Can one use total variation prior for edge-preserving Bayesian inversion?

Matti Lassas[1] **and Samuli Siltanen**[2]

1 Rolf Nevanlinna Institute, P.O.Box 4, FIN-00014 University of Helsinki, Finland

2 Instrumentarium Corp. Imaging Division, Finland, and Gunma University, Department of Mathematics, Tenjin-cho 1-5-1, 376-8515 Kiryu, Japan

Abstract. Estimation of non-discrete physical quantities from indirect linear measurements is considered. Bayesian solution of such an inverse problem involves discretizing the problem and expressing available *a priori* information in the form of a prior distribution in a finite-dimensional space. Since *a priori* information is independent of the measurement, the discretization of the unknown quantity can be arbitrarily fine regardless of the number of measurements. The main result is that total variation prior distribution has certain unfavorable features that appear with very fine discretizations. First, there is no choice of regularization parameter as function of discretization level making Bayesian maximum a posteriori (MAP) and conditional mean (CM) estimates converge simultaneously to useful limits when the discretization is refined arbitrarily. Second, in the case when CM estimates converge, they are not edge-preserving in the limit. Theoretical findings are illustrated by a numerical example with computer simulated data.

Version 45, 5.2.2004

1. Introduction

Consider an indirect noisy measurement m of a physical quantity u:

$$m = Au + \varepsilon, \tag{1.1}$$

where ε is random noise and the linear operator A models the measurement. The corresponding inverse problem is

$$\text{given } m, \text{ find } u. \tag{1.2}$$

We assume that the object u is *a priori* known to be piecewise regular. Our aim is to raise methodological concerns about the solution of (1.2) with Bayesian inversion using total variation prior distribution.

Practical solution of (1.2) with Bayesian inversion requires discretization of the problem and expressing available *a priori* information on u in the form of a *prior distribution* $\pi_{\rm pr}^{(n)}$ in a finite-dimensional subspace $Y_n \subset Y$, where u is *a priori* known to belong to some function space Y. Let m and ε be random vectors taking values in $\mathbb{R}^N$, and denote their distributions by π_m and π_ε, respectively. Given a realization $\hat{m}$ of the measurement $m = Au + \varepsilon$, Bayes' formula yields the *posterior distribution* for the random variable u_n taking values in Y_n:

$$\pi(u_n \,|\, \hat{m}) = \frac{\pi_{\rm pr}^{(n)}(u_n)\,\pi(\hat{m}\,|\,u_n)}{\pi_m(\hat{m})} \sim \pi_{\rm pr}^{(n)}(u_n)\,\pi_\varepsilon(\hat{m} - Au_n). \tag{1.3}$$

The approximate solution of (1.2) is given as some point estimate for (1.3). Such estimates include the *maximum a posteriori* (MAP) and *conditional mean* (CM) estimates defined by

$$u_n^{\mathrm{MAP}} = \arg \max_{u_n \in Y_n} \pi(u_n \,|\, \hat{m}), \qquad u_n^{\mathrm{CM}} = \int_{Y_n} u_n \, \pi(u_n \,|\, \hat{m}) \, du_n,$$

respectively. The crucial step in Bayesian inversion is the construction of Y_n and $\pi_{\mathrm{pr}}^{(n)}$. The probability measure $\pi_{\mathrm{pr}}^{(n)}$ should assign high probability to functions $u_n \in Y_n$ that are typical in light of *a priori* information on u, and low probability to atypical functions. For more details, see [22, 15, 17, 31, 25, 12].

A celebrated solution method for (1.2) called Total Variation (TV) regularization was introduced for edge-preserving noise removal by Rudin, Osher and Fatemi [29] and later successfully applied to other inverse problems [9, 33]. TV regularization is equivalent to determining the MAP estimate for (1.3) with TV prior distribution. This observation inspired consideration of CM and other Bayesian estimates using the TV prior [17, 30, 19]; preservation of edges was achieved with fixed discretization.

However, from the pure Bayesian point of view, *a priori* information and its discrete representation are independent of the measurement, and the dimension n can be freely chosen. In our view, the space Y_n and the distribution $\pi_{\mathrm{pr}}^{(n)}$ should be constructed for all n so that the following two conditions are satisfied:

(i) There is a random function v taking values in Y such that $\lim_{n\to\infty} u_n = v$ (in a sense to be made precise later). This guarantees that the representation of *a priori* information becomes more accurate when n grows.

(ii) There are continuous linear operators $T_n : Y \to Y_n$ such that $u_n = T_n v$. This means that the finite-dimensional approximations to u are achieved systematically from the limit function v. (We restrict ourselves to linear discretizations T_n as we consider linear inverse problems only.)

Any choice of Y_n and $\pi_{\mathrm{pr}}^{(n)}$ having properties (i) and (ii) is called *discretization invariant.*

We show (for a generic one-dimensional problem) that TV prior is not discretization invariant. We take Y to be the space of continuous functions on the interval $[0, 1]$ vanishing at the endpoints and consider a general class of linear measurements. Our choice of $Y_n \subset Y$ is the space of piecewise linear continuous functions specified by their point values at $\{\frac{1}{n+1}, \frac{2}{n+1}, \dots, \frac{n}{n+1}\} \subset [0, 1]$. Further, we take $\pi_{\mathrm{pr}}^{(n)}$ in (1.3) to be the discrete TV prior with regularization parameter $\alpha_n > 0$. Our main theorems 4.1 and 5.2 concern the behavior of MAP and CM estimates for (1.3) when $n \to \infty$. Their proofs are based on epiconvergence of optimization problems and specific types of stochastic convergence, respectively. According to our theorems, there are only two choices of regularization parameter α_n as function of n leading to nontrivial convergence behavior:

If a_n does not depend on n, any sequence $\{u_n^{\mathrm{MAP}}\}$ of minimizers has a subsequence that converges in BV topology as $n \to \infty$. (For fixed n and α_n, the MAP estimate is not necessarily unique.) However, the TV prior distributions (and CM estimates) diverge.

If we choose $\alpha_n = \widetilde{\alpha}\sqrt{n+1}$ with some $\widetilde{\alpha} > 0$, then $\lim_{n\to\infty} u_n^{\mathrm{MAP}} = 0$. Further, the posterior distributions converge to a distribution of a random variable v taking values in Y, and the CM estimates converge. However, v is Gaussian‡ and thus not edge-preserving, and condition (ii) above is not satisfied.

Thus the answer to the question in the title is negative. However, we view our results positively as a quest for researchers to design discretization-invariant prior distributions for edge-preserving inversion of (1.1).

This paper is organized as follows. In section 2 we give basic definitions and show how regularization theory and Bayesian inversion are related. In section 3 we define discretization invariance. In section 4 we prove a result about convergence of MAP estimates. In section 5 we prove results about convergence of prior distributions and CM estimates. In section 6 we illustrate our theoretical findings by numerical computations.

In the sequel, we use the abbreviations i.d., i.p., a.e., and a.s. for the terms *in distribution, in probability, almost every,* and *almost surely*, respectively.

2. The generic posterior distribution

We restrict ourselves in this work to a class of one-dimensional inverse problems. Our choice of spaces Y and Y_n is as follows.

Definition 2.1 (Function spaces Y and Y_n) *Let Y be the space of continuous functions on the interval $[0,1]$ vanishing at the endpoints:*

$$Y = C_0([0,1]) = \{u \in C([0,1]) \, : \, u(0) = u(1) = 0\}.$$

For any integer $n > 0$, let $Y_n \subset Y$ be the following set of piecewise linear functions on the interval $[0,1]$:

$$Y_n = \{u \in Y \, : \, u|_{[x_j^n, x_{j+1}^n]} \text{ is linear for } j = 0, \dots, n\},$$

where

$$x_j^n = \frac{j}{n+1} \qquad \text{for } j = 0, \dots, n+1.$$

Further, consider the roof-top basis $\{\psi_j^n\}_{j=1}^n$ for Y_n, where $\psi_j^n \in Y_n$ satisfy $\psi_j^n(x_k^n) = \delta_{jk}$ for $j = 1, \dots, n$ and $k = 0, \dots, n+1$.

We use the following class of probability distributions as priors.

Definition 2.2 (p-variation distribution) *Let (Ω, Σ, P) be a complete probability space. Let $n > 0$ be an integer, $\alpha_n > 0$, and $1 \le p \le 2$. The Y_n-valued random function*

$$u_n(t) = u_n(t, \omega) = \sum_{j=1}^n u_j^n(\omega) \psi_j^n(t), \quad \omega \in \Omega, \tag{2.1}$$

*has p-*variation distribution in Y_n with regularization parameter α_n *if the $\mathbb{R}^n$-valued random vector $[u_1^n(\omega), \dots, u_n^n(\omega)]^T$ has the probability density function*

$$\begin{aligned} \pi_{p,n}(u_1^n, \dots, u_n^n) &= c_{p,n} \, \exp\Big(-\alpha_n \| \frac{\partial}{\partial t} \sum_{j=1}^n u_j^n \psi_j^n(t) \|_{L^p(0,1)}^p \Big) \\ &= c_{p,n} \, \exp\Big(-\frac{\alpha_n}{(n+1)^{1-p}} \sum_{j=1}^{n+1} |u_j^n - u_{j-1}^n|^p \Big), \end{aligned} \tag{2.2}$$

‡ Numerical evidence and a conjecture was first presented by Markku Lehtinen [21].

where $u_0^n = u_{n+1}^n = 0$ and $c_{p,n}$ is a normalization constant. The special case $p = 1$ is called total variation (TV) probability distribution in Y_n.

Note that $p = 2$ gives a Gaussian distribution. In definition 2.2 we require that $u_0^n = u_{n+1}^n = 0$ for the p-variation distribution to be a probability density function. Without the requirement, a constant could be added to u_n without altering $\pi_{p,n}(u_1^n, \ldots, u_n^n)$.

We are ready to define the posterior distribution analyzed in this paper.

Definition 2.3 (The generic posterior distribution) *Denote by $\mathcal{Z}$ the set of Borel measures $A(dt)$ with finite variation that are supported on compact subsets of $(0,1)$. Let $A_j \in \mathcal{Z}$ for $j = 1, \ldots, N$ and $\sigma > 0$. Given $u \in Y$, let the measurement m be the random vector in $\mathbb{R}^N$ with components*

$$m_j = (Au)_j + \varepsilon_j = \langle A_j, u \rangle + \varepsilon_j = \int_0^1 v(t) A_j(dt) + \varepsilon_j, \tag{2.3}$$

where $\varepsilon_j \sim N(0, \sigma)$ are independent Gaussian errors. Assume given a realization $\hat{m}$ of m. Modelling a priori *knowledge about u with the p-variation distribution in the space Y_n leads to the posterior distribution*

$$\pi(u_n \,|\, \hat{m}) = \widetilde{c} \exp\Big(- \frac{1}{2\sigma^2} \|Au_n - \hat{m}\|_{\mathbb{R}^N}^2 - \alpha_n \|u_n'\|_p^p \Big), \tag{2.4}$$

where $\|\cdot\|_{\mathbb{R}^N}$ is the standard Euclidean norm and $\|u_n'\|_p^p$ is understood in the sense of (2.1) and (2.2).

Recall that $A(dt)$ has finite variation if there are finite non-negative measures $A_1(dt)$ and $A_2(dt)$ such that $A(dt) = A_1(dt) - A_2(dt)$. The norm of $\mathcal{Z}$ is given by

$$\|A\|_{\mathcal{Z}} = \inf\{A_1([0,1]) + A_2([0,1]) \, : \, A(dt) = A_1(dt) - A_2(dt)\}.$$

We close the section by pointing out a connection between regularization theory and Bayesian MAP estimates. Maximizing the posterior distribution (2.4) is equivalent to the minimization problem

$$\arg \min_{u_n \in Y_n} \Big\{ \|Au_n - m\|_{\mathbb{R}^N}^2 + \beta_n \int_0^1 |u_n'(t)|^p dt \Big\}. \tag{2.5}$$

with $\beta_n = 2\sigma^2 \alpha_n$. But (2.5) is by definition the Tikhonov ($p = 2$) or TV ($p = 1$) regularized solution of the inverse problem (1.2). We note that the edge-preserving property of discrete TV regularization is related to the regularization term in (2.5) allowing large values of derivatives.

3. Discretization invariance

It is tempting to consider the p-variation prior distribution (2.2) as discretization of the following formal prior distribution:

$$\pi_u(v) \overset{\text{formally}}{\sim} \exp(-\alpha \|v'\|_p^p), \quad v \in Y. \tag{3.1}$$

Generalizing the successful solution of finite-dimensional inverse problems in Y_n to solution of the continuous problem in Y using (3.1) is a natural idea. This can indeed be done for $p = 2$: see Lehtinen, Päivärinta and Somersalo [22] and Lasanen [20]. To analyze the case $1 \leq p < 2$ we need the following definitions.

Definition 3.1 (Linear discretization of random functions) *Let $u_{n(\ell)}(t, \omega)$ be $Y_{n(\ell)}$-valued random variables with $\ell = 1, 2, 3, \ldots$ and $1 < n(1) < n(2) < n(3) < \cdots$. Assume that* *(i) There is a Y-valued random variable $v(t) = v(t, \omega)$ such that for any $t \in [0, 1]$*

$$\lim_{\ell\to\infty} u_{n(\ell)}(t) = v(t) \qquad i.d.$$

(ii) There are bounded linear operators $T_{n(\ell)} \in L(Y)$ such that for any $t \in [0, 1]$

$$u_{n(\ell)}(t) = (T_{n(\ell)}v)(t).$$

Then $u_{n(\ell)}$ are linear discretizations of a random function. *Further, we say that v can be approximated by finite-dimensional random variables in a discretization invariant manner and $u_{n(\ell)}$ are* proper linear discretizations *of v.*

Note that Definition 3.1 is analogous to that of Lasanen [20] (see also [15, 23]).

Definition 3.2 (Discretization invariant choice of Y_n and $\pi_{pr}^{(n)}$) *Assume given $Y_{n(\ell)}$ and $\pi_{pr}^{(n(\ell))}$ for (1.3) with $\ell = 1, 2, 3, \dots$ and $1 < n(1) < n(2) < n(3) < \cdots$. Let $u_{n(\ell)}$ be random functions taking values in $Y_{n(\ell)}$ and having distribution $\pi_{pr}^{(n(\ell))}$. We say that the choice of $Y_{n(\ell)}$ and $\pi_{pr}^{(n(\ell))}$ is* discretization invariant *if $u_{n(\ell)}$ are linear discretizations of a random function in the sense of Definition 3.1.*

We will show that the discrete random variables distributed according to the generic posterior distribution (2.4) with $1 \le p < 2$ are not linear discretizations of any random function. See Remark 5.1 on page 11.

4. Convergence of MAP estimates

We analyze the convergence of MAP estimates

$$u_n^{\mathrm{MAP}}(t; p, \alpha_n) \in \arg\max_{u_n\in Y_n} \exp\Big(-\frac{1}{2\sigma^2}\|Au_n - \hat{m}\|^2_{\mathbb{R}^N} - \alpha_n\|u_n'\|_p^p\Big) \quad (4.1)$$

for the posterior distribution (2.4) as the discretization is refined arbitrarily: $n = n(\ell) = 2^\ell - 1$ and $\ell \to \infty$. This choice of $n(\ell)$ ensures that $Y_{n(\ell)} \subset Y_{n(\ell+1)}$, which is needed below in the case $p = 1$. In the case $1 < p \le 2$ the object function in (4.1) is strictly convex and there is a unique MAP estimate, whereas in the case $p = 1$ the MAP estimate is not necessarily unique.

We recall two important function spaces. First, the Sobolev space $W^{1,p}(0, 1)$ consists of $L^p(0, 1)$ functions with weak derivatives in $L^p(0, 1)$. By the Sobolev imbedding theorem we know that $W_0^{1,p}$ functions are continuous. Second, the space of functions of bounded variation is defined as follows. Let

$$BV(0, 1) = \{u \in L^1(\mathbb{R}) \ : \ \|u\|_{BV} < \infty, \ \mathrm{supp}\,(u) \subset [0, 1]\},$$
$$BV_0(0, 1) = \{u \in BV(0, 1) : \ u(0) = u(1) = 0\},$$

where

$$\|u\|_{BV} = \sup\{\int_0^1 u(s)\partial_s\phi(s)\, ds \ : \ \phi \in C^\infty([0, 1]), \ \|\phi\|_{L^\infty} \le 1\}.$$

Note that for $u \in BV_0(0, 1)$ the derivative $\partial_s u$, defined in sense of distributions, is a measure. We say that $u_n \in BV(0, 1)$ converge in weak–$*$ topology of BV to $u \in BV(0, 1)$ if $\|u_n - u\|_{L^1} \to 0$ and $\int_{\mathbb{R}} \phi(\partial_s u_n - \partial_s u)(s) \to 0$ for any $\phi \in C_0([0, 1])$, i.e., the measures $\partial_s u_n$ converge weakly to $\partial_s u$. In this case we denote $u_n \xrightarrow{\text{BV-w}^*} u$. Note that the trace $u \mapsto u(s_0)$, $0 \le s_0 \le 1$ is continuous from weak–$*$ topology of $BV(0, 1)$ to $\mathbb{R}$.

Theorem 4.1 *Let $\sigma > 0$, $\hat{m} \in \mathbb{R}^N$ and $n = n(\ell) = 2^\ell - 1$ for $\ell = 2, 3, \ldots$. Assume $A_j \in L^1(0,1) \cap \mathcal{Z}$ for $j = 1, \ldots, N$.*

(i) Let $\alpha_n = \widetilde{\alpha}/(2\sigma^2)$ with some $\widetilde{\alpha} > 0$ and $1 < p \le 2$. Let $u_n^{\mathrm{MAP}} = u_n^{\mathrm{MAP}}(t; p, \alpha_n)$ be the unique MAP estimate given by (4.1) for the posterior distribution (2.4). Then there is a unique limit function $\widetilde{u}(\,\cdot\,; p, \widetilde{\alpha}) \in W_0^{1,p}(0,1)$ such that

$$\lim_{\ell \to \infty} u_{n(\ell)}^{\mathrm{MAP}} = \widetilde{u}(\,\cdot\,; p, \widetilde{\alpha})$$

in the weak topology of $W_0^{1,p}(0,1)$.

(ii) Let $\alpha_n = \widetilde{\alpha}/(2\sigma^2)$ with some $\widetilde{\alpha} > 0$ and $p = 1$. Then for any sequence $\{u_{n(\ell)}^{\mathrm{MAP}}\}_{\ell=2}^\infty$ of maximizers of (4.1) there is a subsequence that converges in weak–$*$ topology of BV to some $\widetilde{u} \in BV_0(0,1)$.

(iii) Let the regularization parameters $\alpha_n > 0$ satisfy $\lim_{\ell \to \infty} \alpha_{n(\ell)} = \infty$. Then

$$\lim_{\ell \to \infty} u_{n(\ell)}^{\mathrm{MAP}} = 0$$

in the norm topology of $W_0^{1,p}(0,1)$ for $1 < p \le 2$ and BV for $p = 1$.

We note that the restriction $A_j \in L^1(0,1)$ is needed only in the case $p = 1$. The proof of Theorem 4.1 consists in part of standard arguments in the field of epiconvergence of minimization problems. However, we present the details for the reader's convenience.

Proof. We define two optimization problems that are limits of optimization problems in finite-dimensional spaces. In the case $1 < p \le 2$ we consider the problem

$$\begin{aligned} \widetilde{u}(t; p, \beta) &= \arg \min_{u \in W_0^{1,p}} S(u), \\ S(u) &= \|Au - \hat{m}\|_{\mathbb{R}^N}^2 + \beta \|u'\|_p^p, \end{aligned} \tag{4.2}$$

where $\beta > 0$ and $W_0^{1,p}$ is the closure of $C_0^\infty(0,1)$ in the $W^{1,p}(0,1)$ norm. In the case $p = 1$ we consider the problem

$$\begin{aligned} \widetilde{u}(t; p, \beta) &\in \arg \min_{u \in BV_0(0,1)} S(u), \\ S(u) &= \|Au - \hat{m}\|_{\mathbb{R}^N}^2 + \beta \|u\|_{BV}. \end{aligned} \tag{4.3}$$

Next we use methods of convex analysis and approximate (4.2, 4.3) with a discrete minimization problem

$$\widetilde{u}_n(t; p, \beta_n) \in \arg\min_u S_n(u), \tag{4.4}$$

where $S_n : W_0^{1,p} \to \overline{\mathbb{R}}_+ = \mathbb{R}_+ \cup \{0, \infty\}$ for $1 < p \le 2$ or $S_n : BV_0(0,1) \to \overline{\mathbb{R}}_+$ for $p = 1$ is the convex non-linear function

$$S_n(u) = \|Au - \hat{m}\|_{\mathbb{R}^N}^2 + \beta_n \|u'\|_p^p + I_{Y_n}(u).$$

Here $\beta_n > 0$ are such that $\beta = \lim_{n \to \infty} \beta_n$ and I_{Y_n} is the convex indicator function: $I_{Y_n}(u) = 0$ if $u \in Y_n$ and $I_{Y_n}(u) = \infty$ if $u \notin Y_n$.

The proof of the convergence of the MAP estimate for the posterior distribution (2.4) is based, using the terminology of [28], to the epiconvergence of S_n to S.

Note that (4.1) is equivalent to

$$u_n^{\mathrm{MAP}}(t; p, \alpha_n) = \arg \min_{u_n \in Y_n} \Big(\|Au_n - \hat{m}\|_{\mathbb{R}^N}^2 + 2\sigma^2 \alpha_n \|u_n'\|_p^p \Big), \tag{4.5}$$

which in turn is equivalent to (4.4) with $\beta_n = 2\sigma^2\alpha_n = \widetilde{\alpha}$.

Consider the operators $T_n : C_0(0,1) \to Y_n$ defined by

$$T_n u(t) = \sum_{j=1}^{n} u(x_j^n)\psi_j^n(t), \quad x_j^n = \frac{j}{n+1},$$

where the functions ψ_j^n are as in Definition 2.1. Thus $T_n u$ is obtained by linear interpolation from the point values $u(x_j^n)$.

Properties of T_n for $1 < p \le 2$**.** Denote $\Delta x^n := x_2^n - x_1^n = (n+1)^{-1}$. Now for $t \in I = I_j^n = [x_j^n, x_{j+1}^n]$ with $j = 0, 1, 2, \ldots, n$, we have

$$T_n u(t) = u(x_j^n) + \frac{t - x_j^n}{\Delta x^n} \int_{x_j^n}^{x_{j+1}^n} u'(t)dt,$$

and, further,

$$(T_n u)'(t) = \frac{1}{\Delta x^n} \int_{x_j^n}^{x_{j+1}^n} u'(t)dt = [u']_I,$$

where we denote by $[u']_I$ the average of u' over I. Note that $|(T_n u)'(t)| \le 2Mu'(t)$ where Mu' is the Hardy-Littlewood maximal function of u'. For any $t \in (0,1)$, let $j(t,n)$ be some index for which $t \in I_{j(t,n)}^n$. Since $u' \in L^1(0,1)$, we have (similarly to the standard theorem of Lebesgue points)

$$\lim_{n\to\infty} u'(t) - [u']_{I_{j(t,n)}^n} = 0$$

for almost every t, see [10]. Since $|(T_n u)'(t) - u'(t)| \le (2Mu'(t) + |u'(t)|) \in L^p(0,1)$, Lebesgue's theorem of dominated convergence yields that

$$\lim_{n\to\infty} \|(T_n u - u)'\|_{L^p(0,1)}^p = \int_0^1 \lim_{n\to\infty} |u'(t) - (u')_{I_{j(t,n)}^n}|^p dt = 0.$$

Hence $\lim_{n\to\infty} T_n = I$ in the strong operator topology of $W_0^{1,p}$. Moreover,

$$\|(T_n u)'\|_{L^p} \le 2\|Mu'\|_{L^p} \le C\|u'\|_{L^p}$$

and hence the norms $\|T_n\|_{W_0^{1,p} \to W_0^{1,p}}$ are uniformly bounded.

Epiconvergence in case $1 < p \le 2$**.** First we consider the case $\beta_n = \beta = \widetilde{\alpha}$. Let $u_n = u_{n(\ell)}, \ell = 2, 3, \ldots$, be a converging sequence in $W_0^{1,p}$ with $\lim_{n\to\infty} u_n = u$. We have shown that $\lim_{n\to\infty} T_n = I$ in the strong operator topology of $W_0^{1,p}$ and that the norms $\|T_n\|_{W_0^{1,p} \to W_0^{1,p}}$ are uniformly bounded. Thus $\lim_{n\to\infty} T_n u_n = u$. Further, since S_n has infinite values in the complement of Y_n and $T_n|_{Y_n} = I$, we see trivially that $S_n(u_n) \ge S_n(T_n u_n)$. Using these facts we estimate

$$\begin{aligned} &\liminf_{n\to\infty} S_n(u_n) \qquad (4.6)\\ &\ge \liminf_{n\to\infty} S_n(T_n u_n) \\ &\ge \liminf_{n\to\infty} \|A(T_n u_n - u) + (Au - \hat{m})\|_{\mathbb{R}^N}^2 + \beta\|(T_n u_n)'\|_p^p \\ &\ge S(u). \end{aligned}$$

Moreover, for $u \in W_0^{1,p}$ there is such a sequence $v_n = T_n u \to u$ in $W_0^{1,p}$ that

$$\begin{aligned} &\limsup_{n\to\infty} S_n(v_n) \qquad (4.7)\\ &= \limsup_{n\to\infty} \|A(T_n u_n - u) + (Au - \hat{m})\|_{\mathbb{R}^N}^2 + \beta_n\|(T_n u)'\|_p^p \\ &= S(u). \end{aligned}$$

By definition (see [3], Prop. 1.14 and also [2], [28]), formulae (4.6) and (4.7) mean that the functions S_n epiconverge to S. Define $s_0 = \inf\{S(v) : v \in W_0^{1,p}\}$ and $s_n = \inf\{S_n(v) : v \in W_0^{1,p}\}$ for $n = n(\ell), \ell = 2, 3, 4, \ldots$, and consider the set

$$\operatorname{argmin}(S_n) = \{u \in W_0^{1,p} : S_n(u) = s_n\}$$

where S_n attains its minimum (actually, this set contains only the function $\widetilde{u}_n(\beta_n)$). Since S_n epiconverge to S, [3], Proposition 2.9 yields that

$$\limsup_{n\to\infty} \operatorname{argmin}(S_n) \subset B_\varepsilon = \{u \in W_0^{1,p} : \ S(u) \leq s_0 + \varepsilon\} \tag{4.8}$$

for any $\varepsilon > 0$. Now, assume that $0 < \beta < \infty$. Since S is strictly convex and weakly lower semicontinuous, S has a unique global minimum at $\widetilde{u}$ (see e.g. [5, Thm. 2.1.4]). Thus (4.8) implies that $\lim_{n\to\infty} S(u^{\text{MAP}}_{n(\ell)}) = \min(S)$ and in particular $\|u^{\text{MAP}}_{n(\ell)}\|_{W_0^{1,p}}$ are uniformly bounded.

Assume now that $u^{\text{MAP}}_{n(\ell)}$ do not converge weakly to $\widetilde{u}$. Then there is $f \in (W_0^{1,p})'$ such that for some subsequence $\lim_{k\to\infty}\langle f, u^{\text{MAP}}_{n(\ell_k)}\rangle = c_f \neq \langle f, \widetilde{u}\rangle$. By the Banach-Alaoglu theorem $\{u^{\text{MAP}}_{n(\ell_k)}\}$ has a weakly converging subsequence. We denote the limit of such a subsequence by $\widetilde{u}_1$. Since $S : W_0^{1,p}(0,1) \to \mathbb{R}$ is lower semicontinuous in the weak topology of $W_0^{1,p}$, we see that $S(\widetilde{u}_1) = \min(S)$. Then $\langle \widetilde{u} - \widetilde{u}_1, f\rangle \neq 0$ is in contradiction with the fact that the minimum $\widetilde{u}$ is unique. This shows that $u^{\text{MAP}}_{n(\ell)}$ converges weakly in $W_0^{1,p}(0,1)$ to $\widetilde{u}$. This proves (i).

Approximation of functions in the case $p = 1$**.** First we consider approximations in $W_0^{1,1}(0,1) \subset BV_0(0,1)$. Let $h \in L^1(0,1)$ and $\mathcal{N}_\ell$ be the σ-algebra generated by intervals $(0, x_j^{n(\ell)})$, $j = 0, 1, \ldots, n(\ell)+1$. Then the σ-algebra $\mathcal{N}$ generated by $\cup_{\ell=2}^\infty \mathcal{N}_\ell$ is the standard Borel σ-algebra of $(0,1)$. Since $\mathcal{N}_{\ell+1} \subset \mathcal{N}_\ell$, it follows from Doob's second martingale theorem ([27], Cor. C.9, [11], Thm. 10.5.7) that

$$\lim_{\ell\to\infty} \|\mathbb{E}(h|\mathcal{N}_\ell) - h\|_{L^1(0,1)} = 0, \tag{4.9}$$

where $\mathbb{E}(h|\mathcal{N}_\ell)$ is conditional expectation with respect to the σ-algebra $\mathcal{N}_\ell$. Now, for $u \in W_0^{1,1}(0,1)$ the function $(T_n u)'$ is piecewise constant and

$$(T_{n(\ell)}u)' = \mathbb{E}(u'|\mathcal{N}_\ell). \tag{4.10}$$

Formulae (4.9) and (4.10) imply

$$\lim_{\ell\to\infty} \|T_{n(\ell)}f - f\|_{W_0^{1,1}(0,1)} = 0. \tag{4.11}$$

Moreover, we see that $\|T_n f\|_{W_0^{1,1}} = \|(T_n f)'\|_{L^1} \leq c\|f\|_{W_0^{1,1}}$.

Second, we apply the properties of operators T_n to approximate functions in $BV_0(0,1)$. By [36, Thm. 5.2.1] for any $u \in BV_0(0,1)$ and any $u_k \overset{\text{BV-w*}}{\longrightarrow} u$ we have

$$\|u\|_{BV} \leq \liminf_{k\to\infty} \|u_k\|_{BV}. \tag{4.12}$$

Let $u \in BV_0(0,1)$. By [8, Thm. 2.3], there are $\Phi_k \in W^{1,1}(0,1) \cap C^\infty((0,1))$ such that Φ_j converge to u in the weak$-*$ topology of $BV(0,1)$ and

$$\lim_{k\to\infty} \|\Phi_k\|_{BV} = \|u\|_{BV}. \tag{4.13}$$

Note that Φ_j are not assumed to vanish at boundary points. However, since the trace $\phi \mapsto (\phi(0), \phi(1))$ is weakly continuous map $BV(0,1) \to \mathbb{R}^2$ we see that

$\phi_j(s) = \Phi_j(s) - (1-s)\Phi_j(0) - s\Phi_j(1) \in W_0^{1,p}(0,1)$ converge to u in the weak–$*$ topology of $BV_0(0,1)$. Moreover, by (4.13)

$$\lim_{j\to\infty} \|\phi_j\|_{BV} = \|u\|_{BV}. \tag{4.14}$$

By (4.11), there are ℓ_j, $\ell_{j+1} > \ell_j$ such that for any j and $\ell \geq \ell_j$ we have

$$\|T_{n(\ell)}\phi_j - \phi_j\|_{W_0^{1,1}(0,1)} \leq \frac{1}{j}.$$

This implies that there the sequence $T_{n(\ell_j)}\phi_j$ converge to u in the weak–$*$ topology of BV and

$$\lim_{j\to\infty} \|T_{n(\ell_j)}\phi_j\|_{BV} = \|u\|_{BV}.$$

In following, we denote $u_{n_j} = T_{n_j}\phi_j$, $n_j = n(\ell_j)$.

Epiconvergence in the case $p = 1$**.** Since $A_k \in L^1$ and by [5, p. 41], the embedding $BV_0 \to L^\infty(0,1)$ from the weak–$*$ topology of BV to norm topology of L^∞ is continuous, we see that $\langle A_k, u_{n_j}\rangle \to \langle A_k, u\rangle$. Thus, for any $u \in BV_0(0,1)$ there are $u_{n_j} \in Y_{n_j}$ such that

$$S(u) = \limsup_{j\to\infty} S_{n_j}(u_{n_j}). \tag{4.15}$$

Next we note that $\lim_{k\to\infty}\langle A_j, u_k\rangle \to \langle A_j, u\rangle$. Combining this with (4.12) we see that for any $u \in BV_0(0,1)$ and any $u_k \overset{\text{BV-w*}}{\longrightarrow} u$ we have

$$S(u) \leq \liminf_{k\to\infty} S_k(u_k). \tag{4.16}$$

Again, (4.15,4.16) imply that S_n epiconverge to S. Thus by [3], Proposition 2.9,

$$\limsup_{n\to\infty} \operatorname{argmin}(S_n) \subset B_\varepsilon = \{u \in BV_0 \ : \ \ S(u) \leq s_0 + \varepsilon\} \tag{4.17}$$

Thus if $u_k = u_{n(k)}^{\text{MAP}} \in \operatorname{argmin}(S_{n(k)})$, we see that $\lim_{k\to\infty} S(u_k) = \inf(S)$. Moreover, $\|u_k\|_{BV}$ are uniformly bounded. The sequence u_k has a subsequence that converges weakly (see [5, p.41]. In more detail, for proof of L^1 convergence, see [36, Cor. 5.3.4] and for the fact that the measures converge weakly, use Riesz representation theory and Banach-Alaoglu theorem). If $\widetilde{u}$ is a limit of such a subsequence we have $S(\widetilde{u}) = \inf(S)$. By (4.16) for any converging subsequence the limit is a minimizer of S. This proves *ii*.

Finally, we consider *iii*. We see that if $\lim_{n\to\infty}\beta_n = \infty$ then $\lim_{n\to\infty}\|u_n^{\text{MAP}}\|_{W_0^{1,p}} = 0$ for $1 < p \leq 2$ and $\lim_{n\to\infty}\|u_n^{\text{MAP}}\|_{BV} = 0$ for $p = 1$ proving the assertion. **Q.E.D.**

5. Convergence of the CM estimate

We analyze the convergence of the posterior distribution (2.4) with regularization parameter α_n as the discretization is refined arbitrarily, or $n \to \infty$. In particular, we are interested in the convergence of the CM estimate

$$u_n^{\text{CM}}(t; p, \alpha_n) = \int_{\mathbb{R}^n} (\sum_{j=1}^n s_j \psi_j^n(t))\pi(s_1, \ldots, s_n|\hat{m})\, ds_1 \ldots ds_n, \tag{5.1}$$

where $\pi(s_1, \ldots, s_n|\hat{m})$ is the conditional probability density function of coefficients of u_n in the basis $\{\psi_j^n\}$.

We introduce some definitions and notations.

Definition 5.1 (Convergence weakly i.d.) *Let v_n and v be $C(0,1)$-valued random variables. We say that v_n converges to v weakly in distribution if $\langle A, v_n\rangle \to \langle A, v\rangle$ i.d. for all $A \in \mathcal{Z}$ when $n \to \infty$.*

Definition 5.2 (Measurement σ-algebras $\mathcal{M}$, $\mathcal{M}_n$) *Denote $m_j = \langle A_j, v\rangle + \varepsilon_j$ and $m_j^n = \langle A_j, v_n\rangle + \varepsilon_j$ with some $A_j \in \mathcal{Z}$ for $j = 1, \dots, N$ and independent errors $\varepsilon_j \sim N(0,1)$. Let the σ-algebras $\mathcal{M}, \mathcal{M}_n \subset \Sigma$ be generated by the sets*

$$\{\omega \in \Omega : \ m_j(\omega) < \lambda_j\}, \quad \lambda_j \in \mathbb{R},$$
$$\{\omega \in \Omega : \ m_j^n(\omega) < \lambda_j\}, \quad \lambda_j \in \mathbb{R},$$

respectively.

Note that the above noise processes ε_j are the same for m_j and m_j^n, and for simplicity we take $\sigma = 1$ in this section.

We denote the conditional expectation of $v(t)$ with respect to σ-algebra $\mathcal{M}$ by $\mathbb{E}(v(t)|\mathcal{M})$. Recall that $\mathbb{E}(v(t)|\mathcal{M}) = \mathbb{E}(v(t)|\mathcal{M})(\omega)$ is the random variable that is measurable with respect to $\mathcal{M}$ and for which

$$\int_S \mathbb{E}(v(t)|\mathcal{M})\, P(d\omega) = \int_S v(t)\, P(d\omega) \quad \text{for all } S \in \mathcal{M}.$$

Since $\mathbb{E}(v(t)|\mathcal{M})$ is a $\mathcal{M}$-measurable random variable, there exists a deterministic function $\hat{m} \mapsto \widetilde{E}(v(t)|\hat{m})$ so that

$$\mathbb{E}(v(t)|\mathcal{M}) = \widetilde{E}(v(t)|m(\omega)) \quad \text{a.s.}$$

(see [11], Thm. 4.2.8). We call $\widetilde{E}(v(t)|\hat{m})$ the conditional mean with measurement $\hat{m}$ and occasionally denote it by $\widetilde{E}(v(t)|m = \hat{m})$. Let $B(\hat{m}, r) \subset \mathbb{R}^N$ be a ball with center $\hat{m}$ and radius r. We can write

$$\widetilde{E}(v(t)|\hat{m}) = \lim_{r \to 0} \frac{1}{P(\{m(\omega) \in B(\hat{m}, r)\})} \int_{\{m(\omega) \in B(\hat{m}, r)\}} v(t)\, P(d\omega) \tag{5.2}$$

for a.e. $\hat{m}$. If a conditional probability density function $\pi_{v(t)}(\cdot\,|\hat{m})$ exists and ε is normally distributed, the Radon-Nikodym derivative of the density of m with respect to Lebesgue measure is C^∞ smooth and we can write

$$\widetilde{E}(v(t)|\hat{m}) = \int_{\mathbb{R}} s\pi_{v(t)}(s|\hat{m})\, ds.$$

We define the conditional expectation $\widetilde{E}(v_n(t)|\hat{m})$ similarly.

Theorem 5.1 *Let v_n and v be random variables taking values in $C(0,1)$. Assume that $v_n \to v$ weakly i.d. when $n \to \infty$ such that for a given $t \in (0,1)$ variables $v_n(t)$, $n = 1, 2, \dots$ are uniformly integrable and $v(t,\omega), v_n(t,\omega) \in L^1(\Omega)$. Then*

$$\lim_{n\to\infty} \widetilde{E}(v_n(t)|\hat{m}) = \widetilde{E}(v(t)|\hat{m}) \ \textit{for a.e. } \hat{m} \in \mathbb{R}^N, \tag{5.3}$$
$$\lim_{n\to\infty} \widetilde{E}(F(v_n(t))|\hat{m}) = \widetilde{E}(F(v(t))|\hat{m}) \ \textit{for a.e. } \hat{m} \in \mathbb{R}^N \tag{5.4}$$

where $F = \chi_{(-\infty,\lambda]}$, $\lambda \in \mathbb{R}$.

Theorem 5.1 is proven later. Formula (5.4) means that the posterior distributions converge in law.

In particular, we consider the case where $v_n = u_n$ are p-variation random variables taking values in Y_n and $v = u_B$ is the Brownian bridge defined below.

Definition 5.3 (Brownian bridge u_B) *Define u_B as stochastic process $u_B(t) = u_B(t,\omega)$, $t \in [0,1]$, $\omega \in \Omega$, having zero expectation and covariance function*

$$\mathbb{E}(u_B(t_1)u_B(t_2)) = \sigma_p^2 |t_1| \cdot |1 - t_2|, \tag{5.5}$$

where $t_1, t_2 \in [0,1]$, $t_1 \le t_2$ and

$$\sigma_p^2 = \int_{\mathbb{R}} x^2 e^{-|x|^p} dx. \tag{5.6}$$

By Kolmogorov's theorem and (5.5) we can choose such a version of u_B that its realizations $t \mapsto u_B(t,\omega)$ are continuous a.s. Here we say that a random variable $a(t,\omega)$ is version of $b(t,\omega)$ if the distributions of $(a(t_1),\dots,a(t_\ell))$ and $(b(t_1),\dots,b(t_\ell))$ in $\mathbb{R}^\ell$ coincide for any $t_1,\dots,t_\ell \in [0,1]$ and $\ell > 0$.

Theorem 5.2 *Let $1 \le p \le 2$ and $\widetilde{\alpha} > 0$. Let u_n, $n = 1,2,\dots$ be p-variation random functions in Y_n with regularization parameter*

$$\alpha_n = \widetilde{\alpha}(n+1)^{1-\frac{p}{2}}.$$

Then for any $t \in [0,1]$ the prior distributions converge, i.e.,

$$\lim_{n\to\infty} u_n(t) = u_B(t) \quad i.d. \tag{5.7}$$

Moreover, the posterior distributions converge in distribution:

$$\lim_{n\to\infty} \widetilde{E}(F(u_n(t))|\hat{m}) = \widetilde{E}(F(u_B(t))|\hat{m}) \qquad \text{for a.e. } \hat{m} \in \mathbb{R}^N,$$

where $F = \chi_{(-\infty,\lambda]}$, $\lambda \in \mathbb{R}$ and the CM-estimates converge:

$$\lim_{n\to\infty} \widetilde{E}(u_n(t)|\hat{m}) = \widetilde{E}(u_B(t)|\hat{m}) \qquad \text{for a.e. } \hat{m} \in \mathbb{R}^N.$$

Further, consider the regularization parameter $\alpha_n = \widetilde{\alpha}(n+1)^q$. If $q > 1 - \frac{p}{2}$ then $\lim_{n\to\infty} u_n(t) = 0$. If $q < 1 - \frac{p}{2}$ then the random variables $u_n(t)$ do not converge even in distribution.

Theorem 5.2 is proven later.

Remark 5.1 *In (5.7) the limit $u_B(t)$ is independent of p (up to the scaling factor σ_p). Thus the choice $p = 1$ represents in the limit $n \to \infty$ the same* a priori *knowledge than the choice $p = 2$ (Gaussian smoothness prior). In particular, this implies that the TV prior distribution is not discretization invariant: since TV prior distributions converge to a Gaussian distribution and any linear discretizations of Gaussian distributions are also Gaussian, we see that TV priors are not discretizations of any random variable in the sense of Definition 3.1.*

5.1. Convergence of TV prior distributions

Here we consider random variables u_n having p-variation distribution in Y_n and show that they converge to the Brownian bridge. This is needed to show that Theorem 5.1 implies Theorem 5.2.

We note that such results are well known in statistical mechanics—indeed, a non-harmonic random field in a one-dimensional lattice (such as u_n) is generally known to converge to a free Gaussian field. For this type of results, see [7, 26]. However, compared to such work, we assume less regularity of the probability density functions and consider the integrals of these variables. For these reasons we think it is appropriate to present a full proof.

Theorem 5.3 *Let $1 \le p \le 2$ and $\widetilde\alpha > 0$. Let u_n be a random variable taking values in Y_n and having p-variation distribution with regularization parameter $\alpha_n = \widetilde\alpha(n+1)^{1-\frac{p}{2}}$. Moreover, let $0 < t_1 < t_2 < \ldots < t_Q < 1$. Then*

(i) We have

$$\lim_{n\to\infty}(u_n(t_1),\ldots,u_n(t_Q)) = (U_{t_1},\ldots,U_{t_Q}) \qquad i.d.$$

Here U_t are Gaussian random variables having Gaussian joint distributions with zero expectation and covariances

$$\mathbb{E}(U_{t_j}U_{t_k}) = \sigma_p^2|t_j|\cdot|1-t_k| \qquad \text{for } t_j \le t_k,$$

where σ_p is given by (5.6).

(ii) The variables $u_n(t_j)$, $j = 1,\ldots,Q$, $n = 1,2,\ldots,$ are uniformly integrable.

(iii) Let $0 < s' < t_1 < \ldots < t_Q \le s'' < 1$. Let $A_j \in \mathcal{Z}$, $j = 1,2,\ldots,N$, be measures supported on $[s',s''] \subset (0,1)$ and denote

$$a_n^j(\omega) = \int_{s'}^{s''} u_n(t,\omega)A_j(dt), \quad a^j(\omega) = \int_{s'}^{s''} u_B(t,\omega)A_j(dt).$$

Then a_n^j converges to a^j i.d. when $n \to \infty$ for $j = 1,\ldots,N$. Also, joint distributions of a_n^j and $u_n(t_k)$ converge i.d. to a^j and $u_B(t_k)$, respectively, for $j = 1,\ldots,N$ and $k = 1,\ldots,Q$. Here u_B is the Brownian bridge of Definition 5.3.

Proof. Let $(h_1^n, h_2^n, \ldots, h_{n+1}^n)$ be a random vector in $\mathbb{R}^{n+1}$ with density

$$\widetilde\pi_n(y_1,y_2,\ldots,y_{n+1}) = c\exp\big\{-\widetilde\alpha(n+1)^{p/2}\big(|y_1|^p + \sum_{j=1}^n |y_{j+1}-y_j|^p\big)\big\}. \qquad (5.8)$$

Comparison of (5.8) and (2.2) shows that $\widetilde\pi_n(y_1,\ldots,y_n,0) = c_n\pi_{p,n}(y_1,\ldots,y_n)$. Define a piecewise linear function

$$h_n(t) = \sum_{j=1}^{n+1} h_j^n\psi_j^n(t), \qquad t \in [0,1], \qquad (5.9)$$

where $\psi_j^n \in Y^n$ are as in Definition 2.1 for $j = 1,\ldots,n$, and ψ_{n+1}^n is the piecewise linear function satisfying $\psi_{n+1}^n(1) = 1$ and $\psi_{n+1}^n(t) \equiv 0$ for $0 \le t \le n/(n+1)$.

Consider the convergence of a single variable $u_n(t)$ for fixed $0 < t < 1$. Let

$$\xi_j^n = \widetilde\alpha^{1/p}(n+1)^{1/2}(h_j^n - h_{j-1}^n), \qquad (5.10)$$

where $h_0^n = 0$ and $j = 1,\ldots,n+1$. Then

$$h_j^n = \frac{1}{\widetilde\alpha^{1/p}(n+1)^{1/2}}\sum_{\ell=1}^{j}\xi_\ell^n \qquad \text{for } j = 1,\ldots,n+1. \qquad (5.11)$$

Now ξ_j^n are identically distributed independent variables with

$$\xi_j^n \sim \pi_\xi(t) := c_p\exp(-|t|^p), \qquad j = 1,\ldots n+1. \qquad (5.12)$$

Let $\xi_\ell \sim \pi_\xi$ be independent variables for $\ell = 1,2,3,\ldots$ and define

$$S_j = \frac{1}{\sqrt{j}}\sum_{\ell=1}^{j}\xi_\ell \quad \text{for } j = 1,2,3,\ldots.$$

Note that S_j does not depend on n. Then h_n can be represented as

$$h_n(t) = k_{n,t} S_{\theta(n,t)} + r_{n,t}\xi^n_{\theta(n,t)+1}, \tag{5.13}$$

where

$$k_{n,t} = \frac{\theta(n,t)^{1/2}}{\widetilde{\alpha}^{1/p}(n+1)^{1/2}}, \qquad r_{n,t} = \frac{(n+1)t - \theta(n,t)}{\widetilde{\alpha}^{1/p}(n+1)^{1/2}},$$

and $\theta(n,t)$ is the largest integer j such that $\frac{j}{n+1} \leq t$.

For clarity, we (somewhat non-standardly) denote the probability density function of $h_n(t)$ at $a \in \mathbb{R}$ with the condition $g = 0$ by

$$\pi(h_n(t) = a \,|\, g = 0) := \pi_{h_n(t)}(a \,|\, g = 0).$$

The Lipschitz continuity and positivity of $u_n(t)$ and $h_n(t)$ justify the use of Bayes' formula for probability density functions, and we get

$$\pi(h_n(t) = a \,|\, h_n(1) = 0) = \frac{\pi(h_n(t) = a)\ \pi(h_n(1) = 0 \,|\, h_n(t) = a)}{\pi(h_n(1) = 0)}. \tag{5.14}$$

Since $h_n(1) - h_n(t)$ has the same distribution as $h_n(1-t)$, we see by (5.14) that

$$\pi(h_n(t) = a \,|\, h_n(1) = 0) = c_n\, \pi(h_n(t) = a)\ \pi(h_n(1-t) = -a). \tag{5.15}$$

We know that $h_n(t)$ converges i.d. to a Gaussian random variable when $n \to \infty$. Namely, we see from (5.12) and the central limit theorem that $\lim_{n\to\infty} h_n(t) = h(t)$, where $h(t)$ is Brownian motion with $\mathbb{E}\, h(t) = 0$, $\mathbb{E}(h(t) - h(s))^2 = |t-s|\sigma_p^2$ and $t \geq s$. This is not quite enough for our purposes and we need to modify the proof of the central limit theorem.

Denote the characteristic function of ξ_1 by $\varphi(s) = \mathbb{E}e^{is\xi_1}$. Fourier transforming $\exp(-|t|^p)$ shows that $\varphi \in C^\infty(\mathbb{R})$. Since $\varphi''(0) = -\sigma_p^2$ by (5.6), there are such $\sigma_0 < \sigma_p$ and $\varepsilon > 0$ that

$$|\varphi(s)| \leq \exp(-\sigma_0^2 s^2/2) \qquad \text{for} \quad |s| < \varepsilon.$$

We write $\varphi = \varphi_1 + \varphi_2$ with $\varphi_1(s) = 0$ for $|s| \geq \varepsilon$ and $\varphi_2(s) = 0$ for $|s| < \varepsilon$. Further, we denote $a = \sup|\varphi_2(s)| < 1$.

It is well known that the characteristic function of the random variable S_j is $\Psi_j(s) = (\varphi(s/\sqrt{j}))^j$. By the central limit theorem,

$$\lim_{j\to\infty} \Psi_j(s) = \exp(-\sigma_p^2\,|s|^2/2) \quad \text{for } s \in \mathbb{R}. \tag{5.16}$$

As the supports of φ_1 and φ_2 are disjoint, we see that

$$\Psi_j(s) = \Psi_j^1(s) + \Psi_j^2(s) = (\varphi_1(\frac{s}{\sqrt{j}}))^j + (\varphi_2(\frac{s}{\sqrt{j}}))^j. \tag{5.17}$$

Now we see that for any $q \geq 1$

$$\|\Psi_j^2\|^q_{L^q} \leq a^{q(j-1)} \int_{\mathbb{R}} |\varphi_2(\frac{s}{\sqrt{j}})|^q ds \leq c a^{q(j-1)} j^{1/2} \to 0. \tag{5.18}$$

Moreover, we have the following estimate for Ψ_j^1:

$$|\Psi_j^1(s)| \leq (\exp(-\sigma_0^2(\frac{s}{\sqrt{j}})^2/2))^j = \exp(-\sigma_0 s^2/2). \tag{5.19}$$

Since $\exp(-\sigma_0 s^2/2) \in L^q(\mathbb{R})$ for any $q \geq 1$, we see by (5.16)–(5.19) that

$$\lim_{j\to\infty} \|\Psi_j - \exp(-\sigma_p^2\,|\cdot|^2/2)\|_{L^q(\mathbb{R})} = 0, \qquad 1 \leq q < \infty. \tag{5.20}$$

By (5.13), the characteristic function of $h_n(t)$ is

$$V_n(s) = \Psi_{\theta(n,t)}(k_{n,t}s)\cdot \varphi(r_{n,t}s). \tag{5.21}$$

Note that

$$\lim_{n\to\infty} \|\varphi(r_{n,t}s) - 1\|_{C^1(-L,L)} = 0 \quad \text{for any } L, \quad |\varphi(r_{n,t}s)| \le 1. \tag{5.22}$$

The characteristic function of $h_n(1-t)$, denoted by $G_n(s)$, has a similar expression. Then by (5.15) the characteristic function of $u_n(t)$ has the form

$$\Phi_n(s) = c_n\,(V_n * G_n)(s).$$

By (5.20)-(5.22) we see that $(V_n * G_n)(s)$ converges for any $s \in \mathbb{R}$:

$$\begin{aligned}
&\lim_{n\to\infty} (V_n * G_n)(s) = (V * G)(s),\\
&V(s) \qquad = \lim_{n\to\infty} V_n(s),\\
&G(s) \qquad = \lim_{n\to\infty} G_n(s).
\end{aligned}$$

Since $V(s)$ and $G(s)$ are Gaussian functions, the condition $\Phi_n(0) = 1$ implies that the normalization constants $c_n = \pi(h_n(1) = 0)^{-1}$ converge to a positive constant when $n \to \infty$. Thus, using (5.20) and (5.21) we see that

$$\begin{aligned}
&\lim_{n\to\infty} \Phi_n(s) = \Phi(s;t) := \exp(-\sigma(t)s^2/2), \qquad (5.23)\\
&\frac{1}{\sigma(t)^2} \qquad = \frac{1}{t\sigma_p^2} + \frac{1}{(1-t)\sigma_p^2}.
\end{aligned}$$

Since the limit (5.23) exists at every s and the limit function $\Phi(s;t)$ is continuous at $s = 0$, it follows from Levy's continuity theorem [11, Thm 9.8.2] that there is such a random variable U_t that

$$\lim_{n\to\infty} \mathbb{E}(h_n(t)\,|\,h_n(1) = 0) = U_t \qquad \text{i.d.}$$

and that the characteristic function of U_t is $\Phi(s;t)$. Claim 1 is proved for $Q = 1$.

To prove claim 2 of the theorem we consider L^2-bounds. We see that

$$\mathbb{E}(|u_n|^2) = \int_{\mathbb{R}} s^2 \pi(u_n = s)ds = -\partial_s^2 \Phi_n(s)|_{s=0} = -c_n(\partial_s V_n) * (\partial_s G_n)(s)|_{s=0}.$$

In view of the definitions of V_n and G_n, let us consider $\partial_s \Psi_j$:

$$\partial_s \Psi_j(s) = \partial_s(\varphi(\frac{s}{\sqrt{j}}))^j = j\frac{1}{\sqrt{j}}(\partial_s\varphi)(\frac{s}{\sqrt{j}})\,(\varphi(\frac{s}{\sqrt{j}}))^{j-1}. \tag{5.24}$$

Fourier transforming $\exp(-|t|^p)$ shows that $\varphi \in C^\infty(\mathbb{R})$ and that

$$\begin{aligned}
&|\varphi(s)| \le c(1+|s|)^{-1-p}, \qquad s \in \mathbb{R},\\
&\varphi(s) \;= 1 - \frac{\sigma_p}{2}s^2 + O(s^3), \qquad s \text{ near } 0.
\end{aligned}$$

Thus $|s^{-1}\partial_s\varphi(s)| \le c'$ for $s \in \mathbb{R}$ and we get the estimate

$$|\partial_s \Psi_j(s)| = |\frac{(\partial_s\varphi)(s/\sqrt{j})}{s/\sqrt{j}}|\,|s(\varphi(\frac{s}{\sqrt{j}}))^{j-1}| \le c'|s(\varphi(\frac{s}{\sqrt{j}}))^{j-1}|.$$

Writing $\varphi = \varphi_1 + \varphi_2$ as before we see that for $1 \le q < \infty$

$$\|s(\varphi_2(\frac{s}{\sqrt{j}}))^{j-1}\|_{L^q}^q \le a^{q(j-3)} \int_{\mathbb{R}} s^q|\varphi_2(\frac{s}{\sqrt{j}})|^{2q}ds \le c\,a^{q(j-3)}j^{(1+q)/2} \to 0$$

as $j \to \infty$. Moreover, we see that

$$|s(\varphi_1(\frac{s}{\sqrt{j}}))^{j-1}| \le s(\exp(-\sigma_0^2(\frac{s}{\sqrt{j}})^2/2))^{j-1} \le s\exp(-\sigma_0^2 s^2/4),$$

which is an integrable bound. Thus using (5.21) and (5.24) we see that

$$|V_n(s)| + |\partial_s V_n(s)| \le C_2 s \exp(-\sigma_0^2 s^2/8), \quad |s| < n^{1/2}\varepsilon, \tag{5.25}$$
$$\lim_{n\to\infty} \int_{\mathbb{R}\setminus[-n^{1/2}\varepsilon, n^{1/2}\varepsilon]} (|V_n(s)|^q + |\partial_s V_n(s)|^q) ds = 0, \quad 1 \le |q| < \infty.$$

Thus we have on interval $[-n^{1/2}\varepsilon, n^{1/2}\varepsilon]$ an uniform integrable bound and in the complement of this interval we can estimate L^q-norms uniformly.

Using (5.25) we see for $V_n(s)$ and $G_n(s)$ that

$$|(\partial_s V_n) * (\partial_s G_n)(0)| \le \|\partial_s V_n\|_{L^2} \|\partial_s G_n\|_{L^2} \le C_3.$$

Thus $\mathbb{E}(|u_n(t)|^2) \le C_4$, and in particular, the family $u_n(t)$ is uniformly integrable.

It remains to prove claim 1 for $Q > 1$ and claim 3. We prove them together by considering the joint distribution of $u_n(t_k)$ and $\langle A_j, u_n\rangle$ for $j = 1, \ldots, N$. For this, we denote $A_{N+k} = \delta(t - t_k)$ for $k = 1, \ldots, Q$ implying that

$$u_n(t_k) = \langle A_{N+k}, u_n\rangle, \quad j = 1, \ldots, Q.$$

Denote

$$c_{n\ell}^j = \int_0^1 \psi_\ell^n(t) A_j(dt), \quad b_{nk}^j = \sum_{\ell=k}^{n+1} c_{n\ell}^j, \tag{5.26}$$

and write using (5.9), (5.11) and (5.26) and changing order of summation

$$\langle A_j, h_n\rangle = \sum_{\ell=1}^{n+1} h_\ell^n c_{n\ell}^j = \sum_{\ell=1}^{n+1} \Big(\frac{\widetilde{\alpha}^{-1/p}}{(n+1)^{1/2}} \sum_{k=1}^{\ell} \xi_k^n\Big) c_{n\ell}^j = \frac{\widetilde{\alpha}^{-1/p}}{(n+1)^{1/2}} \sum_{k=1}^{n+1} b_{nk}^j \xi_k^n, \tag{5.27}$$

for $j = 1, \ldots, N+Q$. Now $Y_{nk}(\omega) := \widetilde{\alpha}^{-1/p}(n+1)^{-1/2}(b_{nk}^j \xi_k^n(\omega))_{j=1}^{N+Q}$ are independent random vectors in $\mathbb{R}^{N+Q}$ and

$$(\langle A_j, h_n\rangle)_{j=1}^{N+Q} = \sum_{k=1}^{n+1} Y_{nk}. \tag{5.28}$$

We prove that (5.28) converges in distribution to a Gaussian variable when $n \to \infty$. By the Fabian-Hannan version of the Lindenberg central limit theorem [13] it is enough to show that

$$\lim_{n\to\infty} \sum_{k=1}^{n+1} \mathbb{E}(|\beta\cdot Y_{nk}|^2\, 1_{|\beta\cdot Y_{nk}|>\varepsilon}) = 0 \tag{5.29}$$

for arbitrary $\beta \in \mathbb{R}^{N+Q} \setminus 0$ and $\varepsilon > 0$. Here $1_{|\beta\cdot Y_{nk}|>\varepsilon}$ is the indicator function equal to 1 if $|\beta\cdot Y_{nk}| > \varepsilon$ and zero otherwise.

Let $\beta \in \mathbb{R}^{N+Q} \setminus 0$ and $\varepsilon > 0$. Set $\gamma_{nk} = \sum_{j=1}^{N+Q} \beta_j b_{nk}^j$ for $k = 1, \ldots, n+1$. Note that $|b_{nk}^j| \le \|A_j\|_{\mathrm{TV}}$ and $|\gamma_{nk}| \le C|\beta|$. Estimate

$$\sum_{k=1}^{n+1} \mathbb{E}(|\beta\cdot Y_{nk}|^2\, 1_{|\beta\cdot Y_{nk}|>\varepsilon}) \le \sum_{k=1}^{n+1} \mathbb{E}(|\frac{\widetilde{\alpha}^{-1/p}\gamma_{nk}\xi_k}{(n+1)^{1/2}}|^2\, 1_{|\gamma_{nk}\xi_k|>\varepsilon\, \widetilde{\alpha}^{1/p}(n+1)^{1/2}})$$

$$\leq \frac{c\,\widetilde{\alpha}^{-2/p}}{n+1}\sum_{k=1}^{n+1}\gamma_{nk}^2\int_{|\gamma_{nk}x|>\varepsilon\,\widetilde{\alpha}^{1/p}(n+1)^{1/2}} x^2 e^{-|x|^p/2}\,dx$$
$$\leq \frac{c\,\widetilde{\alpha}^{-2/p}}{n+1}\sum_{k=1}^{n+1}\gamma_{nk}^2\exp(-|\gamma_{nk}|^{-1}\varepsilon\,\widetilde{\alpha}^{1/p}(n+1)^{1/2}/3).$$

Since $|\gamma_{nk}|^{-1}\geq C^{-1}|\beta|^{-1}$ we see that sum in the limit (5.29) is defined and formula (5.29) is true. Hence $(\langle A_j,h_n\rangle)_{j=1}^{N+Q}$ converges in distribution to a Gaussian variable. The limit is $(\langle A_j,h\rangle)_{j=1}^{N+Q}$, where $h(t)$ is Brownian motion on $t\in[0,1]$, $h(0)=0$.

We have proved pointwise convergence; next we consider convergence in L^q. To simplify notation we replace s'' by the approximation

$$s_n=\frac{r_n}{n},\quad r_n=\theta(n,s'')+1.$$

The characteristic function of the random variable $(\xi_k^n, b_{nk}^1\xi_k^n,\ldots,b_{nk}^{N+Q}\xi_k^n)$ is

$$\mathbb{E}(\exp(i(\xi_k^n\zeta+\sum_{j=1}^{N+Q}b_{nk}^j\xi_k^n\eta^j)))=\varphi(\zeta+\sum_{j=1}^{N+Q}b_{nk}^j\eta^j).$$

Thus the characteristic function of $(h(s_n),\langle A_1,h_n\rangle,\ldots,\langle A_{N+Q},h_n\rangle)$ is

$$\Phi_n(\zeta,\eta^1,\ldots,\eta^{N+Q})=\mathbb{E}(\exp(i(h_n(s_n)\zeta+\sum_{j=1}^{N+Q}a_n^j\eta^j)))=\prod_{k=1}^{r_n}\varphi\Big(\frac{\zeta+\sum_{j=1}^{N+Q}b_{nk}^j\eta^j}{\widetilde{\alpha}^{1/p}(n+1)^{1/2}}\Big),$$

where we used (5.27) and $b_{nk}^j=0$ for $k>r_n$. We have shown that $(\langle A_j,h\rangle)_{j=1}^{N+Q})$ converges i.d. to a Gaussian variable. Thus there is such a Gaussian function $\Phi(\zeta,\eta)$ that

$$\lim_{n\to\infty}\Phi_n(\zeta,\eta)=\Phi(\zeta,\eta)\tag{5.30}$$

for any $\zeta\in\mathbb{R}$ and $\eta=(\eta^1,\ldots,\eta^{N+Q})\in\mathbb{R}^{N+Q}$.

Now, let ε,σ_2 and $a<1$ be such that

$$|\varphi(t+s)|\leq\exp(-\sigma_2|t+s|^2/2)\quad\text{for } |s|<\varepsilon,\ |t|\leq 2\varepsilon,$$
$$|\varphi(t+s)|\leq a\quad\text{for } |s|<\varepsilon,\ |t|>2\varepsilon.$$

Fix $\eta=(\eta^1,\ldots,\eta^{N+Q})$ and set $\beta=\sup_{n,k}|\sum_{j=1}^{N+Q}b_{nk}^j\eta^j|$. Take n so large that $|\beta|\,\widetilde{\alpha}^{-1/p}(n+1)^{-1/2}<\varepsilon$. Then

$$1_{|\zeta|<2\varepsilon\,\widetilde{\alpha}^{1/p}(n+1)^{1/2}}\left|\prod_{k=1}^{r_n}\varphi\Big(\frac{\zeta+\sum_{j=1}^{N+Q}b_{nk}^j\eta^j}{\widetilde{\alpha}^{1/p}(n+1)^{1/2}}\Big)\right|$$
$$\leq 1_{|\zeta|<2\varepsilon\,\widetilde{\alpha}^{1/p}(n+1)^{1/2}}\exp\Big(-\frac{\sigma_2}{2}\sum_{k=1}^{r_n}\Big|\frac{\zeta+\sum_{j=1}^{N+Q}b_{nk}^j\eta^j}{\widetilde{\alpha}^{1/p}(n+1)^{1/2}}\Big|^2\Big)\leq\exp\big(\frac{\sigma_2}{\widetilde{\alpha}^{1/p}}(\beta^2-\frac{1}{6}|\zeta|^2)\big)$$

and

$$\Big\|1_{|\zeta|>2\varepsilon\,\widetilde{\alpha}^{1/p}(n+1)^{1/2}}\prod_{k=1}^{r_n}\varphi\Big(\frac{\zeta+\sum_{j=1}^{N+Q}b_{nk}^j\eta^j}{\widetilde{\alpha}^{1/p}(n+1)^{1/2}}\Big)\Big\|_{L^q}^q$$
$$\leq a^{q(r_n-1)}\int_{\mathbb{R}}\Big|\varphi\Big(\frac{t+\sum_{j=1}^{N}b_{n1}^j\eta^j}{\widetilde{\alpha}^{1/p}(n+1)^{1/2}}\Big)\Big|^q dt$$
$$\leq c\,a^{q(r_n-1)}\widetilde{\alpha}^{1/p}(n+1)^{1/2}\to 0 \text{ as } n\to\infty.$$

This and (5.30) shows that for any fixed η

$$\lim_{n\to\infty} \|\Phi_n(\,\cdot\,,\eta)-\Phi(\,\cdot\,,\eta)\|_{L^q(\mathbb{R})}=0,\quad 1\le q<\infty. \tag{5.31}$$

Consider next $\vec{a}_n=(\langle A_j,u_n\rangle)_{j=1}^{N+Q}$ and $\vec{b}=(b_1,\dots,b_{N+Q})\in\mathbb{R}^{N+Q}$. Then

$$\begin{aligned}
R_n(f,b_1,\dots,b_{N+Q}) &= \pi(h_n(s_n)=f,\ \vec{a}_n=\vec{b}\,|\,h_n(1)=0)\\
&= \frac{1}{\pi(h_n(1)=0)}\pi(h_n(s_n)=f,\ \vec{a}_n=\vec{b},\ h_n(1)=0)\\
&= c_n\pi(h_n(1)=0\,|\,h_n(s_n)=f,\ \vec{a}_n=\vec{b})\,\pi(h_n(s_n)=f,\ \vec{a}_n=\vec{b},)\\
&= c_n\pi(h_n(1)=0\,|\,h_n(s_n)=f)\,\pi(h_n(s_n)=f,\ \vec{a}_n=\vec{b},).
\end{aligned}$$

Here in the last equality we have used the fact that h_j^n is Markov sequence for $j=1,\dots,n$ and A_j are supported on $[s',s'']$ with $s''<s_n$, and thus

$$\pi(h_n(1)=0\,|\,h_n(s_n)=f,\ \vec{a}_n=\vec{b})=\pi(h_n(1)=0\,|\,h_n(s_n)=f).$$

Let us now introduce an auxiliary variable d and a function

$$R_n^1(d,f,b_1,\dots,b_{N+Q})=c_n\pi(h_n(1)=0\,|\,h_n(s_n)=f-d)\,\pi(h_n(s_n)=f,\ \vec{a}_n=\vec{b})$$

for which $R_n(f,b_1,\dots,b_{N+Q})=R_n^1(0,f,b_1,\dots,b_{N+Q})$. Define functions $W^n:\mathbb{R}^{1+N+Q}\to\mathbb{R}$ and $K^n:\mathbb{R}\to\mathbb{R}$ by

$$\begin{aligned}
W^n(f,\vec{b}) &= \pi(h_n(s_n)=f,\ \vec{a}_n=\vec{b}),\\
K^n(f) &= \pi(h_n(1)=0\,|\,h_n(s_n)=f)=\pi(h_n(1-s_n)=-f),
\end{aligned}$$

and denote their Fourier transforms by $\hat{W}^n(\xi,\eta)$ and $\hat{K}^n(\xi)$. If ζ is the Fourier variable corresponding to d, the Fourier transform of $R_n^1(d,f,b_1,\dots,b_{N+Q})$ is by (5.32)

$$\hat{R}_n^1(\zeta,\xi,\eta^1,\dots,\eta^{N+Q})=c_n\hat{W}^n(\zeta,\eta^1,\dots,\eta^{N+Q})\,\hat{K}^n(\xi+\zeta).$$

Then the characteristic function of the variable $(u_n(s_n),\vec{a}_n)$ is

$$\Phi_n(\xi,\eta^1,\dots,\eta^{N+Q})=c_n\int_{\mathbb{R}}R_n^1(\zeta,\xi,\eta^1,\dots,\eta^{N+Q})d\zeta.$$

Since $|\hat{W}^n(\zeta,\eta^1,\dots,\eta^{N+Q})|\le 1$ and $\hat{K}^n$ converge to a Gaussian function in $L^1(\mathbb{R})$ by (5.16) and (5.31) we see that

$$\lim_{n\to\infty}\Phi_n(\xi,\eta^1,\dots,\eta^{N+Q})=(\lim_{n\to\infty}c_n)\int_{\mathbb{R}}(\lim_{n\to\infty}R_n^1(\zeta,\xi,\eta^1,\dots,\eta^{N+Q}))d\zeta.$$

Here, R_n^1 converges to a Gaussian function. Thus the characteristic function $\Phi_n(\xi,\eta)$ with fixed ξ and η converges to the Gaussian function $\Phi(\xi,\eta)$, which implies convergence of joint distributions i.d.. **Q.E.D.**

5.2. Convergence of posterior distributions

Here we prove Theorems 5.1 and 5.2.

Let v_n and v be random variables taking values in $C(0,1)$ and assume $v_n\to v$ weakly i.d. By applying Skorohod's representation theorem (see e.g. [1]), and enlarging the probability space Ω if necessary, we have

Proposition 5.1 *Let $t \in (0,1)$ be given. The random variables $v_n(t,\omega)$, $v(t,\omega)$, $a_j^n = \langle v_n, A_j\rangle$, and $a_j = \langle v, A_j\rangle$, $j = 1,\dots,N$ have such versions $\widetilde{v}_n(t,\omega)$, $\widetilde{v}(t,\omega)$, $\widetilde{a}_j$, and $\widetilde{a}_j^n$ that almost surely*

$$\lim_{n\to\infty} \widetilde{v}_n(t) = \widetilde{v}(t), \qquad \lim_{n\to\infty} \widetilde{a}_j^n = \widetilde{a}_j, \qquad j = 1,\dots,N.$$

Using the versions $\widetilde{a}_j^n$ and $\widetilde{a}_j$ of $\langle A_j, v_n\rangle$ and $\langle A_j, v\rangle$, respectively, we define random variables $\widetilde{m}_j^n = \widetilde{a}_j^n + \varepsilon_j$ and $\widetilde{m}_j = \widetilde{a}_j + \varepsilon_j$. Here the errors $\varepsilon_j \sim N(0,1)$ are the same independent random variables.

Proof.(Theorem 5.1) Let F be either the identity map $F(s) = s$ or $F(s) = \chi_{(-\infty,\lambda]}(s)$ with $\lambda \in \mathbb{R}$. Since the random variables $F(\widetilde{v}_n(t))$ are uniformly integrable, converge a.s. (and thus i.p.), and the limiting variable $F(\widetilde{v}(t))$ is in $L^1(\Omega)$, we have by Vitali's convergence theorem, [11, Thm 10.3.6] that

$$\lim_{n\to\infty} \|F(\widetilde{v}_n(t)) - F(\widetilde{v}(t))\|_{L^1(\Omega)} = 0.$$

Consider the random variables $z_n = (\widetilde{v}_n(t), \widetilde{a}_1^n, \dots, \widetilde{a}_N^n) = (z_n^0, z_n') \in \mathbb{R}^{N+1}$ and $z = (\widetilde{v}(t), \widetilde{a}_1, \dots, \widetilde{a}_N) = (z^0, z') \in \mathbb{R}^{N+1}$. Assume given a realization $\hat{m} \in \mathbb{R}^N$ of the measurement and denote $g(y,\hat{m}) = \pi_\varepsilon(y' - \hat{m})$, $y = (y^0, y') \in \mathbb{R}^{N+1}$.

For clarity, we start our computations with the case where the laws P_{z_n} and P_z of z_n and z are absolutely continuous with respect to the Lebesgue measure in $\mathbb{R}^{N+1}$ and there exist continuous probability density functions $\pi_{z_n}(y)$ and $\pi_z(y)$.

Note that $\widetilde{m}^n$ has a smooth positive probability density function in $\mathbb{R}^N$ given by

$$\pi(\widetilde{m}^n = \hat{m}) = \int_{\mathbb{R}^{N+1}} \pi_\varepsilon(y' - \hat{m})\pi_{z_n}(y^0, y')dy = \int_{\mathbb{R}^{N+1}} g(y', \hat{m})\pi_{z_n}(y^0, y')dy. \quad (5.32)$$

A similar formula holds for $\pi(\widetilde{m} = \hat{m})$ and thus we have

$$\begin{aligned} \pi(\widetilde{m}^n = \hat{m}) &= \mathbb{E}(g(z_n', \hat{m})), \\ \pi(\widetilde{m} = \hat{m}) &= \mathbb{E}(g(z', \hat{m})). \end{aligned}$$

Since g is a bounded continuous function and $z_n' \to z_n$ weakly, we see that

$$\lim_{n\to\infty} \pi(\widetilde{m}^n = \hat{m}) = \pi(\widetilde{m} = \hat{m}). \quad (5.33)$$

Moreover, $\pi_{\widetilde{m}^n}$ is a smooth function and we have by Bayes' formula

$$\begin{aligned} \pi(\widetilde{v}_n(t) = z^0 | \widetilde{m}^n = \hat{m}) &= \frac{1}{\pi(\widetilde{m}^n = \hat{m})} \int_{\mathbb{R}^N} \pi_{z_n,\widetilde{m}}(z^0, y', \hat{m})dy', \\ &= \frac{1}{\pi(\widetilde{m}^n = \hat{m})} \int_{\mathbb{R}^N} g(y', \hat{m})\pi_{z_n}(z^0, y')dy'. \end{aligned}$$

By (5.2), $\widetilde{E}(F(\widetilde{v}_n(t))|\widetilde{m}^n = \hat{m})$ is equal to

$$\lim_{r\to 0} \frac{\int_{B(\hat{m},r)} \int_{\mathbb{R}} F(y^0)\pi(\widetilde{v}_n(t) = y^0, \widetilde{m}^n = w)\, dy^0 dw}{\int_{B(\hat{m},r)} \pi(\widetilde{m}^n = w)\, dw} \quad (5.34)$$

and taking integral over $\mathbb{R}$ outside the limit and letting $r \to 0$ leads to

$$\begin{aligned} \widetilde{E}(F(\widetilde{v}_n(t))|\widetilde{m}^n = \hat{m}) &= \int_{\mathbb{R}} F(y^0) \frac{\pi(\widetilde{v}_n(t) = y^0, \widetilde{m}^n = \hat{m})}{\pi(\widetilde{m}^n = \hat{m})} dy^0 \\ &= \frac{1}{\pi(\widetilde{m}^n = \hat{m})} \int_{\mathbb{R}^{N+1}} F(y^0) g(y, \hat{m})\pi_{z_n}(y^0, y')\, dy. \quad (5.35) \end{aligned}$$

Since a similar formula holds for $u_B(t)$, we have proven

$$\widetilde{E}(F(\widetilde{v}_n(t))|\widetilde{m}_n = \hat{m}) = \frac{1}{\pi(\widetilde{m}_n = \hat{m})}\mathbb{E}(F(z_n^0)g(z_n', \hat{m})), \tag{5.36}$$

$$\widetilde{E}(F(\widetilde{v}(t))|\widetilde{m} = \hat{m}) \quad = \frac{1}{\pi(\widetilde{m} = \hat{m})}\mathbb{E}(F(z^0)g(z', \hat{m})).$$

In the general case, where the laws P_{z_n} and P_z of z_n and z are not absolutely continuous, we have to replace formula (5.32) by

$$\pi(\widetilde{m}^n = \hat{m}) = \int_{\mathbb{R}^{N+1}} g(y, \hat{m})\, P_{z_n}(dy). \tag{5.37}$$

Again, since g is smooth by (5.37), we see that $\pi_{\widetilde{m}^n}$ is smooth. Also, we can replace formula (5.34) by

$$\widetilde{E}(F(\widetilde{v}_n(t))|\widetilde{m}^n = \hat{m}) = \lim_{r\to 0} \frac{\int_{B(\hat{m},r)} \left(\int_{\mathbb{R}^N\times\mathbb{R}} F(y^0)\pi_\varepsilon(y'-w)\, P_{z_n}(dy)\right) dw}{\int_{B(\hat{m},r)} \pi(\widetilde{m}^n = w)\, dw}.$$

Since $\pi(\widetilde{m}^n = w)$ is smooth, there is $C > 0$ such that

$$\left| \frac{\int_{B(\hat{m},r)} \pi_\varepsilon(y'-w)dw}{\int_{B(\hat{m},r)} \pi(\widetilde{m}^n = w)\, dw} \right| \leq C \quad \text{for } y' \in \mathbb{R}^N.$$

Thus Fubini's theorem and Lebesgue's theorem of dominated convergence gives an analog of formula (5.35):

$$\widetilde{E}(F(\widetilde{v}_n(t))|\widetilde{m}^n = \hat{m}) = \frac{1}{\pi(\widetilde{m}^n = \hat{m})} \int_{\mathbb{R}^{N+1}} F(y^0)g(y', \hat{m})\, P_{z_n}(dy).$$

These formulae imply (5.33) and (5.36) also in the general case.

Let $H(y) = F(y^0)g(y', \hat{m})$. Since $F(z_n^0) \to F(z^0)$ in $L^1(\Omega)$ and $|g| \leq 1$, we have $\lim_{n\to\infty} H(z_n) = H(z)$ in $L^1(\Omega)$. This and (5.33) imply that

$$\lim_{n\to\infty} \widetilde{E}(F(\widetilde{v}_n(t))|\widetilde{m}^n = \hat{m}) = \widetilde{E}(F(\widetilde{v}(t))|\widetilde{m} = \hat{m}).$$

Since the distributions of $v_n(t)$ and $\widetilde{v}_n(t)$ as well as those of m^n and $\widetilde{m}^n$ coincide, we have by (5.2)

$$\widetilde{E}(F(v_n(t))|m^n = \hat{m}) = \widetilde{E}(F(\widetilde{v}_n(t))|\widetilde{m}^n = \hat{m}).$$

This proves the assertion. **Q.E.D.**

Proof. (Theorem 5.2) Let us first consider the case $\alpha_n = \widetilde{\alpha}(n+1)^{1-p/2}$. By Theorem 5.3, $u_n \to u_B$ weakly i.d., $u_n(t), t \in (0,1)$, are uniformly integrable and $u_B(t) \in L^1(\Omega)$. Then (5.3) and (5.4) follow from Theorem 5.1.

When $\alpha_n = \widetilde{\alpha}(n+1)^q$, we see that $w_n(t) = \widetilde{\alpha}(n+1)^{(q-1-p/2)/p}\widetilde{u}_n(t)$ converges to the Brownian bridge. Thus $\widetilde{u}_n(t)$ converges to zero in $L^1(\Omega)$ when $q < 1-p/2$. When $q > 1-p/2$, we see that $u_n(t)$ cannot converge even i.d. **Q.E.D.**

6. Computational results

6.1. The model problem

Charge coupled devices (CCD) are commonly used in digital cameras and medical X-ray imaging devices. CCDs typically consist of a two-dimensional array of pixels capable of measuring the amount of visible light illuminating the area of the pixel over a period of time. We give a rough model for the measurement of the intensity distribution of light on one row of CCD pixels.

We take the quantity u in (1.1) to be a real-valued function on the unit interval $[0,1]$. Given $N > 1$, we divide the subinterval $[\frac{1}{N+2}, \frac{N+1}{N+2}] \subset [0,1]$ into N pixels $[x_j^{N+1}, x_{j+1}^{N+1}]$ with $x_j^{N+1} = j/(N+2)$ for $j = 1,\ldots,N$. The measurement of the jth pixel is

$$m_j = \langle A_j, u\rangle + \varepsilon_j = \int_{x_j^{N+1}}^{x_{j+1}^{N+1}} u(t)dt + \varepsilon_j, \tag{6.1}$$

where $A_j = \chi_{(x_j^{N+1}, x_{j+1}^{N+1})}$ as shown in Figure 1, and ε_j are normally distributed random numbers with standard deviation $\sigma > 0$. The numbers ε_j model measurement errors resulting from quantum and electronic noise of the CCD.

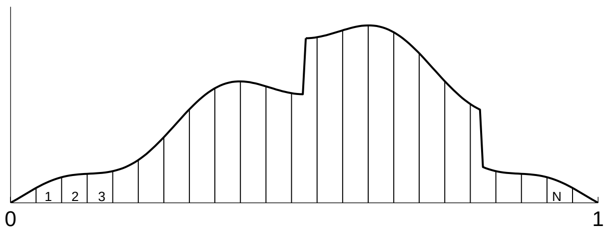

Figure 1. Idealized one-dimensional model for the measurement. The function represents a distribution of light on the interval $[0, 1]$. Pixels are represented by intervals and measurements are integrals of light intensity over those intervals. No measurement is made on the leftmost and rightmost interval.

We use the p-variation distribution in Y_n with $p = 1$ or $p = 2$ for representing *a priori* information on u. For convenience, we take the number of pixels to be of the form $N = 2^L - 2$ with $L > 1$, and the dimension $n = 2^\ell - 1$ is chosen to be greater than the number of measurements: $\ell > L$. In this case the grid $x_1^{N+1},\ldots,x_N^{N+1}$ is a subset of the grid $x_0^n,\ldots,x_{n+1}^n$.

6.2. Computational methods

6.2.1. Computation of MAP estimate with $p = 2$ Denote $U = [u_1^n,\ldots,u_n^n]^T$. Consider the minimization problem

$$\widetilde{U} = \arg\min_{U\in\mathbb{R}^n} \Big(\|\mathcal{A}U - \hat{m}\|_{\mathbb{R}^N}^2 + \frac{2\sigma^2\alpha_n}{\Delta x^n} \|\mathcal{D}U\|_{\mathbb{R}^{n+1}}^2 \Big), \tag{6.2}$$

where $\Delta x^n = (n+1)^{-1}$. The $N \times n$ matrix $\mathcal{A}$ implements the measurement:

$$(\mathcal{A}U)_k = \langle A_k, \sum_{j=1}^n u_j^n \psi_j^n \rangle, \qquad k = 1,2,\ldots,N, \tag{6.3}$$

where the roof-top basis functions ψ_j^n for the space Y_n are as in Definition 2.1. Integration of the piecewise linear functions in (6.3) over the intervals $[x_j^{N+1}, x_{j+1}^{N+1}]$ is implemented simply and exactly by the classical trapezoidal rule. Thus the jth row of $\mathcal{A}$ takes the form

$$[\underbrace{0,0,\ldots,0}_{j\,2^{\ell-L}-1}, \frac{1}{2}\Delta x^n, \underbrace{\Delta x^n, \Delta x^n, \ldots, \Delta x^n}_{2^{\ell-L}-1}, \frac{1}{2}\Delta x^n, 0, 0, \ldots, 0].$$

Prior information is coded into $\mathcal{D}$, the $(n+1)\times n$ matrix defined by

$$(\mathcal{D}U)_k = u_k^n - u_{k-1}^n, \quad k = 1,\dots,n+1, \quad u_0^n = 0 = u_{n+1}^n.$$

Following Varah [35] we write (6.2) in stacked form

$$\begin{bmatrix} \mathcal{A} \\ (\frac{2\sigma^2\alpha_n}{\Delta x^n})^{1/2}\mathcal{D} \end{bmatrix} U = \begin{bmatrix} \hat{m} \\ 0 \end{bmatrix} \tag{6.4}$$

and compute $\widetilde{U} = [\widetilde{u}_1^n,\dots,\widetilde{u}_n^n]^T$ as the least-squares solution of (6.4) using the Moore-Penrose pseudoinverse. In view of (2.2), (4.1), (4.5) and (6.2) we have

$$u_n^{\mathrm{MAP}}(t;2,\alpha_n) = \sum_{j=1}^{n} \widetilde{u}_j^n \psi_j^n(t).$$

6.2.2. Computation of MAP estimate with TV prior Consider the non-unique minimization problem

$$\widetilde{U} \in \arg\min_{U\in\mathbb{R}^n} \Big(\frac{1}{2\sigma^2}\|\mathcal{A}U - \hat{m}\|_{\mathbb{R}^N}^2 + \alpha_n \sum_{j=1}^{n+1} |(\mathcal{D}U)_j| \Big), \tag{6.5}$$

where the matrices $\mathcal{A}$ and $\mathcal{D}$ are as in Section 6.2.1. We write $\mathcal{D}U$ in the form

$$V_+, V_- \in \mathbb{R}_+^{n+1}, \qquad V_+ - V_- = \mathcal{D}U, \tag{6.6}$$

where $\mathbb{R}_+ := \{x\in\mathbb{R} : x \ge 0\}$. Now problem (6.5) is equivalent to

$$\widetilde{\mathcal{U}} = \arg\min \Big(\frac{1}{2\sigma^2}\|\mathcal{A}U\|_2^2 - \frac{1}{\sigma^2}\hat{m}^T\mathcal{A}U + \alpha_n\underline{\mathbf{1}}^T V_+ + \alpha_n\underline{\mathbf{1}}^T V_- \Big), \tag{6.7}$$

where $\widetilde{\mathcal{U}} = [\widetilde{U}^T\ \widetilde{V}_+^T\ \widetilde{V}_-^T]^T$ and $\underline{\mathbf{1}} = [1,1,...,1]^T \in \mathbb{R}^{n+1}$; the minimum is taken over $U\in\mathbb{R}^n, V_+, V_- \in \mathbb{R}^{n+1}$ satisfying (6.6). The solution of (6.7) with constraints (6.6) satisfies

$$(\widetilde{V}_+)_j = \max((\mathcal{D}\widetilde{U})_j, 0), \qquad (\widetilde{V}_-)_j = \max((-\mathcal{D}\widetilde{U})_j, 0).$$

Write now $\mathcal{U} = [U^T\ V_+^T\ V_-^T]^T$ and

$$H = \begin{bmatrix} \frac{1}{\sigma^2}\mathcal{A}^T\mathcal{A} & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}, \quad f = \begin{bmatrix} -\frac{1}{\sigma^2}\mathcal{A}^T\hat{m} \\ \alpha_n\underline{\mathbf{1}} \\ \alpha_n\underline{\mathbf{1}} \end{bmatrix}.$$

Now problem (6.5) is converted to a standard quadratic minimization problem

$$\widetilde{\mathcal{U}} = \arg\min\{\frac{1}{2}\mathcal{U}^T H\mathcal{U} + f^T\mathcal{U}\} \tag{6.8}$$

with linear constraints (6.6). We assume that we can find algorithmically an approximation to one of the possibly many solutions to (6.8) and (6.6). In view of (2.2), (4.1), (4.5) and (6.5) we define

$$u_n^{\mathrm{MAP}}(t;1,\alpha_n) = \sum_{j=1}^{n} \widetilde{u}_j^n \psi_j^n(t).$$

6.2.3. Computation of CM estimates Monte Carlo Markov Chain (MCMC) methods can be used to generate a collection $U^{(1)}, \dots, U^{(K)} \in \mathbb{R}^n$ of samples asymptotically distributed according to the posterior distribution

$$\widetilde{\pi}(U \,|\, \hat{m}) = \widetilde{c}_{n,p} \exp\Big(- \frac{1}{2\sigma^2} \|\mathcal{A}U - \hat{m}\|^2_{R^N} - \frac{\alpha_n}{(n+1)^{1-p}} \sum_{\nu=1}^{n+1} |(\mathcal{D}U)_\nu|^p \Big). \quad (6.9)$$

If K is large, we have

$$\int_{\mathbb{R}^n} U\, \widetilde{\pi}(U \,|\, \hat{m})\, dU \approx \frac{1}{K-k_0} \sum_{k=k_0+1}^{K} U^{(k)} =: \widetilde{U}, \quad (6.10)$$

where the first $k_0 > 0$ samples have been discarded because MCMC algorithms typically need such a *burn-in period* before the samples start to explore the posterior distribution representatively.

We use the Metropolis-Hastings (MH) algorithm [14, 16]. To implement the MH algorithm we define the *proposal distribution* $Q(V, \cdot)$ on $\mathbb{R}^n$, parameterized by $V \in \mathbb{R}^n$, as follows. Fix $1 \le N_{\text{update}} \le n$ and $\kappa > 0$. Pick randomly N_{update} distinct numbers from the set $1, 2, \dots, n$ according to uniform probability distribution. Order the numbers and denote them by $j_1, j_2, \dots, j_{N_{\text{update}}}$. Then a candidate vector $U \in \mathbb{R}^n$ is picked according to $Q(V, \cdot)$ if $U = V + \mathcal{E}_\kappa$, where

$$\mathcal{E}_\kappa = [0, \dots, 0, \varepsilon'_{j_1}, 0, \dots, 0, \varepsilon'_{j_2}, 0, \dots, 0, \varepsilon'_{N_{\text{update}}}, 0, \dots, 0]^T$$

with $\varepsilon'_{j_\ell} \sim N(0, \kappa)$ independent random numbers. Note that, if $\pi_Q(V, U)$ denotes the density of $Q(V, \cdot)$, the *transition probabilities* are symmetric: $\pi_Q(V, U) = \pi_Q(U, V)$.

Due to the above symmetry, the MH algorithm takes the simple form

1 Set $k := 0$ and initialize $U^{(0)}$ by e.g. $U^{(0)} := [0, \dots, 0]^T$.

2 Set $U := U^{(k)} + \mathcal{E}_\kappa$.

3 If $\widetilde{\pi}(U \,|\, \hat{m}) \ge \widetilde{\pi}(U^{(k)} \,|\, \hat{m})$ then set $U^{(k+1)} = U$ and go to 5.

4 Draw a random number s from the uniform distribution on $[0, 1]$. If $s \le \frac{\widetilde{\pi}(U \,|\, \hat{m})}{\widetilde{\pi}(U^{(k)} \,|\, \hat{m})}$ then set $U^{(k+1)} := U$; else set $U^{(k+1)} := U^{(k)}$.

5 If $k = K$ then stop; else set $k \leftarrow k + 1$ and go to 2.

We close this section by defining the *acceptance rate* of the Markov chain produced by the MH algorithm (discarding k_0 first samples):

$$r = \frac{\text{number of accepted candidates}}{K - k_0}. \quad (6.11)$$

6.3. Results

In our numerical examples we take u to be the step function satisfying $u(t) = 1$ for $t \in [1/3, 2/3]$ and $u(t) = 0$ otherwise. We consider a measurement with $N = 2^5 - 2 = 30$ pixels and random errors ε_j with standard deviation $\sigma = 0.001$. See Figure 2 for a plot of a realization of the measurement.

We perform all the computations with Matlab 6.5 running in a desktop PC computer equipped with a 2.8 GHz Intel Pentium 4 processor and 1 GB of RAM.

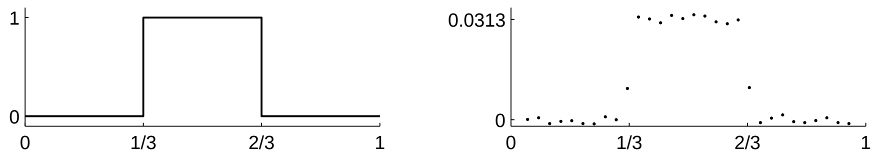

Figure 2. Left: Simulated intensity distribution $u(t)$. Right: Simulated noisy measurement $\hat{m}$. The dots are plotted at center points of pixels.

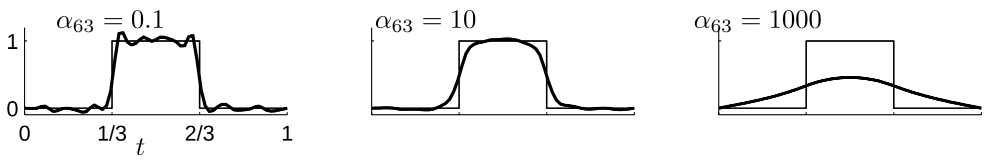

Figure 3. Gaussian MAP estimates with three different choices of α_{63}. The function $u(t)$ is plotted with a thin line. Left: too small α_{63} fails to regularize the solution. Middle: satisfactory regularized solution. Right: too large regularization parameter.

6.3.1. The Gaussian case We start by determining a suitable regularization parameter α_{63} for MAP estimates with 2-variation prior and fixed discretization level $\ell = 6, n = 63$. See Figure 3 for the least-squares solutions of (6.4) with various regularization parameters. Based on visual inspection, we choose $\alpha_{63} = 10$.

We turn to computing MAP estimates with varying levels of discretization. We solve (6.4) with n =63, 127, 255, 511, 1023, 2047, 4095, with $\alpha_n = 10$ for all n. The solutions agree with good precision at the coarsest discretization level:

$$\max_{j=1,\ldots,63}\{|u_{63}^{\mathrm{MAP}}(x_j^{63}; 2, \widetilde{\alpha}_2) - u_n^{\mathrm{MAP}}(x_j^{63}; 2, \widetilde{\alpha}_2)|\} \leq 0.003. \tag{6.12}$$

The computation takes less than a second for $n < 512$ and 295 seconds for $n = 4095$.

6.3.2. MAP estimates for TV prior We determine a suitable regularization parameter α_{63} for MAP estimates with TV prior at fixed discretization level $\ell = 6, n = 63$ by numerical experimentation. As the result we choose $\alpha_{63} = 135$.

We compare two ways of choosing the regularization parameter as function of n, both satisfying $\alpha_{63} = 135$:

$$\text{(i) } \alpha_n = 135, \qquad \text{(ii) } \alpha_n = 16.875\sqrt{n+1}. \tag{6.13}$$

We use MOSEK Optimization Toolbox's `quadprog` routine (available from `www.mosek.com`) to solve (6.8) with the constraints (6.6) in dimensions $n = 255, 1023, 4095$. Each computation takes less than 60 seconds of CPU time. See the left column in Figure 4 for plots of the MAP estimates with choices (i) and (ii).

We note that small changes in parameter values changed the computation considerably, sometimes even resulting in divergence of the algorithm. We presume this is due to the non-uniqueness of the solution of the optimization problem. However, the presented results did not exhibit these problems and we believe them to be approximations to some functions in the set of solutions to the optimization problem.

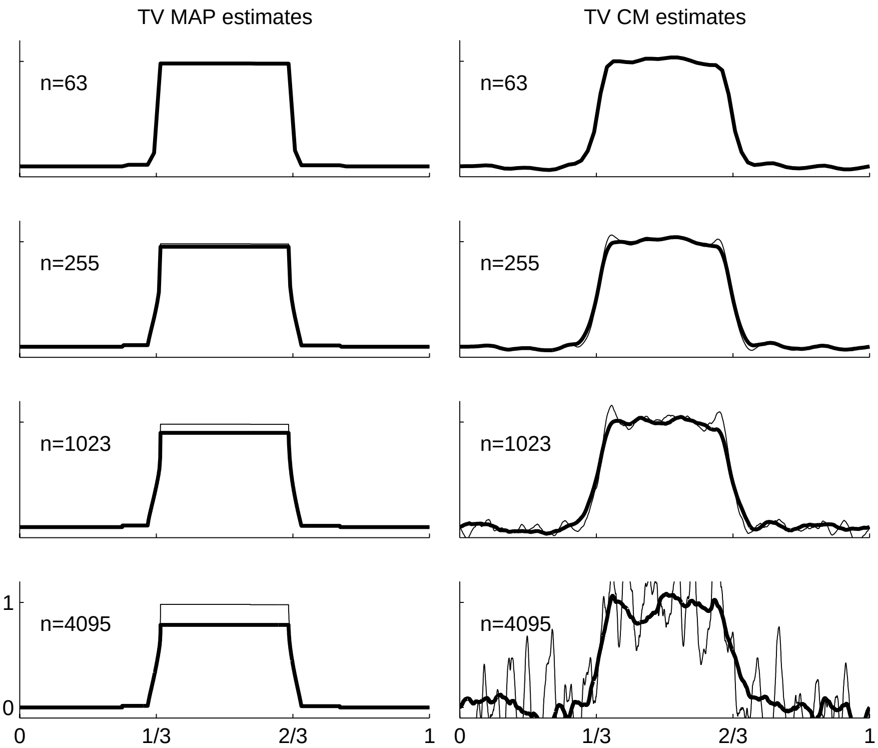

Figure 4. In all the plots in this figure, the coordinate axes limits are the same to allow easy comparison. Left column: MAP estimates for the TV prior with regularization parameter $\alpha_n = 135$ (thin line) and $\alpha_n = 16.875\sqrt{n+1}$ (thick line). Right column: CM estimates for the TV prior with regularization parameter $\alpha_n = 135$ (thin line) and $\alpha_n = 16.875\sqrt{n+1}$ (thick line).

6.3.3. CM estimates for TV prior We compute CM estimates using the MH algorithm for $n = 63, 255, 1023, 4095$. The choices (i) and (ii) of α_n given in (6.13) are compared. Parameters of the MCMC computations are given Table 1; in each case we take the zero vector as initial guess. See the right column in Figure 4 for plots of the CM estimates.

We actually use the MH algorithm slightly differently than explained in Section 6.2.3. Denote by r_{1000} the *local acceptance rate* of the last 1000 samples. Choosing too large κ to start with leads to $r_{1000} = 0$ and the chain does not move. On the other hand, choosing a very small κ results in a positive r_{1000} that, however, keeps growing until reaching a value close to 1; then the candidates are always accepted and the chain moves very slowly. To overcome this problem we introduce automatic doubling of κ whenever $r_{1000} > .35$, but then the resulting chain is not Markov. However, after running for a while, r_{1000} becomes nearly constant and κ is not changed any more. An interpretation of our strategy is that we use the κ-doubling scheme to find a good initial guess for the $K - k_0$ samples in the end of the chain that were drawn with constant κ. Those $K - k_0$ most recent samples form a Markov chain.

n	α_n	$K-k_0$	r	N_{update}	κ	Time (hours)
63	135	80 000 000	0.23	10	0.082	3
255	135	80 000 000	0.24	10	0.082	6
255	270	80 000 000	0.24	10	0.041	6
1023	135	80 000 000	0.25	10	0.082	18
1023	540	80 000 000	0.25	10	0.021	16
4095	135	160 000 000	0.24	100	0.024	153
4095	1080	610 000 000	0.23	100	0.004	520

Table 1. Parameters of MCMC computations. The number n is the dimension of the problem, $K-k_0$ is the number of samples used for computing the CM estimate, r is the acceptance rate defined in (6.11), N_{update} and κ are parameters of the proposal distribution and the last column indicates how many CPU hours the computations took.

6.4. Discussion

We have computed the following statistical estimates for the posterior distribution of the model problem with varying levels of discretization:

p	Estimate	α_n	Limit function
2	MAP	10	$\widetilde{u}(\,\cdot\,;2,10)$
2	CM	10	$\widetilde{u}(\,\cdot\,;2,10)$
1	MAP	135	$\widetilde{u}(\,\cdot\,;1,135)$
1	MAP	$16.875\sqrt{n+1}$	0
1	CM	135	Does not exist
1	CM	$16.875\sqrt{n+1}$	Smooth

The column "Limit function" above indicates the expected result of each computation in light of Theorems 4.1 and 5.2; the function $\widetilde{u}$ is defined by (4.2). How well do our computations agree with the theory?

The Gaussian computations in Section 6.3.1 together with the error estimate (6.12), illustrate the convergence of the Gaussian MAP estimates (and CM estimates as well, since the two coincide in the Gaussian case).

In case of the TV MAP estimates, the choice (ii) of regularization parameter gives the zero estimate at the limit $n \to \infty$. This is evident from Figure 4. Choice (i), or constant α_n, should lead to convergence to a limit function $\widetilde{u}(\,\cdot\,;1,135)$. This is clear from the superposition of those estimates for $n = 63, 255, 1023, 4095$ shown in Figure 5. It is interesting to note that, in spite of the apparent edge-preserving nature of the MAP estimates in Figure 5, the limit function $\widetilde{u}(\,\cdot\,;1,135)$ is known to belong to the Sobolev space $W_0^1(0,1)$ and thus be continuous!

The right column in Figure 4 shows the CM estimates for the TV prior. We can see that the choice (i) of regularization parameter leads to more and more oscillatory, divergent CM estimates, as expected (although the plot with $n = 4095$ is not to be completely trusted due to the possibly insufficient number of samples used). On the

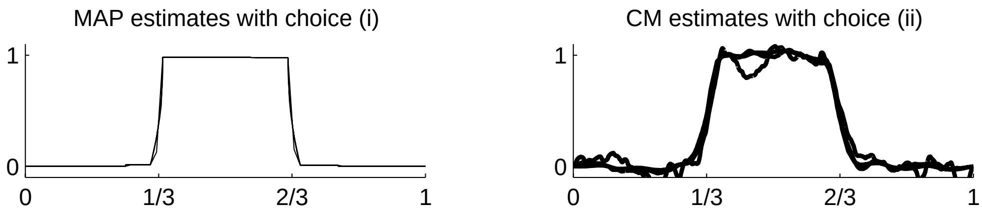

Figure 5. Left: superposition of TV MAP estimates with $n = 63, 255, 1023, 4095$ and choice (i), or $\alpha_n = 135$. Right: superposition of the TV CM estimates with $n = 63, 255, 1023, 4095$ and choice (ii), or $\alpha_n = 16.875\sqrt{n+1}$.

other hand, the CM estimates with choice (ii) are supposed to converge to a limit function. As the superposition in Figure 5 reveals, the CM estimate for $n = 4095$ is not of best possible quality. This is due to the very slow convergence of the chain; the computation took 520 hours. However, in our view the degree of convergence is enough to conclude that the limit function is not edge-preserving.

7. Acknowledgements

This work was supported by the Finnish Academy, the Finnish Technology Agency (TEKES), Instrumentarium Corporation, Imaging Division, Finland, and Invers Ltd., Finland. The second author was funded in part by the Grant-in-Aid for JSPS Fellows (No. 0002757) of the Japan Society for the Promotion of Science. The authors are grateful for fruitful discussions with Jean Bricmont, Jari Kaipio, Ville Kolehmainen, Antti Kupiainen, Sari Lasanen, Geoff Nicholls, Teemu Pennanen and Erkki Somersalo. Finally, the authors thank Markku Lehtinen for pointing out the most essential observation that total variation priors converge to a Gaussian variable.

References

[1] Arcudi O 1998 Convergence of conditional expectations given the random variables of a Skorohod representation. *Statist. Probab. Lett.* 40 1–8.

[2] Attouch H and Wets R 1987 Epigraphical analysis. Analyse non lineaire (Perpignan, 1987). Ann. Inst. H. Poincare Anal. Non Lineaire 6 (1989), suppl., 73–100.

[3] Attouch H 1984 Variational convergence for functions and operators. Applicable Mathematics Series. Pitman, Boston, MA. xiv+423 pp.

[4] Aubert G, Barlaud M, Blanc-Fraud M L and Charbonnier P 1997 *Deterministic edge-preserving regularization in computed imaging*, IEEE Transaction on Image Processing **5**(12)

[5] Aubert G, Kornprobst, P. (2000) *Mathematical Problems in Image Processing*, Applied Mathematical Sciences **147**, Springer Verlag.

[6] Boylan E 1971 Equiconvergence of martingales. *Ann. Math. Statist.* 42 552–559

[7] Brascamp H J, Lieb E H and Lebowitz J L 1976 The statistical mechanics of anharmonic lattices. Proceedings of the 40th Session of the International Statistical Institute (Warsaw, 1975), Vol. 1. Invited papers. *Bull. Inst. Internat. Statist.* 46 393–404

[8] Demengel F and Temam R 1984 Convex functions of a measure and applications. *Indiana Univ. Math. J.* **33**, 673–709.

[9] Dobson D C and Santosa F 1996 Recovery of blocky images from noisy and blurred data. *SIAM J. Appl. Math.* **56**, 1181–1198.

[10] Duoandikoetxea J 2001 Fourier analysis. Graduate Studies in Mathematics, 29. AMS. xviii+222 pp.
[11] Dudley R M 1989 Real analysis and probability. The Wadsworth & Brooks/Cole Mathematics Series. Pacific Grove, CA. xii+436 pp.
[12] Evans S N and Stark P B 2002 Inverse problems as statistics. *Inverse Problems* **18** R55–R97
[13] Fabian V and Hannan J 1985 Introduction to Probability and Mathematical Statistics. Wiley, New York..
[14] Gilks W R, Richardson S and Spiegelhalter D J 1996 Markov Chain Monte Carlo in Practice Chapmann & Hall
[15] Haario H, Laine M, Lehtinen M, Saksman S and Tamminen J. MCMC methods for high dimensional inversion in remote sensing. To appear in *J. Roy. Statist. Soc. Ser. B.*
[16] Hastings W K 1970 Monte Carlo sampling methods using Markov Chains and their applications *Biometrika* **57** 97–109
[17] Kaipio J P, Kolehmainen V, Somersalo E, and Vauhkonen M 2000 Statistical inversion and Monte Carlo sampling methods in electrical impedance tomography *Inverse Problems* 16 1487–1522.
[18] Kaipio J P, Kolehmainen V, Vauhkonen M, and Somersalo E 1999 Inverse problems with structural prior information. *Inverse Problems* 15 713–729
[19] Kolehmainen V, Siltanen S, Järvenpää S, Kaipio J P, Koistinen P, Lassas M, Pirttilä J and Somersalo E 2003 Statistical inversion for X-ray tomography with few radiographs II: Application to dental radiology *Physics in Medicine and Biology* 48 1465–1490
[20] Lasanen S 2002 Discretizations of generalized random variables with applications to inverse problems. Dissertation, University of Oulu. Ann. Acad. Sci. Fenn. Math. Diss. No. 130, 64 pp.
[21] Lehtinen, M. Personal communications.
[22] Lehtinen, Markku S.; Päivärinta, Lassi; Somersalo, Erkki Linear inverse problems for generalised random variables. Inverse Problems 5 (1989), no. 4, 599–612.
[23] D'Ambrogi B, Mäenpää S and Markkanen M 1999 Discretization independent retrieval of athmospheric ozone profile *Geophysica* **35**(1-2) 87-99.
[24] Meyer Y 2001: Oscillating patterns in image processing and nonlinear evolution equations, University lecture series, Volume 22, AMS, ISBN 0-8218-2920-3
[25] Mosegaard K and Sambridge M 2002 Monte Carlo analysis of inverse problems. *Inverse Problems* **18**, R29–R54.
[26] Naddaf A and Spencer T 1997 On homogenization and scaling limit of some gradient perturbations of a massless free field. Comm. Math. Phys. 183, no. 1, 55–84
[27] Oksendal B 1998 Stochastic differential equations. An introduction with applications. Universitext. Springer-Verlag. xx+324 pp.
[28] Rockafellar R T, Wets R J-B 1998 Variational analysis. Grundlehren der Mathematischen Wissenschaften, 317. Springer-Verlag, Berlin, xiv+733 pp.
[29] Rudin L I, Osher S and Fatemi E 1992 Nonlinear total variation based noise removal algorithms. *Physica D* **60** 259–268
[30] Siltanen S, Kolehmainen V, Järvenpää S, Kaipio J P, Koistinen P, Lassas M, Pirttilä J and Somersalo E 2003 Statistical inversion for X-ray tomography with few radiographs I: General theory *Physics in Medicine and Biology* 48 1437–1463
[31] Tamminen J 1999 MCMC methods for inverse problems *Geophysical publications* **48**, Finnish Meteorological Institute, ISBN 951-697-496-1
[32] Tikhonov A N 1963 Solution of incorrectly formulated problems and the regularization method, *Soviet Mathematics — Doklady*, vol. 4, pp.1035–1038.
[33] Vogel C and Oman M 1998 Fast, robust total variation-based reconstruction of noisy, blurred images. *IEEE Trans. Image Process.* **7** 813–824.
[34] Vakhania N, Tarieladze V I and Chobanyan S A 1987 Probability distributions on Banach spaces. Mathematics and its Applications (Soviet Series), 14. D. Reidel Publishing, xxvi+482 pp.
[35] Varah J M 1979 A practical examination of some numerical methods for linear discrete ill-posed problems *SIAM Rev.* 21, pp. 100–111.
[36] Ziemer, W. 1989 *Weakly differentiable functions. Sobolev spaces and functions of bounded variation.* Graduate Texts in Mathematics, 120. Springer-Verlag, xvi+308 pp.