \alpha \in \mathbb{R} \) and \( n \in \mathbb{N} \).

**Lemma 8.** For any \( 0 < \epsilon < \frac{1}{4} \), \( n \in \mathbb{N} \) and \( b \geq 0 \), there are constants \( C \) and \( C(b) \) such that

\[
 -C(b)\epsilon N^2 \leq \inf_{v \in PL(n)} G_{\epsilon,n}(v, b) \leq -C(\sqrt{\epsilon}N - 1).
\]

**Proof.** The upper bound for the infimum follows by setting \( v = 1 + \sqrt{\epsilon} \) and using the inequality \( \log(1 + x) \geq \frac{x}{2} \) for small \( x > 0 \). For the lower bound, first note that

\[
 -\log(\epsilon^2 + (Q_nv)^2) \geq -2\log(\epsilon^2 + (Q_nv)^2) \geq -2\log(\epsilon + |Q_nv|) \geq -2(\epsilon + |Q_nv| - 1).
\]

Since \( \int T |Q_nv| \, dx \leq \int T |v| \, dx \), it also holds that

\[
 \int T -N \log((\epsilon^2 + (Q_nv)^2) \, dx \geq \int T -2N(\epsilon + |v| - 1) \, dx.
\]
Now denote \( h_t(t) = -2N(\epsilon + |t| - 1) + \frac{b}{4\epsilon}(1-t)^2 \) for any \( t \in \mathbb{R} \). Clearly we have \( h_+(-t) \geq h_+(t) \) for \( t \geq 0 \). For positive values of \( t \), the function \( h_+ \) is the quadratic function \( h_+(t) = -2N\epsilon - 2N(t-1) + \frac{b}{4\epsilon}(t-1)^2 \) with respect to the variable \( t-1 \geq 0 \). The minimum of this function is obtained when \( t-1 = \frac{4\epsilon}{b}N\epsilon \) and thus \( h_+(t) \geq -2N\epsilon - \frac{4\epsilon}{b}N^2\epsilon \). It is now easy to verify that
\[
\int_{-T}^T -2N(\epsilon + |t| - 1) + \frac{b}{4\epsilon}(1-t)^2 \, dt \geq -\frac{4\epsilon}{b}N^2\epsilon - 2N\epsilon \geq -C(b)N^2\epsilon.
\]
Together with inequality (A.2), this yields the claim.

\[ \square \]

Lemma 9. Assume that a sequence \( v_\epsilon \in H^1(T; [0, 1 + C\epsilon]) \) satisfies \( \int_T (1 - v_\epsilon)^2 \, dt \leq C'\epsilon \) for some constants \( C, C' > 0 \). Then it follows that \( \lim_{\epsilon \to 0} L_\epsilon(v_\epsilon) = 0 \).

**Proof.** Let us denote \( E_\epsilon = \{ t \in T | v_\epsilon(t) \leq \frac{1}{2} \} \) for \( \epsilon > 0 \). The Lebesgue measure of \( E_\epsilon \) is bounded by \( |E_\epsilon| \leq C\epsilon \) and thus \( \int_{E_\epsilon} |\log(\epsilon + v_\epsilon^2)| \, dt \leq 2C\epsilon \log \epsilon \) which converges to zero as \( \epsilon \to 0 \). Denote \( \tilde{v}_\epsilon = \max(v_\epsilon, \frac{1}{2}) \). Clearly also \( v_\epsilon \to 1 \) in \( L^2(T) \) and hence by the Lebesgue dominated convergence theorem \( \lim_{\epsilon \to 0} L_\epsilon(v_\epsilon) \leq \lim_{\epsilon \to 0} (L_\epsilon(\tilde{v}_\epsilon) + 2C\epsilon \log \epsilon) = 0 \). This proves the statement.

The following lemma is proved in [54] in more detail.

Lemma 10. For any \( 0 < s < \frac{1}{2} \), \( u \in L^1(a, b) \cap H^{-t}(a, b) \) with \( a, b \in \mathbb{R} \) such that \( b > a \) we have
\[
\left| \int_a^b u \, dt \right| \leq C \| b - a \|^{\frac{1}{2} - t} \| u \|_{H^s(a, b)}.
\]

**Proof.** By [41] the dual space of \( H^s(a, b) \) is \( \tilde{H}^s(a, b) = \{ f \in H^s(\mathbb{R}) | \text{supp}(f) \subset [a, b] \} \) with norm \( \| f \|_{\tilde{H}^s(a, b)} = \| f \|_{H^s(\mathbb{R})} \). Furthermore, the mapping \( T : f \mapsto f \mathbf{1}_{[a, b]} \) is continuous in \( H^s(\mathbb{R}) \) for any \(-1/2 < t < 1/2\). In particular, we have that \( \mathbf{1}_{[a, b]} \in \tilde{H}^{s'}(a, b) \). Without losing any generality we assume that \( a = -b \). The Fourier transform of \( \mathbf{1}_{[-b, b]} \) satisfies
\[
\hat{\mathbf{1}}_{[-b, b]}(\xi) = C \frac{\sin(b\xi)}{\xi} \] and thus
\[
\left| \int_{-b}^b u \, dt \right| \leq \| \mathbf{1}_{[-b, b]} \|_{\tilde{H}^s(-b, b)} \| u \|_{H^{-t}(-b, b)}
\]
\[
= C \left( \int_{-\infty}^\infty \left| \frac{\sin(b\xi)}{\xi} \right|^2 (1 + |\xi|)^{2s} \, d\xi \right)^{1/2} \| u \|_{H^{-t}(-b, b)}
\]
\[
\leq C' b^{\frac{1}{2} - t} \| u \|_{H^{-t}(-b, b)}
\]
for some constant \( C' > 0 \). \( \square \)

**References**

[1] Ambrosio L 1989 A compactness theorem for a new class of functions of bounded variations *Rend. Mat. Ital.* B 3 857–81

[2] Ambrosio L 1990 Existence theory for a new class of variational problems *Arch. Ration. Mech. Anal.* 111 291–322

[3] Ambrosio L 1995 A new proof of the SBV compactness theorem *Calc. Var.* 3 127–37

[4] Ambrosio L and Dal Maso G 1990 A general chain rule for distributional derivatives *Proc. Am. Math. Soc.* 108 691–702

[5] Ambrosio L, Fusco N and Pallara D 2000 *Functions of Bounded Variation and Free Discontinuity Problems* (Oxford: Oxford University Press)

[6] Ambrosio L and Tortorelli V M 1990 Approximation of functionals depending on jumps by elliptic functional via \( \Gamma \)-convergence *Commun. Pure Appl. Math.* 43 999–1036
[39] Lassas M, Saksman E and Siltanen S 2009 Discretization-invariant Bayesian inversion and Besov space priors
Inverse Problems Imaging 3 87–122
[40] Lehman M, Päivärinta L and Somersalo E 1989 Linear inverse problems for generalized random variables
Inverse Problems 5 599–612
[41] Lions J L and Magenes E 1972 Non-Homogeneous Boundary Value Problems and Applications I (Berlin: Springer)
[42] Mandelbaum A 1984 Linear estimators and measurable linear transformations on a Hilbert Space Z. Wahrsch. Verw. Gebiete 62 385–97
[43] Mumford D and Shah J 1989 Optimal approximations by piecewise smooth functions and associated variational
problems Commun. Pure Appl. Math. 42 577–685
[44] Neubauer A and Pikkarainen H K 2008 Convergence results for the Bayesian inversion theory J. Inverse Ill-posed
Probl. 16 601–13
[45] Nikolova M and Ng M K 2005 Analysis of half-quadratic minimization methods for signal and image recovery
SIAM J. Sci. Comput. 27 937–66
[46] Perona P and Malik J 1990 Scale-space and edge detection using anisotropic diffusion IEEE Trans. Pattern
Anal. Mach. Intell. 12 629–39
[47] Piiroinen P 2005 Statistical measurements, experiments and applications Ann. Acad. Sci. Fenn. Math.
Diss. 143
[48] Raj A, Singh G, Zabih R, Kressler B, Wang Y, Schuff N and Weiner M 2007 Bayesian parallel imaging with
edge-preserving priors Magn. Reson. Med. 57 821
[49] Ramlau R and Ring W 2007 A Mumford-Shah level-set approach for the inversion and segmentation of X-ray
tomography data J. Comput. Phys. 221 539–57
[50] Rondi L 2008 On the regularization of the inverse conductivity problem with discontinuous conductivities
Inverse Problems Imaging 2 397–409
[51] Rondi L and Santosa F 2001 Enhanced electrical impedance tomography via the Mumford-Shah functional
ESAIM Control Optim. Calculus Variations 6 517–38
[52] Rozanov Yu A 1982 Markov Random Fields (Applications of Mathematics) (Berlin: Springer)
[53] Rudin L I, Osher S and Fatemi E 1992 Nonlinear total variation based noise removal algorithms Physica
D 60 259–68
[54] Saksman E 1988 Sobolev spaces $H^s(\Omega)$ and the asymptotics of the norm $\|1\|_{H^s(\Omega)}$ (in finnish) Licenciate
Thesis University of Helsinki
[55] Sethian J A 1999 Level Set Methods and Fast Marching Methods (Cambridge: Cambridge University Press)
[56] Strong D M and Chan T F 2003 Edge-preserving and scale-dependent properties of total variation regularization
Inverse Problems 19 165–87
[57] Voutilainen A, Kolehmainen V and Kaipio J P 2001 Statistical inversion of aerosol size measurement data
Inverse Problems Eng. 9 67–94