

ADAPTIVE TRAVEL TIME TOMOGRAPHY WITH LOCAL SPARSITY

Michael Bianco and Peter Gerstoft

UCSD Scripps Institution of Oceanography
La Jolla, CA 92093-0238

ABSTRACT

We develop a 2D travel time tomography method which regularizes the inversion by modeling sparsely patches of slowness pixels from discrete slowness map, and adapts sparse dictionaries to the slowness data. This locally-sparse travel time tomography (LST) approach considers global and local behavior of slowness, whereas conventional regularization methods consider only global covariance of pixels. We develop a *maximum a posteriori* formulation of LST, and further exploit the sparsity of patches using dictionary learning. We demonstrate the LST method on densely, but irregularly sampled synthetic slowness maps.

Index Terms— Sparse modeling, machine learning, dictionary learning, geophysics, seismics

1. INTRODUCTION

Travel time tomography methods attempt to estimate complex Earth structure, which contains smooth and discontinuous features at multiple spatial scales, using seismic and acoustic wave travel times between recording stations [1, 2]. The inversion of the travel times for a slowness model (inverse of speed) is ill-posed, with often dense but irregular ray coverage of environments. Conventional tomography techniques regularize the inversion by restricting the models to be only smooth or discontinuous, which include 1st and 2nd order Tikhonov regularization [1, 2]. Other regularization methods have been proposed which employ of wavelet functions [3–6], total variation (TV) [7, 8], or adaptive discretizations of slowness [9].

Recent works in acoustics have utilized sparse modeling and compressive sensing (CS) [10–12] to improve performance in beamforming [13, 14] and inversion for ocean acoustic properties [15–18]. Similarly, the more recent wavelet-based methods in seismic tomography, e.g. [4, 6, 19], assume sparse wavelet coefficients. In sparse modeling and CS, inverse problems are regularized by modeling signals as sparse combinations of vectors or *atoms* from set or *dictionary* of atoms, which can be prescribed or learned [11, 12, 20]. This

paradigm is ubiquitous in signal processing for image denoising and inpainting [11, 12], and medical imaging [21, 22], to name a few examples. Learned dictionaries can improve reconstruction performance over prescribed dictionaries and recently, inversion methods with dictionary learning have been developed in the geosciences. Applications include denoising seismic traces [23] and ocean acoustic recordings [24], full waveform inversion [25], and estimation of ocean sound speed profiles [15, 16].

In this paper, we develop a sparse and adaptive approach to 2D travel time tomography, which we refer to as locally-sparse travel time tomography (LST). The LST sparsely models local behaviors of overlapping groups of pixels from a discrete slowness map, called *patches*. Large scale features in the slowness map are constrained using least squares. This approach is similar to works in image denoising [26] and CS magnetic resonance imaging (MRI) [22]. We develop a *maximum a posteriori* (MAP) formulation to the problem and use the iterative thresholding and signed K-means (ITKM) dictionary learning algorithm to improve the slowness models over prescribed dictionaries. We demonstrate the performance of LST considering 2D surface wave tomography with synthetic slowness maps and travel time data. The results are compared with conventional tomography. More details of the approach and further experimental results are available in a forthcoming paper [27].

2. OVERVIEW OF LST

Given travel time perturbations $\mathbf{t} \in \mathbb{R}^M$ from M ray paths through a discrete slowness map (see Fig. 1(a)), and tomography matrix $\mathbf{A} \in \mathbb{R}^{M \times N}$, LST estimates the *sparse slowness* \mathbf{s}_s . We first estimate the *global slowness* \mathbf{s}_g , and then obtain the *patch slowness* $\mathbf{D}\mathbf{x}_i$ for patch i of \mathbf{s}_g . Here $\mathbf{D} \in \mathbb{R}^{n \times Q}$ is a dictionary of Q atoms, and $\mathbf{x}_i \in \mathbb{R}^n$ is the sparse coefficients with n the number of pixels in a patch. Finally slownesses $\{\mathbf{D}\tilde{\mathbf{x}}_i \forall i\}$ are averaged with \mathbf{s}_g to obtain \mathbf{s}_s .

2.1. Global slowness and travel time

We discretize a 2D slowness map as a $W_1 \times W_2$ pixel image, shown in Fig. 1(a), where each pixel has constant slowness. The slowness pixels are represented by the vector

This work is supported by the Office of Naval Research, Grant No. N00014-11-1-0439.

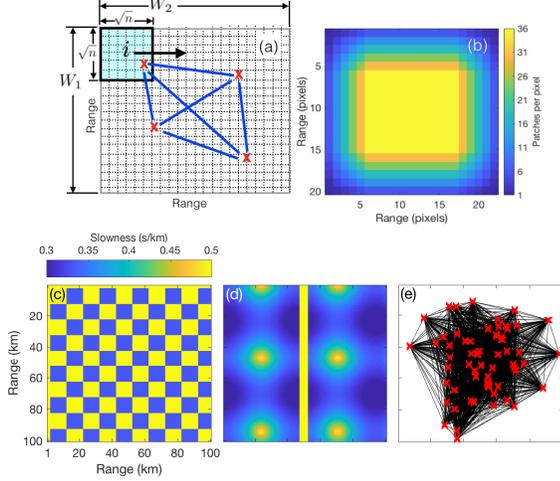


Fig. 1: (a) 2D slowness patches and slowness map parameters, with (b) example patch distribution. Synthetic slowness s' for (c) checkerboard map and (d) smooth-discontinuous map ($W_1 = W_2 = 100$ pixels (km)). (e) 2016 straight ray paths (surface wave) from 64 seismic stations (red X's).

$s' = s_g + s_0 \in \mathbb{R}^N$, where s_0 is reference slownesses and s_g is perturbations from the reference, with $N = W_1 W_2$. We assume travel time observations $t' = t + t_0$ from M straight ray paths, where t_0 and t are the reference travel time and perturbations. Since s_0 and $t_0 = \mathbf{A}s_0$ are known, we estimate the perturbations

$$t = \mathbf{A}s_g + \epsilon, \quad (1)$$

where $\epsilon \in \mathbb{R}^M$ is Gaussian noise $\mathcal{N}(\mathbf{0}, \sigma_\epsilon^2 \mathbf{I})$, with mean $\mathbf{0}$ and covariance $\sigma_\epsilon^2 \mathbf{I}$, in the travel time observations. We call (1) the *global model*, as it captures the large-scale features that span the discrete map and generates t . We assume dense ray coverage, and do not explicitly account for varying ray density (see Sec. 3).

2.2. Local sparse model

Each patch is a $\sqrt{n} \times \sqrt{n}$ group of pixels from s_s , see Fig. 1(a). The patches are selected from s_s by the binary matrix $\mathbf{R}_i \in \{0, 1\}^{n \times N}$. Hence the slownesses in patch i are $\mathbf{R}_i s_s$. Each patch is indexed by the row w_1 and column w_2 of its top-left pixel in the 2D image as $(w_{1,i}, w_{2,i})$. We consider all overlapping patches, with $w_{1,i}$ and $w_{2,i}$ differing from their neighbor by ± 1 (*stride* of one). Thus, for a $W_1 \times W_2$ pixel image, the number of patches is $I = (W_1 - \sqrt{n} + 1)(W_2 - \sqrt{n} + 1)$.

$\mathbf{R}_i s_s$ is approximated by sparse combinations atoms from \mathbf{D} . The coefficients \mathbf{x}_i are estimated using the ℓ_0 pseudo-norm (see (7)), which penalizes the number of non-zero coefficients [12]. We call (7) the local model, as it captures the smaller scale, localized features contained by patches.

The atoms in \mathbf{D} are considered ‘‘elemental patches’’, where only a small number of atoms from $Q \ll I$ are

necessary to adequately approximate $\mathbf{R}_i s_s$. Atoms can be prescribed functions, e.g. wavelets or the discrete cosine transform (DCT), or learned from the data (see Sec. 3.2).

3. DERIVATION OF LST MAP OBJECTIVE

Starting with Bayes’ rule, we derive the LST MAP objective for s_s , incorporating both local sparse prior and global constraints. For the derivation, we assume the dictionary \mathbf{D} and sensing matrix \mathbf{A} known. In Sec. 3.2, dictionary learning is included in the algorithm.

The posterior density is formulated as

$$p(s_g, s_s, \mathbf{X} | t) \propto p(t | s_g, s_s, \mathbf{X}) p(s_g | s_s, \mathbf{X}) p(s_s | \mathbf{X}) p(\mathbf{X}), \quad (2)$$

where $\mathbf{X} = [\mathbf{x}_1, \dots, \mathbf{x}_I] \in \mathbb{R}^{Q \times I}$ are the coefficients describing all patches. If s_g (s_s) is known, so is t (s_g), whereby

$$p(s_g, s_s, \mathbf{X} | t) \propto p(t | s_g) p(s_g | s_s) p(s_s | \mathbf{X}) p(\mathbf{X}). \quad (3)$$

We assume $p(t | s_g)$, $p(s_g | s_s)$, and $p(s_s | \mathbf{X})$ are Gaussian, which results in a simple LST objective. Hence, for the global model, $p(t | s_g) = \mathcal{N}(\mathbf{A}s_g, \Sigma_\epsilon)$ and $p(s_g | s_s) = \mathcal{N}(s_g, \Sigma_g)$ where $\Sigma_\epsilon \in \mathbb{R}^{K \times K}$ is the covariance of ϵ and $\Sigma_g \in \mathbb{R}^{N \times N}$ is the covariance of s_g .

For the local model, the patch slownesses $\{\mathbf{R}_i s_s \forall i\}$ are considered independent, giving the local likelihood $p(s_s | \mathbf{X})$

$$p(s_s | \mathbf{X}) = \prod_i p(\mathbf{R}_i s_s | \mathbf{x}_i) = \prod_i \mathcal{N}(\mathbf{D}\mathbf{x}_i, \Sigma_{p,i}) \quad (4)$$

where $\Sigma_{p,i} \in \mathbb{R}^{n \times n}$ is the covariance of the patch slownesses for each patch i . Assuming the coefficients \mathbf{x}_i independent and sparse, $\ln p(\mathbf{X}) = \sum_i \ln p(\mathbf{x}_i)$, with $\ln p(\mathbf{x}_i) \propto \|\mathbf{x}_i\|_0$. We further assume the number of non-zero coefficients T is the same for every patch (for which the ℓ_0 -norm penalty is well suited), and errors iid with $\Sigma_\epsilon = \sigma_\epsilon^2 \mathbf{I}$, $\Sigma_g = \sigma_g^2 \mathbf{I}$, and $\Sigma_{p,i} = \sigma_{p,i}^2 \mathbf{I}$, where \mathbf{I} is the identity matrix. Hence the MAP estimate $\{\hat{s}_g, \hat{s}_s, \hat{\mathbf{X}}\}$ is from (3)

$$\{\hat{s}_g, \hat{s}_s, \hat{\mathbf{X}}\} = \arg \min_{s_g, s_s, \mathbf{X}} \left\{ \frac{1}{\sigma_\epsilon^2} \|\mathbf{t} - \mathbf{A}s_g\|_2^2 + \frac{1}{\sigma_s^2} \|s_g - s_s\|_2^2 + \frac{1}{\sigma_{p,i}^2} \sum_i \|\mathbf{D}\mathbf{x}_i - \mathbf{R}_i s_s\|_2^2 \right\} \text{ subject to } \|\mathbf{x}_i\|_0 = T \forall i. \quad (5)$$

3.1. Solving for the MAP estimate

We find the MAP estimates $\{\hat{s}_g, \hat{s}_s, \hat{\mathbf{X}}\}$ solving (5) via block-coordinate minimization, similar to [22, 26]. The global objective is written from (5)

$$\hat{s}_g = \arg \min \|t - \mathbf{A}s_g\|_2^2 + \lambda_1 \|s_g - s_s\|_2^2, \quad (6)$$

where $\lambda_1 = (\sigma_\epsilon / \sigma_g)^2$ is a regularization parameter.

The local objective from (5) for each patch is solved with $s_s = \hat{s}_g$ (decoupling the local and global objectives), giving

$$\hat{\mathbf{x}}_i = \arg \min_{\mathbf{x}_i} \|\mathbf{D}\mathbf{x}_i - \mathbf{R}_i \hat{s}_g\|_2^2 \text{ subject to } \|\mathbf{x}_i\|_0 = T. \quad (7)$$

Given: $\mathbf{t} \in \mathbb{R}^M$, $\mathbf{A} \in \mathbb{R}^{M \times N}$, $\mathbf{s}_s^0 = \mathbf{0} \in \mathbb{R}^N$, $\mathbf{D}^0 = \text{Haar}$,
DCT (or) noise $\mathcal{N}(0, 1) \in \mathbb{R}^{n \times Q}$, λ_1, λ_2, T , and $j = 1$

Repeat until convergence:

1. Global estimate: solve (6) using LSQR [28],

$$\hat{\mathbf{s}}_g^j = \arg \min_{\mathbf{s}_g^j} \|\mathbf{A}\mathbf{s}_g^j - \mathbf{t}\|_2^2 + \lambda_1 \|\mathbf{s}_g^j - \mathbf{s}_g^{j-1}\|_2^2.$$

2. Local estimate

- a: Setting $\mathbf{s}_s^j = \hat{\mathbf{s}}_g^j$, center patches $\{\mathbf{R}_i \hat{\mathbf{s}}_g^j \forall i\}$ and
- i. (Dictionary learning) Find \mathbf{D}^j using ITKM [20].
 - i. (Prescribed dictionary) Set $\mathbf{D}^j = \mathbf{D}^0$.
 - ii. Solve (7) using OMP,

$$\hat{\mathbf{x}}_i^j = \arg \min_{\mathbf{x}_i^j} \|\mathbf{D}^j \mathbf{x}_i^j - \mathbf{R}_i \hat{\mathbf{s}}_g^j\|_2^2 \text{ subject to } \|\mathbf{x}_i^j\|_0 = T.$$

b: Obtain $\hat{\mathbf{s}}_{s,n}^j$ by (10) as

$$\hat{\mathbf{s}}_{s,n}^j = \frac{\lambda_2 \hat{\mathbf{s}}_{g,n}^j + b_n s_{p,n}^j}{\lambda_2 + b_n}$$

$$j = j + 1$$

Table 1: Sparse travel time tomography (LST) algorithm with fixed or adaptive dictionaries

With $\hat{\mathbf{X}} = [\hat{\mathbf{x}}_1, \dots, \hat{\mathbf{x}}_I]$ from (7) and $\hat{\mathbf{s}}_g$ from (6), we find $\hat{\mathbf{s}}_s$ from (5), assuming $\sigma_{p,i}^2 = \sigma_p^2$

$$\hat{\mathbf{s}}_s = \arg \min_{\mathbf{s}_s} \lambda_2 \|\hat{\mathbf{s}}_g - \mathbf{s}_s\|_2^2 + \sum_i \|\mathbf{D}\hat{\mathbf{x}}_i - \mathbf{R}_i \mathbf{s}_s\|_2^2, \quad (8)$$

where $\lambda_2 = (\sigma_p/\sigma_g)^2$ is a regularization parameter. The estimate $\hat{\mathbf{s}}_s$ is obtained analytically from (8) by

$$\hat{\mathbf{s}}_s = \left(\lambda_2 \mathbf{I} + \sum_i \mathbf{R}_i^T \mathbf{R}_i \right)^{-1} \left(\lambda_2 \hat{\mathbf{s}}_g + \sum_i \mathbf{R}_i^T \mathbf{D}\hat{\mathbf{x}}_i \right), \quad (9)$$

which averages the pixels from the patch estimates $\{\mathbf{D}\hat{\mathbf{x}}_i \forall i\}$, with weight given to $\hat{\mathbf{s}}_g$ by λ_2 . From (9), the average patch slowness is $\mathbf{s}_p = \left(\sum_i \mathbf{R}_i^T \mathbf{R}_i \right)^{-1} \left(\sum_i \mathbf{R}_i^T \mathbf{D}\hat{\mathbf{x}}_i \right)$, with $\mathbf{b} = \text{diag}(\sum_i \mathbf{R}_i^T \mathbf{R}_i) \in \mathbb{Z}^N$ the number of patches per pixel. Hence, (9) is expressed as an operation at pixel n by

$$\hat{\mathbf{s}}_{s,n} = \frac{\lambda_2 \hat{\mathbf{s}}_{g,n} + b_n s_{p,n}}{\lambda_2 + b_n}. \quad (10)$$

3.2. LST algorithm with dictionary learning

The results (6), (7), and (9) give the LST algorithm for estimating \mathbf{s}_s , shown in Table 1, as a MAP estimate with local sparse priors using a prescribed dictionary \mathbf{D} . Dictionary learning via the ITKM [20] is added to the LST in the solution to the local objective (7). The global objective (6) is solved using the sparse least squares program LSQR [28]. The local objective (7) is solved using OMP after the slowness patches $\{\mathbf{R}_i \hat{\mathbf{s}}_g \forall i\}$ are centered [11].

The complexity of each LST iteration is determined primarily by LSQR computation in the global estimate, $O(2MN)$, and by ITKM $O(knQI)$ and OMP $O(TnQI)$ in the local estimate, where k is the ITKM iterations (see

Table 1). For large slowness maps, we expect the LST complexity to be dominated by LSQR. In our simulations we obtain reasonable run times (see Sec. 4.1).

3.3. Conventional tomography

We illustrate conventional tomography with a Bayesian approach [29], which enforces smoothness regularization with a global (non-diagonal) covariance. Considering the measurements (1), the MAP estimate of the slowness is

$$\hat{\mathbf{s}}_g = (\mathbf{A}^T \mathbf{A} + \eta \boldsymbol{\Sigma}_L^{-1})^{-1} \mathbf{A}^T \mathbf{t}, \quad (11)$$

where $\eta = (\sigma_\epsilon/\sigma_c)^2$ is a regularization parameter, σ_c is the conventional slowness variance, and smoothness $\boldsymbol{\Sigma}_L(i, j) = \exp(-D_{i,j}/L)$. Here, $D_{i,j}$ is the distance between cells i and j , and L is the length scale [29, 30].

4. SIMULATION

We demonstrate the performance of LST (Sec. 3, Table 1) relative to a conventional tomography (Sec. 3.3). Experiments are conducted using simulated travel times from two synthetic 2D slowness maps (Fig. 1(c,d)) with dimensions $W_1 = W_2 = 100$ pixels (km). The checkerboard pattern (Fig. 1(c)) contains only discontinuous slowness whereas the smooth-discontinuous map (Fig. 1(d)) contains a fault-like discontinuity in a smooth map. The slowness estimates from LST are plotted as $\hat{\mathbf{s}}'_s = \hat{\mathbf{s}}_s + \mathbf{s}_0 \in \mathbb{R}^N$, and for conventional $\hat{\mathbf{s}}'_g = \hat{\mathbf{s}}_g + \mathbf{s}_0 \in \mathbb{R}^N$.

The slownesses are sampled by $M = 2016$ straight-rays between 64 seismic stations (see Fig. 1(e)). The travel time vector \mathbf{t} is found by integrating along these ray paths. We consider only straight ray propagation, to focus on our proposed inversion approach. The reference slowness is calculated from the mean travel time using the tomography matrix \mathbf{A} . The LST inversion valid-region is obtained with a dilation operation with a patch template along the outermost ray paths. The conventional valid region is the outermost pixels along the ray paths. The conventional valid region is used for error calculations for both methods.

We consider the noise-free case ($\sigma_\epsilon = 0$). The results of the LST and conventional tomography are shown in Figs. 2 and 3. RMSE (s/km) of the estimates $\hat{\mathbf{s}}'_s$ and $\hat{\mathbf{s}}'_g$ relative to the true slowness \mathbf{s}' is printed on the 2D estimates. We invert using LST with and without dictionary learning. We consider two prescribed dictionaries \mathbf{D} , overcomplete Haar wavelet and DCT dictionaries (both $Q = 169$, $n = 64$, since Haar wavelet dimensions power of 2).

4.1. Inversion parameters and results

The regularization parameter values for LST and conventional tomography were selected to minimize RMSE (s/km). For LST, the best parameters were: for both prescribed dictionary

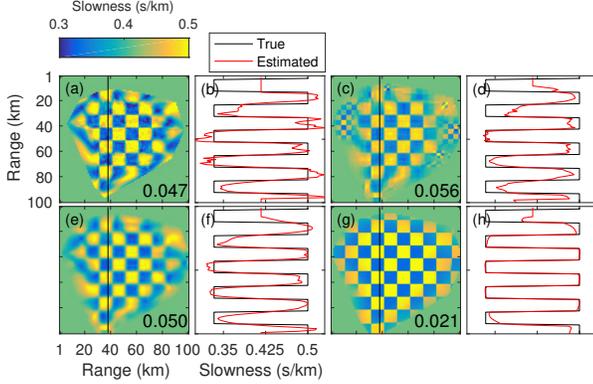


Fig. 2: LST and conventional tomography for checkerboard map (Fig. 1(c)). 2D and 1D (from black line in 2D) slowness estimates against true slowness for: (a,b) conventional \hat{s}'_g ; LST \hat{s}'_s with (c,d) Haar dictionary and (e,f) with DCT dictionary \mathbf{D} ; and (g,h) with dictionary learning. RMSE (s/km) is printed on each 2D image.

and dictionary learning, $\lambda_1 = 0 \text{ km}^2$ (in (6)) and $\lambda_2 = 0$ (in (8)); for prescribed dictionaries $T = 2$ non-zero coefficients in (7); for dictionary learning, $T = 1$, $n = 100$, and for the checkerboard (smooth-discontinuous) $Q = 166$ ($Q = 268$ atoms). Since for the noise free case $\sigma_\epsilon = 0$, we expect $\lambda_1 = (\sigma_\epsilon/\sigma_g)^2 = 0 \text{ km}^2$ to be best. We assume the slowness patches are well approximated by the sparse model (7), and expect $\sigma_p \ll \sigma_g$. Hence, we expect the best value of $\lambda_2 = (\sigma_p/\sigma_g)^2$ in (8) to be small. For conventional tomography (Sec. 3.3), the best parameters were $L = 10 \text{ km}$ and $\eta = 0.1 \text{ km}^2$ (in (11)) for the both the checkerboard and smooth-discontinuous maps, which deviates from the expected value of $\eta = (\sigma_\epsilon/\sigma_c)^2 = 0$.

While the discontinuous shapes in the Haar dictionary are similar to the discontinuous content of the checkerboard image, the local features in the higher order Haar wavelets overfit the ray sampling where sampling is poor (near the edges of the inversion, Fig. 2(c,d)). The performance of the Haar wavelets is better for the smooth-discontinuous slowness map (Fig. 3(d-f)) than for the checkerboard. As shown in Fig. 3(d-f), the Haar wavelets add false high frequency structure to the slowness reconstruction but the trends in the smooth-discontinuous features are well preserved. The inversion performance of the DCT transform (Fig. 2(e,f) and Fig. 3(g,j)) is better than the Haar wavelets for both cases, but matches less closely the discontinuous slowness features, as the DCT atoms are smooth. The smoothness of the DCT atoms better preserve the smooth slowness structure.

The LST with dictionary learning (Fig. 2(g,h) and Fig. 3(j-l)) achieves the best RMSE relative to the true slowness s' . As in the other cases, the performance degrades near the edges of the ray sampling, where the ray coverage is poor, but high resolution is maintained across a large part of the sampling region. The RMSE of the Haar wavelet inversion for the checkerboard is greater than for the conventional method,

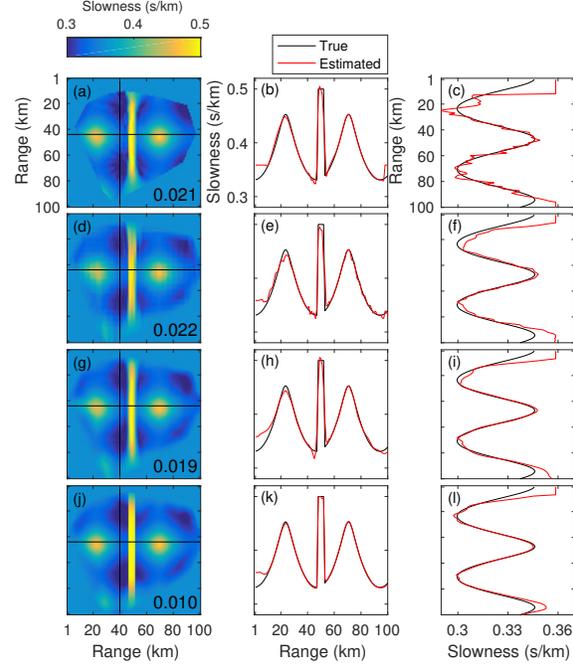


Fig. 3: LST and conventional tomography for smooth-discontinuous map (Fig. 1(d)). 2D and 1D (horizontal and vertical, from black lines in 2D) slowness estimates against true slowness for: (a-c) conventional \hat{s}'_g ; LST \hat{s}'_s with (d-f) Haar dictionary and (g-i) with DCT dictionary \mathbf{D} ; and (j-l) with dictionary learning. RMSE (s/km) is printed on each 2D image.

although resolution is lost in the conventional MAP inversion near the more densely sampled region of the wave speed maps. The RMSE for the DCT is less than the that of the Haar wavelets and also the conventional MAP inversion. A better qualitative fit to the true slowness is also observed.

The LST algorithm (Table 1) used 100 iterations for all cases and the ITKM used 50 iterations. In Matlab, the inversion with dictionary learning took ~ 5 min on a Macbook Pro 2.5 GHz Intel Core i7.

5. CONCLUSIONS

We derived a travel time tomography method which incorporates a sparse prior on patches of the slowness image, which we refer to as the LST algorithm. The LST uses prescribed or learned dictionaries, though the learned dictionaries improve performance. The local sparse prior and dictionary learning provide an improved slowness model, which is capable of modeling simultaneously smooth discontinuous features.

We considered 2D surface wave tomography, and for densely sampled slowness maps obtained superior results from the LST over conventional tomography. The LST is relevant to other tomography scenarios where slowness structure is irregularly sampled, in for instance ocean [16] and terrestrial [31] acoustics.

6. REFERENCES

- [1] R.C. Aster, B. Borchers, and C.H. Thurber, *Parameter estimation and inverse problems*, Elsevier, San Diego, 2nd edition, 2013.
- [2] N. Rawlinson, S. Pozgay, and S. Fishwick, “Seismic tomography: a window into deep earth,” *Phys. Earth and Planetary Interiors*, vol. 178, no. 3, pp. 101–135, 2010.
- [3] L.Y. Chiao and B.Y. Kuo, “Multiscale seismic tomography,” *Geophys. J. Int.*, vol. 145, no. 2, pp. 517–527, 2001.
- [4] I. Loris, G. Nolet, I. Daubechies, and F.A. Dahlen, “Tomographic inversion using ℓ_1 -norm regularization of wavelet coefficients,” *Geophys. J. Int.*, vol. 170, no. 1, pp. 359–370, 2007.
- [5] R. Hawkins and M. Sambridge, “Geophysical imaging using trans-dimensional trees,” *Geophys. J. Int.*, vol. 203, no. 2, pp. 972–1000, 2015.
- [6] H. Fang and H. Zhang, “Wavelet-based double-difference seismic tomography with sparsity regularization,” *Geophys. J. Int.*, vol. 199, no. 2, pp. 944–955, 2014.
- [7] Y. Lin and L. Huang, “Quantifying subsurface geophysical properties changes using double-difference seismic-waveform inversion with a modified total-variation regularization scheme,” *Geophys. J. Int.*, vol. 203, no. 3, pp. 2125–2149, 2015.
- [8] X. Zhang and J. Zhang, “Model regularization for seismic traveltimes tomography with an edge-preserving smoothing operator,” *J. Appl. Geophys.*, vol. 138, pp. 143–153, 2017.
- [9] T. Bodin, M. Sambridge, and K. Gallagher, “A self-parametrizing partition model approach to tomographic inverse problems,” *Inverse Problems*, vol. 25, no. 5, 2009.
- [10] E. Candés, “Compressive sampling,” *Proc. Int. Congr. Math.*, vol. 3, pp. 1433–1452, 2006.
- [11] J. Mairal, F. Bach, and J. Ponce, “Sparse modeling for image and vision processing,” *Found. Trends Comput. Graph. Vis.*, vol. 8, no. 2–3, pp. 85–283, 2014.
- [12] M. Elad, *Sparse and Redundant Representations*, Springer, New York, 2010.
- [13] A. Xenaki, P. Gerstoft, and K. Mosegaard, “Compressive beamforming,” *J. Acoust. Soc. Am.*, vol. 136, no. 1, pp. 260–271, 2006.
- [14] P. Gerstoft, C.F. Mecklenbräuker, A. Xenaki, and S. Nannuru, “Multisnapshot sparse Bayesian learning for DOA,” *IEEE Signal Process. Lett.*, vol. 23, no. 10, pp. 1469–1473, 2016.
- [15] T. Wang and W. Xu, “Sparsity-based approach for ocean acoustic tomography using learned dictionaries,” *OCEANS 2016 Shanghai IEEE*, pp. 1–6, 2016.
- [16] M. Bianco and P. Gerstoft, “Compressive acoustic sound speed profile estimation,” *J. Acoust. Soc. Am.*, vol. 139, no. 3, pp. EL90–EL94, 2016.
- [17] M. Bianco and P. Gerstoft, “Dictionary learning of sound speed profiles,” *J. Acoust. Soc. Am.*, vol. 141, no. 3, pp. 1749–1758, 2017.
- [18] M. Bianco and P. Gerstoft, “Regularization of geophysical inversion using dictionary learning,” *IEEE Int. Conf. Acoust., Speech, and Signal Process. (ICASSP)*, 2017.
- [19] J. Charléty, S. Voronin, G. Nolet, I. Loris, F.J. Simons, K. Sigloch, and I.C. Daubechies, “Global seismic tomography with sparsity constraints: comparison with smoothing and damping regularization,” *J. Geophys. Research: Solid Earth*, vol. 118, no. 9, pp. 4887–4899, 2013.
- [20] K. Schnass, “Local identification of overcomplete dictionaries,” *J. Mach. Learning Research*, vol. 16, pp. 1211–1242, 2015.
- [21] M. Lustig, D. Donoho, and J.M. Pauly, “Sparse MRI: the application of compressed sensing for rapid MR imaging,” *Magn. Reson. Med.*, vol. 58, no. 6, pp. 1182–1195, 2007.
- [22] S. Ravishankar and Y. Bresler, “MR image reconstruction from highly undersampled k-space data by dictionary learning,” *IEEE Trans. Med. Imag.*, vol. 30, no. 5, pp. 1028–1041, 2011.
- [23] Y. Chen, “Fast dictionary learning for noise attenuation of multidimensional seismic data,” *Geophys. J. Int.*, vol. 209, no. 1, pp. 21–31, 2017.
- [24] M. Taroudakis and C. Smaragdakis, “De-noising procedures for inverting underwater acoustic signals in applications of acoustical oceanography,” *EuroNoise 2015 Maastricht*, pp. 1393–1398, 2015.
- [25] L. Zhu, E. Liu, and J.H. McClellan, “Sparse-promoting full waveform inversion based on online orthonormal dictionary learning,” *Geophys.*, vol. 82, no. 2, pp. R87–R107, 2017.
- [26] M. Elad and M. Aharon, “Image denoising via sparse and redundant representations over learned dictionaries,” *IEEE Trans. Image Process.*, vol. 15, no. 12, pp. 3736–3745, 2006.
- [27] M. Bianco and P. Gerstoft, “Sparse travel time tomography with adaptive dictionaries,” *ArXiv e-prints*, Jan. 2018, <https://arxiv.org/abs/1712.08655v2>.
- [28] C.C. Paige and M.A. Saunders, “LSQR: an algorithm for sparse linear equations and sparse least squares,” *ACM Trans. Math. Software*, vol. 8, no. 1, pp. 43–71, 1982.
- [29] C.D. Rodgers, *Inverse methods for atmospheric sounding: theory and practice*, World Sci. Pub. Co., 2000.
- [30] A. Tarantola, *Inverse problem theory*, Elsevier Sci. Pub. Co., Inc., 1987.
- [31] R. Rabenstein and P. Annibale, “Acoustic source localization under variable speed of sound conditions,” *Wireless Commun. Mob. Comp.*, 2017.