Estimation and uncertainty quantification for piecewise smooth signal recovery

Victor Churchill · Anne Gelb

Abstract This paper presents a sparse Bayesian learning (SBL) algorithm for linear inverse problems with a high order total variation (HOTV) sparsity prior. For the problem of sparse signal recovery, SBL often produces more accurate estimates than maximum *a posteriori* estimates, including those that rely on ℓ_1 regularization. Moreover, rather than a single signal estimate, SBL yields a full posterior density estimate which can be used for uncertainty quantification. However, SBL is only immediately applicable to problems having a *direct* sparsity prior, or to those that can be formed via synthesis. This paper demonstrates how a problem with an HOTV sparsity prior can be formulated via synthesis, and then develops a corresponding Bayesian learning method. This expands the class of problems available to Bayesian learning to include, e.g., inverse problems dealing with the recovery of piecewise smooth functions or signals from data. Numerical examples are provided to demonstrate how this new technique is effectively employed.

1 Introduction

Many real-world phenomena give rise to piecewise smooth signals, [18]. As such, inverse problems to recover them from measurement data is a well-studied problem, [25]. Particular attention has been paid to piecewise smooth signal or function

V. Churchill
Department of Mathematics, Dartmouth College, Hanover, NH, USA
E-mail: Victor.A.Churchill.GR@dartmouth.edu

A. Gelb
Department of Mathematics, Dartmouth College, Hanover, NH, USA
E-mail: Anne.E.Gelb@dartmouth.edu

recovery from Fourier or spectral data, [12–14]. A now standard approach to piecewise smooth signal recovery is to minimize a least squares cost function with ℓ_1 norm based high order total variation (HOTV) regularization, [6,21]. This is well-known to encourage sparsity in the approximate edge domain of the function. In practice, this can be achieved by penalizing the gradient domain of the signal using the HOTV operator $\boldsymbol{T}_m \in \mathbb{R}^{(N-m)\times N}$, a finite difference approximation to the mth gradient. In particular, this paper deals with HOTV orders $m = 1, 2, 3$. While we do not explicitly consider $m \geq 4$, the methods developed are easily adapted. Using such an operator is common for inverse problems in image processing when one has a prior belief that the signal of interest being recovered is approximately piecewise polynomial of order $m - 1$, [1]. This technique has been useful in applications to improve robustness in synthetic aperture radar imaging, [1,22], and to recover fine details in electron tomography imaging, [23].

The main contribution of this paper is an alternative Bayesian learning based method for inverse problems where an HOTV sparsity prior is appropriate. This expands the class of problems available to this clearly strong method which provides a full posterior density estimate rather than a single point estimate. Because the sparsity assumption for piecewise smooth signal reconstruction is typically viewed in the analysis formulation, i.e. $\boldsymbol{T}_m \boldsymbol{x} = \boldsymbol{s}$ with $\boldsymbol{x}$ the signal of interest and $\boldsymbol{s}$ the sparse representation, Bayesian learning is not immediately applicable. In particular, since $\boldsymbol{T}_m$ is not square and therefore not invertible, more work is required. In what follows, our approach is to form an equivalent synthesis formulation of the form $\boldsymbol{x} = \boldsymbol{V}\boldsymbol{s}$ in order to effectively reduce the problem to sparse signal recovery. Since sparse Bayesian learning (SBL), [27], is applicable and has been shown to be superior to many other methods for sparse signal recovery, [15,17], then one can expect superiority in this synthesis construction as well. Our procedure involves a modification from [19] to the analysis operators $\boldsymbol{T}_m$ to make these operators full rank and therefore invertible. This ultimately enables the formulation of a Bayesian learning algorithm for inverse problems with a HOTV sparsity prior like piecewise smooth signal recovery.

This paper is organized as follows. Section 2 reviews sparse signal recovery using a maximum a posteriori estimate, and describes how both the synthesis and analysis approaches are typically employed to recover signals that are sparse in a transform domain (e.g. the HOTV domain). Section 3 explains how to formulate a synthesis approach for the HOTV analysis operators via the technique introduced in [19]. Since SBL typically provides superior performance for the synthesis approach, in Section 4 we demonstrate how SBL specifically can be applied to synthetic HOTV. Numerical examples are implemented in Section 5, where we demonstrate that our new approach, which we call high order total variation Bayesian learning (HOTVBL), outperforms the standard ℓ_1 norm based HOTV regularization (analysis approach). Some concluding remarks and ideas for future investigations are provided in Section 6.

2 Background

2.1 Sparse signal recovery

Let $\boldsymbol{x} \in \mathbb{R}^N$ be a sparse signal with $k \ll N$ of its elements nonzero. We seek to recover $\boldsymbol{x}$ from measurements

$$\boldsymbol{b} = \boldsymbol{A}\boldsymbol{x} + \boldsymbol{n}, \tag{1}$$

where $\boldsymbol{A} \in \mathbb{R}^{J \times N}$ is a given forward measurement matrix and the given data is $\boldsymbol{b} \in \mathbb{R}^J$. The vector $\boldsymbol{n} \in \mathbb{R}^J$ is a noise vector accounting for model and measurement error. In this paper we will assume that $\boldsymbol{n}$ is zero-mean white Gaussian with variance ν^2, and under the assumption that the entries of $\boldsymbol{b}$ are independent as in [17], we have the likelihood model

$$p(\boldsymbol{b}|\boldsymbol{x}) = (2\pi\nu^2)^{-J/2} \exp\left(-\frac{1}{2\nu^2}||\boldsymbol{A}\boldsymbol{x} - \boldsymbol{b}||_2^2\right). \tag{2}$$

A straightforward way to estimate $\boldsymbol{x}$ is to maximize this likelihood:

$$\begin{aligned}
\boldsymbol{x}_{ML}^* &= \arg\max_{\boldsymbol{x}} \{p(\boldsymbol{b}|\boldsymbol{x})\} \\
&= \arg\max_{\boldsymbol{x}} \left\{(2\pi\nu^2)^{-J/2} \exp\left(-\frac{1}{2\nu^2}||\boldsymbol{A}\boldsymbol{x} - \boldsymbol{b}||_2^2\right)\right\} \\
&= \arg\min_{\boldsymbol{x}} \left\{||\boldsymbol{A}\boldsymbol{x} - \boldsymbol{b}||_2^2\right\}.
\end{aligned} \tag{3}$$

However, solving (3) frequently yields a solution that is not sparse, i.e. with many greater than k nonzero elements. To see this, consider the denoising problem where $J = N$ and $\boldsymbol{A}$ is the identity matrix. In this case the estimate is just the noisy collected signal, $\boldsymbol{x}_{ML}^* \equiv \boldsymbol{b}$. To improve on this result, a prior on $\boldsymbol{x}$ is often incorporated to encourage sparsity. For example, the Laplace density function

$$p(\boldsymbol{x}) = \left(\frac{\mu}{2}\right)^N \exp\left(-\mu||\boldsymbol{x}||_1\right), \tag{4}$$

is frequently chosen because it corresponds to the ℓ_1 regularization often used in compressive sensing. Here μ determines the spread of the distribution and can intuitively be associated with how sparse $\boldsymbol{x}$ is. We note, however, that there are many sparsity-encouraging priors characterized by sharp peaks at zero. E.g., $||\boldsymbol{x}||_1$ in (4) can also be replaced with $||\boldsymbol{x}||_p^p$ for $p \in (0, 1]$ which would correspond to ℓ_p regularization. Using Bayes' theorem we can now compute a maximum *a posteriori* (MAP) estimate by maximizing the posterior

$$\begin{aligned}
\boldsymbol{x}_{MAP}^* &= \arg\max_{\boldsymbol{x}} \{p(\boldsymbol{x}|\boldsymbol{b})\} \\
&= \arg\max_{\boldsymbol{x}} \left\{\frac{p(\boldsymbol{b}|\boldsymbol{x})p(\boldsymbol{x})}{p(\boldsymbol{b})}\right\} \\
&= \arg\max_{\boldsymbol{x}} \left\{-\log p(\boldsymbol{b}|\boldsymbol{x}) - \log p(\boldsymbol{x})\right\} \\
&= \arg\min_{\boldsymbol{x}} \left\{||\boldsymbol{A}\boldsymbol{x} - \boldsymbol{b}||_2^2 + \mu\nu^2||\boldsymbol{x}||_1\right\}.
\end{aligned} \tag{5}$$

The first term in (5), often referred to as the fidelity term, is minimized when the solution aligns most closely with the given data. The second term is the imposed

sparsity penalty on $\boldsymbol{x}$. In the field of compressive sensing the sparsity prior parameter μ and noise variance ν^2 are often combined as $\lambda = \mu\nu^2$ and relabeled as the regularization parameter, which balances the fidelity term, the sparsity penalty, and noise reduction. Even though the inversion can be ill-posed, if certain conditions are met, then with high probability $\boldsymbol{x}$ can be exactly recovered from many fewer than N measurements using this method, [4]. From (5) it is also evident that without prior information of μ or ν^2, it can be difficult to choose a suitable regularization parameter for any given application.

Moreover, even if the prior parameters are known, the maximum is not categorically representative of the posterior density. Hence it may be favorable to estimate the entire posterior density and *then* derive statistics. In SBL, [27], a flexible, hierarchical prior whose parameters are learned from the data is used to encourage sparsity and estimate the entire posterior density. Confidence intervals can be derived from this posterior to aid in uncertainty quantification, [17]. For sparse signal recovery, in terms of accuracy at a given sparsity level k, SBL has outperformed a variety of other methods, [15,17], including the ℓ_1 regularization scheme in (5), [26], and more advanced reweighting algorithms, [5,7]. For more evidence see, e.g., Figs. 6, 7, and 8 in [15], and Figs. 2 and 4 in [17].

2.2 Synthesis and analysis

It is important to note that while ℓ_1 and reweighted regularization schemes are readily adapted to signal processing applications, i.e. where the sparsity occurs in some related domain (e.g. the gradient or wavelet domain), SBL is specifically designed for sparse signals. Because of its clear advantages in sparse signal recovery, we would like to adapt SBL for these other problems as well. Specifically, we are interested in applying SBL when the vector of interest $\boldsymbol{x}$ is not sparse, but a known transformation of $\boldsymbol{x}$ is. There are two approaches: synthesis and analysis.

In the *synthesis* approach, which is typically associated with compressed sensing, we formulate a method based on the assumption that

$$\boldsymbol{x} = \boldsymbol{V}\boldsymbol{s}, \tag{6}$$

where $\boldsymbol{V} \in \mathbb{R}^{N \times M}$ is called a synthesis operator and $\boldsymbol{s} \in \mathbb{R}^M$ is a sparse vector. Sparse signal recovery is used to obtain $\boldsymbol{s}$ from measurements in (1), now written as $\boldsymbol{b} = \boldsymbol{A}\boldsymbol{V}\boldsymbol{s} + \boldsymbol{n}$. Using ℓ_1 regularization as in (5), a synthesis approach recovery for the signal of interest $\boldsymbol{x}$ is given by

$$\boldsymbol{x}^*_{synthesis} = \boldsymbol{V}\left[\arg\min_{\boldsymbol{s}}\left\{||\boldsymbol{A}\boldsymbol{V}\boldsymbol{s} - \boldsymbol{b}||_2^2 + \lambda||\boldsymbol{s}||_1\right\}\right]. \tag{7}$$

Because synthesis effectively reduces the problem to sparse signal recovery, the SBL method is directly applicable to problems formed as (7) by recovering $\boldsymbol{s}$ and then synthesizing via (6). In the corresponding *analysis* approach, we formulate a method based on the assumption that

$$\boldsymbol{T}\boldsymbol{x} = \boldsymbol{s}, \tag{8}$$

where $\boldsymbol{T} \in \mathbb{R}^{M \times N}$ is called an analysis operator and $\boldsymbol{s} \in \mathbb{R}^M$ is a sparse vector. In an ℓ_1 regularization scheme, the signal of interest $\boldsymbol{x}$ is directly estimated by

regularizing on the sparsity of $\boldsymbol{Tx}$ as

$$\boldsymbol{x}^*_{analysis} = \arg\min_{\boldsymbol{x}} \left\{ ||\boldsymbol{Ax} - \boldsymbol{b}||_2^2 + \lambda||\boldsymbol{Tx}||_1 \right\}. \tag{9}$$

Choosing to use (7) or (9) may simply depend on whether it is more natural to view the sparsifying transformation as $\boldsymbol{Tx} = \boldsymbol{s}$ or $\boldsymbol{x} = \boldsymbol{Vs}$. The choice of analysis versus synthesis and the differences and similarities between them are further analyzed in [10]. A particular result of interest is that if $\boldsymbol{V} := \boldsymbol{T}^{-1}$, then (7) and (9) are equivalent. As discussed in more detail in Section 4.1, because of the conjugate prior structure used in the SBL method, SBL is not readily applicable to problems viewed in the analysis approach.

In this paper we focus on a particular problem that is typically viewed in the analysis approach, that is when the underlying signal can be viewed as a piecewise smooth function. We will mainly consider the case for which the analysis operator $\boldsymbol{T}$ is the high order total variation (HOTV) operator $\boldsymbol{T}_m \in \mathbb{R}^{(N-m)\times N}$, a finite difference approximation to the mth gradient. Using such an analysis operator is used in inverse problems when one has a prior belief that the signal of interest being recovered is approximately piecewise polynomial of order $m-1$, [1]. In particular, we are interested in $m = 1, 2, 3$. While we do not explicitly consider $m \geq 4$ in this paper, we provide general formulae for these cases.

Our goal is to formulate a Bayesian learning method for piecewise smooth signal recovery, or more generally inverse problems with a HOTV sparsity prior. This will expand the class of problems available to Bayesian learning. However, as mentioned above, the HOTV problem is viewed in the analysis approach with $\boldsymbol{T}_m\boldsymbol{x} = \boldsymbol{s}$ as in (8), and $\boldsymbol{T}_m$ is not square and therefore not invertible. Hence SBL is not immediately applicable. In Section 3, we demonstrate how to form an equivalent synthesis operator for HOTV in order to reduce the problem to sparse signal recovery, after which we can directly apply SBL. As a consequence, this approach should yield the same benefits as SBL does for standard sparse signal recovery. Our procedure involves a modification from [19] to make the analysis operators full rank and therefore invertible. This ultimately enables us to then formulate a Bayesian learning algorithm for inverse problems with a HOTV sparsity prior in Section 4.

3 Synthesis Operators for HOTV Regularization

In part because of its edge-preserving properties, HOTV regularization is a common technique for inverse problems in image processing, [1,22,23]. The corresponding HOTV operator, $\boldsymbol{T}_m \in \mathbb{R}^{(N-m)\times N}$ in (9), is a scaled finite difference approximation of the mth gradient.[1] For example, when $N = 6$ we have

$$\boldsymbol{T}_1 = \begin{bmatrix} -1 & 1 & 0 & 0 & 0 & 0 \\ 0 & -1 & 1 & 0 & 0 & 0 \\ 0 & 0 & -1 & 1 & 0 & 0 \\ 0 & 0 & 0 & -1 & 1 & 0 \\ 0 & 0 & 0 & 0 & -1 & 1 \end{bmatrix}; \boldsymbol{T}_2 = \begin{bmatrix} 1 & -2 & 1 & 0 & 0 & 0 \\ 0 & 1 & -2 & 1 & 0 & 0 \\ 0 & 0 & 1 & -2 & 1 & 0 \\ 0 & 0 & 0 & 1 & -2 & 1 \end{bmatrix}; \boldsymbol{T}_3 = \begin{bmatrix} -1 & 3 & -3 & 1 & 0 & 0 \\ 0 & -1 & 3 & -3 & 1 & 0 \\ 0 & 0 & -1 & 3 & -3 & 1 \end{bmatrix}.$$

$$\tag{10}$$

---

[1] As mentioned, we only consider $m \leq 3$, which sufficiently captures the signal variation in our examples. Higher order gradients may be more suitable in other applications, or when resolution is insufficient, [23].

Clearly $\boldsymbol{T}_1\boldsymbol{x} = \boldsymbol{s}$ is sparse whenever the underlying signal $\boldsymbol{x}$ is piecewise constant, since $\boldsymbol{T}_1$ is an exact transformation to the edge domain. High order gradients are useful when it is assumed that the smooth regions of the signal are better approximated by piecewise polynomials.

It has been demonstrated that SBL is more effective for sparse signal recovery than many other algorithms, including ℓ_1 regularization and many variants, [15, 17]. In the synthesis approach, problems with transform sparsity priors are essentially reduced to sparse signal recovery. Hence, since SBL is more effective for sparse signal recovery, it may be advantageous to use SBL whenever an analysis approach can be replaced by a synthesis approach. In addition, Bayesian learning is able to estimate a posterior distribution for the signal as opposed to a single signal estimate, which can aid in uncertainty quantification. In what follows we demonstrate this idea. Specifically, we employ SBL to estimate the HOTV sparsity representation of the signal and subsequently synthesize the piecewise smooth signal of interest. A density estimate for both the sparse representation and the signal of interest are obtained.

Since SBL is available *only* to problems formed via synthesis, we must first find a corresponding synthesis operator for $\boldsymbol{T}_m$. A problem quickly arises in developing a synthesis approach for HOTV, however. Notably, $\boldsymbol{T}_m$ is not invertible (or square), so the required synthesis operator $\boldsymbol{V}$ such that $\boldsymbol{x} = \boldsymbol{V}\boldsymbol{s}$ is not immediately apparent. Hence as in [19], we "complete" $\boldsymbol{T}_m$, which we will denote as $\tilde{\boldsymbol{T}}_m$, by adding rows in its null space. As in Appendix D of [19], rows corresponding to the 0 through $(m-1)$th forward difference coefficients are added. For example, when $N = 6$ and $m = 1, 2, 3$ we have

$$
\tilde{\boldsymbol{T}}_1 = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ -1 & 1 & 0 & 0 & 0 & 0 \\ 0 & -1 & 1 & 0 & 0 & 0 \\ 0 & 0 & -1 & 1 & 0 & 0 \\ 0 & 0 & 0 & -1 & 1 & 0 \\ 0 & 0 & 0 & 0 & -1 & 1 \end{bmatrix} ; \tilde{\boldsymbol{T}}_2 = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ -1 & 1 & 0 & 0 & 0 & 0 \\ 1 & -2 & 1 & 0 & 0 & 0 \\ 0 & 1 & -2 & 1 & 0 & 0 \\ 0 & 0 & 1 & -2 & 1 & 0 \\ 0 & 0 & 0 & 1 & -2 & 1 \end{bmatrix} ; \tilde{\boldsymbol{T}}_3 = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ -1 & 1 & 0 & 0 & 0 & 0 \\ 1 & -2 & 1 & 0 & 0 & 0 \\ -1 & 3 & -3 & 1 & 0 & 0 \\ 0 & -1 & 3 & -3 & 1 & 0 \\ 0 & 0 & -1 & 3 & -3 & 1 \end{bmatrix} .
$$

(11)

Observe that $\tilde{\boldsymbol{T}}_m \in \mathbb{R}^{N \times N}$ has rank N and yields a new sparse representation

$$
\tilde{\boldsymbol{T}}_m \boldsymbol{x} = \boldsymbol{t} = \begin{bmatrix} r \\ \boldsymbol{s} \end{bmatrix} , \tag{12}
$$

where $r \in \mathbb{R}^m$ and $\boldsymbol{s} \in \mathbb{R}^{N-m}$, thus $\boldsymbol{t} \in \mathbb{R}^N$. Moreover, the matrix completion is constructed in a sensible way since the points added to the sparse representation are simply a finite difference approximation to the derivative. For example, if the previous stencil contained three points, such as the case corresponding to $\boldsymbol{T}_2$, to construct the completion matrix $\tilde{\boldsymbol{T}}_2$ the coefficients of two point centered differencing are used in the second row. Since we can not approximate a derivative with one grid point, we simply use 1 in the first row. The process is similar for generating general $\tilde{\boldsymbol{T}}_m$, with each of the top m rows (except the first) having values corresponding to the coefficients of the mth finite difference derivative approximations.

The synthesis operators for HOTV analysis operators in (11) are subsequently defined by $\boldsymbol{V}_m := \tilde{\boldsymbol{T}}_m^{-1}$. For example, for $N = 6$ and $m = 1, 2, 3$ we have

$$\boldsymbol{V}_1 = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 & 0 \\ 1 & 1 & 1 & 0 & 0 & 0 \\ 1 & 1 & 1 & 1 & 0 & 0 \\ 1 & 1 & 1 & 1 & 1 & 0 \\ 1 & 1 & 1 & 1 & 1 & 1 \end{bmatrix}, \boldsymbol{V}_2 = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 & 0 \\ 1 & 2 & 1 & 0 & 0 & 0 \\ 1 & 3 & 2 & 1 & 0 & 0 \\ 1 & 4 & 3 & 2 & 1 & 0 \\ 1 & 5 & 4 & 3 & 2 & 1 \end{bmatrix}, \boldsymbol{V}_3 = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 & 0 \\ 1 & 2 & 1 & 0 & 0 & 0 \\ 1 & 3 & 3 & 1 & 0 & 0 \\ 1 & 4 & 6 & 3 & 1 & 0 \\ 1 & 5 & 10 & 6 & 3 & 1 \end{bmatrix}. \tag{13}$$

In general $\boldsymbol{V}_m$ is lower triangular. A general formula for $\boldsymbol{V}_m$ for $m \geq 2$ in terms of $\boldsymbol{V}_{m-1}$ is

$$\boldsymbol{V}_m[i,j] = \begin{cases} \boldsymbol{V}_{m-1}[i,j] & \text{if } j < m \\ \sum_{k=1}^{i} \boldsymbol{V}_{m-1}[k,j] & \text{if } j \geq m \end{cases}. \tag{14}$$

The main result of [19] of use in this paper is Lemma 3.2, which asserts that assuming there is sparsity in the latter $N - m$ elements of the new sparse vector (i.e. the original sparse representation), the ℓ_1 regularized estimate using $\tilde{\boldsymbol{T}}_m$ is consistent with the original problem using $\boldsymbol{T}$ since the added rows are in the null space of $\boldsymbol{T}$. In particular, the following two estimates via analysis and synthesis are shown to be equivalent:

$$\begin{aligned} \boldsymbol{x}_{\ell_1}^* &= \arg\min_{\boldsymbol{x}} \left\{ ||\boldsymbol{A}\boldsymbol{x} - \boldsymbol{b}||_2^2 + \lambda ||\boldsymbol{T}_m \boldsymbol{x}||_1 \right\} \\ &= \boldsymbol{V}_m \left[\arg\min_{\boldsymbol{t}} \left\{ ||\boldsymbol{A}\boldsymbol{V}_m \boldsymbol{t} - \boldsymbol{b}||_2^2 + \lambda ||\boldsymbol{t}[m+1:N]||_1 \right\} \right], \end{aligned} \tag{15}$$

where $\boldsymbol{t}[m+1:N]$ denotes the latter $N - m$ elements of $\boldsymbol{t}$. That is, in the synthesis form the sparsity-encouraging ℓ_1 norm only regularizes with respect to the elements of the original analysis operation. We use this equivalency in the next section to inspire the use of $\boldsymbol{V}_m$ as a synthesis operator in a Bayesian learning procedure. In addition, our approach does not require the sparsity in the remaining $N - m$ elements of the new sparse vector (as assumed in Lemma 3.2 in [19]), since the *data* directly dictate which elements in the sparsity domain have non-zero value.

4 High Order Total Variation Bayesian Learning (HOTVBL)

The MAP estimate provided in (5) is typically aligned with the compressive sensing approach for sparse signal recovery, and forms the basis for the approximation in (9) when the signal is sparse in some transform domain. As noted previously, the MAP estimate is not categorically representative of the posterior density. Because of this limitation, a better approach is needed.

In Bayesian learning, instead of a fixed sparsity-inducing prior on $\boldsymbol{t} = \tilde{\boldsymbol{T}}_m \boldsymbol{x}$, an empirical prior characterized by flexible parameters that must be estimated from the data is used. In this investigation we focus on sparse Bayesian learning (SBL), [27], which has also been used in Bayesian compressed sensing, [17]. It is important to note that SBL is only available to problems formed via synthesis or directly sparse problems. Indeed, this is what motivated our derivation of the HOTV synthesis operator in Section 3.

Recall that we seek to employ SBL since in many cases it has been shown empirically and theoretically to be superior in terms of accuracy to MAP estimates, [11,15,29,30]. Theoretical analysis in [20] and [30] shows that SBL provides a closer approximation to the ℓ_0 norm of the sparse signal than the ℓ_1 norm. For the noiseless case, it was shown in [29] that the global minimum of the effective SBL cost function is achieved at a solution such that the posterior mean equals the maximally sparse solution. Furthermore, local minima are achieved at sparse solutions, regardless of noise. Empirically, [15] shows that SBL achieves superior sparse signal recovery results compared to ℓ_1, reweighted ℓ_1, and reweighted ℓ_2 regularization (see [4,5,7], respectively). This is further supported by multi-run testing in [17]. In addition, SBL provides a full posterior distribution and confidence intervals versus a point estimate, and automatically estimates all parameters from the given data.

Hence SBL will be used in an attempt to more accurately detect the sparse HOTV (or approximate edge) representation by recovering the sparse signal $\boldsymbol{t} = \tilde{\boldsymbol{T}}_m \boldsymbol{x}$ from noisy measurements

$$\boldsymbol{b} = \boldsymbol{A}\boldsymbol{x} + \boldsymbol{n} = \boldsymbol{A}\boldsymbol{V}_m\boldsymbol{t} + \boldsymbol{n} := \boldsymbol{H}_m\boldsymbol{t} + \boldsymbol{n}, \tag{16}$$

where $\boldsymbol{n}$ is distributed zero-mean Gaussian with unknown variance ν^2. The piecewise smooth signal is then recovered via synthesis by $\boldsymbol{x} = \boldsymbol{V}_m\boldsymbol{t}$.

4.1 Sparse Bayesian Learning (SBL)

Below is a brief review of how SBL is formulated. More details can be found in [17, 27]. First we develop a parametrized prior on $\boldsymbol{t}$. Because Gaussian noise is assumed in (16), we define a conjugate zero-mean Gaussian prior on each element of $\boldsymbol{t}$

$$p(\boldsymbol{t}|\boldsymbol{a}) = \prod_{i=1}^{N} \mathcal{N}(\boldsymbol{t}_i|0, \boldsymbol{a}_i^{-1}),$$

where $\boldsymbol{a}_i$ is the precision or inverse variance. We then define a conjugate Gamma prior over $\boldsymbol{a}$

$$p(\boldsymbol{a}|a, b) = \prod_{i=1}^{N} \Gamma(\boldsymbol{a}_i|a, b).$$

Finally, we marginalize over the hyperparameters $\boldsymbol{a}$ to obtain the overall prior on $\boldsymbol{t}$ as

$$p(\boldsymbol{t}|a, b) = \prod_{i=1}^{N} \int_0^\infty \mathcal{N}(\boldsymbol{t}_i|0, \boldsymbol{a}_i^{-1}) \Gamma(\boldsymbol{a}_i|a, b) d\boldsymbol{a}_i. \tag{17}$$

Each integral being multiplied in (17) is distributed via the Student's t-distribution, which, for suitable a and b, is strongly peaked at $\boldsymbol{t}_i = 0$. Therefore this prior favors $\boldsymbol{t}_i$ being zero, hence encouraging sparsity. We also impose a conjugate Gamma prior $\Gamma(\beta|c, d)$ on $\beta = \frac{1}{\nu^2}$. Only point estimates are needed for $\boldsymbol{a}$ and β, so we simply set $a, b, c, d = 0$ implying uniform hyperpriors on a logarithmic scale for $\boldsymbol{a}$

and β, [27]. Because of the conjugate priors used above, the posterior distribution for $\boldsymbol{t}$ can be solved for analytically as a multivariate Gaussian distribution

$$p(\boldsymbol{t}|\boldsymbol{b}, \boldsymbol{a}, \beta) = \mathcal{N}(\boldsymbol{t}|\boldsymbol{\mu}, \boldsymbol{\Sigma}),$$

with mean and covariance matrix given by

$$\boldsymbol{\mu} = \beta\boldsymbol{\Sigma}\boldsymbol{H}_m^T\boldsymbol{b}, \tag{18}$$

$$\boldsymbol{\Sigma} = \left(\beta\boldsymbol{H}_m^T\boldsymbol{H}_m + \boldsymbol{\Lambda}\right)^{-1}, \tag{19}$$

where $\boldsymbol{\Lambda} = \mathrm{diag}(\boldsymbol{a})$, [2].

Marginalizing over $\boldsymbol{t}$, the marginal log-likelihood for $\boldsymbol{a}$ and β is

$$\begin{aligned}
\log p(\boldsymbol{y}|\boldsymbol{a}, \beta) &= \log \int p(\boldsymbol{y}|\boldsymbol{g}, \beta)p(\boldsymbol{g}|\boldsymbol{a})d\boldsymbol{g} \\
&= -\frac{1}{2}\left(J\log 2\pi + \log|\boldsymbol{C}| + \boldsymbol{y}^t\boldsymbol{C}^{-1}\boldsymbol{y}\right),
\end{aligned} \tag{20}$$

with $\boldsymbol{C} = \beta^{-1}\boldsymbol{I} + \boldsymbol{H}_m\boldsymbol{\Lambda}^{-1}\boldsymbol{H}_m^T$, [2]. Note that (20) cannot be maximized in closed form. In [27], a maximum likelihood approximation is employed that uses the point estimates for $\boldsymbol{a}$ and β to maximize (20), and is implemented via an *expectation-maximization* (EM) algorithm, [9]. In particular, the update for $\boldsymbol{a}$ to maximize (20) is

$$\boldsymbol{a}_i^{(\mathrm{new})} = \frac{\gamma_i}{\boldsymbol{\mu}_i^2} \tag{21}$$

for each i, with $\boldsymbol{\mu}_i$ the ith posterior mean weight from (18) and $\gamma_i = 1 - \boldsymbol{a}_i\boldsymbol{\Sigma}_{ii}$ with $\boldsymbol{\Sigma}$ from (19). For β the update is

$$\beta^{(\mathrm{new})} = \frac{M - \sum_i \gamma_i}{||\boldsymbol{b} - \boldsymbol{H}_m\boldsymbol{\mu}||_2^2}. \tag{22}$$

Appendix A of [27] gives details on the derivation of these terms. Observe that $\boldsymbol{a}^{(\mathrm{new})}$ and $\beta^{(\mathrm{new})}$ are functions of $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$, and vise versa. The EM algorithm iterates between (18) and (19), and (21) and (22) until a convergence criterion is satisfied. Due to the properties of the EM algorithm, SBL is globally convergent, i.e. each iteration is guaranteed to reduce the cost function, [29]. It has been observed that most $\boldsymbol{a}_i \to \infty$, corresponding to a sparse result with many $\boldsymbol{t}_i \approx 0$.

Note that after the convergence criterion has been satisfied, the final $\boldsymbol{\mu}^*$ and $\boldsymbol{\Sigma}^*$ are the mean and covariance matrix, respectively, of the Gaussian approximation to the posterior density function for $\boldsymbol{t} = \tilde{\boldsymbol{T}}_m\boldsymbol{x}$, *not* $\boldsymbol{x}$. This density can perhaps be useful for tasks typically accomplished by edge detection such as boundary identification, scale separation, or other downstream processes such as determining the support of the signal, or which regions of the signal may need further investigation. While this approximate edge density may be of some use in and of itself, the approximate Gaussian posterior density for the piecewise smooth signal of interest $\boldsymbol{x}$ is defined by the statistics

$$\begin{aligned}
\mathrm{E}(\boldsymbol{x}) &= \boldsymbol{V}_m\boldsymbol{\mu}^*, \\
\mathrm{Cov}(\boldsymbol{x}) &= \boldsymbol{V}_m\boldsymbol{\Sigma}^*\boldsymbol{V}_m^T.
\end{aligned} \tag{23}$$

Similar to the case of sparse signal recovery, $\text{Cov}(\boldsymbol{x})$ can be used to develop confidence intervals for the estimated values of $\boldsymbol{x}$ as will be shown in Section 5.

Note that each iteration of the described EM algorithm requires the inversion of an $N \times N$ matrix to compute the covariance matrix $\boldsymbol{\Sigma}$. This scales to $\mathcal{O}(N^3)$ operations – clearly inefficient for large N. Fast algorithms based on the cost function (20) have been developed, [11, 28], and are used in our numerical experiments. For signals of the size implemented in Section 5 (e.g., $N = 128$ and $N = 250$), we observed no difference between HOTVBL and the minimization of (9) in terms of runtime.

5 Numerical Results

We now perform a variety of tests comparing HOTV ℓ_1 estimates $\mathbf{x}^*_{\ell_1}$ given by (15) and the HOTVBL procedure described in Section 4.1. In particular we use $\mathbf{x}^*_{BL} = \text{E}(\mathbf{x})$ from (23) as the point estimate associated with HOTVBL. The noise level in the collected data is measured by signal-to-noise ratio defined

$$SNR = 20 \cdot \log_{10} \left(\sqrt{\frac{\sum_{i=1}^{N} \boldsymbol{x}_i^2}{\sum_{i=1}^{N} \boldsymbol{n}_i^2}} \right). \tag{24}$$

We compare the reconstructions using the relative error defined

$$RelErr(\boldsymbol{x}^*) = \frac{||\boldsymbol{x}^* - \boldsymbol{x}||_2}{||\boldsymbol{x}||_2}, \tag{25}$$

where $\boldsymbol{x}^*$ is the recovered signal and $\boldsymbol{x}$ is the ground truth. This provides a total error measure for each experiment. We also use the maximum error defined

$$MaxErr(\boldsymbol{x}^*) = \arg \max_i |\boldsymbol{x}_i^* - \boldsymbol{x}_i| \tag{26}$$

to quantify the worst case pointwise error.

Test 1: Probability of success at a given sparsity level with underdetermined Gaussian forward model and no noise. In this test, first a sparse signal $\boldsymbol{t} \in \mathbb{R}^{250}$ with k nonzero elements is generated with the height of the spikes drawn from a standard Normal distribution. This signal is then transformed by $\boldsymbol{V}_m$ in order to generate a piecewise $(m-1)$ order polynomial $\boldsymbol{x} \in \mathbb{R}^{250}$ with k jumps or edges whose heights are standard Normal distributed. This signal of interest is then obfuscated by a matrix $\boldsymbol{A} \in \mathbb{R}^{50 \times 250}$ where the entries of $\boldsymbol{A}$ have also been drawn from a standard Normal distribution. No noise is added, such that the data is modeled exactly by $\boldsymbol{b} = \boldsymbol{A}\boldsymbol{x} = \boldsymbol{H}_m\boldsymbol{t}$. The signal $\boldsymbol{x}$ is then reconstructed using HOTVBL with appropriate m. The above process represents one trial. A trial is considered a success if $MaxErr(\boldsymbol{x}^*_{BL}) \leq 10^{-3}$. For each $k = 1, 2, 3, \ldots, 25$, we ran 500 trials, with the success probability determined as the number of successes divided by 500. This is a standard test of sparse signal recovery methods, [15]. Figure 1 shows the results for HOTVBL using $m = 1, 2, 3$. Plots for the noise-free variant of (15), i.e. basis pursuit [8], are omitted as there were *no successes* registered in any of the 500 trials for any k value. Note that this lack of success is likely due at least in part to the inherent regularization parameter $\lambda = 1$ used in basis pursuit, [3]. In addition, the stringent success definition in terms of pointwise error likely ruled out many reconstructions whose relative error would be acceptable.
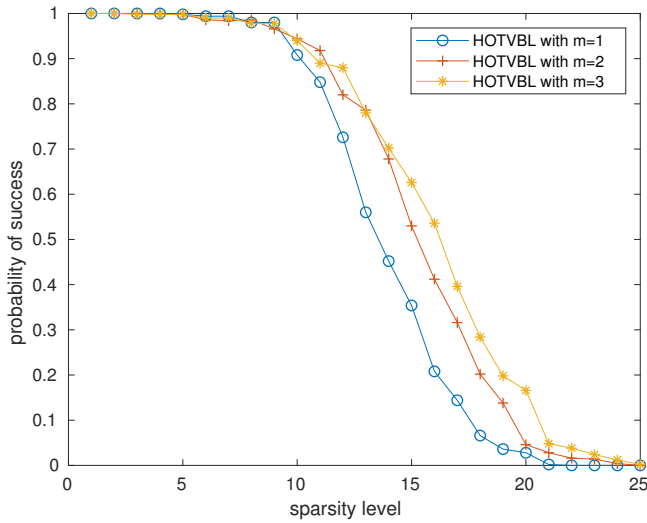
Fig. 1 Sparsity level versus probability of success.

Test 2: Denoising reconstruction of ideal signals with varying noise level. In this test, we consider the classical denoising problem, which epitomizes the difficulty in balancing fidelity, sparsity, and noise reduction. In denoising, $\boldsymbol{A} = \boldsymbol{I}$ the identity, meaning we collect a noisy signal $\boldsymbol{b} = \boldsymbol{x} + \boldsymbol{n}$, and regularize by the HOTV sparsity of the signal to return a result more faithful to the unknown ground truth signal. We compare the resulting reconstructions $\mathbf{x}^*_{\ell_1}$ from (15) and the proposed HOTVBL procedure $\mathbf{x}^*_{BL}$. We test first on *ideal* signals, that is ground truth signals that are exactly piecewise polynomial with only a single jump. In these cases, $\tilde{\boldsymbol{T}}_m$ is an appropriate sparsifying transform. Since the ground truth in this case is known, we can optimize the regularization parameter λ in (15) to minimize the relative error. We show this best-case scenario while noting that without oracle knowledge of the signal, this optimal result may be difficult to obtain in real-world examples.

Figures 2, 3, and 4 show comparisons of $\mathbf{x}^*\ell_1$ and $\mathbf{x}^*_{BL}$ for denoising piecewise constant, linear, and quadratic functions with one jump with $SNR = 0$ dB. Tables 1, 2, and 3, show the error statistics for $SNR = 0$ dB as well as other experiments on the same signals at various lower noise levels $SNR = 10, 20, 30$ dB. Bold in these tables indicates the superior performance. There is a significant improvement in accuracy both near edges and in smooth regions.

Finally, note the error bars in these plots and the significance they have with respect to uncertainty quantification. In signal recovery, typically only a single signal estimate is the final result. However, the data collected typically holds more information. In HOTVBL a posterior density is estimated rather than a single point estimate. This allows us to form error bars for the signal of interest as well as its sparse representation. These error bars indicate the certainty of the estimate. They represent the 99% confidence interval associated with the point estimate. These intervals are computed from the diagonal elements of the covariance matrix (i.e., the variance at each point). In addition to potential utility in downstream process-

Table 1 Comparison of relative and maximum errors with varying noise level for piecewise constant function with $m = 1$.

SNR	$MaxErr(\boldsymbol{x}_{BL}^*)$	$RelErr(\boldsymbol{x}_{BL}^*)$	$MaxErr(\boldsymbol{x}_{\ell_1}^*)$	$RelErr(\boldsymbol{x}_{\ell_1}^*)$
30 dB	**0.0004**	**0.0004**	0.0226	0.0108
20 dB	**0.0005**	**0.0005**	0.0968	0.0208
10 dB	**0.0197**	**0.0189**	0.1788	0.0639
0 dB	**0.0563**	**0.0556**	0.3580	0.1710

ing, one general observation these intervals yield is that uncertainty is typically higher in edge regions than in smooth regions.
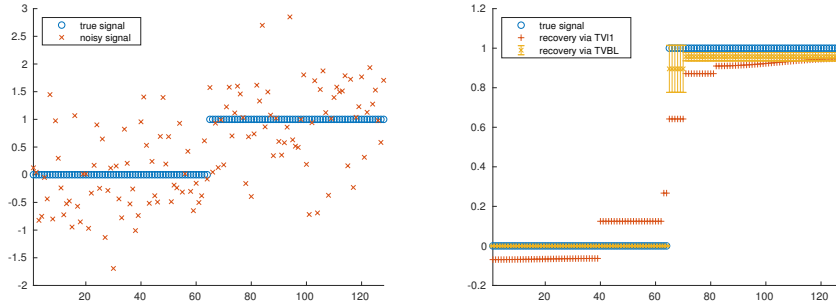

Fig. 2 Noisy, true, and recovered piecewise constant signals from noisy image data with SNR=0 dB using $m = 1$.
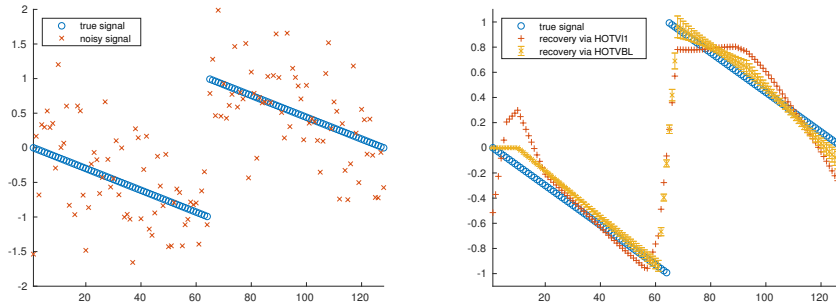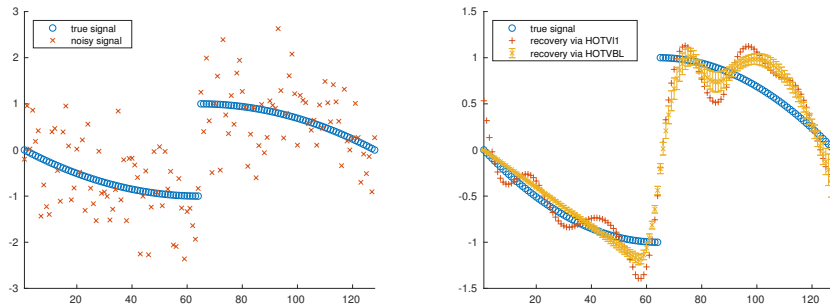

Fig. 3 Noisy, true, and recovered piecewise linear signals from noisy image data with SNR=0 dB using $m = 2$.

Test 3: Fourier reconstruction of non-ideal function from noisy complex Fourier data. Complex data can also be used with HOTVBL. E.g., if the signal is real and complex data with complex Gaussian noise is collected, then the model (16)

Table 2 Comparison of relative and maximum errors with varying noise level for piecewise linear function with $m = 2$.

SNR	$MaxErr(\boldsymbol{x}_{BL}^*)$	$RelErr(\boldsymbol{x}_{BL}^*)$	$MaxErr(\boldsymbol{x}_{\ell_1}^*)$	$RelErr(\boldsymbol{x}_{\ell_1}^*)$
30 dB	**0.0059**	**0.0054**	0.0693	0.0164
20 dB	**0.0094**	**0.0086**	0.0993	0.0516
10 dB	**0.3863**	**0.0636**	0.7247	0.1533
0 dB	**0.8681**	**0.2649**	0.9273	0.3850

Fig. 4 Noisy, true, and recovered piecewise quadratic signals from noisy image data with SNR=0 dB using $m = 3$.

Table 3 Comparison of relative and maximum errors with varying noise level for piecewise quadratic function with $m = 3$.

SNR	$MaxErr(\boldsymbol{x}_{BL}^*)$	$RelErr(\boldsymbol{x}_{BL}^*)$	$MaxErr(\boldsymbol{x}_{\ell_1}^*)$	$RelErr(\boldsymbol{x}_{\ell_1}^*)$
30 dB	**0.0128**	**0.0050**	0.0905	0.0227
20 dB	**0.0209**	**0.0142**	0.2025	0.0630
10 dB	**0.1226**	**0.0583**	0.7387	0.1801
0 dB	1.2022	**0.3286**	**1.1926**	0.3688

simply needs to be modified to

$$\begin{bmatrix} \text{Re}(\boldsymbol{y}) \\ \text{Im}(\boldsymbol{y}) \end{bmatrix} = \begin{bmatrix} \text{Re}(\boldsymbol{H}_m) \\ \text{Im}(\boldsymbol{H}_m) \end{bmatrix} \boldsymbol{t} + \begin{bmatrix} \text{Re}(\boldsymbol{n}) \\ \text{Im}(\boldsymbol{n}) \end{bmatrix}. \tag{27}$$

The problem of reconstructing piecewise smooth signals from spectral or Fourier data, i.e. where $\mathbf{A}$ is the discrete Fourier transform, is a well-studied problem, [12–14]. In this problem, discrete Fourier data is collected with SNR $= 10$ dB. The Bayesian learning procedure operates exactly as in Section 4.1. Signals $\mathbf{x}_{BL}^*$ and $\mathbf{x}_{\ell_1}^*$ recovered using $m = 2$ and $m = 3$ are shown in Figure 5. In opposition to *Test 2*, the signal used here is piecewise smooth with no m value perfectly sparsifying the signal. The values $m = 2$ and $m = 3$ were chosen because there are fewer nonzero coefficients in the sparsity representation compared with using $m = 1$. In particular, $\tilde{\boldsymbol{T}}_1 \boldsymbol{x}$ had $k = 52$, $\tilde{\boldsymbol{T}}_2 \boldsymbol{x}$ had $k = 34$, and $\tilde{\boldsymbol{T}}_3 \boldsymbol{x}$ had $k = 39$. The maximum and relative errors for $m = 1, 2, 3$ are given in Table 4.

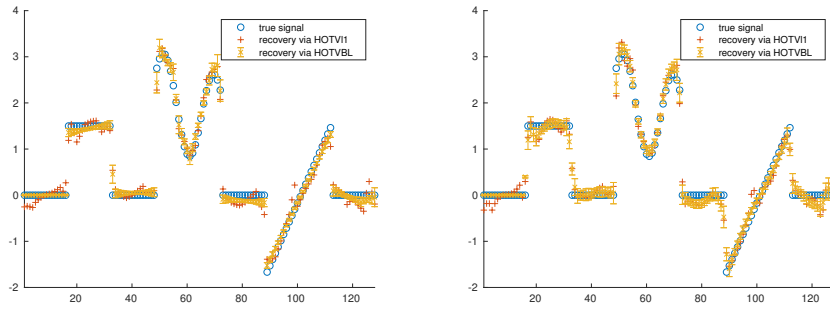**Fig. 5** Noisy, true, and recovered signals from Fourier data with SNR=10 dB on the left with $m = 2$ and on the right with $m = 3$.

Table 4 Comparison of relative and maximum errors with varying m for piecewise smooth function from Fourier data with SNR=10 dB.

m	$MaxErr(\boldsymbol{x}^*_{BL})$	$RelErr(\boldsymbol{x}^*_{BL})$	$MaxErr(\boldsymbol{x}^*_{\ell_1})$	$RelErr(\boldsymbol{x}^*_{\ell_1})$
1	0.5680	**0.1262**	**0.4893**	0.1329
2	**0.4542**	**0.0904**	0.5396	0.1491
3	**0.5674**	**0.1250**	0.5998	0.1670

6 Conclusion

This paper presented a Bayesian learning method for inverse problems with an HOTV sparsity prior, including the problem of piecewise smooth function recovery. The standard analysis form HOTV-regularized problem was reformulated by completing the rank of the HOTV analysis operator and inverting it to retrieve an equivalent synthesis operator. This allowed the creation of a Bayesian learning algorithm for piecewise smooth signal recovery that is typically only available for directly sparse problems. Our numerical experiments show that these methods show promise because of their accuracy, the provision of a full posterior density estimate including confidence intervals, and data-driven parameter estimation. HOTVBL is in particular much better suited than the standard HOTV regularized problem in low SNR environments.

Future investigations will include efforts to improve efficiency, perhaps by preprocessing with prior information, which may help to mitigate the cost of implementing HOTVBL for two-dimensional imaging problems. Another potential application for HOTVBL is in effective shock tracking for numerical conservation laws, where the number of grid points are typically much smaller than the number of pixels in a two dimensional image. HOTVBL may potentially increase the accuracy of the ℓ_1 regularization techniques for solving conservation laws discussed in [24, 16] for solving numerical conservation laws. Another benefit in extending the use of HOTVBL to conservation laws is that it will provide a full posterior density estimate as well.

Conflict of interest

The authors declare that they have no conflict of interest.

References

1. Archibald, R., Gelb, A., Platte, R.B.: Image reconstruction from undersampled fourier data using the polynomial annihilation transform. Journal of Scientific Computing **67**(2), 432–452 (2016)
2. Bishop, C.: Pattern Recognition and Machine Learning. Springer (2006)
3. Boyd, S., Parikh, N., Chu, E., Peleato, B., Eckstein, J., et al.: Distributed optimization and statistical learning via the alternating direction method of multipliers. Foundations and Trends® in Machine learning **3**(1), 1–122 (2011)
4. Candès, E.J., Romberg, J., Tao, T.: Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information. IEEE Transactions on information theory **52**(2), 489–509 (2006)
5. Candes, E.J., Wakin, M.B., Boyd, S.P.: Enhancing sparsity by reweighted ℓ_1 minimization. Journal of Fourier analysis and applications **14**(5), 877–905 (2008)
6. Chan, T., Marquina, A., Mulet, P.: High-order total variation-based image restoration. SIAM Journal on Scientific Computing **22**(2), 503–516 (2000)
7. Chartrand, R., Yin, W.: Iteratively reweighted algorithms for compressive sensing. In: Acoustics, Speech and Signal Processing, 2008. ICASSP 2008. IEEE International Conference on, pp. 3869–3872. IEEE (2008)
8. Chen, S.S., Donoho, D.L., Saunders, M.A.: Atomic decomposition by basis pursuit. SIAM review **43**(1), 129–159 (2001)
9. Dempster, A.P., Laird, N.M., Rubin, D.B.: Maximum likelihood from incomplete data via the em algorithm. Journal of the royal statistical society. Series B (methodological) pp. 1–38 (1977)
10. Elad, M., Milanfar, P., Rubinstein, R.: Analysis versus synthesis in signal priors. Inverse problems **23**(3), 947 (2007)
11. Faul, A.C., Tipping, M.E.: Analysis of sparse Bayesian learning. In: Advances in neural information processing systems, pp. 383–389 (2002)
12. Gelb, A.: A hybrid approach to spectral reconstruction of piecewise smooth functions. Journal of Scientific Computing **15**(3), 293–322 (2000)
13. Gelb, A.: Reconstruction of piecewise smooth functions from non-uniform grid point data. Journal of Scientific Computing **30**(3), 409–440 (2007)
14. Gelb, A., Tadmor, E.: Spectral reconstruction of piecewise smooth functions from their discrete data. ESAIM: Mathematical Modelling and Numerical Analysis **36**(2), 155–175 (2002)
15. Giri, R., Rao, B.D.: Type I and type II bayesian methods for sparse signal recovery using scale mixtures. IEEE Trans. Signal Processing **64**(13), 3418–3428 (2016)
16. Glaubitz, J., Gelb, A.: High order edge sensors with ℓ^1 regularization for enhanced discontinuous galerkin methods. SIAM Journal on Scientific Computing **41**(2), A1304–A1330 (2019)
17. Ji, S., Xue, Y., Carin, L.: Bayesian compressive sensing. IEEE Transactions on Signal Processing **56**(6), 2346–2356 (2008)
18. Mallat, S.: A wavelet tour of signal processing. Elsevier (1999)
19. Ortelli, F., van de Geer, S.: Synthesis and analysis in total variation regularization. arXiv preprint arXiv:1901.06418 (2019)
20. Rao, B.D., Wipf, D.P.: Comparing the effects of different weight distributions on finding sparse representations. In: Advances in Neural Information Processing Systems, pp. 1521–1528 (2006)
21. Rudin, L.I., Osher, S., Fatemi, E.: Nonlinear total variation based noise removal algorithms. Physica D: Nonlinear Phenomena **60**(1-4), 259–268 (1992)
22. Sanders, T., Gelb, A., Platte, R.B.: Composite sar imaging using sequential joint sparsity. Journal of Computational Physics **338**, 357–370 (2017)
23. Sanders, T., Gelb, A., Platte, R.B., Arslan, I., Landskron, K.: Recovering fine details from under-resolved electron tomography data using higher order total variation ?1 regularization. Ultramicroscopy **174**, 97–105 (2017)

24. Scarnati, T., Gelb, A., Platte, R.B.: Using $l1$ regularization to improve numerical partial differential equation solvers. Journal of Scientific Computing **75**, 225–252 (2018)
25. Stefan, W., Renaut, R.A., Gelb, A.: Improved total variation-type regularization using higher order edge detectors. SIAM Journal on Imaging Sciences **3**(2), 232–251 (2010)
26. Tibshirani, R.: Regression shrinkage and selection via the lasso. Journal of the Royal Statistical Society. Series B (Methodological) pp. 267–288 (1996)
27. Tipping, M.E.: Sparse bayesian learning and the relevance vector machine. Journal of machine learning research **1**(Jun), 211–244 (2001)
28. Tipping, M.E., Faul, A.C., et al.: Fast marginal likelihood maximisation for sparse bayesian models. In: AISTATS (2003)
29. Wipf, D.P., Rao, B.D.: Sparse bayesian learning for basis selection. IEEE Transactions on Signal processing **52**(8), 2153–2164 (2004)
30. Wipf, D.P., Rao, B.D.: ℓ_0-norm minimization for basis selection. In: Advances in Neural Information Processing Systems, pp. 1513–1520 (2005)