Transform-domain sparsity regularization for inverse problems in geosciences

GEOPHYSICS, VOL. 74, NO. 5 (SEPTEMBER-OCTOBER 2009); P. R69–R83, 13 FIGS.
10.1190/1.3157250

Transform-domain sparsity regularization for inverse problems in geosciences

Behnam Jafarpour[1], Vivek K. Goyal[2], Dennis B. McLaughlin[3], and William T. Freeman[2]

ABSTRACT

We have developed a new regularization approach for estimating unknown spatial fields, such as facies distributions or porosity maps. The proposed approach is especially efficient for fields that have a sparse representation when transformed into a complementary function space (e.g., a Fourier space). Sparse transform representations provide an accurate characterization of the original field with a relatively small number of transformed variables. We use a discrete cosine transform (DCT) to obtain sparse representations of fields with distinct geologic features, such as channels or geologic formations in vertical cross section. Low-frequency DCT basis elements provide an effectively reduced subspace in which the sparse solution is searched. The low-dimensional subspace is not fixed, but rather adapts to the data. The DCT coefficients are estimated from spatial observations with a variant of compressed sensing. The estimation procedure minimizes an l_2-norm measurement misfit term while maintaining DCT coefficient sparsity with an l_1-norm regularization term. When measurements are noise-dominated, the performance of this procedure might be improved by implementing it in two steps — one that identifies the sparse subset of important transform coefficients and one that adjusts the coefficients to give a best fit to measurements. We have proved the effectiveness of this approach for facies reconstruction from both scattered-point measurements and areal observations, for crosswell traveltime tomography, and for porosity estimation in a typical multiunit oil field. Where we have tested our sparsity regularization approach, it has performed better than traditional alternatives.

INTRODUCTION

The problem of estimating patterns and structures is encountered in many engineering and science applications, from earth sciences to computer vision and medical imaging. These problems are often ill-posed (having more unknowns than measurements), which can result in nonunique solutions. For example, the problem of estimating subsurface structures such as channels and faults from limited point observations is severely ill-posed when these structures are described in terms of many independent pixel values. In such situations, structural assumptions about the solutions are usually built into the solution algorithm to favor solutions in close proximity to a prior description of the unknown structure. However, it is often difficult to specify the spatial patterns a priori, especially when they are poorly defined and have irregular geometry. In such cases, it is better to introduce implicit structural information through constraints on smoothness or sparsity to produce solutions that naturally reveal patterns present in the data.

A common approach for constraining the solutions of ill-posed inverse problems is through regularization (Tikhonov and Arsenin, 1977). As stated by Constable et al. (1987), regularization serves two primary purposes: to stabilize the solution of an ill-posed inverse problem and to give a solution that fits measurements adequately without producing artifacts that are unjustifiably complex. Tikhonov regularization methods try to achieve these objectives by modifying the inverse problem objective function so a sum of two terms is minimized: a data misfit term and a regularization term that penalizes low-order spatial derivatives of the solution. This results in spatially-smooth solutions that are consistent with the observed data. Because of their simplicity and desirable smoothness properties, Tikhonov regularization methods have enjoyed widespread applica-

Manuscript received by the Editor 30 May 2008; revised manuscript received 28 February 2009; published online 1 September 2009.
[1]Texas A&M University, Department of Petroleum Engineering, College Station, Texas, U.S.A. E-mail: behnam@pe.tamu.edu.
[2]Massachusetts Institute of Technology, Department of Electrical Engineering and Computer Science, Cambridge, Massachusetts, U.S.A. E-mail: vgoyal@mit.edu; billf@mit.edu.
[3]Massachusetts Institute of Technology, Department of Civil and Environmental Engineering, Cambridge, Massachusetts, U.S.A. E-mail: dennism@mit.edu.

tion in inverse theory, particularly in the geosciences.

Although smoothness is a desirable property in many applications, the choice of a regularization technique ultimately should take into account the underlying physics that has generated the observed quantities. Consequently, the form of the regularization constraint should be consistent with and promote the expected properties of the physical system. For instance, compactness measures are preferred to smoothness when working with models that are believed to have sharp local features (Last and Kubik, 1983; Portniaguine and Zhdanov, 1999; Ajo-Franklin et al., 2007).

Another approach to regularization is to use compression techniques to reduce the number of unknowns in the inverse problem. In spatially-discretized models, these techniques typically transform the original finite set of problem variables into another set of variables that provide an equivalent description of the original spatial field. If the transformation is appropriately chosen, a relatively small subset of the new variables can provide a very good approximation. In such cases, the remaining transformed variables can be neglected in the inversion procedure. This truncation approximation yields a new inverse problem that has fewer unknowns and is better posed. Furthermore, the solutions provided often are better able to capture spatial patterns of particular interest in many applications, including geophysics.

One challenge of a compression approach to regularization is the need to identify which transformed variables should be retained and which should be neglected. This problem can be addressed from the perspective of sparse reconstruction if a substantial number of the transformed variables can be assumed to be small enough to be approximated as zeros. In this case, the transformed representation is said to be sparse because only a few of the transformed variables are nonzero. Then the inverse problem reduces to the identification of the nonzero transformed variable values that give the best fit to observations. Consequently, regularization is achieved through transformation and truncation of the problem variables rather than through incorporation of smoothness terms in the performance function.

Minimization of the l_1-norm of the low-order spatial derivatives of the model (Claerbout and Muir, 1979; Bube and Langan, 1997) is one of several regularization techniques that have been proposed for stabilizing inverse problems. An important property of the l_1-norm minimization, compared to its l_2-norm counterpart, is its smaller penalty for large outliers and larger penalty for small values. These properties lend themselves to better detection of piecewise continuous features with sparse spatial derivatives. A particular version of l_1-norm regularization techniques is known as total variation, which has been proposed for edge detection problems and reconstruction of piecewise smooth properties (Rudin et al., 1992; Acar and Vogal, 1994; Yu and Dougherty, 2000). Although the use of l_1-norm minimization is not new, the use of the l_1-norm as a proxy for sparsity has only been formalized recently under the compressed sensing paradigm, which also provides certain theoretical guarantees on reconstruction quality (Candès and Tao, 2006; Donoho, 2006a). Lin and Herrmann (2007) give a recent geophysical application of this approach.

Here we consider l_1-norm minimization for subsurface inverse problems. A key assumption is that either the original model is sparse or an appropriately transformed version of the model has a sparse approximation. For purposes of this paper, we presume that techniques inspired by image compression can be used to generate transformed inverse problems likely to have sparse representations. Choice of the sparsifying transform is application specific and usually requires information about the expected structure of the underlying field. Here we consider the discrete cosine transform (DCT), which provides sparse approximation for correlated geological patterns (Jafarpour and McLaughlin, 2009). Appendix A describes the discrete cosine transform (DCT).

We will also describe the methodology and problem formulation, summarize results for relevant examples, and discuss our conclusions. The methodology section includes an overview of compressed sensing to introduce our transform-domain, sparsity-based approach to regularization and inversion. Our results include examples that illustrate the effectiveness of l_1-norm minimization for identifying sparsity in the DCT domain. In these examples, the inverse solution is derived from either scattered observations or areal measurements. We conclude with a summary of the results and a consideration of their implications.

METHODOLOGY AND PROBLEM FORMULATION

Here (and in Appendix A), we review the basic concepts of linear transforms and show how they yield sparse representations of spatial fields. We also describe compressed-sensing and basis-pursuit methods, which exploit sparsity to provide accurate regularized estimates. Then we propose a two-step version of the basis-pursuit method that successfully addresses problems that can arise with noisy measurements. The sparsity assumption is key to the new formulation and stems from the ubiquitous role of spatially continuous features in the earth sciences. In particular, the geologic continuity of subsurface channels in the spatial domain translates into a nearly sparse representation in the transformed domain. The sparseness of the channel-facies characterization problem fits well with our approach.

Image compression and linear transforms

The use of truncated transform methods to solve inverse problems can be viewed as a type of low-rank model representation, also known as model parameterization. The basic concept is to replace the original formulation with one that captures the essential features of the problem with a smaller number of unknowns. Several image-compression techniques (Jain, 1989; Gonzalez and Woods, 2002) exploit this mechanism for reducing the number of variables needed to describe natural images. The discrete cosine transform used in this paper is a typical example of a linear transform suitable for image compression (see Appendix A).

Effectiveness of the DCT for compressing natural images is well established in the signal processing and image compression literature (Jain, 1989; Gonzalez and Woods, 2002). As an example, Figure 1a shows sample images of low frequency 2D DCT basis functions that can be used to represent 45 by 45 data on a grid. The basis images are arranged according to their orientation and level of detail in a descending order from upper left to lower right. Figure 1b shows an idealized 45 by 45 porosity field (left) and its corresponding 45 by 45 set of DCT coefficients plotted with a logarithmic scale (right). DCT coefficients in Figure 1b are rotated to be consistent with the basis orientation shown in Figure 1a and for better visualization. We can see from Figure 1b that many DCT coefficients outside the upper left corner have small values and do not contribute much to the image. When these small coefficients are truncated (set equal to zero) and only the largest ($S = 15$) coefficients are retained, the result is the sparse DCT representation in Figure 1c (right). The corresponding

compressed image (left) looks much like the original, even though it is constructed from a much smaller number of variables.

Sparse estimation, compressed sensing, and basis-pursuit methods

Compressed sensing is a recently introduced technique for estimating sparse signals from partial (or undersampled) observations in a complementary, "incoherent" domain. It has attracted researchers' attention in several disciplines, including signal processing and statistics. We give a simple introduction to this approach to illustrate the concept. We present a more general formulation in the next section, although further mathematical details are left to the original publications on this topic (Candès et al., 2006; Candès and Tao, 2006; Donoho, 2006a).

To introduce compressed sensing, we consider the estimation of channel facies in a reservoir model with N grid blocks. Suppose we assemble all the block porosity values in an N-dimensional spatial-domain vector $\mathbf{u}$ and denote the discrete cosine transform of $\mathbf{u}$ by an N-length vector $\mathbf{v}$, so $\mathbf{u} = \mathbf{\Phi v}$ where $\mathbf{\Phi}$ is the $N \times N$ inverse discrete cosine transformation matrix. The transform-domain vector $\mathbf{v}$ is unknown but is expected to be (nearly) sparse.

We assume that the unknown sparse signal $\mathbf{v}$ has sparsity S (i.e., that the vector $\mathbf{v}$ has S nonzero coefficients, which we denote as $\|\mathbf{v}_N\|_0 = S$), and we attempt to estimate $\mathbf{v}$ from noiseless observations of $\mathbf{u}$ assembled in the M-length measurement vector $\bar{\mathbf{u}}$. (Vector $\bar{\mathbf{u}}$ is a subsampled version of $\mathbf{u}$). Thus we seek to solve $\bar{\mathbf{u}} = \bar{\mathbf{\Phi}}\mathbf{v}$, where $\bar{\mathbf{\Phi}}$ is the M-by-N matrix containing the appropriate rows of $\mathbf{\Phi}$, for estimates $\tilde{\mathbf{v}}$. Because $M < N$, this is an underdetermined system of equations with an infinite number of solutions. However, the fact that $\mathbf{v}$ only has a small number of nonzero coefficients can be exploited, often resulting in a unique solution. The l_0-norm minimization problem

$$\min_{\tilde{\mathbf{v}} \in \mathrm{R}^N} \|\tilde{\mathbf{v}}\|_0 \text{ subject to } \bar{\mathbf{u}} = \bar{\mathbf{\Phi}}\tilde{\mathbf{v}} \qquad (1)$$

formalizes the search for solutions with minimum support (the $\mathbf{v}_{N\times 1}$ with the smallest possible number of nonzero elements that satisfy the constraint). The solution of equation 1 will be the sparse signal $\mathbf{v}$ unless there exists an equally sparse or sparser solution that satisfies the measurements equations, which is not likely (Candès and Tao, 2006).

Unfortunately, solving the l_0-norm problem in equation 1 essentially requires a combinatorial search, with very high computational complexity, over sparsity patterns (known as a nondeterministic polynomial-time hard problem (Natarajan, 1995). A computationally manageable alternative to equation 1 is the basis-pursuit problem, which uses an l_1-norm minimization (Chen et al., 2001; Donoho, 2006a) and is expressed as

$$\min_{\tilde{\mathbf{v}} \in \mathrm{R}^N} \|\tilde{\mathbf{v}}\|_1 \text{ subject to } \bar{\mathbf{u}} = \bar{\mathbf{\Phi}}\tilde{\mathbf{v}}. \qquad (2)$$

This basis-pursuit formulation is a convex optimization problem that always has a solution if the measurement equations are consistent. The minimization problem given by equation 2 can easily be posed as a linear program (Bloomfield and Steiger, 1983). Donoho (2006b) derives conditions under which the problems described by equations 1 and 2 are equivalent.

Whether recovery of a signal $\mathbf{u}$ from $\bar{\mathbf{u}}$ succeeds depends on the full signal $\mathbf{u}$ (particularly the sparsity pattern of its discrete cosign transform) and the choice of which M components are observed. A qualitative understanding can be obtained from analyses involving a random matrix with independent and identically distributed Gaussian entries that relates the signal in a sparsifying basis to the observed quantities. (In contrast, the matrix in our problem consists of rows of the inverse DCT matrix.) As the problem dimensions become large, the signal is reconstructed asymptotically when $M > 2 \cdot S \cdot log\ (N/M)$, where S is the number of nonzero entries in the sparse signal (Donoho and Tanner, 2008). Effectiveness of sparsity-based interpolation is shown with a simple example below.

Sparse estimation with prior information and noisy data

The preceding example suggests that sparsity of the solution in the transform domain can be exploited to develop a more efficient and better-posed estimation scheme. When reliable prior knowledge of

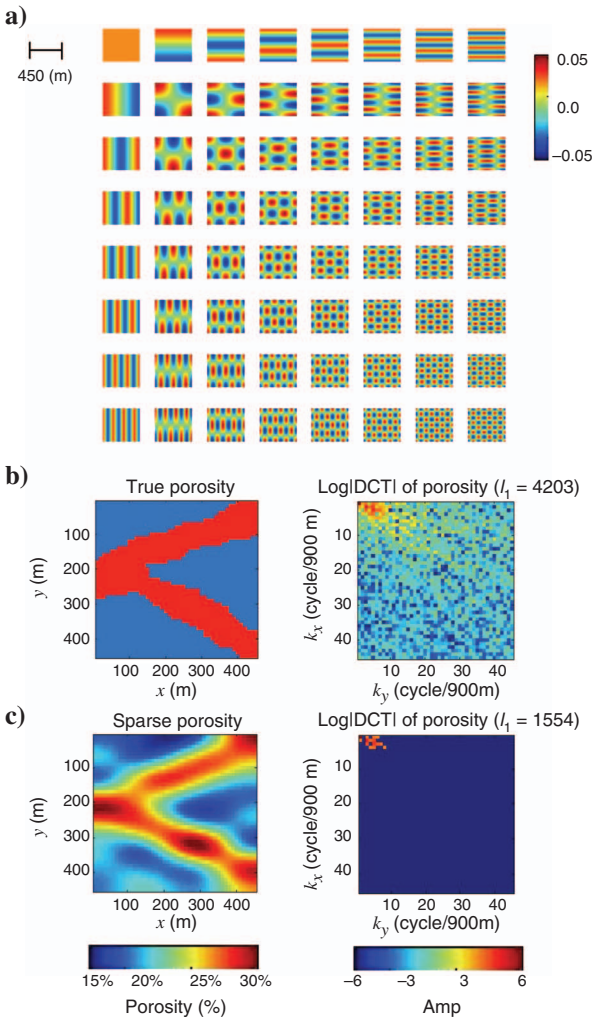

Figure 1. Transform domain (DCT) representation of the true facies model with its sparse approximation in the DCT domain: (a) sample DCT basis images; (b) The true 45×45 porosity distribution (left) with high (red) and low (blue) porosity and its DCT coefficient magnitudes in logarithm scale (right); (c) The sparse representation of true facies distribution (left) with $S = 15$ nonzero DCT coefficients (right) are displayed in the second row (l_1 refers to the l_1-norm of the estimated DCT coefficients). Plot of DCT coefficients in (b) is rotated for better visualization and consistency with DCT basis in (a).

the unknown parameters is available, it can be used to constrain the inverse problem further and improve the reconstruction algorithm. In addition, observation errors need to be taken into account by including a misfit term in the optimization problem. A general formulation of the above problem that incorporates prior information and allows for observation errors can be written as

$$\min_{\mathbf{v}\in R^N} \|\mathbf{C}_\mathbf{u}^{-1/2}(\mathbf{u}-\mathbf{\Phi}\tilde{\mathbf{v}})\|_p + \gamma\|\mathbf{W}\tilde{v}\|_q, \quad (3)$$

where the first term in the objective function penalizes deviations from the observations with an l_p-norm, and the second term enforces the desire for a sparse solution by penalizing the weighted magnitude of the DCT coefficients with an l_q-norm. Covariance matrix $\mathbf{C}_\mathbf{u}$ contains information about the quality (weight) of each individual observation. Throughout this paper, we assume $\mathbf{C}_\mathbf{u}^{-1/2}$ is the identity matrix; which is equivalent to independent and identically distributed measurement errors. This assumption can be readily relaxed if the errors are correlated. Weighting matrix $\mathbf{W}$ can be computed from prior information, using the image-training approach described in Appendix B. Weighting coefficients in this matrix effectively include (or exclude) relevant (or irrelevant) basis vectors in the approximation and promote the selection of basis vectors that are more consistent with geologic prior information. Regularization parameter γ adjusts the relative importance of the prior and observation terms.

Here we consider three alternatives for the objective function norms l_q and l_p in equation 3: the linear least squares (LLS) $[p = q = 2]$, the least absolute deviation (LAD) $[p = q = 1]$, and the least mixed norm (LMN) $[p = 2, q = 1]$ solutions. When the problem is viewed from a Bayesian perspective, these alternatives correspond to adoption of either Gaussian or Laplacian probability distributions for observation noise and for vector $\mathbf{v}$ of DCT coefficients (Bloomfield and Steiger, 1983; Tikhonov and Arsenin, 1977; Schweppe, 1973; Alliney and Ruzinsky, 1994). Then the alternative solutions provide the maximum a posteriori (MAP) estimates associated with the corresponding combinations of measurement noise and DCT coefficient probability distributions.

Several studies on the distribution of DCT coefficients have been conducted for various types of images, including natural and medical images. As a result, different probabilistic assumptions have been proposed for the norm in equation 3 (Reiniger and Gibson, 1983; Eggerton and Srinath, 1986; Muller, 1993; Eude et al.; 1994). Reiniger and Gibson (1983) find the Laplacian distribution to be more appropriate for DCT coefficients, although others report that no single distribution can be adopted. Eude et al. (1994) use the generalized Gaussian function (GGF), which includes the uniform, Laplacian, and Gaussian distributions as special cases, for the DCT coefficient distribution. Lack of agreement in these studies suggests that the choice of a DCT coefficient distribution depends on the type of images and data sources encountered in a particular application.

Two-step solution of the noisy estimation problem

Donoho (2006b) shows that, under some mild conditions, the solution of the original l_0-norm problem in equation 1 and the solution of the l_1-norm problem in equation 2 are identical for sparse noise-free systems. However, when observations are noisy and the formulation in equation 3 is used, identification of the sparsity pattern through the l_1-norm minimization might no longer give the same result as the desired l_0-norm minimization. Such solution variability reflects an important difference between the l_0- and l_1-norm formulations in equations 1 and 2. Although the l_1-norm of vector $\mathbf{v}$ can be decreased by reducing the magnitude of its elements, the l_0-norm depends only on the number of nonzero elements and is not sensitive to their magnitudes. In general, l_1-norm minimization of the DCT coefficients has two effects: first, elimination of the insignificant DCT coefficients leading to increased sparsity (which is desirable in our application), and second, reduction in the magnitude of larger DCT coefficients (undesirable because it can lead potentially to underestimated DCT coefficients). When sparse solutions are sought by assigning a large value of the sparsity regularization weight γ, in equation 3, the second of these effects can result in underestimation of the unknown features especially in the presence of noise.

For this reason, we propose a two-step LMN solution, which divides the l_1-norm minimization reconstruction approach. First, we identify the important DCT basis vectors and sparsity structure in the solution using the l_1-norm minimization with a large γ value, which is equivalent to the noise-free, basis-pursuit approach of equation 2. Next, we estimate DCT coefficients corresponding to the sparsity pattern obtained from step one through l_2-norm minimization of the observation misfits, as in the LMN approach with $\gamma = 0$.

This two-step implementation exploits the fact that the l_1-norm minimization successfully identifies the sparsity pattern in the solution but may underestimate the magnitudes of the sparse coefficients. The first step of the procedure uses an l_1-norm minimization only to identify DCT coefficients that will be assigned nonzero values. In the second step, we find more accurate estimates for nonzero coefficients by minimizing a least-squares data-misfit objective function. The least-squares minimization in the second step maintains the sparsity structure obtained in the first step while improving quality of the coefficient estimates, which we illustrate with an example in the next section.

RESULTS AND DISCUSSION

Here we present and discuss inversion and estimation results for several example problems. We start with a simple noise-free scattered-point-measurement example to demonstrate the use of structural assumptions in the compressed sensing framework. In this example, we consider each of the three alternative formulations to compare the performance of l_1-norm regularization $(q = 1)$ used in LAD and LMN to impose sparsity constraints versus the l_2-norm regularization $(q = 2)$ used in LLS. In our next example, we apply our two-step estimation procedure to the problem of estimating channel features from noisy areal measurements (i.e., denoising). Our third example examines transform domain l_1-norm regularization (through the one-step LMN approach) in a traveltime tomography problem. Finally, we apply the one-step LMN formulation to a field example dealing with porosity estimation.

Interpolation using scattered point measurements

The examples described in this section illustrate how the compressed sensing formulation presented earlier can be used to reconstruct channelized facies in a subsurface environment. For the following examples, we assume that the original signal is the sparse signal with $S = 15$ (Figure 1c). We make this simplifying assumption (which will be relaxed later) for this illustrative example only.

Figure 2 demonstrates results obtained when a limited number of observations $(M = 40)$ is used to constrain the reconstruction. The first column in Figure 2 shows the signal recovery when all 2025

DCT coefficients are permitted to take nonzero values. It can be seen from the resulting DCT coefficients (second row) that many high-frequency coefficients are assigned nonzero values. The second and third columns of Figure 2b show the same experiment when the optimization is restricted to the span of $N = 210$ and $N = 120$ low-frequency basis elements. The solutions are more reasonable from a geologic perspective, and the "true" DCT coefficients are perfectly reconstructed for $N = 120$. This suggests that when prior knowledge of the facies distribution is available, it should be used to constrain the search subspace.

The above example used a perfectly sparse signal (in the DCT domain) with noise-free observations to illustrate sparse-signal reconstruction with compressed sensing. In practice, many examples can be found in which the underlying features are not perfectly sparse

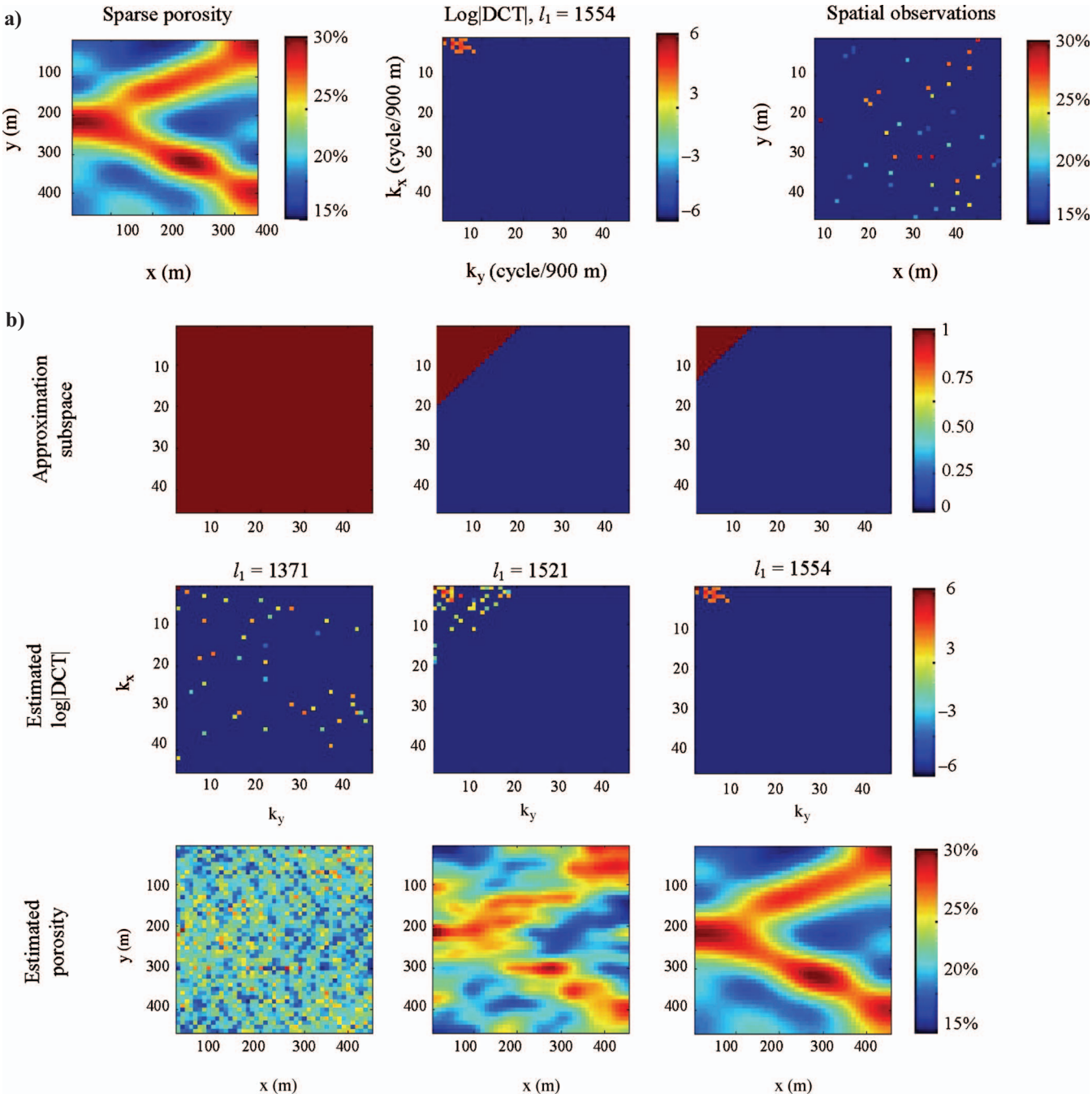

Figure 2. Sparse facies reconstruction from randomly located measurements using low-frequency subspaces: (a) The true sparse (in DCT domain) facies distribution (left), its corresponding sparse ($S = 15$)log-|DCT| coefficients (middle), and ($M = 40$) randomly located observations in space (right); (b) Three masks (first row) that are used to define low-frequency search subspaces in the DCT domain with dimensions $N = 2025$, 210, and 78 (left, middle, right). The corresponding solutions in the DCT and spatial domain are shown in the second and third rows, respectively. The l_1 notation refers to the l_1-norm of the estimated DCT coefficients.

but have a nearly sparse compressed representation. Furthermore, real observations usually are corrupted with noise. In the next example, we consider the interpolation problem in equation 3 to account for noise in the observations, assume that the unknown transformed parameter **v** is only approximately sparse (it can have nonzero but relatively insignificant coefficients), and use prior structural information about the solution to arrive at more consistent solutions.

In our next example, again we assume the observation points are located randomly in space, but we consider the effects of using different ways to incorporate prior information. We set the dimension of the DCT coefficient search subspace to $N = 78$ and number of observations to $M = 30$. When prior information from image training is used, the prior weighting matrix **W** is obtained from the procedure in Appendix B. When prior training is not used, **W** is the identity matrix and equal weights are given to all DCT coefficients. The value of parameter γ is varied to impose different levels of sparsity on the DCT coefficients of the solution.

Figure 3a-c shows the true porosity distribution, corresponding DCT coefficient magnitudes (logarithm scale), and observation locations and values used for the randomly scattered measurement example. Figure 4 shows reconstructed porosity fields obtained with LLS, LAD, and LMN inversion methods. Figure 4a illustrates results for the LLS formulation. The first two rows of Figure 4a give estimates obtained when no assumption is made about the channel orientation (this section is labeled as Approximation subspace 1). The first and second columns give results for two cases with small and large γ values, corresponding to less and more weight, respectively, given to the regularization term in the objective function in equation 3. The third and fourth rows in Figure 4a (Approximation subspace 2) show the LLS estimates when qualitative prior information about channel orientation is available. In this case, more DCT basis elements with left-to-right variability are selected (see basis images in the top portion of Figure 1a). Results are similar to the previous example where the channel orientation was not known except for a slight directional bias in the estimated features, implying that qualitative orientation information might not constrain the LLS solution sufficiently. The last two rows of Figure 4a (Approximation subspace 3) display the results when prior information from basis training is used. As seen in these figures, LLS estimates perform well only when an accurate prior is used to weigh the appropriate basis vectors.

Weighting matrix **W** provides important quantitative information about the expected significant basis elements by specifying smaller penalties for expected significant coefficients (see the basis training discussion and Figure B-1). In the absence of a good prior, the l_2-norm constraint used for regularization does not yield the channelized structure present in the true model. Closer examination of the results reveals that estimated DCT coefficients are not sparse, confirming that the l_2-norm is not a good choice to preserve sparsity.

Figure 4b and c illustrate solutions to the LAD and LMN formulations for the same experiment. Results with increasing prior information (for different approximation subspaces) and sparsity regularization terms are shown. The similarity between results for the LMN and LAD formulations implies that the l_1-norm minimization of DCT coefficients (rather than in the misfit term of the objective function) is responsible for the superior reconstruction they provide as compared to LLS solutions. Reconstruction improves when additional prior information is included. In the case of qualitative directional information (third and fourth rows in Figure 4b and c), the estimated channel orientation is more biased toward the left-to-right direction. Additional improvements are obtained by incorporating consistent prior information inferred from a correct training image (the last two rows in these figures), as expected. Results of these experiments suggest that for chosen values of S, N, and M, the untrained basis is sufficient for finding an approximate sparse solution to the true facies model.

Overall, LLS results suggest that the l_2-norm of the DCT coefficients fails to identify the sparse structure of the original facies distribution and can provide only a reasonable estimate when accurate prior information is supplied. This can limit the use of this formulation in practice, where prior information might be inaccurate or unavailable. In general, the l_2-norm is very sensitive to large deviations (outliers) and tends to neglect smaller misfit terms. This behavior tends to produce small-scale artifacts at unobserved locations, especially when observation noise is present. Given the similarity between the LAD and LMN formulations for inducing sparsity in the solution, we compare only the LMN and LLS solutions in our next example.

Reconstruction using noisy areal measurements (denoising)

The previous examples use relatively accurate but scattered observations in space. Another important application of sparse estimation is to extract features from a complete grid of noisy measurements. An example is a seismic image of porosity. Significant uncertainty can be present in seismic imaging data because of interactions

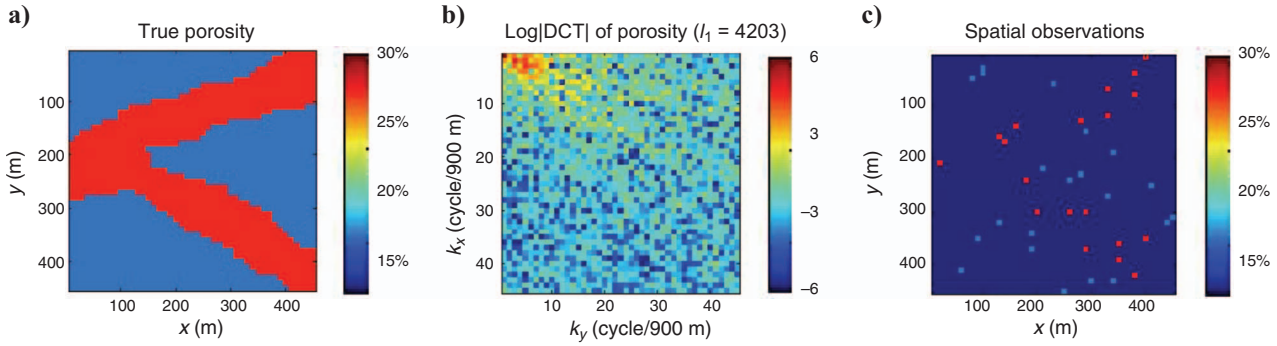

Figure 3. (a) True channel facies porosity distribution, (b) the corresponding DCT coefficients in logarithm scale, and (c) $M = 30$ observations that are randomly located in space (c).

between the signals and geologic layers between the sensor and reservoir.

Figure 5a and b shows the true porosity model and its corresponding DCT coefficient magnitudes (in logarithm scale), respectively, in this example. We assume that all model grid blocks are observed (Figure 5c), but measurement noise has corrupted the data. The sparsity constraint is imposed on DCT coefficients to preserve channel continuity and suppress the effect of measurement noise. The number of observations is $M = 2025$ and the dimension of the approximation subspace (Figure 5d), is $N = 210$ (other cases were tried and yielded similar results). The true model used to generate the observations is not perfectly sparse. Instead, the level of sparsity is varied in each experiment to evaluate its effect on the solution.

Measurement uncertainty is simulated by adding observation errors synthetically to the true facies model. Noisy measurements are generated by adding an uncorrelated Gaussian random error with zero mean and standard deviation of 10% to the original facies values for each grid block in the true porosity image. For comparison, porosity values inside and outside the channels in the true model are ~27% and ~18%, respectively.

The compressed sensing formulation presented earlier is intended for reconstruction of high-dimensional sparse signals from a relatively small number of measurements ($M \ll N$). However, the examples in this section use a large number of noisy measurements. Therefore, the primary goal of our l_1-norm minimization is to extract a feature from the noisy measurements. Rather than perfect signal

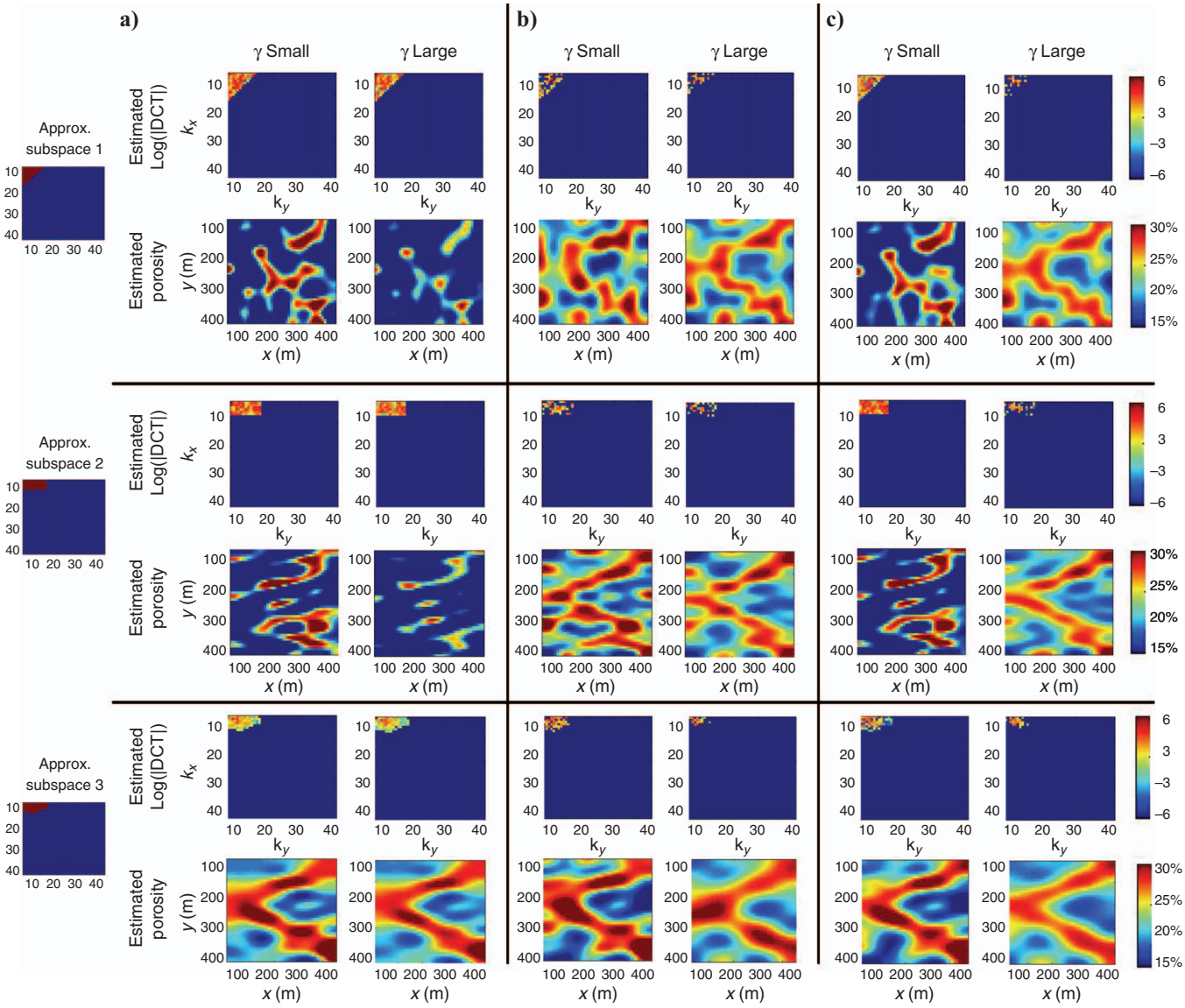

Figure 4. Reconstruction of porosity model in Figure 3a from $M = 30$ scattered measurements (Figure 3c) using (a) LAD, (b) LLS, and (c) LMN formulations. In each case, reconstruction results using three approximation subspaces are shown (subspaces 1, 2, and 3 refer, respectively, to small wavelength DCT basis elements with no orientation preference, left-to-right orientation preference, and quantitative prior training as shown in Figure B-1. Reconstruction results with a small (left column) and large (right column) regularization weights γ are shown for each case.

recovery, we focus on l_1-norm regularization to find a sparse approximate solution that best describes the feature.

We perform a suite of tests to evaluate the performance of l_1-norm minimization by comparing the LLS, ones-step LMN, and two-step LMN formulations. These experiments are designed to examine the sensitivity of the reconstruction to prior basis training, the weight given to the sparsity constraint term in the objective function, and the search dimension.

Figure 6 illustrates reconstruction results for the noisy areal measurements in Figure 5d using the LLS, one-step LMN, and two-step LMN formulations. The importance of the regularization term relative to the measurement mismatch term in the objective function equation 3 is adjusted with the weighting parameter γ. Figure 6a displays reconstruction results for the DCT coefficients (first row) and porosity values (second row) with the LLS formulation using no training, i.e., where $\mathbf{W}$ is the identity matrix. The panels from left to right in Figure 6a show the reconstructed solutions with increasing γ values. For $\gamma = 0$, the reconstructions simply minimize the l_2-norm of the data fitting error. This results in nonzero values for almost all DCT coefficients. Increasing γ reduces the energy of the DCT coefficients at the expense of greater data fitting error. It is evident from Figure 6a that this behavior results in underestimation of porosity values in the spatial domain. These results confirm that the LLS estimator is not appropriate for extracting significant DCT basis elements and tends to underestimate the magnitudes of these significant DCT coefficients.

Figure 6b depicts the one-step LMN estimation results without prior training. The columns from left to right show solutions with increasing γ values to put more emphasis on sparsity. As seen in Figure 6c, the behavior of the one-step LMN solution is different than that of the LLS solution. With increasing values of γ, more insignificant DCT coefficients are eliminated than in the LLS approach (compare Figure 6a and b). Furthermore, the magnitudes of the significant DCT coefficients do not decrease as sharply with increasing value of γ as with the LLS solution. This selective behavior of the l_1-norm minimization is exploited to systematically identify the significant basis elements of the underlying unknown feature from the data. The third and fourth columns in Figure 6b show the one-step LMN solutions for large γ values. It can be observed from these figures that magnitudes of larger DCT coefficients begin to decrease with increasing values of γ. One major distinction between the LLS and one-step LMN solutions at large γ values is that in the LLS approach, increasing γ values mainly reduces larger DCT coefficients (per sensitivity of l_2-norm to larger values). However, in the one-step LMN solution, increasing γ values mainly results in selective elimination of the insignificant DCT coefficients.

Figure 6c displays results obtained with the two-step LMN procedure. The first step of this procedure is equivalent to the one-step LMN procedure with a large γ value. The sparsity structure of the two-step LMN solution after step one, DCT coefficients obtained from step one, and corresponding porosity facies structure obtained after the least-squares minimization of step two are shown, respectively, in the first, second, and third columns of Figure 6c. Improved results obtained from the two-step procedure suggests that step one l_1-norm minimization can identify sparse structure of the true field adequately. However, the step two l_2-norm minimization can use this step one result to facilitate estimation of important DCT coefficients.

Synthetic traveltime tomography example

Next, we apply our one-step LMN sparse estimation procedure with l_1-norm regularization to a simple straight-ray cross-well traveltime tomography example. Our objective is to reconstruct the slowness of a porous medium in the interval between two wells. We use a smooth Gaussian slowness field as the true model and compare the slowness derived from l_1-norm regularized estimates of DCT coefficients with slowness derived from least-squares solutions with first- and second-order Tikhonov regularization terms. Tikhonov regularization is a preferred reconstruction technique when recovering fields that are believed to have mainly smooth features (Tikhonov and Arsenin, 1977).

Figure 7a shows the configuration of the cross-well tomography example. A uniformly spaced system of ten sources is located on the left end of the domain and a symmetric array of ten receivers is located on the right end of the interval. The resulting 100 arrival-time measurements are used to infer the slowness structure of the medium. Figure 7b illustrates the true slowness used to generate the synthetic inversion data. The slowness values in the true model vary between 0.66 μs/m and 321.52 μs/m. Two sets of experiments are performed; one with exact (noise-free) measurements, and one with additive Gaussian noise with a standard deviation corresponding to 10% of the actual measurement magnitudes.

Figure 8 summarizes results obtained with noise-free measurements. Each row in this figure corresponds to a separate reconstruction method.

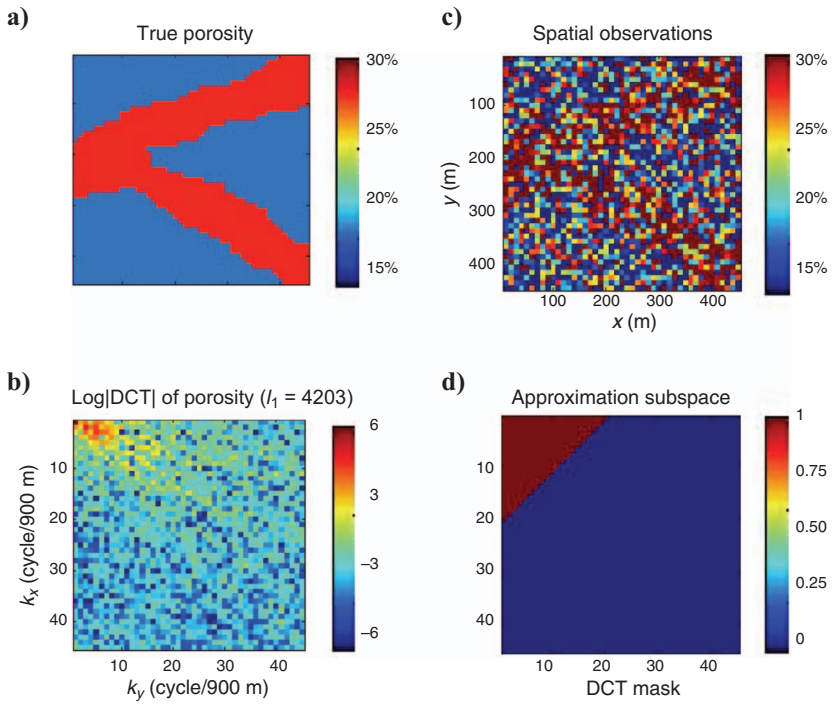

Figure 5. (a) True channel facies porosity distribution, (b) the corresponding DCT coefficient magnitudes in logarithm scale, (c) grid observations that are corrupted by noise, and (d) the approximation subspace without directional preference.

Figure 8a and b depicts solutions for first- and second-order Tikhonov regularizations (corresponding to minimization of first- and second-order spatial derivatives, respectively). Figure 8c illustrates estimates obtained when l_1-norm DCT coefficient regularization is used. In each case, four reconstruction solutions corresponding to zero, small, medium, and large regularization weights (left to right) are shown. The l_1-norm regularization in Figure 8c was applied to a 465-dimensional ($N = 465$) low-frequency DCT approximation subspace. From Figure 8a and b, it is evident that when the underlying slowness model is smooth and observations are perfect, first- and second-order Tikhonov regularization methods provide similar and reasonably accurate reconstructions. By increasing the contribution of the regularization term in the objective function, increasingly smoother solutions are obtained (Figure 8a and b, left to

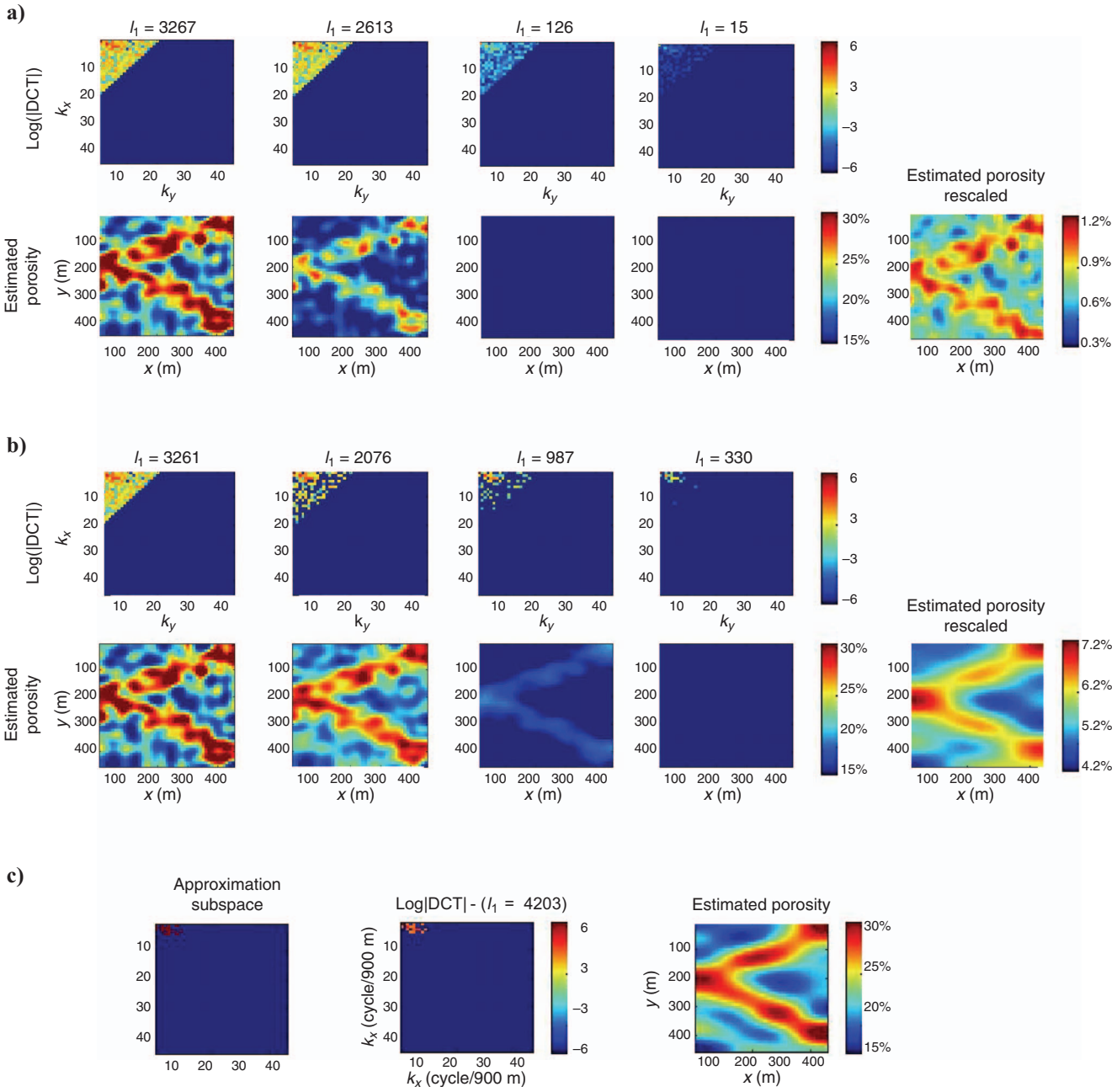

Figure 6. Reconstruction results for the true model in Figure 5a using noise-corrupted areal measurements (Figure 5c) with (a) LLS and (b) and (c) first and second steps, respectively, of the LMN formulations with increasing value of regularization weights γ (left to right). In each case, estimated DCT coefficients (first row) and the corresponding estimated porosity field (second row) are shown; the last column shows rescaled reconstruction result for the largest γ used; (c) Reconstruction results for two-step implementation of the LMN formulation. The sparse approximation subspace (left), estimated DCT coefficients magnitude in logarithm scale (middle), and facies distributions (right) are shown. The l_1 notation refers to the l_1-norm of the estimated DCT coefficients.

right). Consequently, the quality of the solution is improved initially until reconstruction results are overly smooth (last column of Figure 8a-c).

The first column in Figure 8c contains the reconstruction estimate without including the l_1-norm regularization in the one-step LMN method (i.e., there is only least-squares minimization of the misfit term). It can be inferred from this figure that although the nonregularized reconstruction in the DCT domain is capable of identifying the general low- and high-slowness regions, it fails to capture the smoothness and exact value of the slowness features, particularly near the source and receiver locations. Clearly, in the absence of l_1-norm regularization the DCT parameterization is outperformed by the Tikhonov regularization methods. Even though Tikhonov methods solve the problem in the $N = 2025$ dimensional spatial domain (compared to $N = 465$ in the DCT domain), effectively they can eliminate fine-scale (short-wavelength) features. As seen in the second through fourth columns of Figure 8c, when l_1-norm regularization is included, reconstruction outcomes in the DCT domain are improved markedly and outperform Tikhonov regularization solutions. This encouraging outcome is attributed to the fact that the DCT representation of the true model is approximately sparse and that the structure of the existing sparsity is captured effectively by solving an l_1-norm minimization problem in the low-dimensional, discrete-cosine-transform subspace.

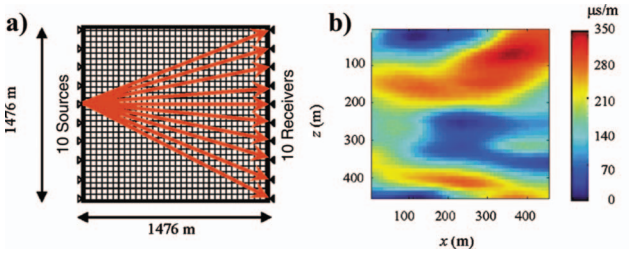

Figure 7. Cross-well traveltime tomography example. (a) Configuration of the example with an array of ten sources (10 m apart) and an array of ten receivers (10 m apart); (b) the true slowness model.

Figure 9 depicts the results from a similar set of experiments, in which Gaussian noise with standard deviation equal to 10% of the data values is added to the measurements. The noise in the data should have the most noticeable effect when a small regularization

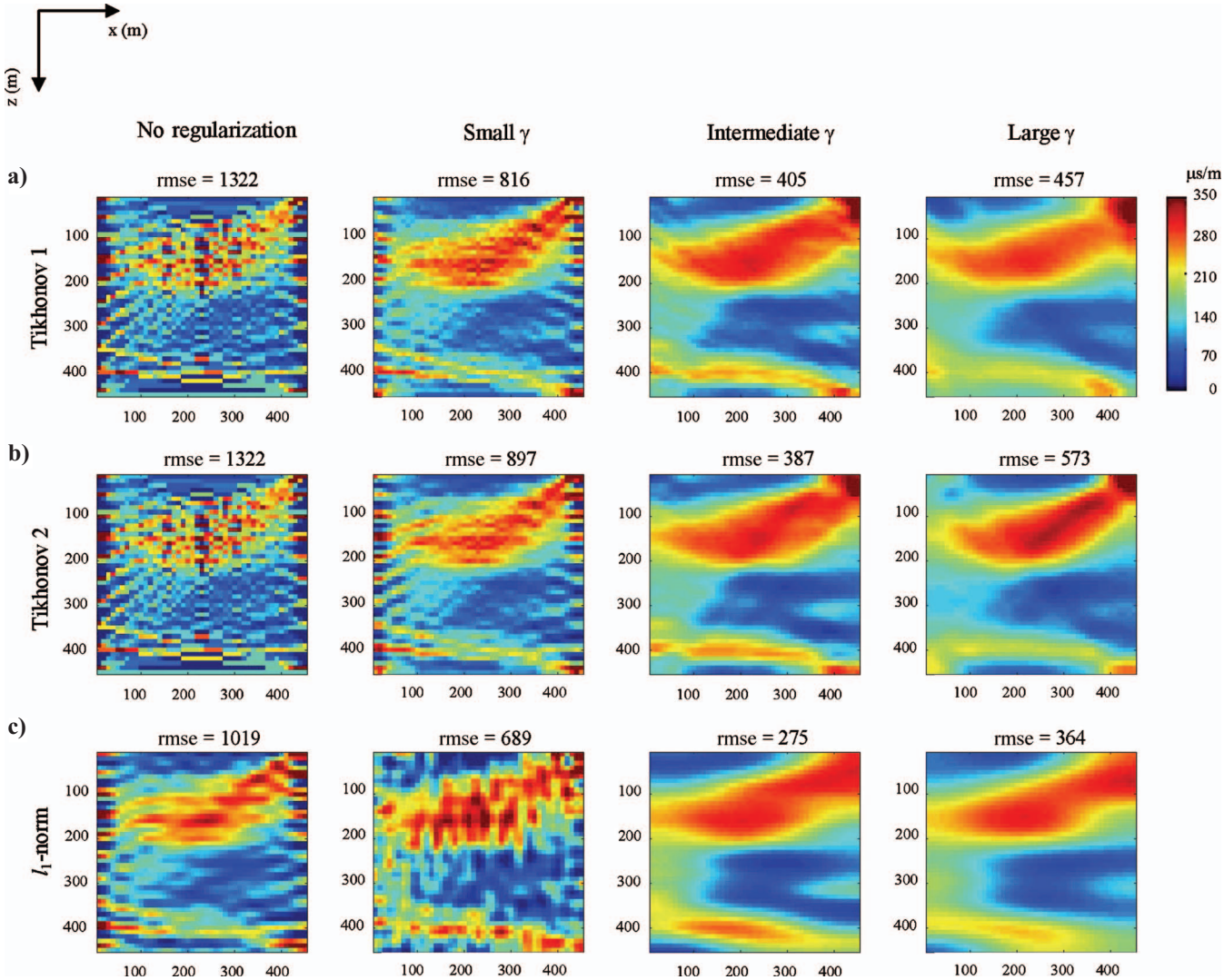

Figure 8. Cross-well traveltime tomography of the model shown in Figure 7b without adding noise to data. The reconstruction results for Tikhanov regularization with (a) first order, and (b) second order spatial derivative minimization, and (c) l_1-norm regularization (LMN) in the DCT domain are shown with increasing regularization weight γ (from left to right). The root mean squared error (rmse) is in (μs/m) unit.

weight is used. Inspection of the results in Figure 9 suggests that in all methods (Figure 9a-c), a large regularization weighting coefficient is needed to suppress the effect of noise. We make an interesting observation by comparing results pertaining to l_1-norm DCT coefficient regularization (i.e., Figures 8c and 9c). Except when a large regularization weight is used, relatively high-frequency elements tend to appear as the procedure attempts to fit the noise in the measurement. We can reduce this effect by incorporating prior knowledge about the underlying model (e.g., by eliminating the irrelevant high-frequency basis vectors).

Field example

As our final example, we apply the one-step LMN procedure to sparse reservoir porosity from the Burbank oil field in Oklahoma, United States. The reservoir is made of sandstone and is divided into several flow units. Hird (1993) provides a more detailed description of the field. Here we use porosity data from a section of the sixth flow unit to construct a porosity map of the entire field. We use three methods to reconstruct the underlying porosity map: a spline interpolation technique documented by Sandwell (1987), a geostatistical kriging procedure, and l_1-norm regularization in the DCT domain. The kriging estimator uses a variogram model (or a stationary covariance) to interpolate porosity values from point measurements. It considers only data near each estimation point (usually within a few correlation lengths) when deriving the interpolated value. Therefore, it can be expected to be a locally accurate interpolator (Goovaerts, 1999).

To include the smoothness assumption of splines and kriging in l_1-norm regularization, we used a low-frequency approximation subspace with $N = 78$. Because the information in the variogram model describes the spatial persistence expected in the field, a separate case was considered in which an ensemble of 100 random realizations was generated randomly from the kriging variogram. These realizations were used to weight the basis elements in the l_1-norm regularization following the procedure outlined in Appendix B.

Figure 10a shows the data set and the configuration of the estimation domain that is considered. The porosity data vary within the $[3.45\%, 31\%]$ range. The domain of interest is discretized into a 100 by 50 array of cells of size 1312 m by 1640 m in x- and y-directions, respectively. Sixty observations were used to construct the porosity map. Anisotropic exponential variograms with the range of

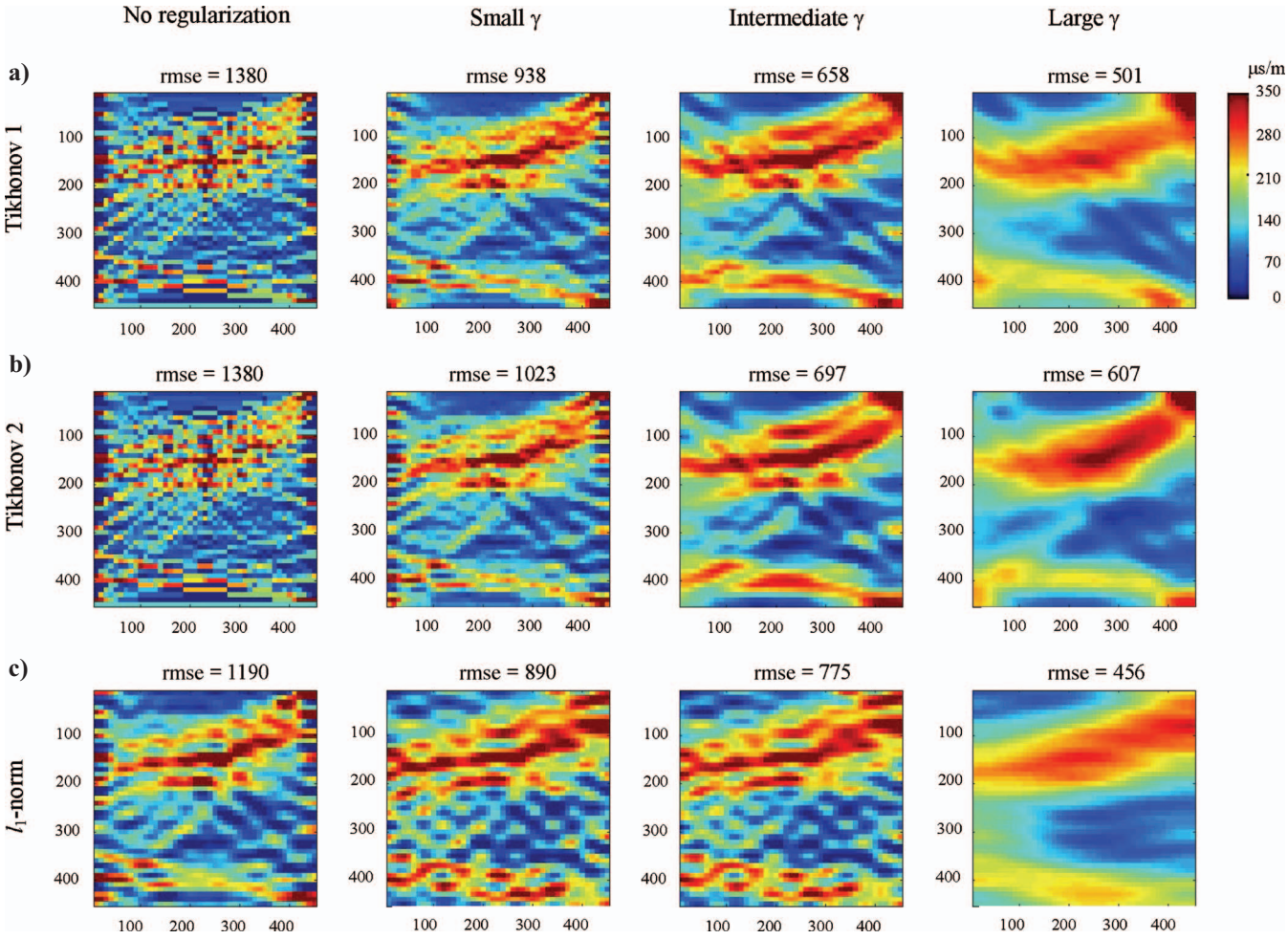

Figure 9. Cross-well traveltime tomography of the model shown in Figure 7b after adding 10% noise to data. Reconstruction results for Tikhonov regularization with (a) first-order and (b) second-order spatial derivative minimization, and (c) l_1-norm regularization (LMN) in the DCT domain are shown with increasing regularization weight γ (from left to right). The root mean squared error (rmse) is in (μs/m) unit.

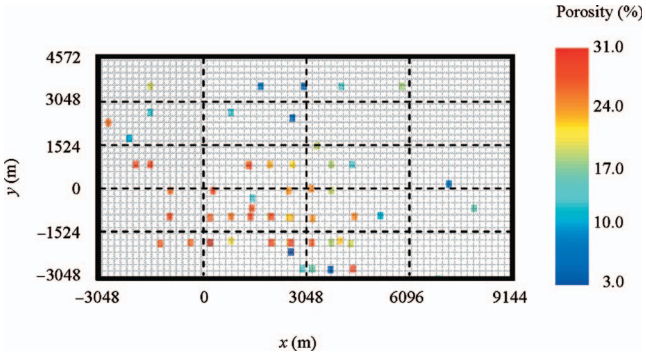

Figure 10. Porosity data from Burbank field, Oklahoma, U.S.A. (a section of the sixth flow unit is shown).

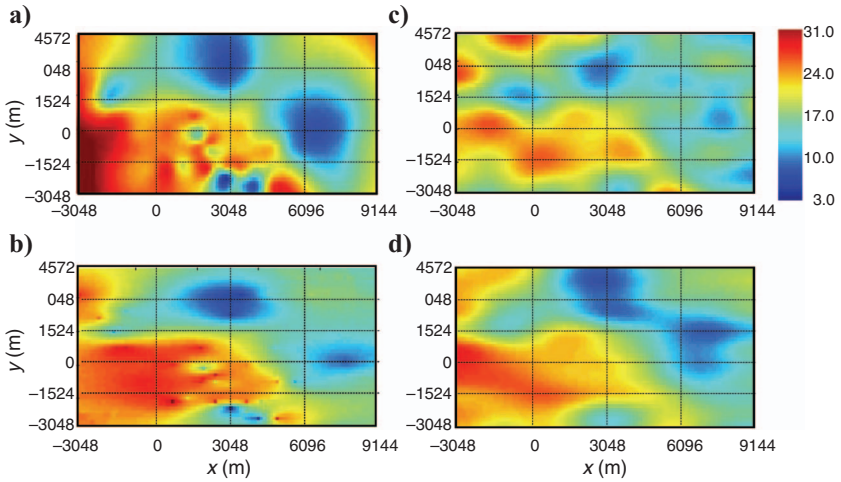

Figure 11. Reconstruction results for the Burbank field porosity data using (a) the spline methods of Sandwell (1987), (b) geostatistical kriging, (c) l_1-norm regularization in a low frequency DCT domain (LMN), and (d) l_1-norm regularization in a low frequency DCT domain conditioned on the kriging variogram model (LMN-Var).

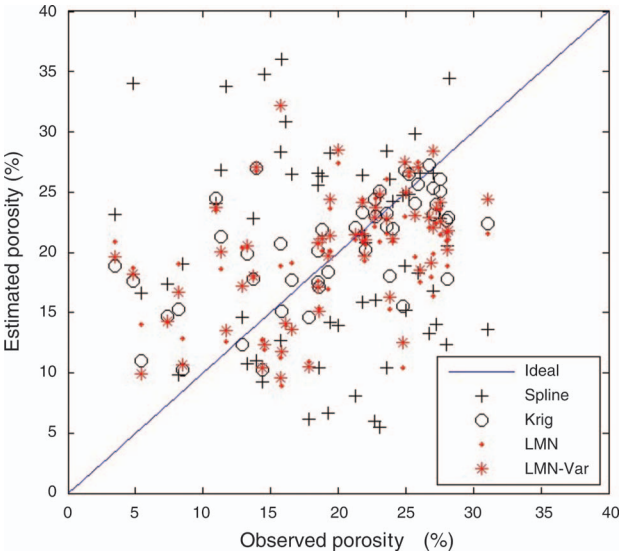

Figure 12. Cross-validation results for Burbank field example. LMN represents the solution with l_1-norm regularization in a low frequency DCT domain and LMN-Var refers to l_1-norm regularization in a low frequency DCT domain conditioned on the kriging variogram model. Root-mean-squared errors are 87 for the spline, 60 for kriging, 54 for l_1-norm, and 51 for l_1-norm and with conditioning on the variogram model.

$a_x = 20{,}000$ ft and $a_y = 12{,}500$ ft, a sill value of $C = 60$, and a nugget $C_0 = 5$ described the spatial correlation in the data adequately. This variogram was used in the kriging algorithm. Kriging calculations were performed with the KT3D program included in the GSLIB software package (Deutsch and Journel, 1998). Sandwell (1987) provides details of the spline interpolation method.

Burbank field porosity reconstruction results are summarized in Figures 11 and 12. Figure 11a-d depicts the map of porosity estimates that are obtained from spline interpolation, kriging, one-step LMN with l_1-norm minimization, and one-step LMN with l_1-norm minimization conditioned on the variogram model, respectively. Interpolation results from these methods capture the general trends of low and high porosity. However, the structural continuity of each method is slightly different. A significant difference between the regularization and interpolation methods (spline and kriging) is that the interpolation methods reproduce the observed data exactly, and the regularization method trades off measurements misfit and prior structural assumptions (expressed in terms of sparsity in the DCT domain and/or spatial persistence as described by the variogram). Figure 11c and d shows the effect of using continuity of the variogram as an additional constraint. The resulting estimate in Figure 11d bears some resemblance to the kriging result in Figure 11b. The main difference between the two estimates is in the connectivity within the low-porosity region. Minimization of the l_1-norm in the DCT domain tends to result in more continuous features with smaller values. Potentially, this can lead to underestimation of porosity values and overestimation of connectivity in the reservoir. The two-step procedure described in the previous section (and in Figure 6c) can be used if a smaller data misfit is desired.

A standard cross-validation procedure was used to evaluate the performance of each method in estimating an observed value. For each of 60 estimation problems, one measurement was removed from the data set and the remaining 59 measurements were used to generate a porosity map. Estimated and observed values for the removed measurement were then compared to evaluate the accuracy of the estimation method. Figure 12 shows a scatter plot of estimated versus observed values for each method. The results suggest that in an rms-error sense, l_1-norm regularization provides a better estimate for this example. A simple explanation for this behavior is that interpolation methods are more sensitive to local observations; therefore, removing a data point has more impact on the overall interpolation result (especially in regions with scattered observations). However, l_1-norm-regularized methods are less sensitive to local data and are designed to find global trends in the field.

CONCLUSIONS

Here we introduce a new regularization method for inverse problems in geophysical applications. This method combines a linear compression transform procedure (we use the DCT in our examples) with an inversion approach that seeks a sparse solution composed of a small subset of the transformed basis-function coefficients. This approach to inversion has its origin in basis-pursuit and sparse-reconstruction algorithms that rely on l_1-norm minimization. In gener-

al, we have shown that promoting sparsity through l_1-norm minimization can be combined effectively with structural prior assumptions to reconstruct geologic features in an appropriate transform domain. We used several subsurface-characterization examples to demonstrate the effectiveness of our regularization approach when the underlying assumptions are applicable.

A key element in the development and application of the method is identification of a transform that can be expected to yield an approximately sparse representation of the original field. General compression transforms such as DCT or discrete wavelet transforms are obvious sparse basis choices that can be suitable for sparsity promoting with l_1-norm minimization. By incorporating structural prior knowledge about the properties of the underlying model (e.g., geologic setting and depositional environment) it is possible to identify an approximation subspace in the transform domain that encourages solutions with specific properties (as suggested by prior information). Structural prior information can become increasingly useful (even necessary) when sufficient data are not available to constrain the solution or when observations are corrupted with noise. Because the low-frequency DCT basis vectors provide reasonable approximations of smooth fields, continuous geologic features are likely to be compressible and have an approximately sparse DCT representation. In this case, most of the transformed coefficients have small magnitudes and can be truncated (as structural prior assumption) with marginal effect on the quality of the solution.

Successful use of the transform domain l_1-norm regularization in the presence of noise may necessitate the use of a large regularization weight to suppress fine-scale artifacts that can emerge falsely because of observation noise. However, adopting too large an l_1-norm regularization weight can lead to excessive smoothing and underestimation of property values. We propose a heuristic two-step procedure to remedy this problem. In the first step, we identify a sparse solution structure (i.e., the subspace spanned by the set of most significant DCT basis elements) using l_1-norm minimization with a large regularization weight in the objective function. In the second step, we minimize the data misfit by adjusting magnitudes of the basis-function coefficients selected in step one. This two-step approach was more effective and resulted in improved solutions when the data were corrupted by noise and a large regularization parameter was needed to dampen the noise effect.

In summary, this inversion approach seems promising for geoscience applications where, because of spatial correlations commonly observed in facies distributions, a sparse approximation to the underlying model is believed to exist in a transform domain (often through decorrelating transforms). The potential of this approach across a range of applications (including geophysical inverse problems, subsurface characterization, and medical imaging) merits further investigation, including tests on realistic large-scale problems.

APPENDIX A

THE DISCRETE COSINE TRANSFORM

The discrete cosine transform is a linear transform that is widely used for image coding because of its compression power for smooth and correlated data. The DCT basis that we have used in this paper is the most commonly used from a family of discrete cosine transforms (see Püschel and Moura, 2003). The 1D (forward) DCT of a discrete signal $\mathbf{u}$ of length N can be expressed as (Ahmed et al., 1974; Rao and Yip, 1990):

$$v(k) = \alpha(k) \sum_{n=0}^{N-1} u(n) \cos\left[\frac{\pi(2n+1)k}{2N}\right] \quad 0 \le k \le N-1 \tag{A-1}$$

$$\alpha(k = 1{:}N-1) = \sqrt{2}\alpha(0) = \sqrt{\frac{2}{N}} \tag{A-2}$$

Equations A-1 and A-2 describe an orthogonal transformation, so its inverse has a very similar form:

$$u(n) = \sum_{k=0}^{N-1} \alpha(k) v(k) \cos\left[\frac{\pi(2n+1)k}{2N}\right] \quad 0 \le n \le N-1 \tag{A-3}$$

The DCT can be interpreted as a counterpart to the discrete Fourier transform (DFT) that provides real-valued outputs for real-valued inputs. Compared to DFT, the DCT has better energy-compaction properties and approaches the performance of the Karhunen-Loeve transform (which is optimal in the mean-squared-error sense) for certain Morkov processes. Hence, it has enjoyed widespread application in image compression (Rao and Yip, 1990). Many of the properties of the DFT extend in some way to the DCT. In particular, the transform coefficients have a natural order from low- to high-frequency, and the DCT is extended to multidimensional signals by Cartesian product. Applying the transform separately in each dimension and using algorithms like fast Fourier transforms make these computations efficient.

Here we make extensive use of the 2D DCT. Figure 1a shows the 64 basis elements that arise in a 2D DCT with size 8 in each dimension. Smooth 2D signals can be approximated well by a small number of DCT coefficients (Jain, 1989; Gonzalez and Woods, 2002). More specifically, if the horizontal and vertical frequency content are significantly different, then the best approximation will use a different number of coefficients horizontally and vertically. Such asymmetric sampling is illustrated in Figure 1b and is exploited in the porosity estimation shown in Figure 4.

APPENDIX B

BASIS TRAINING PROCEDURE

A training procedure can be developed to obtain the weighting coefficients (elements of $\mathbf{W}$) from the available training data. A simple training procedure is described briefly in this section for our synthetic example. We assume that prior information takes the form of a training library that includes representative examples of spatial patterns that can be expected in a given application. Such a library can be used to determine, on average, the significance of each basis vector in the reconstruction. Here, we compute the weighting matrix $\mathbf{W}$ in equation 3 using the following procedure:

1) The DCT of each image in the library is computed.
2) Absolute values of the DCT coefficients are averaged across the library.
3) Means of the absolute values of the resulting DCT ensemble are truncated by specifying a threshold to remove basis vectors unlikely to have a significant contribution (i.e., vectors corresponding to basis coefficients with small mean magnitudes).

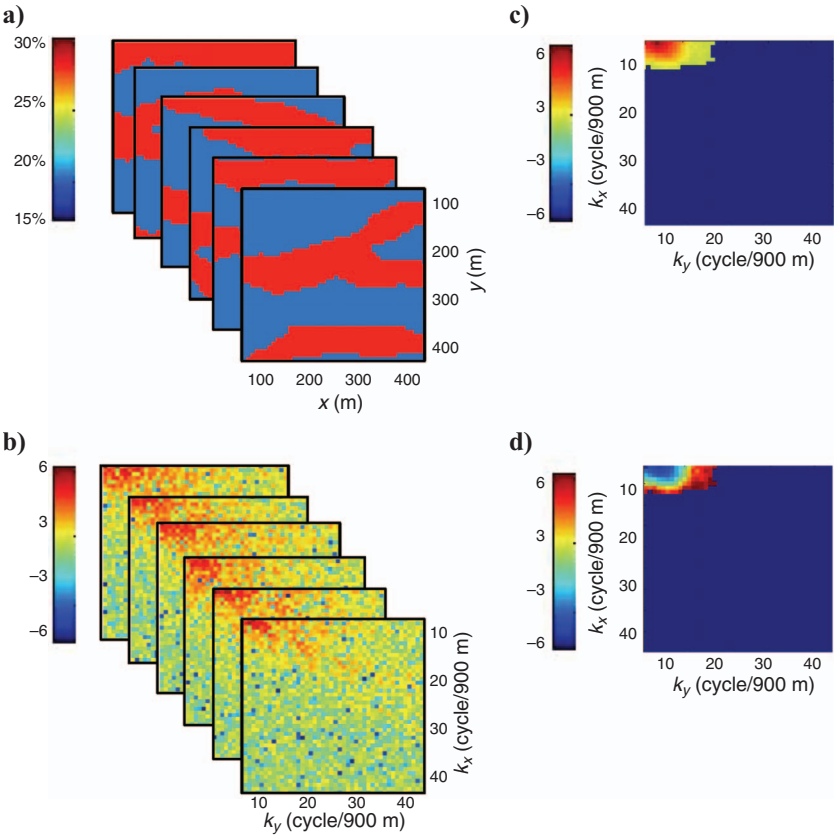

Figure B-1. Training DCT basis with a prior ensemble of porosity. (a) Sample facies realizations generated by SGeMS using multipoint geostatistics (Strebelle and Journel, 2001) and used for training the DCT basis; (b) magnitude of the DCT coefficients for the prior ensemble in (a), shown in logarithm scale; (c) the largest (in magnitude) 78 DCT coefficients (averaged over prior ensemble); and (d) the reciprocal of the results in (c) used to build a diagonal weighting matrix **W**.

(4) Reciprocals of the remaining coefficients are normalized to yield **W**.

Matrix **W** (obtained from the above procedure) ensures that in the reconstruction algorithm, significant basis vectors (as suggested by the prior library) are given small penalties so they can take large values, whereas large penalties are associated with less significant basis vectors to keep their corresponding coefficients small. Insignificant (high-frequency) basis vectors are eliminated in the estimation problem.

Figure B-1 illustrates the training procedure used to obtain **W**. Figure B-1a presents sample 45×45 channelized facies realizations generated from a 250×250 training image borrowed from Caers and Zhang (2004) and using multipoint geostatistical simulation software SGeMS (Strebelle and Journel, 2001; Remy, 2004). Figure B-1b and c shows, respectively, the DCT coefficients associated with each member of the library and the result of the thresholding operation, which retains only the basis vectors corresponding to the 78 DCT coefficients that are largest on average. Weight matrix **W** is a diagonal matrix with its diagonal elements corresponding to the reciprocal of the remaining averaged DCT magnitudes as shown in Figure B-1d. The unknown coefficients of the thresholded basis vectors are the parameters to be estimated in the inversion problem. It is important to note that if the prior image library is not consistent with the main features in the unknown parameters, prior training can be misleading and reconstruction results can be affected adversely (Jafarpour and McLaughlin, 2009).

REFERENCES

Acar, R., and C. Vogel, 1994, Analysis of bounded variation penalty methods for ill-posed problems: Inverse Problems, **10**, 1217–1229.

Ahmed, A. T., T. Natarajan, and K. R. Rao, 1974, Discrete cosine transform: IEEE Transactions on Biomedical Engineering, **C23**, 90–93.

Ajo-Franklin, J. B., B. J. Minsley, and T. M. Daley, 2007, Applying compactness constraints to differential seismic traveltime tomography: Geophysics, **72**, no. 4, R67–R75.

Alliney, S., and S. Ruzinsky, 1994, An algorithm for the minimization of mixed l_1 and l_2 norms with application to Bayesian estimation: IEEE Transactions on Signal Processing, **42**, 618–627.

Bloomfield, P., and W. Steiger, 1983, Least absolute deviations: Theory, applications, and algorithms (Progress in probability and statistics): Birkhäuser.

Bube, K., and R. Langan, 1997, Hybrid l_1/l_2 minimization with applications to tomography: Geophysics, **62**, 1183–1195.

Caers, J., and T. Zhang, 2004, Multiple-point geostatistics: A quantitative vehicle for integrating geologic analogs into multiple reservoir models: AAPG Memoir, **80**, 383–394.

Candès, E., J. Romberg, and T. Tao, 2006, Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information: IEEE Transactions on Information Theory, **52**, 489–509.

Candès, E., and T. Tao, 2006, Near optimal signal recovery from random projections: Universal encoding strategies: IEEE Transactions on Information Theory, **52**, 5406–5425.

Chen, S. S., D. L. Donoho, and M. A. Saunders, 2001, Atomic decomposition by basis pursuit: SIAM Review, **43**, 129–159.

Claerbout, J., and F. Muir, 1979, Robust modeling with erratic data: Geophysics, **38**, 826–844.

Constable, S., R. Parker, and C. Constable, 1987, Occam's inversion: a practical algorithm for generating smooth models from electromagnetic sounding data: Geophysics, **52**, 289–300.

Deutsch, C. V., and A. G. Journel, 1998, GSLIB: Geostatistical software library and user's guide: Oxford University Press.

Donoho, D. L., 2006a, Compressed sensing: IEEE Transactions on Information Theory, **52**, 1289–1306.

———, 2006b, For most large underdetermined systems of linear equations the minimal l_1-norm solution is also the sparsest solution: Communications on Pure and Applied Mathematics, **59**, 797–829.

Donoho, D. L., and J. Tanner, 2009, Counting faces of randomly-projected polytopes when the projection radically lowers dimension: Journal of the American Mathematical Society, **22**, 1–53.

Eggerton, J. D., and M. D. Srinath, 1986, Statistical distribution of image DCT coefficients: Computers and Electrical Engineering, **12**, 137–145.

Eude, T., R. Grisel, H. Cherifi, and R. Debrie, 1994, On the distribution of the DCT coefficients: International Conference on Acoustics, Speech, and Signal Processing, IEEE, Proceedings, **5**, 365–368.

Gonzalez, R. C., and R. E. Woods, 2002, Digital image processing, 2nd edition: Prentice Hall.

Goovaerts, P., 1999, Geostatistics in soil science: State-of-the-art and perspectives: Geoderma, **89**, 1–45.

Hird, K. B., 1993, A conditional simulation method for reservoir description using geological and well performance constraints: Ph.D. dissertation, University of Tulsa.

Jafarpour, B., and D. B. McLaughlin, 2009, Reservoir characterization with the discrete cosine transform, Part 1: Parameterization, and Part 2: History matching: SPE Journal, **14**, no. 1, 182–201.

Jain, A. K., 1989, Fundamentals of digital image processing: Prentice Hall.

Last, B., and K. Kubik, 1983, Compact gravity inversion: Geophysics, **48**, 713–721.

Lin, T., and F. Herrmann, 2007, Compressed wavefield extrapolation: Geophysics, **72**, no. 5, SM77–SM93.

Muller, F., 1993, Distribution shape of two-dimensional DCT coefficients of

natural images: Electronics Letters, **29**, 1935–1936.
Natarajan, B. K., 1995, Sparse approximate solutions to linear systems: SIAM Journal of Computing, **24**, 227–234.
Püschel, M., and J. Moura, 2003, The algebraic approach to the discrete cosine and sine transforms and their fast algorithms: SIAM Journal of Computing, **32**, 1280–1316.
Portniaguine, O., and M. Zhdanov, 1999, Focusing geophysical inversion images: Geophysics, **64**, 874–887.
Rao, K. R., and P. Yip, 1990, Discrete cosine transform: algorithms, advantages, applications: Academic Press.
Reiniger, R., and J. Gibson, 1983, Distribution of the two-dimensional DCT coefficients of images: IEEE Transactions on Communications, **31**, 835–839.
Remy, N., 2004, S-GeMS: A geostatistical earth modeling library and software: Ph.D. thesis, Stanford University.
Rudin, L., S. Osher, and E. Fatemi, 1992, Nonlinear total variation based noise removal algorithms: Physica D, **60**, 259–268.
Sandwell, D. T., 1987, Biharmonic Spline interpolation of GEOS-3 and SEASAT altimeter data: Geophysical Research Letters, **2**, 139–142.
Schweppe, F. C., 1973, Uncertain dynamic systems: Prentice Hall.
Strebelle, S., and A. G. Journel, 2001, Reservoir modeling using multiple-point statistics: SPE, 71324.
Tikhonov, A. N., and V. I. Arsenin, 1977, Solution of ill-posed problems: John-Wiley and Sons.
Yu, M., and D. Dougherty, 2000, Modified total variation methods for three dimensional electrical resistance tomography inverse problems: Water Resources Research, **36**, 1653–1664.