

# Regularization of linear and non-linear geophysical ill-posed problems with joint sparsity constraints

A. Gholami and H. R. Siahkoobi

*Institute of Geophysics, University of Tehran, Tehran, Iran. E-mail: agholami@ut.ac.ir*

Accepted 2009 November 11. Received 2009 October 27; in original form 2008 November 10

## SUMMARY

In this paper, we deal with the solution of linear and non-linear geophysical ill-posed problems by requiring the solution to have sparse representations in two appropriate transformation domains, simultaneously. Geological structures are often smooth in properties away from sharp discontinuities (i.e. jumps in 1-D and edges in 2-D). Thus, an appropriate ‘regularizer’ function should be constructed so that recovers the smooth parts as well as the sharp discontinuities. Sparsity inversion techniques which require the solution to have a sparse representation with respect to a pre-selected basis or frames (e.g. wavelets), can recover the regions of smooth behaviour in model parameters well, but the solution suffers from the pseudo-Gibbs phenomenon, and is smoothed around discontinuities. On the other hand, requiring sparsity in Haar or finite-difference (FD) domain can lead to a solution without generating smoothed edges and the pseudo-Gibbs phenomenon. Here, we set up a regularizer function which can be benefited from the advantages of both wavelets and Haar/FD operators in representation of the solution. The idea allows capturing local structures with different smoothness in the model parameters and recovering smooth/constant pieces of the solution together with discontinuities. We also set up an information function without requiring the true model for selecting optimum wavelet and parameter  $\beta$  which controls the weight of the two sparsifying operators in the inverse algorithm.

For both linear and non-linear geophysical inverse problems, the performance of the method is illustrated with 1-D and 2-D synthetic examples and a field example from seismic traveltime tomography. In all of the examples tested, the proposed algorithm successfully estimated more credible and high-resolution models of the subsurface compared to those of the smooth and traditional sparse reconstructions.

**Key words:** Wavelet transform; Inverse theory; Tomography.

## INTRODUCTION

Many attempts have been made in the geophysical context to find a unique and stable solution for geophysical inverse problems having the following main disadvantages: (1) *Underdetermination*: the problem is of course that an under-determined inverse problem usually has infinitely many solutions. (2) *Instability*: the inverse process causes unstable solutions because of the eigenvalues of the forward operator that are close to zero and because the observed data are usually contaminated with noise (which here we assume to be Gaussian distributed random). Therefore, it is apparently impossible to identify which of the candidate solutions is indeed the correct one and how to stabilize it without some prior information/assumption about the solution. Regularization methods are used to integrate the prior information/assumption into the solution of such inverse problems. What is obvious is that each regularization method has its own advantages and disadvantages and this arises from the fact that geological images are more complicated than can be explained

by a specific *a priori* information. Smooth inversion (Tikhonov & Arsenin 1977) produces a solution that is stable against noise, it recovers the regions of smooth behaviour properly but not the edges of the model (Klann *et al.* 2006; Youzwishen & Sacchi 2006; Gholami & Siahkoobi 2009a).

Non-smooth inversion is a suitable approach to overcome such difficulty with more computational complexities (Sacchi & Ulrych 1995; Ajo-Franklin *et al.* 2007; Loris *et al.* 2007; Routh *et al.* 2007). Edge-preserving inversion is sensitive to presence of the noise in data (Charbonnier *et al.* 1997; Routh *et al.* 2007), but well preserves sharp discontinuities in model parameters. CHA-based sparsity inversion is stable against noise (Donoho 1995a; Donoho & Johnstone 1998; Daubechies *et al.* 2004; Loris *et al.* 2007; Routh *et al.* 2007). The method is computationally efficient (Kim *et al.* 2007; Goldstein & Osher 2008) and enables one to apply regularization separately at different scales (Li *et al.* 1996; Loris *et al.* 2007; Gholami & Siahkoobi 2009a). It recovers the regions of smooth behaviour properly but leads to somewhat smoothed

edges and the solution will be oscillatory in the vicinity of sharp discontinuities. Therefore, a combination of different regularization methods integrates more priori information into the solution of inverse problems and hence can lead to a more plausible solution (Durand & Froment 2001; Candes & Guo 2002; Lustig *et al.* 2007; Herrmann *et al.* 2008a; Gholami & Siahkoobi 2009b).

In this paper, we impose another sparsity constraint on the solution of the commonly used wavelet-based sparsity inversion to furthermore get benefited by the advantages of edge-preserving smoothing techniques. Both Haar and FD operators are examined as additional sparsifying operators to achieve this goal. Also, based on the measure of smoothness and roughness of the recovered solution, we introduce a method for choosing optimum wavelet function and parameter  $\beta$  which controls the weight of the sparsifying operators. The performance of the proposed method is illustrated by applying on a variety of synthetic and field data sets from both linear and non-linear seismic traveltime tomography; the results are then compared with those of smooth and sparsity inversions.

In what follows, we first review the material on linear inverse problem including Tikhonov and sparsity-based regularizations. Then, the joint-sparsity inversion is introduced and iteratively obtaining of a solution by split Bregman technique is explained. It is continued by explaining how to select the optimum wavelet function and regularization parameters. Next, the performance of the method is evaluated by linear inversion of traveltime data. Finally, the method is extended to heterogeneous media and applied on data sets from a synthetic offset VSP and a field cross-well survey.

## LINEAR INVERSE PROBLEM

In this section, we consider the solution of a general linear inverse problem which arises frequently in geophysical applications (e.g. Sacchi & Ulrych 1995; van Wijk *et al.* 2002; Wang 2003; Herrmann *et al.* 2008b). Non-linear problems are considered later in this paper. A discrete linear inverse problem in general form can be written as follows:

$$\mathbf{d} = \mathbf{G}\mathbf{m} + \mathbf{e}, \quad (1)$$

where the observed data  $\mathbf{d} \in \mathbb{R}^N$  consist of unknown desired model  $\mathbf{m} \in \mathbb{R}^M$  first transformed by the ill-posed matrix operator  $\mathbf{G} \in \mathbb{R}^{N \times M}$  and then corrupted by Gaussian distributed random noise  $\mathbf{e} \in \mathbb{R}^N$  of zero-mean and finite variance  $\sigma^2$ . Given  $\mathbf{d}$  and  $\mathbf{G}$ , we want to solve the system of linear equations (1) for unknown model  $\mathbf{m}$ . The first approach for solving such system of equations is to minimize the data misfit  $\frac{1}{2} \|\mathbf{d} - \mathbf{G}\mathbf{m}\|_2^2$ . Even though  $\mathbf{G}$  has finite dimensional range, some of its eigenvalues can cluster close to zero making it an ill-posed problem. So, the minimizer  $\hat{\mathbf{m}} = \mathbf{G}^\dagger \mathbf{d}$  of the data misfit, where  $\mathbf{G}^\dagger$  stands for the Pseudo-inverse of  $\mathbf{G}$ , may not exist or the variance of the coloured noise  $\mathbf{G}^\dagger \mathbf{e}$  will be large, resulting in an unstable estimate. To overcome this, *a priori* information must be used to discard the models, which fit the noise in order to render the solution unique and stable (Gholami & Siahkoobi 2009b). This can be done by regularization techniques where the solution satisfies the supplied prior information while optimizing a suitably defined cost function (Tarantola 2005),

$$\begin{cases} \hat{\mathbf{m}} = \arg \min_{\mathbf{m}} f(\mathbf{m}), \\ f(\mathbf{m}) = \frac{1}{2} \|\mathbf{d} - \mathbf{G}\mathbf{m}\|_2^2 + \alpha \varphi(\mathbf{m}), \end{cases} \quad (2)$$

where  $f(\mathbf{m})$  is the cost function, the quadratic term is the data misfit, and  $\alpha > 0$  is the regularization parameter; we call  $\varphi(\mathbf{m})$

the *regularizer* function. The Tikhonov regularizer is defined as  $\varphi_{\text{Tikh}}(\mathbf{m}) = \|\mathbf{L}\mathbf{m}\|_2^2 = \sum_i |[\mathbf{L}\mathbf{m}]_i|^2$ , where  $\mathbf{L}$  is a discrete derivative operator (Tikhonov & Arsenin 1977). This regularizer penalizes large coefficients of  $\mathbf{L}\mathbf{m}$  much more than smaller ones. Hence, when attempting to penalize the small coefficients due to the noise, the large coefficients which are associated with the signal are more strongly penalized, leading to the distortion of the recovered solution (Youzwishen & Sacchi 2006; Gholami & Siahkoobi 2009a). The sparsity regularizer  $\varphi_s(\mathbf{m}) = \|\mathbf{L}\mathbf{m}\|_1 = \sum_i |[\mathbf{L}\mathbf{m}]_i|$  can be used to less penalize the outlier coefficients and hence capturing small-scale details in the solution (Bertete-Aguirre *et al.* 2002; Routh *et al.* 2007; Loris *et al.* 2007; Herrmann *et al.* 2008b). It can yield a solution which is sparse under the regularization operator  $\mathbf{L}$ . The performance of the sparsity regularizer requires that the operator efficiently (in the sparsity sense) represents the true model, while it inefficiently represents the noise corrupted version of it. The sparser the model coefficient vector, the lower the reconstruction error. Since no specific operator can provide a perfect sparse representation of complicated geological models, then none of the above mentioned regularizer functions will recover the smooth parts and sharp discontinuities of the model, simultaneously (Durand & Froment 2001; Lustig *et al.* 2007; Herrmann *et al.* 2008a). In the next section we provide an approach to do this appropriately.

## JOINT SPARSITY INVERSION

Our motivation for joint-sparsity inversion is two-fold. First, the existence of some multiscale transformation operators (e.g. wavelets) which provide sparse representations of geological models. Second, the existence of some specific operators (e.g. Haar and FD) which do provide sparse representations of models not suffering from smoothed edges and the pseudo-Gibbs phenomenon. Consequently, by imposing sparsity constraints jointly on model parameters in such domains, we aim to combine the advantages of both operators to obtain a high-resolution solution for geophysical inverse problems. In the field of image processing and medical imaging, several authors have observed that superior reconstructions occur when the advantages of such operators are combined (e.g. Starck *et al.* 2001; Candes & Guo 2002; Lustig *et al.* 2007). In seismic imaging, Herrmann *et al.* (2008a) used non-linear anisotropic diffusion jointly with sparsity inversion. Here, we propose following regularizer function

$$\varphi_{\text{JSC}}(\mathbf{m}) = (1 - \beta) \sum_i |[\mathbf{S}_1 \mathbf{m}]_i| + \beta \sum_i |[\mathbf{S}_2 \mathbf{m}]_i|, \quad (3)$$

where  $0 \leq \beta \leq 1$  is a parameter that controls the relative weight of the two sparsity terms and  $\mathbf{S}_1$  and  $\mathbf{S}_2$  are two suitable sparsifying operators. We choose  $\mathbf{S}_1$  as a suitable wavelet operator and  $\mathbf{S}_2$  as one of the Haar or FD operators. Taking  $\beta = 0$  reduces eq. (3) to the commonly used wavelet-based sparsity regularizer (Daubechies *et al.* 2004; Loris *et al.* 2007). The wavelet transform of geological models is generally sparse with many small coefficients, so the sparsity regularizer will remove small coefficients based on the amount of regularization applied by  $\alpha$  which is determined by the noise level. This leads to smoothed edges and puts some oscillations near the discontinuities of the final solution. Loris *et al.* (2007) used overcomplete basis functions with a specific thresholding algorithm to mollify the oscillatory artefact. Here we assigned a non-zero value for  $\beta$ ,  $0 < \beta < 1$ , to enforce the model to be jointly sparse in the domain defined by  $\mathbf{S}_2$ . This will typically recover small coefficients to eliminate such ripples while preserving edges. This

arises from the fact that  $\mathbf{S}_2$  does provide sparse representations of models without generating smoothed edges and the pseudo-Gibbs phenomenon. Finally, eq. (3) with  $\beta = 1$  can be used to obtain a piecewise-constant solution. In this case, however, depending on the noise level some small edges may be lost and some unrealistic edges may be interpreted as signal (Charbonnier *et al.* 1997; Routh *et al.* 2007). Therefore, the regularizer defined in eq. (3) really benefits from the advantages of both operators and can be used to capture local structures with different smoothness in the model parameters and to recover smooth/constant pieces of the solution together with the discontinuities. We use the split Bregman technique, described in the next section, to find minimizer of the cost function in eq. (2) with regularizer function given by eq. (3).

## SOLUTION BY THE SPLIT BREGMAN METHOD

Following Goldstein & Osher (2008) and ideas from (Bregman 1967; Osher *et al.* 2005), eq. (2) with joint-sparsity constraints is solved iteratively by Bregman splitting method which is derived using the Bregman distance (Bregman 1967) and functional splitting. The method is very efficient in the sense of convergence and numerical stabilities. Its algorithm is easy to code and can be used for solving large-scale geophysical problems because it does not need any stored matrices and is easily parallelizable.

The efficacy of split Bregman method is that it splits the l2-norm and l1-norm components of the cost function to generate a sequence of unconstrained problems which can be solved easily. Then, the procedure of optimizing the original cost function alternates between the main steps of solving the generated subproblems. The subproblem associated to the l2-norm term is quadratic and leads to the following linear system of equations (see Appendix A)

$$\mathbf{m}^{k+1} = (\mathbf{G}^T \mathbf{G} + \gamma \mathbf{S}_1^T \mathbf{S}_1 + \lambda \mathbf{S}_2^T \mathbf{S}_2)^{-1} r h s^{k+1}, \quad (4)$$

where

$$r h s^{k+1} = \mathbf{G}^T \mathbf{d} + \gamma \mathbf{S}_1^T (\mathbf{p}_1^k - \mathbf{q}_1^k) + \lambda \mathbf{S}_2^T (\mathbf{p}_2^k - \mathbf{q}_2^k), \quad (5)$$

$\gamma$  and  $\lambda$  are positive numbers which are the split parameters, and  $\mathbf{m}^{k+1}$  is the constructed model parameters at iteration  $k+1$  which is uniquely determined from the vectors  $\mathbf{p}_i^k, \mathbf{q}_i^k$ , where  $i = 1, 2$ . The algorithm is started by setting  $\mathbf{p}_i^0 = \mathbf{q}_i^0 = 0$ , then at the next iterations they are constructed from the solution of l1-norm subproblems computed easily using shrinkage operators applied on the coefficients of  $\mathbf{m}^{k+1}$ . For the computations of  $\mathbf{p}_i$  and  $\mathbf{q}_i$ , we refer to Bregman splitting algorithm summarized in Table 1 for more details. In the algorithm,  $shrink(x, \tau)$  is the so-called soft thresholding operation (Donoho 1995b)

$$shrink(x, \tau) = \begin{cases} x - \tau & x > \tau, \\ 0 & |x| \leq \tau, \\ x + \tau & x < -\tau. \end{cases} \quad (6)$$

Because of the component-wise character of eq. (6), it is easy to use scale dependent regularization if desired, that is, different thresholds for different scales in wavelet domain. Here, we use the same threshold level for all coefficients. The algorithm is also able to solve 2-D inverse problems with isotropic TV norm,  $\|\mathbf{m}\|_{TV} = \sum_i \sqrt{[\mathbf{D}_1 \mathbf{m}]_i^2 + [\mathbf{D}_2 \mathbf{m}]_i^2}$ , as the second sparsity constraint, where  $\mathbf{D}_1$  and  $\mathbf{D}_2$  are the first-order derivative operators in directions 1 and 2, respectively. This can easily be done by using  $shrink(x, \tau)$  as the generalized shrinkage operator (Wang *et al.* 2007) applied on the corresponding coefficients.

**Table 1.** The Bregman splitting algorithm for minimization of the cost function  $f(\mathbf{m})$  in eq. (2) with regularizer function given by eq. (3).

Chose
$\varepsilon, \lambda$ and $\gamma$
set:
$\tau_1 = \gamma^{-1} \alpha (1 - \beta), \quad \tau_2 = \lambda^{-1} \alpha \beta;$
$\mathbf{B} = \mathbf{G}^T \mathbf{G} + \gamma \mathbf{I} + \lambda \mathbf{S}_2^T \mathbf{S}_2;$
Initialize
$\mathbf{m}^0 = \mathbf{G}^T \mathbf{d}; \mathbf{p}_1^0 = \mathbf{p}_2^0 = \mathbf{q}_1^0 = \mathbf{q}_2^0 = 0;$
$k = 0;$
while $\ \mathbf{m}^k - \mathbf{m}^{k-1}\  / \ \mathbf{m}^k\  > \varepsilon$
$\mathbf{m}^{k+1} = \mathbf{B}^{-1} \{ \mathbf{G}^T \mathbf{d} + \gamma \mathbf{S}_1^T (\mathbf{p}_1^k - \mathbf{q}_1^k) + \lambda \mathbf{S}_2^T (\mathbf{p}_2^k - \mathbf{q}_2^k) \}$
$\mathbf{c}_1^{k+1} = \mathbf{S}_1 \mathbf{m}^{k+1} + \mathbf{q}_1^k$
$\mathbf{c}_2^{k+1} = \mathbf{S}_2 \mathbf{m}^{k+1} + \mathbf{q}_2^k$
$\mathbf{p}_1^{k+1} = shrink(\mathbf{c}_1^{k+1}, \tau_1)$
$\mathbf{p}_2^{k+1} = shrink(\mathbf{c}_2^{k+1}, \tau_2)$
$\mathbf{q}_1^{k+1} = \mathbf{c}_1^{k+1} - \mathbf{p}_1^{k+1}$
$\mathbf{q}_2^{k+1} = \mathbf{c}_2^{k+1} - \mathbf{p}_2^{k+1}$
$k = k + 1;$
end

The computational cost of the algorithm is largely dependent on how fast we can solve the system of equations (4), the application of the sparsifying operators, and their transposes. In practice, none of the matrices for multiscale sparsifying operators is constructed because they can be so large that cannot even be explicitly computed or stored. Instead, in the case of transformation operators with  $\mathbf{S}^T \mathbf{S} \mathbf{x} = \mathbf{I} \mathbf{x}$ , the application of the operator and its transpose simply are the discrete decomposition and reconstruction schemes. For the wavelet basis operators, these can be done using the fast discrete wavelet transform algorithm, the pyramid algorithm, given in Mallat (1998).

It is interesting to note that, unlike other optimization techniques such as Iteratively Reweighted Least Squares (IRLS) where at each iteration the left-hand side matrix is updated, in eq. (4) the right-hand side vector is updated. This brings useful advantages for the algorithm: the structure of  $\mathbf{B}$  remains constant and all tries to invert it amortize over all of the iterations and multiple problems with the same forward operator matrix and different data sets. When  $\mathbf{S}_2$  is the Haar operator, the left-hand side matrix reduces to  $\mathbf{B} = \mathbf{G}^T \mathbf{G} + (\gamma + \lambda) \mathbf{I}$  and when it is FD operator  $\mathbf{B} = \mathbf{G}^T \mathbf{G} + \gamma \mathbf{I} + \lambda \Delta$ , where  $\mathbf{I}$  is the identity matrix and  $\Delta = \mathbf{S}_2^T \mathbf{S}_2$  is a sparse matrix. Therefore, for small-scale problems, where direct solvers can be used to invert or factor  $\mathbf{B}$ , the speed of the algorithm will be very fast.

For large-scale problems a few steps of a pre-conditioned conjugate gradient (PCG) method (Kelley 1999) can be used to approximately solve eq. (4). The pre-conditioner used in the PCG algorithm approximates matrix  $\mathbf{G}^T \mathbf{G}$  with its diagonal entries. Thus, when  $\mathbf{S}_2$  is Haar operator, the diagonal pre-conditioner  $\mathbf{P}$  is

$$\mathbf{P} = diag(\mathbf{G}^T \mathbf{G}) + (\gamma + \lambda) \mathbf{I}, \quad (7)$$

and when  $\mathbf{S}_2$  is FD operator,  $\mathbf{P}$  is

$$\mathbf{P} = diag(\mathbf{G}^T \mathbf{G}) + \gamma \mathbf{I} + \lambda \Delta, \quad (8)$$

where  $diag(\mathbf{G}^T \mathbf{G})$  is the diagonal matrix obtained by setting the off-diagonal entries of the matrix  $\mathbf{G}^T \mathbf{G}$  to zero. Each iteration of the PCG algorithm has a step of inverting  $\mathbf{P}$  in order to compute  $\mathbf{P}^{-1} \mathbf{x}$ , for pre-conditioner (7) this is not an important issue because  $\mathbf{P}$  is

diagonal. Pre-conditioner (8) is sparse symmetric positive definite and so can be efficiently decomposed as  $C^T PC = R^T R$ . Where  $C$  is a permutation matrix,  $R$  is an upper triangular matrix and both of them are also significantly sparse (see Meyer 2000). Once the decomposition is performed,  $P^{-1}x$  can be computed efficiently as  $P^{-1}x = CR^{-1}R^{-T}C^T x$ , where  $R^{-1}$  and  $R^{-T}$  are performed by forward and backward substitutions, respectively. To further improve the speed of PCG, the recovered solution at the previous iteration is used as initial solution for current iteration. Numerical examples have shown that, by doing so the number of required CG iterations decreases rapidly as the final solution is approached.

The total number of iterations and the running time of algorithm depend on parameters  $\gamma$  and  $\lambda$ . In our calculations, we chose  $\lambda = \gamma$  and took into account the following considerations: (1) The condition number of  $B$  improves as  $\gamma$  increases,  $\text{cond}(B) \propto \gamma^{-1}$ , so eq. (4) will be solved efficiently by very few iterations of PCG. (3) A higher value for  $\gamma$  will lead to slower convergence of the split Bregman technique. We used  $\gamma = \sigma_{\max}$ , where  $\sigma_{\max}$  is the largest singular value of  $G$  and found that this usually results to reasonable convergence of both CG and split Bregman method.

### Selection of wavelet function and regularization parameters

The main goal of the joint-sparsity constraints is to recover the smooth and the rough part (edges) of a model. Requiring sparsity of the model in the wavelets domain will capture much of the model energy, especially the smooth part of it, while requiring sparsity in Haar or FD domain will capture the edges of the model. Both wavelet function and  $\beta$  affect the performance of the joint-sparsity regularizer function and must be chosen appropriately. Generally, the wavelet scaling function,  $\phi(x)$ , which leads to the orthonormal wavelet operator  $S_1$ , should have properties similar to the original model (continuity, continuity in derivatives, etc.) to provide a sparse coefficient vector (Daubechies 1992; Mallat 1998). In addition, the second constraint should be integrated into the algorithm carefully to minimize its effect on the regions of smooth behaviour while preserving edges and suppressing oscillations.  $\beta$  is the key parameter that balances the effect of the two constraints in trying to recover the smooth as well as the rough part of the solution. It controls the relative weight of the coefficients in two sparsity domains.

Assuming that one can select  $\alpha$  properly, then the recovered model can be considered as a function of  $\beta$  and wavelet scaling function,  $\hat{m}(\beta, \phi)$ . Often, there is not enough information to choose  $\beta$  and  $\phi(x)$  for an unknown model; yet an appropriate  $(\beta, \phi)$  is required to apply the joint-sparsity constraints. To overcome the problem, one needs to find an algorithm without requiring the true model. One possibility is to define a measure of information (e.g. smoothness and roughness), so called information function, which is optimized by the desired regularized solution. We use the l2-norm of the second derivative of the regularized solution,  $\|D^{(2)}\hat{m}(\beta, \phi)\|_2$ , as a measure of smoothness and l1-norm of its first derivative,  $\|D^{(1)}\hat{m}(\beta, \phi)\|_1$  as a measure of roughness [where  $D^{(i)}$  is the  $i$ th-order derivative operator], and define an information function as follow

$$I(\beta, \phi) = \|D^{(1)}\hat{m}(\beta, \phi)\|_1 + \|D^{(2)}\hat{m}(\beta, \phi)\|_2, \quad (9)$$

where  $\hat{m}$  is the solution obtained by the joint-sparsity inversion. The selected  $\beta$  and  $\phi(x)$  should approximately minimize  $I(\beta, \phi)$ . Later on, the performance of this approach is evaluated through 1-D and 2-D numerical experiments.

Parameter  $\alpha$  controls the relative importance of regularization and data misfit by scaling all of the coefficients in sparse domains. The proper  $\alpha$  can be determined using the  $\chi^2$ -distribution of the  $\|e\|_2^2$  which leads to the misfit condition  $\|G\hat{m} - d\|_2^2 \leq \xi$ , where the tolerance  $\xi$  depends on the noise level in the data as determined by the standard deviation of the noise. By letting  $\xi = \sigma^2 N$ , where  $N$  is the number of data, the corresponding  $\hat{m}$  fits the data on average to one standard deviation. An important part of the optimization is thus the determination of the largest successful  $\alpha$  for which the specified misfit condition holds. It can be found using a bisection method when the desired  $\alpha$  is bracketed by a pair of known points. The bracketing points are found by solving the problem for an  $\alpha$  that is decreased from a sufficiently large starting value. We are aware of the fact that *a priori* estimates of data variance are not available for most geophysical measurements making direct application of this method difficult. Since the goal of this paper is to show the performance of the joint-sparsity inversion scheme, here we assumed that data variance is available. However, other techniques like the L-curve method (Hansen & O'Leary 1993) and the generalized cross-validation (GCV) method (Wahba 1990) can be used for selecting  $\alpha$  without requiring noise estimates.

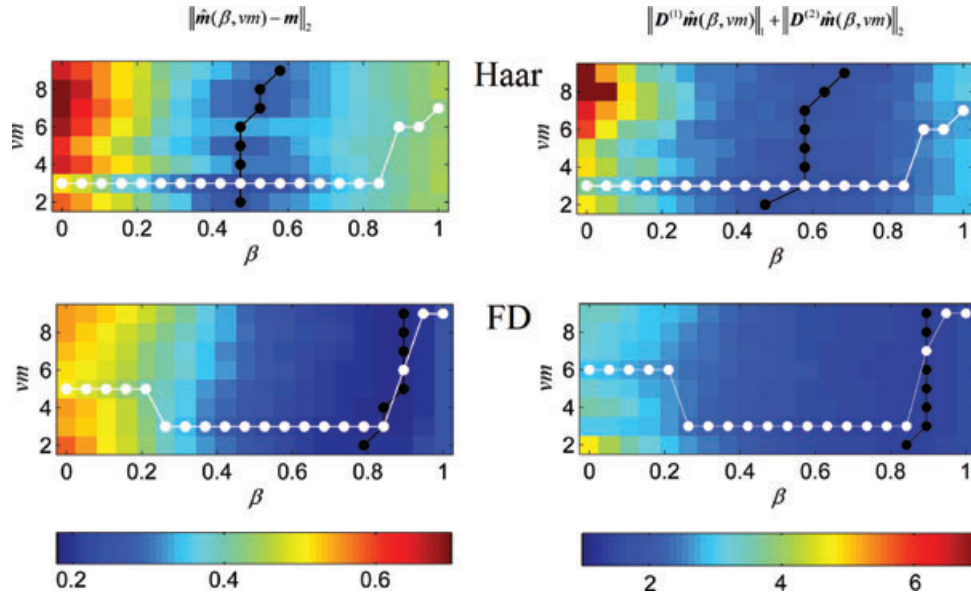
### APPLICATION TO LINEAR EXPERIMENT PROBLEM

In this section, we study the performance of the proposed scheme for solving linear ill-posed problems. We applied the method on data sets from two synthetic 1-D and 2-D numerical experiments. For a 1-D experiment we considered a zero-offset VSP traveltimes inversion and as a 2-D experiment a straight-ray cross-well traveltimes tomography is considered. In both cases we assumed  $S_1$  to be the compactly supported orthonormal Daubechies' wavelets operator (Daubechies 1992), and  $S_2$  to be Haar and FD operators.

#### Zero-offset VSP traveltimes tomography experiment

In this experiment we considered the problem of reconstructing the earth's  $P$ -wave velocity field as a function of depth using zero-offset VSP traveltimes data. Generally this is an ill-posed problem with the forward operator as integration acting on the earth's slowness model along the borehole to produce the traveltimes data. The experiment consists of a well of 2000 m depth where the velocity model along the borehole is made of smooth, constant and linear pieces with sharp discontinuities. Equally spaced 512 geophones are positioned from 0 to 2000 m within the borehole. For each geophone the corresponding traveltimes is the integration of the slowness model from 0 to the depth of the geophone. Random uncertainties in the data are simulated by contaminating the generated traveltimes with Gaussian distributed random noise of zero mean and standard deviation  $\sigma = 0.2$ . Other uncertainties are ignored assuming a perfect modelling. To invert the traveltimes, the medium below the surface is discretized into 1024 horizontal homogeneous layers of equal thickness and unknown  $P$ -wave velocities.

To select the optimum wavelet function and  $\beta$ , we inverted the data set by joint-sparsity inversion using different choices of compactly supported Daubechies wavelets and  $0 \leq \beta \leq 1$  with small increments of  $\beta$ . We then plotted the information  $\|D^{(1)}\hat{m}(\beta, vm)\|_1 + \|D^{(2)}\hat{m}(\beta, vm)\|_2$  as a function of  $\beta$  and increasing orders of the vanishing moments ( $vm$ ) of wavelet function. To test the performance of the information function, we plotted the norm of model error  $\|\hat{m}(\beta, vm) - m\|_2$  for the reconstructions.



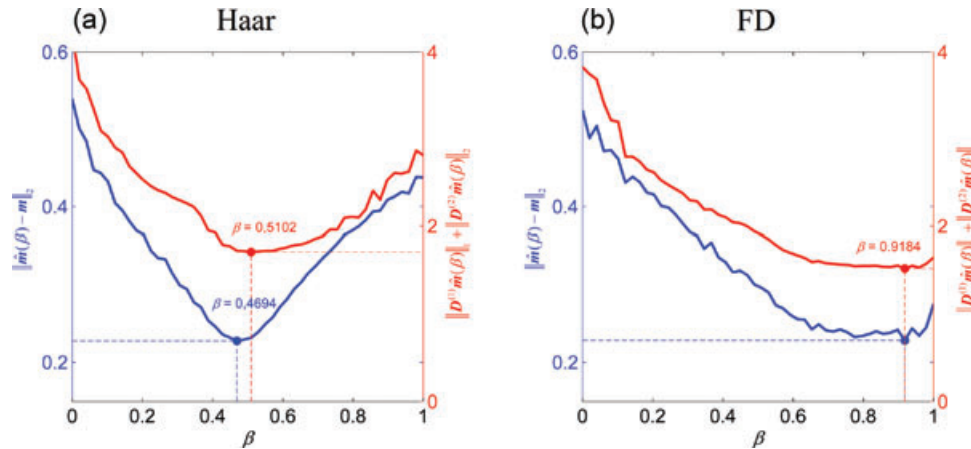
**Figure 1.** Maps of the norm of model error  $\|\hat{m}(\beta, vm) - m\|_2$  (left-hand column) and information function  $\|D^{(1)}\hat{m}(\beta, vm)\|_1 + \|D^{(2)}\hat{m}(\beta, vm)\|_2$  (right-hand column) for inverted model from joint-sparsity inversion using Haar (top row) and FD (bottom row) operators. In the panels, the location of the minimum of each column of the image is indicated by a white circle and the location of the minimum of each row is indicated by a black circle.

The results for  $S_2$  as Haar and FD operators are shown in top and bottom rows of Fig. 1, respectively. In the panels, the location of the minimum of each column of the image is indicated by a white circle and the location of the minimum of each row is indicated by a black circle. The trends of black and white circles are helpful in deciding about the choice of optimum wavelet function and  $\beta$ . For the model under consideration it is important to note that in Fig. 1:

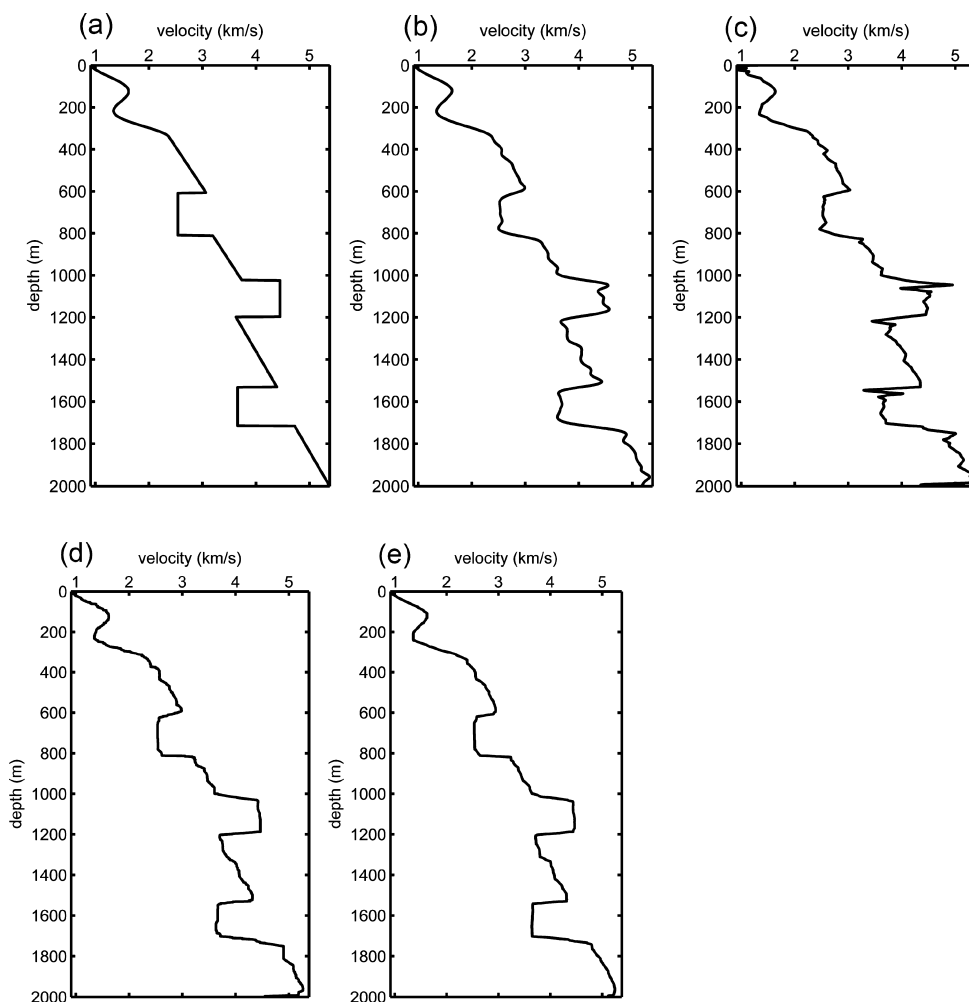
- (1) The trends of black and white circles in the maps of  $\|\hat{m}(\beta, vm) - m\|_2$  and  $\|D^{(1)}\hat{m}(\beta, vm)\|_1 + \|D^{(2)}\hat{m}(\beta, vm)\|_2$  are very similar.
- (2) For both Haar and FD operators, Daubechies wavelet with three vanishing moments (db3) showed the best performance for most of the  $\beta$  values.
- (3) For both Haar and FD operators, increasing of the number of vanishing moments in the wavelet function, increase the optimum  $\beta$  value. This is probably due to the fact that the regularity (i.e.

smoothness) of the Daubechies wavelets increases linearly with its support width (i.e. on the number of vanishing moments).

For most of the wavelet functions in combination with the Haar operator, the optimum  $\beta$  value ranged from 0.4 to 0.6, but in the case of FD operator it varied from 0.8 to 0.95. For further improvement of the choice of optimum  $\beta$ , we used smaller increments of  $\beta$  and performed series of inversion using db3 wavelet and  $0 \leq \beta \leq 1$ . The resulted norm of model error  $\|\hat{m}(\beta) - m\|_2$  as well as information function  $\|D^{(1)}\hat{m}(\beta)\|_1 + \|D^{(2)}\hat{m}(\beta)\|_2$  for Haar and FD operators are shown in Figs 2(a) and (b), respectively. As seen from Fig. 2,  $\beta$  value obtained from optimization of the information function is very close to, if not the same as, its optimum value. Note also that for properly selected regularization parameters, the Haar and FD operators performed approximately the same in the combination with db3 wavelet. The figures also indicate that joint-sparsity inversion performed much better (in the sense of model error) than commonly used sparsity inversions (i.e. when  $\beta = 0$  or 1).



**Figure 2.** Norm of model error  $\|\hat{m}(\beta) - m\|_2$  (the blue coloured curves) and information function  $\|D^{(1)}\hat{m}(\beta)\|_1 + \|D^{(2)}\hat{m}(\beta)\|_2$  (the red coloured curves) calculated for inverted model from joint-sparsity inversion using db3 wavelet together with Haar (left-hand panel) and FD (right-hand panel) operators. The circles indicate minimum of the curves.



**Figure 3.** The results of the synthetic zero-offset VSP traveltime tomography experiment. The true velocity model (a), and the inverted model from (b) second-order regularization, (c) sparsity inversion by promoting sparsity in the db3 wavelet domain, (d) joint-sparsity inversion using the db3 wavelet with Haar operator and (e) joint-sparsity inversion using the db3 wavelet with FD operator.

We are now about to compare the reconstruction from joint-sparsity inversion with those of the smooth and common sparsity inversions. Fig. 3(a) represents the true velocity model. The smooth model inversion using second-order regularization is shown in Fig. 3(b). As seen, the smooth part of the velocity model (from 0 to depth 600 m) was substantially recovered, but the rest of the model was not recovered well, and sharp discontinuities were blurred. A less smooth velocity model from the sparsity inversion by promoting sparsity in db3 wavelet domain is shown in Fig. 3(c). Yet, this reconstruction suffers from unrealistic oscillations. Figs 3(d) and (e) represent the velocity models from joint-sparsity scheme with db3 wavelet in the combination with Haar and FD operator respectively. As seen in the figure, the quality of the reconstructed velocity model is greatly improved and the pseudo-Gibbs phenomenon are suppressed for both cases of the joint-sparsity inversion (Figs 3d and e) comparing to those of smooth (Fig. 3b) or sparsity (Fig. 3c) inversion.

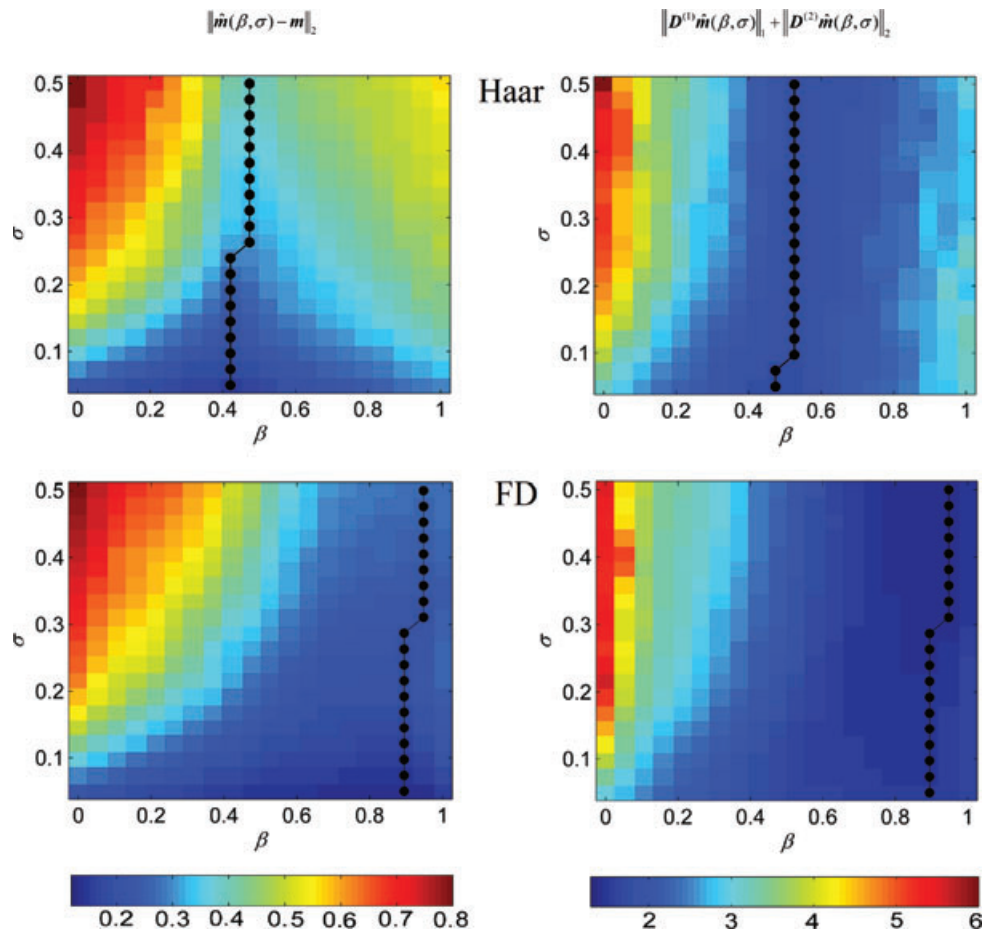
For this synthetic example, we also used a variety of noises with large range of variances to understand the role of noise on inverted model in the context of joint-sparsity constraint and to see the behaviour of  $\beta$  in the presence of different noise levels. For both Haar and FD operators in the combination with db3 wavelet we generated a map of model error  $\|\hat{\mathbf{m}}(\beta, \sigma) - \mathbf{m}\|_2$  with respect to  $\beta$  and

noise standard deviation (Fig. 4, left-hand column). We also calculated the information function  $\|\mathbf{D}^{(1)}\hat{\mathbf{m}}(\beta, \sigma)\|_1 + \|\mathbf{D}^{(2)}\hat{\mathbf{m}}(\beta, \sigma)\|_2$  for inverted models (Fig. 4, right-hand column). In the panels, the location of the minimum of each row of the image is indicated by a black circle which show the approximate optimum  $\beta$  corresponding to each  $\sigma$ . Here, again, obtained  $\beta$  value by the map of  $\|\mathbf{D}^{(1)}\hat{\mathbf{m}}(\beta, \sigma)\|_1 + \|\mathbf{D}^{(2)}\hat{\mathbf{m}}(\beta, \sigma)\|_2$  is very close to, if not the same as, its optimum value. It can also be seen that as noise variance increases, the optimum  $\beta$  tends to get higher value and so the edge-preserver penalty gets higher weight in the inversion.

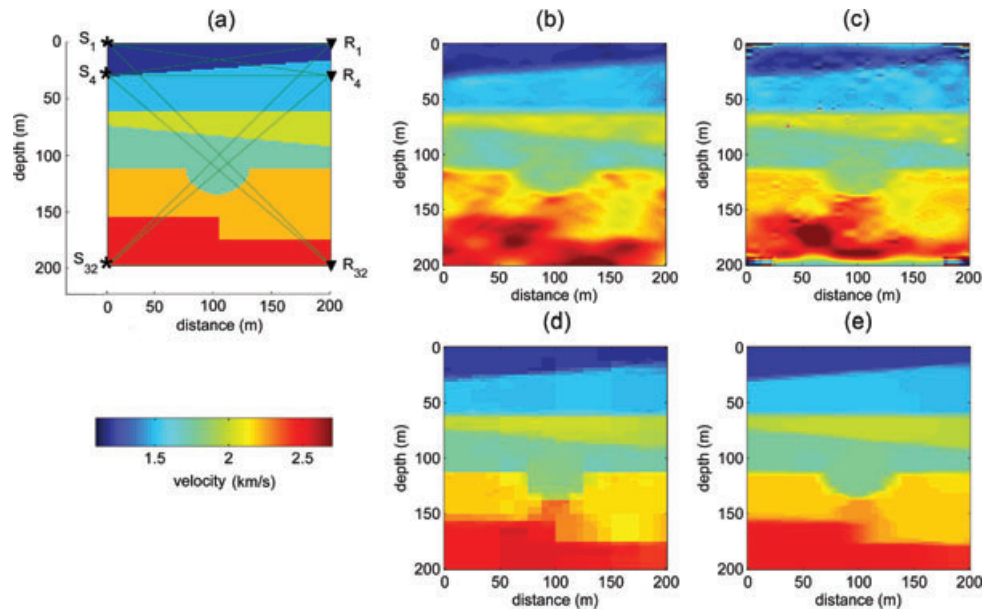
### Straight-ray cross-well traveltime tomography experiment

In this experiment, as a 2-D linear inversion, we considered the problem of determining the earth's  $P$ -wave velocity distribution over a cross-well section using a straight-ray traveltime tomography. The true velocity field between two boreholes (200 m deep and 200 m away from each other) with the source–receiver geometry is shown in Fig. 5(a). The true velocity model contains a number of layers with different thickness and velocities including horizontal and dipping interfaces, a single channel feature and a fault. The traveltimes of the  $P$ -waves transmitted from 32 equally spaced sources in one borehole

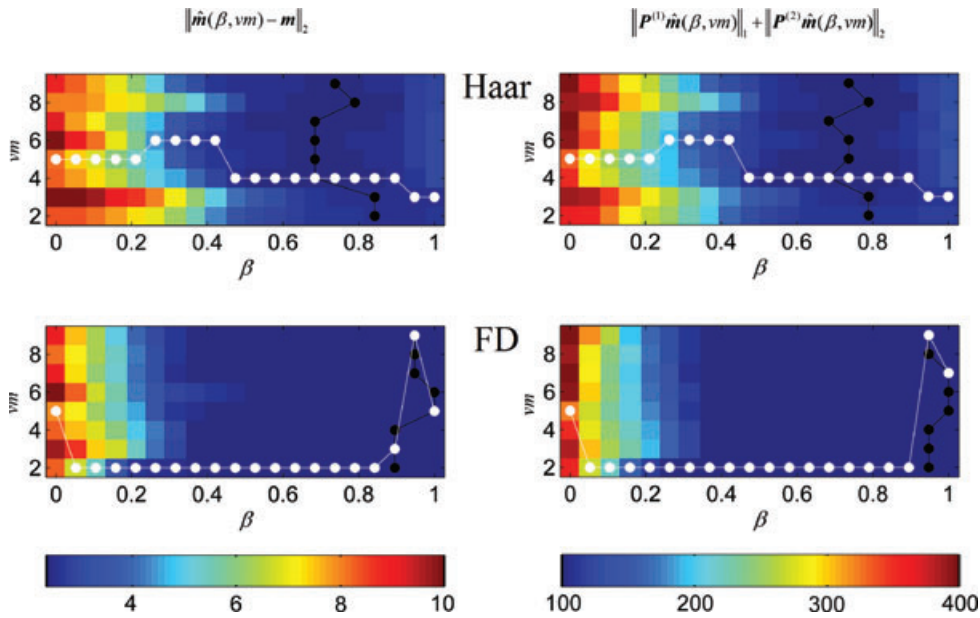




**Figure 4.** Maps of the norm of model error  $\|\hat{m}(\beta, \sigma) - m\|_2$  (left-hand column) and information function  $\|D^{(1)}\hat{m}(\beta, \sigma)\|_1 + \|D^{(2)}\hat{m}(\beta, \sigma)\|_2$  (right-hand column) for the joint-sparsity inversion of zero-offset VSP data using db3 wavelet together with Haar (top row) and FD (bottom row) operators. In the panels, the location of the minimum of each row of the image is indicated by a black circle.



**Figure 5.** (a) The true velocity model and the source–receiver geometry of the synthetic straight-ray cross-well experiment. The inverted models from (b) second-order regularization, (c) sparsity inversion by promoting sparsity in db5 wavelet domain, (d) joint-sparsity inversion using db4 wavelet with Haar operator and (e) joint-sparsity inversion using db2 wavelet with FD operator.



**Figure 6.** Maps of the norm of model error  $\|\hat{\mathbf{m}}(\beta, \mathbf{vm}) - \mathbf{m}\|_2$  (left-hand column) and information function  $\|\mathbf{P}^{(1)}\hat{\mathbf{m}}(\beta, \mathbf{vm})\|_1 + \|\mathbf{P}^{(2)}\hat{\mathbf{m}}(\beta, \mathbf{vm})\|_2$  (right-hand column) for inverted model from joint-sparsity inversion using Haar (top row) and FD (bottom row) operators. In each panel, the black and white circles indicate the location of the minimum of the corresponding columns and rows.

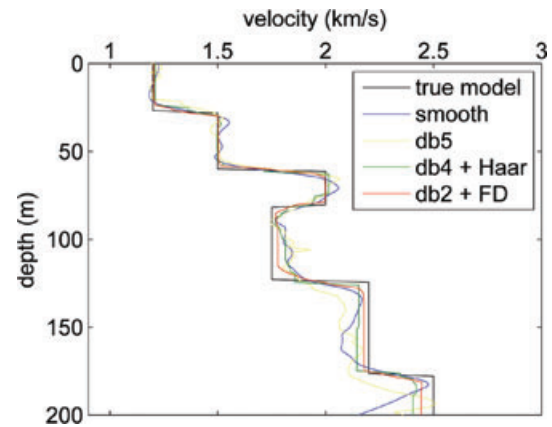
are measured by 32 equally spaced receivers in the neighbouring borehole. In both, the forward operation and the inverse calculation, we assumed straight-rays, which make this problem linear. The exact traveltimes are contaminated by Gaussian distributed random noise of zero mean and standard deviation  $\sigma = 1$  and other uncertainties are ignored assuming a perfect modelling. To invert the traveltimes, the medium was discretized into  $128 \times 128$  grids of equal size and unknown  $P$ -wave velocities.

Fig. 6 shows the maps of norm of model error  $\|\hat{\mathbf{m}}(\beta, \mathbf{vm}) - \mathbf{m}\|_2$  and information function  $\|\mathbf{P}^{(1)}\hat{\mathbf{m}}(\beta, \mathbf{vm})\|_1 + \|\mathbf{P}^{(2)}\hat{\mathbf{m}}(\beta, \mathbf{vm})\|_2$ , where  $\mathbf{P}^{(i)} = \begin{bmatrix} \mathbf{D}_1^{(i)} \\ \mathbf{D}_2^{(i)} \end{bmatrix}$ , for recovered model from joint-sparsity inversion using different choices of compactly supported Daubechies wavelet functions and  $0 \leq \beta \leq 1$  with small increments of  $\beta$ . The results for Haar and FD operators are shown in top and bottom rows of Fig. 6, respectively. Discussions about the trends of black and white circles are similar to those in Fig. 1. For the model under consideration it is important to note that:

- (1) The trends of black and white circles in the maps of  $\|\hat{\mathbf{m}}(\beta, \mathbf{vm}) - \mathbf{m}\|_2$  and  $\|\mathbf{P}^{(1)}\hat{\mathbf{m}}(\beta, \mathbf{vm})\|_1 + \|\mathbf{P}^{(2)}\hat{\mathbf{m}}(\beta, \mathbf{vm})\|_2$  are very similar.
- (2) Haar operator performed well in the combination with db4 wavelet, but db2 wavelet was better combined with FD operator.
- (3) Since the original model was blocky, in both cases the roughness penalty was weighted more than the wavelet operator (optimum  $\beta$  is larger than 0.7).

For most of the wavelet functions in combination with the Haar operator the optimum  $\beta$  value ranged from 0.7 to 0.8, and in the case of FD operator it was larger than 0.8. Due to the flat behaviour of the  $\|\hat{\mathbf{m}}(\beta, \mathbf{vm}) - \mathbf{m}\|_2$  and  $\|\mathbf{P}^{(1)}\hat{\mathbf{m}}(\beta, \mathbf{vm})\|_1 + \|\mathbf{P}^{(2)}\hat{\mathbf{m}}(\beta, \mathbf{vm})\|_2$  at the vicinity of optimum  $\beta$ , it was not necessary to test the  $\beta$  values with smaller increment, so we selected  $\beta = 0.75$  for Haar operator and  $\beta = 0.9$  for FD operator.

We again compare the results of joint-sparsity inversion with those of smooth and common sparsity inversions. The recovered ve-



**Figure 7.** Plot of the velocity models shown in Fig. 5 along the main diagonal. The black, blue, yellow, green and red curves correspond to models in Figs 5(a) to (e), respectively.

locity model from second-order regularization is shown in Fig. 5(b). This image exhibits the expected velocity variations, but with a poor resolution (i.e. the sharp discontinuities are blurred). Fig. 5(c) represents the velocity image obtained from the sparsity inversion by promoting sparsity in the db5 wavelet domain. We chose this particular wavelet because in the absent of roughness penalty ( $\beta = 0$ ) indicated the best performance (see Fig. 6). Comparing to Fig. 5(b), the reconstructed image is less smooth but because of the pseudo-Gibbs phenomenon has a poor resolution. Fig. 5(d) indicates the reconstruction from joint-sparsity inversion using db4 wavelet in the combination with Haar system. The inverted model from joint-sparsity inversion using db2 wavelet in the combination with FD operator is represented in Fig. 5(e). For comparison, velocity models recovered with different inverse methods along the main diagonal are depicted in Fig. 7. Regarding Figs 5 and 7, the joint-sparsity inversion greatly improved the quality of the reconstructions. It suppressed pseudo-Gibbs phenomenon while preserving the sharp



discontinuities compared to the conventional smooth and sparsity inversions.

## NON-LINEAR INVERSE PROBLEM

In this section we deal with the solution of discretized non-linear ill-posed inverse problems generally defined as

$$\mathbf{d} = \mathbf{T}(\mathbf{m}) + \mathbf{e}, \quad (10)$$

where  $\mathbf{d}$ ,  $\mathbf{m}$  and  $\mathbf{e}$  are defined as in eq. (1) and the discrete transformation  $\mathbf{T} : \mathbb{R}^M \rightarrow \mathbb{R}^N$  is the forward operator. Because of the non-linearity of the forward operator, eq. (10) can be solved iteratively (see e.g. Constable *et al.* 1987; Farquharson & Oldenburg 1998; Ramlau & Teschke 2005, 2006; Teschke & Ramlau 2007). We solved it using traditional linearized iterative approach and modified the cost function defined in eq. (2) for discretized non-linear problems as follows:

$$\begin{cases} \hat{\mathbf{m}}_n = \underset{\mathbf{m}_n}{\operatorname{argmin}} f(\mathbf{m}_n), \\ f(\mathbf{m}_n) = \frac{1}{2} \|\mathbf{d} - \mathbf{T}(\mathbf{m}_n)\|_2^2 + \alpha_n \varphi(\mathbf{m}_n), \end{cases} \quad (11)$$

where  $\hat{\mathbf{m}}_n$  is the recovered solution at  $n$ th iteration and  $n = 0, 1, 2, \dots$ . It is assumed that in the case of convergence, such iterative procedure solves the minimization of the original cost function and therefore the solution is independent of the starting model,  $\mathbf{m}_0$ . To solve eq. (11) we followed the standard linearized approach (Constable *et al.* 1987). Assuming  $\mathbf{T}(\mathbf{m})$  is differentiable at  $\mathbf{m}_n$ ,  $\mathbf{T}(\mathbf{m}_n)$  can be approximated with the first two terms in a Taylor's series expansion as

$$\mathbf{T}(\mathbf{m}_n) \approx \mathbf{T}(\mathbf{m}_{n-1}) + \mathbf{J}[\mathbf{m}_n - \mathbf{m}_{n-1}], \quad (12)$$

where  $\mathbf{J} = \mathbf{J}(\mathbf{m}_{n-1})$  is the Jacobian matrix calculated from the model at  $(n - 1)$ th iteration and is defined as  $J_{ij} = \frac{\partial T(\mathbf{m}_{n-1})_i}{\partial m_j}$ . Considering above, our problem becomes

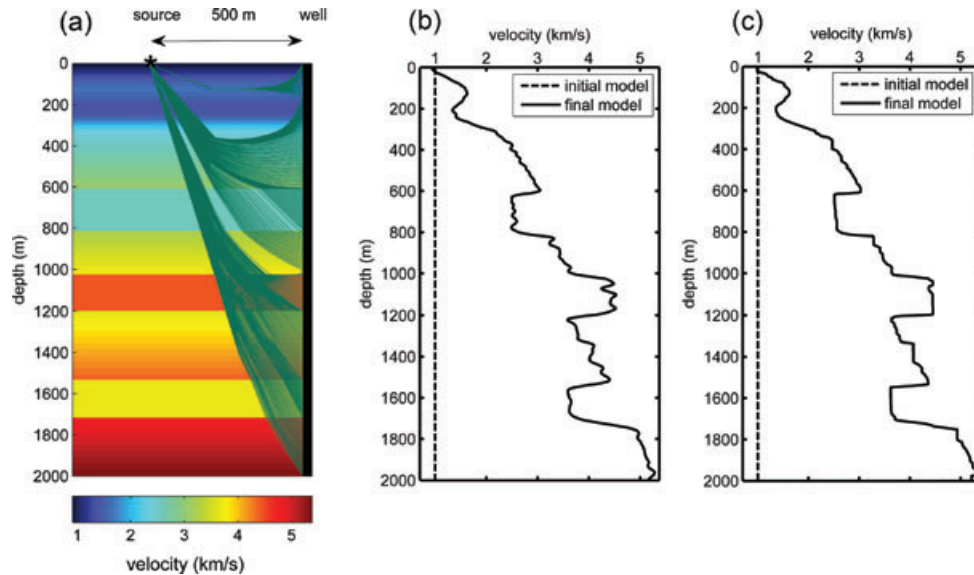
$$\begin{cases} \hat{\mathbf{m}}_n = \underset{\mathbf{m}_n}{\operatorname{argmin}} f(\mathbf{m}_n), \\ f(\mathbf{m}_n) = \frac{1}{2} \|\bar{\mathbf{d}} - \mathbf{J}\mathbf{m}_n\|_2^2 + \alpha_n \varphi(\mathbf{m}_n), \end{cases} \quad (13)$$

where  $\bar{\mathbf{d}} = \mathbf{d} + \mathbf{J}\mathbf{m}_{n-1} - \mathbf{T}(\mathbf{m}_{n-1})$ . Problem (13) is linear and can be solved for  $\hat{\mathbf{m}}_n$  using the algorithms for linear inverse problem. Here the main task which has to be considered is the proper selection of the regularization parameter  $\alpha_n$  at each iteration. Note that we require the model resulting from  $n$ th iteration,  $\hat{\mathbf{m}}_n$ , which also is the minimizer of the cost function (11), to satisfy the desired misfit condition (note also that, here, the misfit depends on the iteration). To do that, we allow  $\alpha_n$  to vary slowly at successive iterations according to  $\alpha_n = \max(\kappa\alpha_{n-1}, \alpha')$  with a sufficiently large starting value at the first iteration. Where  $\alpha'$  is the minimizer of the misfit for the current iteration and can be determined by a 1-D search algorithm. The positive constant  $\kappa$  determines the maximum allowable changes of  $\alpha_n$  at successive iterations and our algorithm uses  $\kappa = 0.5$ . We run the algorithm until the desired misfit condition is reached.

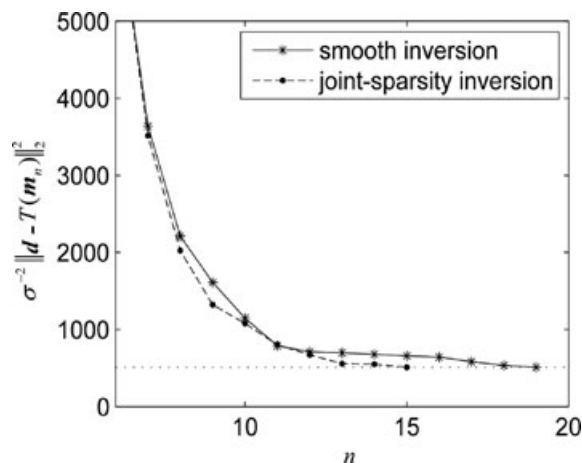
## APPLICATION TO NON-LINEAR EXPERIMENTS DATA

This section is devoted to show the performance of the joint-sparsity inversion for solving non-linear inverse problems. Here, we considered an offset VSP traveltimes tomography and assumed the ray trajectories be sensitive to velocity gradients. To invert the data using joint-sparsity inversion, we used the wavelet function and  $\beta$  as those used for the zero-offset VSP experiment and the problem was solved only for FD operator (the procedure for Haar system is the same). The forward modelling for traveltimes calculations and construction of Jacobian matrices was done using the First Arrival Seismic Tomography (FAST) code described in Zelt & Barton (1998).

The velocity model with source–receiver geometry and ray coverage for an offset VSP experiment is shown in Fig. 8(a). The velocity is constant in the horizontal direction and changes in the vertical direction as in the velocity model used for zero-offset VSP experiment (Fig. 3a). For each geophone, the corresponding traveltimes is the integration of the slowness model along the ray path from a source located 500 m away from the well head to the location of



**Figure 8.** (a) The true velocity model and source–receiver geometry of the synthetic offset VSP experiment. Where the rays are depicted for every other receiver. (b) Smooth inversion output after 19 iterations using second-order Tikhonov regularization. (c) The output of joint-sparsity inversion scheme after 15 iterations. The algorithms were started with a constant velocity ( $1 \text{ km s}^{-1}$ ) initial model (dashed lines).



**Figure 9.** Variation of the data misfit with respect to iteration number for non-linear inversion with smoothing constraint (solid line) and joint-sparsity constraints scheme (dashed line), of the synthetic offset VSP traveltime data set. The horizontal dotted line (value of 512) shows the number of observations for this experiment.

geophone. Other parameters including geophone number and spacing, well depth, the number of layers, and noise level are the same as those in zero-offset VSP experiment.

We used both smooth and joint-sparsity inversion schemes to invert the generated offset VSP traveltime data set. For both methods we started the algorithm with a constant velocity ( $1 \text{ km s}^{-1}$ ) initial model. The final solutions from smooth and joint-sparsity inversions are shown in Figs 8(b) and (c), respectively. As seen from Fig. 8, smooth inversion recovers the smooth parts of the model well, but fails to recover the sharp discontinuities. In contrast, joint-sparsity inversion recovers both the smooth and rough parts of the model appropriately. The variation of the data misfit for both inversion methods with respect to iteration are shown in Fig. 9.

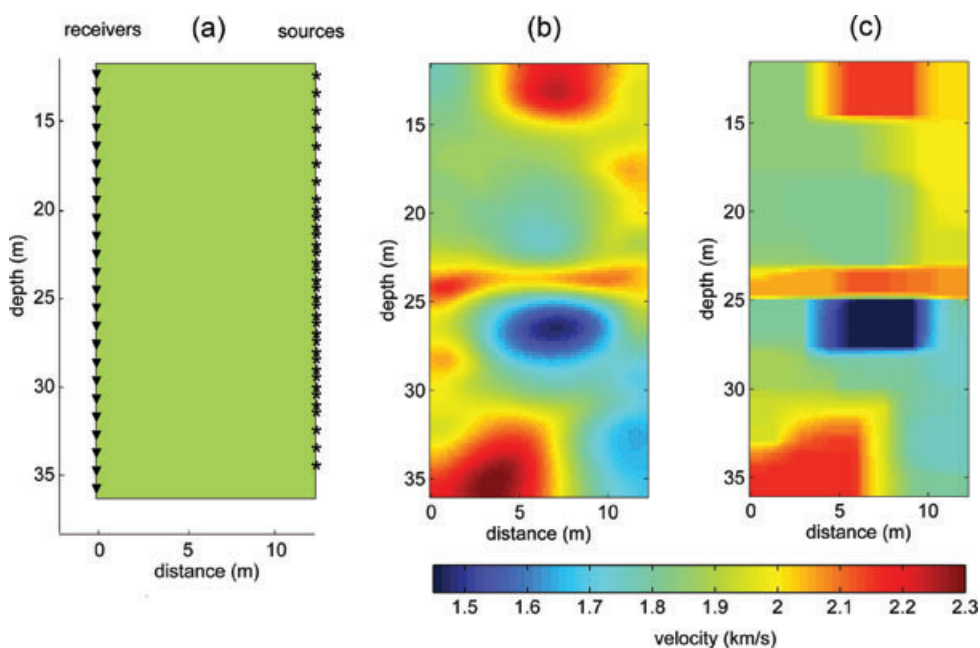
## APPLICATION TO FIELD DATA

As a final illustration of the solution of a non-linear inverse problem using proposed joint-sparsity inversion scheme, we present the results of an inversion of field cross-well data set made available by International Geophysical Technology (IGT). According to the report of IGT, the survey was conducted in the Madrid-Barajas airport to locate some old mining galleries. To perform the inversion, we selected a region of the survey including a total of 35 source positions and 24 receiver positions (Fig. 10a). The target region was discretized into  $128 \times 64$  square grids of unknown  $P$ -waves velocities.

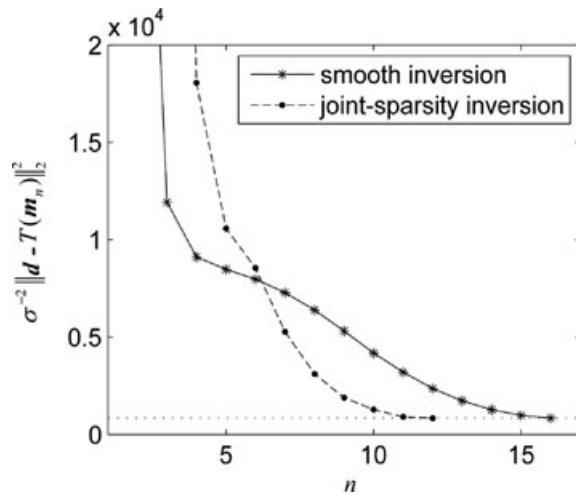
The picked traveltime data were inverted using both smooth and joint-sparsity inversion schemes for the  $P$ -waves velocity field. The inversions were started with a constant velocity ( $1 \text{ km s}^{-1}$ ) initial model. The tomogram in Fig. 10(b) is obtained from first-order regularization. Although we see the indication of some velocity anomalies, but due to smoothing operator the bodies are not sharply defined. Fig. 10(c) indicates the inverted model from joint-sparsity inversion. As expected, the image exhibits the velocity variations at a high resolution. The recovered models in Figs 10(b) and (c) indicate a low velocity ( $\sim 1500 \text{ m s}^{-1}$ ) anomaly at depths between 25 and 28 m. The observed velocity anomaly caused by the presence of water within the one of the survey objective galleries which was later found through a drilling. The variation of the data misfit for successive iterations of the inversion is shown in Fig. 11.

## CONCLUSION

In this paper, we developed a new regularization scheme for a high-resolution solution of linear and non-linear inverse problems which benefits from the advantages of two different sparsifying operators in representation of the desired model. The results of numerical examples from linear and non-linear seismic traveltime tomography demonstrated that: (1) Smooth inversion recovers smooth local structures while blurring the sharp discontinuities in the solution.



**Figure 10.** (a) The source–receiver geometry of the field cross-well tomography survey. (b) Smooth inversion output after 16 iterations using first-order Tikhonov regularization. (c) The output of joint-sparsity inversion scheme after 12 iterations. The algorithms were started with a constant velocity ( $1 \text{ km s}^{-1}$ ) initial model.



**Figure 11.** Variation of the data misfit with respect to iteration number for non-linear inversion with smoothing constraint (solid line) and joint-sparsity constraints scheme (dashed line), of the field cross-well tomography data set. The horizontal dotted line (value of 840) shows the number of observations for this survey.

(2) Orthonormal wavelet-based sparsity inversion provides a solution having the pseudo-Gibbs phenomenon and being smoothed around discontinuities. (3) Joint-sparsity inversion using a combination of wavelets with Haar or FD operator captures local structures with different smoothness in the model parameters and recovers smooth/constant pieces of the solution together with sharp discontinuities.

We have found that the smooth and rough parts of a model can be reconstructed accurately when a mixture of wavelets and FD/Haar are used. Also indicated that, an optimum wavelet function and regularization parameter  $\beta$  can be chosen appropriately using the information function derived from a measure of smoothness and roughness of the recovered solution. Our method can be extended easily to multiple sparsity constraints and can be used for a variety of geophysical inverse problems with three-dimensions.

## ACKNOWLEDGMENTS

The authors would like to thank the GJI editor and the anonymous reviewers for their valuable comments and suggestions. We also thank Prof. C. Zelt and the authors of Wavelab for making their codes publically available. The International Geophysical Technology (IGT) for providing the field cross-well data is appreciated.

## REFERENCES

- Ajo-Franklin, J.B., Minsley, B.J. & Daley, T.M., 2007. Applying compactness constraints to differential traveltime tomography, *Geophysics*, **72**(4), R67–R75.
- Bertete-Aguirre, H., Cherkav, E. & Oristaglio, M., 2002. Non-smooth gravity problem with total variation penalization functional, *Geophys. J. Int.*, **149**, 499–507.
- Bregman, L., 1967. The relaxation method of finding the common points of convex sets and its application to the solution of problems in convex optimization, *USSR Comput. Math. Math. Phys.*, **7**, 200–217.
- Candes, E.J. & Guo, F., 2002. New multiscale transforms, minimum total variation synthesis: applications to edge-preserving image reconstruction, *Signal Process*, **82**, 1519–1543.

- Chambolle, A., DeVore, R.A., Lee, N.Y. & Lucier, B.J., 1998. Nonlinear wavelet image processing: variational problems, compression, and noise removal through wavelet shrinkage, *IEEE Trans. Image Process*, **7**, 319–335.
- Charbonnier, P., Blanc-Feraud, L., Aubert, G. & Barlaud, M., 1997. Deterministic edge-preserving regularization in computed imaging, *IEEE Trans. Image Process*, **6**, 298–311.
- Constable, S.C., Parker, R.L. & Constable, C.G., 1987. Occam's inversion: a practical algorithm for generating smooth models from electromagnetic sounding data, *Geophysics*, **52**, 289–300.
- Daubechies, I., 1992. *Ten Lectures on Wavelets*, SIAM, Philadelphia.
- Daubechies, I., Defriese, M. & De Mol, C., 2004. An iterative thresholding algorithm for linear inverse problems with a sparsity constraint, *Commun. Pure appl. Math.*, **LVII**, 1413–1457.
- Donoho, D.L., 1995a. Nonlinear solution of linear inverse problems by wavelet-vaguelette decomposition, *Appl. Comput. Harmon. Anal.*, **2**, 101–126.
- Donoho, D.L., 1995b. De-noising by soft thresholding, *IEEE, Trans. Inf. Theory*, **41**, 613–627.
- Donoho, D.L. & Johnstone, I.M., 1998. Minimax estimation via wavelet shrinkage, *Ann. Statist.*, **26**, 879–921, as technical report, Department of Stat., Stanford University, 1992.
- Durand, S. & Froment, J., 2001. Artifact free signal denoising with wavelets, in *Proceedings of the IEEE Int. Conf. on Acoust., Speech, & Signal Proc.*, May 7–11, Salt Lake City, Utah.
- Farquharson, C.G. & Oldenburg, D.W., 1998. Non-linear inversion using general measures of data misfit and model structure, *Geophys. J. Int.*, **134**, 213–227.
- Gholami, A. & Siahkoohi, H.R., 2009a. A two-step wavelet-based regularization for linear inversion of geophysical data, *Geophys. Prospect.*, **57**, 847–862.
- Gholami, A. & Siahkoohi, H.R., 2009b. Simultaneous constraining of model and data smoothness for regularization of geophysical inverse problems, *Geophys. J. Int.*, **176**, 151–163.
- Goldstein, T. & Osher, S., 2008. The Split Bregman method for L1 regularized problems, UCLA CAAM Report 08–29.
- Hansen, P. & O'Leary, D., 1993. The use of the L-curve in the regularization of discrete ill-posed problems, *SIAM J. Scient. Comput.*, **14**, 1487–1503.
- Herrmann, F.J., Moghaddam, P.P. & Stolk, C.C., 2008a. Sparsity- and continuity-promoting seismic image recovery with curvelet frames, *Appl. Comput. Harmonic Anal.*, **24**, 150–173.
- Herrmann, F.J., Wang, W., Hennenfent, G. & Moghaddam, P.P., 2008b. Curvelet-based seismic data processing: a multiscale and nonlinear approach, *Geophysics*, **73**, A1.
- Kelley, C.T., 1999. *Iterative Methods for Optimization*, SIAM, Philadelphia.
- Kim, S.J., Koh, K., Lustig, M., Boyd, S. & Gorinevsky, D., 2007. An interior-point method for large-scale  $l_1$ -regularized least squares, *IEEE Trans. Sel. Topics Signal Proc.*, **1**, 606–617.
- Klann, E., Maass, P. & Ramlau, R., 2006. Two-step regularization methods for linear inverse problems, *J. Inverse Ill-Posed Problems*, **14**, 583–607.
- Li, X., Sacchi, M.D. & Ulrych, T.J., 1996. Wavelet transform inversion with a priori scale information, *Geophysics*, **61**, 1379–1385.
- Loris, I., Nolet, G., Daubechies, I. & Dahlen, F.A., 2007. Tomographic inversion using  $l_1$ -norm regularization of wavelet coefficients, *Geophys. J. Int.*, **170**, 359–370.
- Lustig, M., Donoho, D.L. & Pauly, J.M., 2007. Sparse MRI: the application of compressed sensing for rapid MR imaging, *Magn. Resonan. Med.*, **58**, 1182–1195.
- Mallat, S.G., 1998. *A Wavelet Tour of Signal Processing*, Academic Press, San Diego.
- Meyer, C.D., 2000. *Matrix Analysis and Applied Linear Algebra*, SIAM, Philadelphia.
- Osher, S., Burger, M., Goldfarb, D., Xu, J. & Yin, W., 2005. An iterative regularization method for total variation-based image restoration. *Multiscale Model. Simulat.*, **4**, 460–489.
- Ramlau, R. & Teschke, G., 2005. Tikhonov replacement functionals for iteratively solving nonlinear operator equations. *Inverse Problems*, **21**, 1571–1592.

Ramlau, R. & Teschke, G., 2006. A projection iteration for nonlinear operator equations with sparsity constraints, *Num. Math.*, **104**, 177–203.

Routh, P.S., Qu, L., Sen, M.K. & Anno, P.D., 2007. Inversion for non-smooth models with physical bounds, *SEG, Expanded Abstracts*, **26**(1), 1795–1799.

Sacchi, M.D. & Ulrych, T.J., 1995. High resolution velocity gathers and offset space reconstruction, *Geophysics*, **60**, 1169–1177.

Starck, J.L., Donoho, D.L. & Candes, E., 2001. Very high quality image restoration by combining wavelets and curvelets, in *Signal and Image Processing: Wavelet Applications in Signal and Image Processing IX*, SPIE Vol. 4478, eds Laine, A., Unser, M.A. & Aldroubi, A., SPIE, San Diego, CA.

Tarantola, A., 2005. *Inverse Problem Theory and Methods for Model Parameter Estimation*, SIAM, Philadelphia.

Teschke, G. & Ramlau, R., 2007. An iterative algorithm for nonlinear inverse problems with joint sparsity constraints in vector valued regimes and an application to color image inpainting, *Inverse Problems*, **23**, 1851–1870.

Tikhonov, A.N. & Arsenin, V.Y., 1977. *Solutions of Ill-posed Problems*, V.H. Winston and Sons, Washington, DC.

van Wijk, K., Scales, J.A., Navidi, W. & Tenorio, L., 2002. Data and model uncertainty estimation for linear inversion, *Geophys. J. Int.*, **149**, 625–632.

Wahba, G., 1990. Spline models for observational data, in *CBMS-NSF Regional Conference Series in Applied Mathematics*, Vol. 59, SIAM, Philadelphia, PA.

Wang, Y., 2003. Sparseness-constrained least-squares inversion: application to seismic wave reconstruction, *Geophysics*, **68**, 1633–1638.

Wang, Y., Yin, W. & Zhang, Y., 2007. A fast algorithm for image deblurring with total variation regularization, CAAM Technical Reports TR07-10.

Youzwhisen, C.F. & Sacchi, M.D., 2006. Edge preserving imaging, *J. Seism. Explor.*, **15**, 45–58.

Zelt, C.A. & Barton, P.J., 1998. 3D seismic refraction tomography: a comparison of two methods applied to data from the Faeroe Basin, *J. geophys. Res.*, **103**, 7187–7210.

## APPENDIX A

The regularized model is the minimizer of the cost function in eq. (2) with regularizer function given by eq. (3)

$$\arg \min_{\mathbf{m}} \left\{ \frac{1}{2} \|\mathbf{d} - \mathbf{G}\mathbf{m}\|_2^2 + \alpha(1 - \beta) \sum_i |[\mathbf{S}_1\mathbf{m}]_i| + \alpha\beta \sum_i |[\mathbf{S}_2\mathbf{m}]_i| \right\}. \quad (\text{A1})$$

Replacing  $\mathbf{S}_1\mathbf{m}$  by  $\mathbf{p}_1$  and  $\mathbf{S}_2\mathbf{m}$  by  $\mathbf{p}_2$  in eq. (A1), results to the constrained problem

$$\arg \min_{\mathbf{m}} \left\{ \frac{1}{2} \|\mathbf{d} - \mathbf{G}\mathbf{m}\|_2^2 + \alpha(1 - \beta) \sum_i |[\mathbf{p}_1]_i| + \alpha\beta \sum_i |[\mathbf{p}_2]_i| \right\} \quad \text{s.t.} \quad \begin{cases} \mathbf{p}_1 = \mathbf{S}_1\mathbf{m}, \\ \mathbf{p}_2 = \mathbf{S}_2\mathbf{m}. \end{cases} \quad (\text{A2})$$

Converting eq. (A2) into an unconstrained problem, yields

$$\mathbf{m}, \mathbf{p}_1, \mathbf{p}_2 = \arg \min_{\mathbf{m}, \mathbf{p}_1, \mathbf{p}_2} \left\{ \frac{1}{2} \|\mathbf{d} - \mathbf{G}\mathbf{m}\|_2^2 + \alpha(1 - \beta) \sum_i |[\mathbf{p}_1]_i| + \alpha\beta \sum_i |[\mathbf{p}_2]_i| + \frac{\gamma}{2} \|\mathbf{p}_1 - \mathbf{S}_1\mathbf{m}\|_2^2 + \frac{\lambda}{2} \|\mathbf{p}_2 - \mathbf{S}_2\mathbf{m}\|_2^2 \right\}, \quad (\text{A3})$$

the constraints can be enforced strictly by applying the Bregman iteration (Goldstein & Osher 2008), to get a three step iterative algorithm

$$\begin{aligned} \text{step 1 : } \mathbf{m}^{k+1}, \mathbf{p}_1^{k+1}, \mathbf{p}_2^{k+1} &= \arg \min_{\mathbf{m}, \mathbf{p}_1, \mathbf{p}_2} \left\{ \frac{1}{2} \|\mathbf{d} - \mathbf{G}\mathbf{m}\|_2^2 + \alpha(1 - \beta) \sum_i |[\mathbf{p}_1]_i| + \alpha\beta \sum_i |[\mathbf{p}_2]_i| \right. \\ &\quad \left. + \frac{\gamma}{2} \|\mathbf{p}_1 - \mathbf{S}_1\mathbf{m} - \mathbf{q}_1^k\|_2^2 + \frac{\lambda}{2} \|\mathbf{p}_2 - \mathbf{S}_2\mathbf{m} - \mathbf{q}_2^k\|_2^2 \right\} \\ \text{step 2 : } \mathbf{q}_1^{k+1} &= \mathbf{q}_1^k - (\mathbf{p}_1^{k+1} - \mathbf{S}_1\mathbf{m}^{k+1}) \\ \text{step 3 : } \mathbf{q}_2^{k+1} &= \mathbf{q}_2^k - (\mathbf{p}_2^{k+1} - \mathbf{S}_2\mathbf{m}^{k+1}). \end{aligned} \quad (\text{A4})$$

It is shown that step 1 can be performed efficiently by only one iteration of minimization with respect to  $\mathbf{m}$ ,  $\mathbf{p}_1$ , and  $\mathbf{p}_2$  separately. doing so, results to the desired five step algorithm

$$\begin{aligned} \text{step 1 : } \mathbf{m}^{k+1} &= \arg \min_{\mathbf{m}} \left\{ \frac{1}{2} \|\mathbf{d} - \mathbf{G}\mathbf{m}\|_2^2 + \frac{\gamma}{2} \|\mathbf{p}_1^k - \mathbf{S}_1\mathbf{m} - \mathbf{q}_1^k\|_2^2 + \frac{\lambda}{2} \|\mathbf{p}_2^k - \mathbf{S}_2\mathbf{m} - \mathbf{q}_2^k\|_2^2 \right\} \\ \text{step 2 : } \mathbf{p}_1^{k+1} &= \arg \min_{\mathbf{p}} \left\{ \alpha(1 - \beta) \sum_i |p_i| + \frac{\gamma}{2} \|\mathbf{p} - \mathbf{S}_1\mathbf{m}^{k+1} - \mathbf{q}_1^k\|_2^2 \right\} \\ \text{step 3 : } \mathbf{p}_2^{k+1} &= \arg \min_{\mathbf{p}} \left\{ \alpha\beta \sum_i |p_i| + \frac{\lambda}{2} \|\mathbf{p} - \mathbf{S}_2\mathbf{m}^{k+1} - \mathbf{q}_2^k\|_2^2 \right\} \\ \text{step 4 : } \mathbf{q}_1^{k+1} &= \mathbf{q}_1^k - (\mathbf{p}_1^{k+1} - \mathbf{S}_1\mathbf{m}^{k+1}) \\ \text{step 5 : } \mathbf{q}_2^{k+1} &= \mathbf{q}_2^k - (\mathbf{p}_2^{k+1} - \mathbf{S}_2\mathbf{m}^{k+1}). \end{aligned} \quad (\text{A5})$$

Now, step 1 is a quadratic optimization problem, minimization of it yields eqs (4) and (5). Steps 2 and 3 are l1-norm denoising problems and the exact minimizers of them is found component-wise by eq. (6) with thresholding parameters  $\tau_1 = \gamma^{-1}\alpha(1 - \beta)$  and  $\tau_2 = \lambda^{-1}\alpha\beta$ , respectively (see Chambolle *et al.* 1998). Summarizing this steps leads to the algorithm in Table 1.