Broadband dispersion extraction using simultaneous sparse penalization

Shuchin Aeron*, Sandip Bose, Henri-Pierre Valero and Venkatesh Saligrama

Abstract

In this paper we propose a broadband method to extract the dispersion curves for multiple overlapping dispersive modes from borehole acoustic data under limited spatial sampling. The proposed approach exploits a first order Taylor series approximation of the dispersion curve in a band around a given (center) frequency in terms of the phase and group slowness at that frequency. Under this approximation, the acoustic signal in a given band can be represented as a superposition of broadband propagators each of which is parameterized by the slowness pair above. We then formulate a sparsity penalized reconstruction framework as follows. These broadband propagators are viewed as elements from an overcomplete dictionary representation and under the assumption that the number of modes is small compared to the size of the dictionary, it turns out that an appropriately reshaped *support* image of the coefficient vector synthesizing the signal (using the given dictionary representation) exhibits *column sparsity*. Our main contribution lies in identifying this feature and proposing a complexity regularized algorithm for *support recovery* with an ℓ_1 type simultaneous sparse penalization. Note that support recovery in this context amounts to recovery of the broadband propagators comprising the signal and hence extracting the dispersion, namely, the group and phase slownesses of the modes. In this direction we present a novel method to select the regularization parameter based on Kolmogorov-Smirnov (KS) tests on the distribution of residuals for varying values of the regularization parameter. We evaluate the performance of the proposed method on synthetic as well as real data and show its performance in dispersion extraction under presence of heavy noise and strong interference from time overlapped modes.

I. Introduction

In this work we consider the problem of dispersion extraction from borehole acoustic data. Dispersion refers to a systematic variation of the propagation slowness (inverse of velocity) of waves. Analysis of dispersion reveals

Shuchin Aeron is with the department of ECE at Tufts University, Medford, MA 02155. email: aerons@gmail.com

Sandip Bose is with the Math and Modeling department at Schlumberger Doll Research, Cambridge, MA - 02139. email: bose1@slb.com

Henri-Pierre Valero is with Schlumberger Kabushiki Kaisha Center Schlumberger KK 2-2-1 Fuchinobe, Chuo, Sagamihara, Kanagawa 252-0206. email: hvalero@slb.com

Venkatesh Saligrama is with the Department of ECE at Boston University, Boston, MA 02215, email: srv@bu.edu.

This research was supported by a Schlumberger Doll Research grant. This work was presented in part at ICASSP 2010 and constitutes a part of the PhD thesis of the first author.

important mechanical properties of the rock formation surrounding the borehole and thus dispersion analysis is of considerable interest to the geophysical and oilfield services community. A brief survey of borehole acoustic waves and their use in mechanical characterization of the rock properties can be found in [1], [2]. Acoustic waves are excited in the borehole by firing a source, e.g. monopole, dipole or quadrupole source, which produces *modes* with the borehole as a waveguide. A schematic is shown in Figure 1. These borehole modes are dispersive in nature, i.e. the phase slowness is a function of frequency. This function characterizes the mode and is referred to as a dispersion curve. In the problem of dispersion extraction one is interested in estimating the dispersion curves of each of the generated modes.

For the readers who are unfamiliar with the area it is useful to mathematically formalize the set-up and introduce relevant notions and definitions. The relation between the received waveforms and the wavenumber-frequency $(f-k)$ response of the borehole to the source excitation is expressed via the following equation,

$$s(l,t) = \int_0^\infty \int_0^\infty S(f)Q(k,f)e^{i2\pi ft}e^{-ikz_l}df dk \tag{1}$$

for $l = 1, 2, ..., L$ and where $s(l,t)$ denotes the recorded pressure at time t at the l-th receiver located at a distance z_l from the source; $S(f)$ is the source spectrum and $Q(k,f)$ is the wavenumber-frequency response of the borehole. Typically the data is acquired in presence of noise (environmental noise and receiver noise) and interference which we collectively denote by $w(l,t)$. In this work we assume that the noise is *incoherent* additive noise process. Then the noisy observations at the set of receivers can be written as,

$$y(l,t) = s(l,t) + w(l,t) \tag{2}$$

It has been shown that the complex integral in the wavenumber (k) domain in Equation(1) can be approximated by contribution due to the residues of the poles of the system response, [3],[1],[4]. Specifically,

$$\int_0^\infty Q(k,f)e^{-ikz_l}dk \sim \sum_{m=1}^{M(f)} q_m(f)e^{-(i2\pi k_m(f)+a_m(f))z_l} \tag{3}$$

where $q_m(f)$ is the residue of the m-th pole and respectively where $k_m(f)$ and $a_m(f)$ are the corresponding real and imaginary parts of the poles and can be interpreted as the wavenumber and the attenuation as functions of frequency. This means that we have

$$s(l,t) = \int_0^\infty \sum_{m=1}^{M(f)} S_m(f)e^{-(i2\pi k_m(f)+a_m(f))z_l}e^{i2\pi ft}df \tag{4}$$

where $S_m(f) = S(f)q_m(f)$. Under this signal model, the data acquired at each frequency across the receivers can be written as,

$$\underbrace{\begin{bmatrix} Y_1(f) \\ Y_2(f) \\ \vdots \\ \vdots \\ Y_L(f) \end{bmatrix}}_{\mathbf{Y}(f)} = \begin{bmatrix} e^{-(i2\pi k_1(f)+a_1(f))z_1} & \dots & e^{-(i2\pi k_M(f)+a_M(f))z_1} \\ e^{-(i2\pi k_1(f)+a_1(f))z_2} & \dots & e^{-(i2\pi k_M(f)+a_M(f))z_2} \\ \vdots & \dots & \vdots \\ e^{-(i2\pi k_1(f)+a_1(f))z_L} & \dots & e^{-(i2\pi k_M(f)+a_M(f))z_L} \end{bmatrix} \underbrace{\begin{bmatrix} S_1(f) \\ S_2(f) \\ \vdots \\ \vdots \\ S_M(f) \end{bmatrix}}_{\mathbf{S}(f)} + \underbrace{\begin{bmatrix} W_1(f) \\ W_2(f) \\ \vdots \\ \vdots \\ W_L(f) \end{bmatrix}}_{\mathbf{W}(f)} \tag{5}$$

In other words the data at each frequency is a superposition of M exponentials sampled with respect to the receiver locations $z_1, .., z_L$. We refer to the above system of equations as corresponding to a sum of exponentials model at each frequency. $M(f)$ is the effective number of exponentials at frequency f.

At this point we would like to formalize the notion of *mode* and *dispersion curve*.

Definition 1.1: Mode: A mode is a waveform propagating along the borehole (acting as a waveguide) that is characterized by its wavenumber $k_m(f)$ and attenuation $a_m(f)$ which are the real and imaginary parts of a continuous locus of poles in the $f-k$ domain, also known as the dispersion curve or characteristic of the mode, governing the waveguide response to the mode. When excited by a source pulse a mode is *transient* and therefore often *time compact*. It is customary to express the dispersion characteristics in terms of *propagation parameters*, viz., phase slowness given by $s_m^\phi = \frac{k_m(f)}{f}$ at frequency f and group slowness which is the derivative $s_m^g = k_m^{'}(f)$ at a frequency f.

In other words the modal dispersion is completely characterized in the $(f-k)$ domain in terms of the wavenumber response $k(f)$ as a function of frequency. Note that the associated propagation parameters, viz., phase slowness and group slowness are not independent and obey the following relationship,

$$s_m^g = s_m^\phi + f\frac{\partial s_m^\phi}{\partial f} \tag{6}$$

In theory the number of modes can be infinite but in practice only a few are significant, i.e. the model order $M(f)$ is finite and is small. A typical example of the array waveforms and corresponding dispersion curves of the modes are shown in Figure 2.

Traditionally the problem of dispersion extraction has been treated in a narrow-band setting under the sum of exponential model outlined above. For this model an algorithm to estimate the $(f-k)$ response using a variation of Prony's polynomial method for estimation of signal poles at each frequency was proposed in [3]. In [5] a modified Matrix Pencil method of [6] for harmonic retrieval was proposed therein for estimating the wavenumber response at each frequency. These two methods, viz., Matrix Pencil method and Prony's polynomial method, and their variations exhibit similar performance. Under the sum of exponential model at each frequency the problem

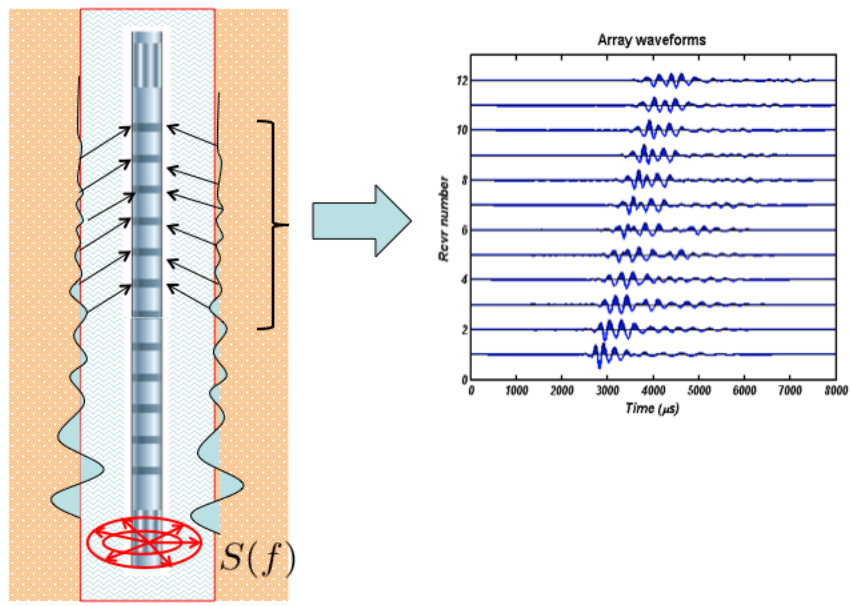


Fig. 1. Schematic showing the physical set-up for acquisition of waves in a borehole in a typical logging scenario. The window of acquisition is usually limited by time or memory constraints and is chosen to incorporate the main signals of interest.

can also be viewed as the problem of estimating the source directions and subspace based array processing methods like ESPRIT and MUSIC, [7],[8] can be employed here. It is worth pointing out that the Matrix Pencil (MP) based method is similar to ESPRIT and hence is essentially a subspace based method. All of these methods are sensitive to *model order*, the knowledge of signal subspace and to the noise, see [Chapter 16, [8]]. Schemes such as in [9] based on Minimum Description Length (MDL), [10] were proposed to estimate the signal subspace and the model order. Statistical tests based on Weighted Subspace Fitting (WSF) were proposed in [8] for estimating the number of sources when the signal subspace is estimated using MDL principle. However these methods work fine under some stationarity assumptions on the signal. In borehole acoustic applications however the signal is transient and highly non-stationary and these methods are not applicable in general.

With respect to dispersion extraction the narrowband methods clearly have some disadvantages. First and foremost these methods do not use the continuity of the dispersion curves and hence produce a scatter like plot in the $(f-k)$ plane. These dots require supervised labeling to obtain final dispersion curves of interest. Also they suffer from the problem of aliasing, see figure 3. Moreover being narrowband these methods are less robust to noise as the continuity of the dispersion curves, i.e. of the wavenumber response across frequencies is not exploited. In this work, in contrast to the narrowband methods our aim will be to propose a broadband method for dispersion extraction. First let us briefly review some previous broadband approaches in this context.

Among the broadband approaches several methods have been proposed for dispersion (or slowness) estimation.

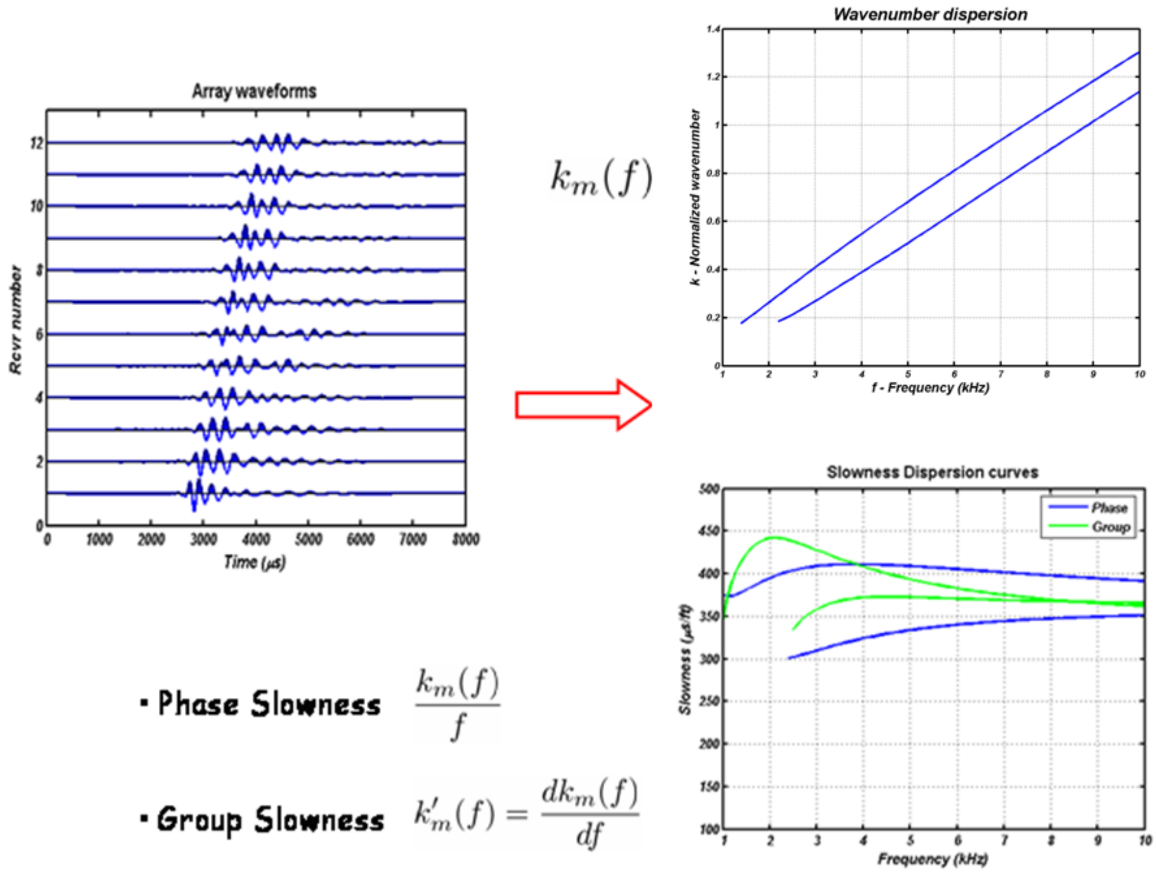


Fig. 2. Figure showing the array waveforms and modal dispersion curves governing the propagation of the acoustic energy. The plot on the upper right shows the phase slowness as a function of frequency given by $\frac{k_m(f)}{f}$ and the lower right plot shows the corresponding group slowness plots given by $\frac{dk_m(f)}{df}$. The task is to estimate these dispersion curves, in particular the phase and the group slowness given the received waveforms.

For non-dispersive modes, i.e. where the slowness is constant across frequencies in [11] a semblance based method was proposed in the space-time domain for slowness estimation. This method exploited the time compactness of the modes across the array and provides good results for time separated modes. The performance is poor when the modes are overlapping in time. Under a similar set-up Capon method of [12] was applied in [13] for sonic velocity logging. One can also use ML methods but the computational requirements are severe as one searches over all the parameters. Moreover model order selection is a problem. Fast ML was proposed in [14] in the context of harmonic retrieval problem for a narrowband set-up. There the idea is to perform successive estimation of the components starting with extracting the component with the highest energy and so on. Fast ML method in principle can be adapted to the broadband set-up in [13], however it is not suitable when the sources have comparable energy. Homomorphic processing of borehole acoustic waves was proposed in [15] and was shown to yield smoother estimates of the dispersion curves compared to Prony's polynomial method. However the method proposed there in is suitable only for single dominant mode. In [16] a broadband method based on linear approximation of the

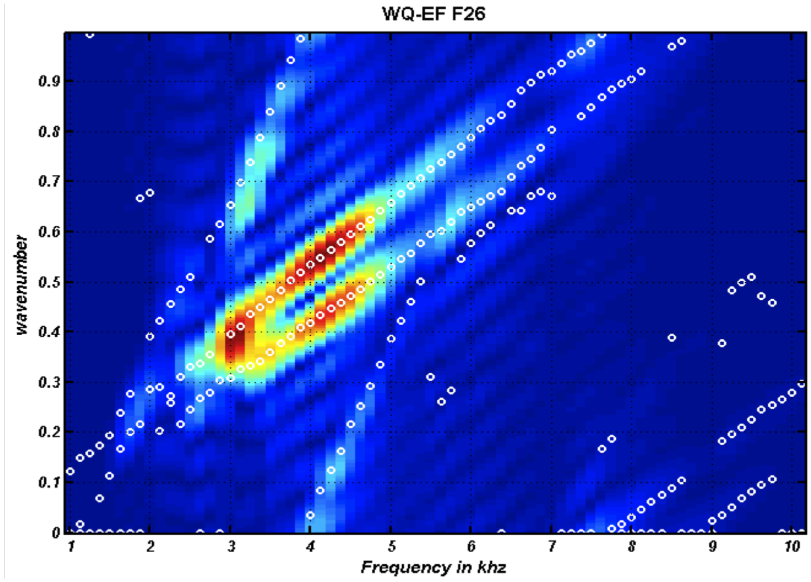


Fig. 3. Figure showing an example of narrowband processing of the array data in the f-k domain (2D FFT). Note the aliasing in wavenumber (normalized to $[0\,1]$) and other artifacts in dispersion extraction.

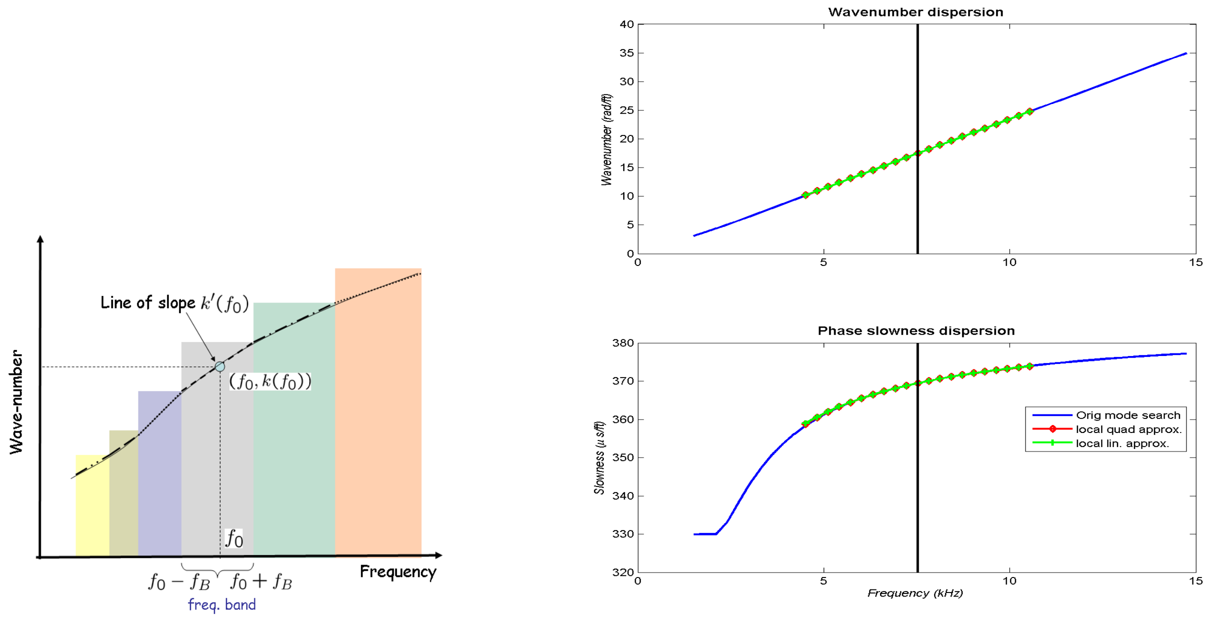
dispersion curves in the $(f-k)$ domain was proposed. The method requires solving a high dimensional non-linear, non-convex optimization problem which has a huge computational complexity. Special purpose dispersion extraction methods such SPI [17] have been proposed but are applicable to one particular kind of mode, e.g. flexural waves. More recently new broadband methods have been proposed that exploit the time-frequency localization of the modes for dispersion extraction, see [18], [19]. Similar ideas that exploit the time frequency localization have been considered before, e.g., the energy reassignment method of [20] that has been applied in [21] for analysis of dispersion of Lamb waves and in [22] for improving the group velocity measurements. However most of these methods are suitable only for cases where there is a significant time frequency separation of modes. In this work we will move away from these approaches and present a broadband dispersion strategy that also works when there is no time frequency separation of the modes.

A. *Organization of the paper*

The rest of the paper is organized as follows. In section II we will present the problem formulation in detail. In section III we will present the broadband dispersion extraction framework and methodology in the frequency-wavenumber (f-k) domain. In section IV we will show the performance of the method on a synthetic data set. In section V we show the performance of the proposed methodology on a real data set. Finally we conclude and provide future research directions in section VI.

II. Problem set-up

In this section we will present the problem formulation for dispersion extraction based on the sum of exponentials model of Equation (4). Here we will make use of the local linear approximation to the dispersion curves and pose



(a) (b)

Fig. 4. (a) Schematic showing the piecewise linear approximation to the dispersion curve. Shown also are the dispersion parameters corresponding to a first order Taylor series approximation in a band around f_0. (b) Shown in this figure is the effect of linear approximation on the slowness dispersion compared with a quadratic approximation.

a broadband problem for dispersion extraction.

To make the ideas mathematically precise we assume that the wavenumber as a function of frequency, i.e. the function(s) $k_m(f)$, can be well approximated by a first order Taylor series expansion in a certain band around a given frequency of interest, i.e. the following approximation,

$$k_m(f) \approx k_m(f_0) + k_m^{'}(f_0)(f - f_0), \; \forall f \in \mathcal{F} = [f_0 - f_B, f_0 + f_B] \tag{7}$$

is valid in a band f_B around a given *center frequency* f_0. In other words one can approximate the dispersion curve as composed of piecewise linear segments over disjoint frequency bands. For each frequency band the dispersion curve is parameterized by the phase and the group slowness. For attenuation we assume that it is constant over the frequency band of interest

$$a_m(f) \approx a_m(f_0), \; \forall f \in \mathcal{F} \tag{8}$$

A schematic depiction of local linear approximation to a dispersion curve over disjoint frequency bands is shown in Figure 4(a). In Figure 4(b) we compare the local linear approximation vs a local quadratic approximation. It can be seen that the linear approximation holds quite good.

In the following, without loss of generality assume that the number of modes $M(f)$ is the same for all frequencies in the band of interest. For sake of brevity we will denote this number by M in the following. Now note that under the linear approximation of the dispersion curve(s) for the modes, the exponential at a frequency f (sampled at receiver locations $z_i, i \in \{1, 2, .., L\}$) corresponding to a mode can be written in a parametric form as,

$$\mathbf{v}_m(f) = \begin{bmatrix} e^{-i2\pi(k_m + k_m'(f-f_0))z_1} e^{-a_m(f_0)z_1} \\ e^{-i2\pi(k_m + k_m'(f-f_0))z_2} e^{-a_m(f_0)z_2} \\ \vdots \\ e^{-i2\pi(k_m + k_m'(f-f_0))z_L} e^{-a_m(f_0)z_L} \end{bmatrix}, m = 1, 2, ..., M \ f, f_0 \in \mathcal{F} \tag{9}$$

One can immediately see that over the set of frequencies $f \in \mathcal{F}$, the collection of sampled exponentials (for a fixed m) $\{\mathbf{v}_m(f)\}_{f \in \mathcal{F}}$ as defined above corresponds to a line segment in the f-k domain thereby parametrizing the *wavenumber dispersion* of the mode in the band in terms of phase and group slowness. In the following we will represent the band $\mathcal{F}$ by F which is a finite set of frequencies contained in $\mathcal{F}$,

$$F = \left\{f_1, f_2, ..., f_{N_f}\right\} \subset \mathcal{F} : f_0 \in F \tag{10}$$

Essentially in a time sampled system with finite samples the discrete set of frequencies in F are taken to be the set of frequencies in the DFT of data $y(l,t)$. Note that under the assumption that the noise $w(l,t)$ is average white Gaussian noise (AWGN) with variance σ^2, it implies that the vector $\mathbf{W}_F$ is distributed as zero mean Gaussian with variance $\sigma^2\mathbf{I}$. Then under the linear approximation of dispersion curves in the band F, a broadband system of equations in the band can be written as,

$$\mathbf{Y}_F = \begin{bmatrix} \mathbf{Y}(f_1) \\ \mathbf{Y}(f_2) \\ \vdots \\ \mathbf{Y}(f_{N_f}) \end{bmatrix} = \begin{bmatrix} \mathbf{V}_M(f_1) & & & \\ & \mathbf{V}_M(f_2) & & \\ & & \ddots & \\ & & & \mathbf{V}_M(f_{N_f}) \end{bmatrix} \begin{bmatrix} \mathbf{S}(f_1) \\ \mathbf{S}(f_2) \\ \vdots \\ \mathbf{S}(f_{N_f}) \end{bmatrix} + \mathbf{W}_F \tag{11}$$

where $\mathbf{Y}(f_i)$ and $\mathbf{S}(f_i)$, $i = 1, 2, ..., N_f$ are defined in equation 5, where $\mathbf{V}_M(f) = [\mathbf{v}_1(f), \mathbf{v}_2(f)...\mathbf{v}_M(f)]$, $\forall f \in F$ and $\mathbf{W}_F = [\mathbf{W}^T(f_1)\mathbf{W}^T(f_2), ..., \mathbf{W}^T(f_{N_f})]^T$ is the noise in band F. In words the data in a frequency band is a superposition of exponentials at each frequency where the parameters of the complex exponentials are *linked* by phase and group slowness and attenuation factor across the frequencies. We can therefore propose an alternative representation based on collection of all the exponentials across the frequency for each mode like so,

$$\mathbf{P}_F(m) = \begin{bmatrix} \mathbf{v}_m(f_1) & & & \\ & \mathbf{v}_m(f_2) & & \\ & & \ddots & \\ & & & \mathbf{v}_m(f_{N_f}) \end{bmatrix} ; \ \mathbf{S}_F(m) = \begin{bmatrix} S_m(f_1) \\ S_m(f_2) \\ \vdots \\ S_m(f_{N_f}) \end{bmatrix} \tag{12}$$

The matrix $\mathbf{P}_F(m)$ corresponds to the matrix of exponentials for the wavenumber dispersion for mode m in the band F, and the vector $\mathbf{S}_F(m)$ is the vector of the mode spectrum for mode m in the band F. Then the signal (without noise) in band F is a superposition of M_F modes and can be written as

$$\mathbf{S}_F = \sum_{m=1}^{M} \mathbf{P}_F(m)\mathbf{S}_F(m) \tag{13}$$

Now note that under the piecewise linear approximation of the dispersion curves over disjoint sets of frequency bands it is sufficient to consider the problem of dispersion extraction in each band with the dispersion curves in the band parameterized by corresponding phase and group slowness. The entire dispersion curve can in turn be obtained by combining the dispersion curves obtained in each of these bands. The problem of combining the dispersion curves across the bands is important and is two fold; (a) It involves appropriate matching and smoothing of the estimates of dispersion curves at the boundaries of the disjoint bands and; (b) Labeling of the modes across bands, i.e. data (mode) association across bands. This is non-trivial in general and will be addressed in a separate work. Therefore, in this work in the following we will focus on dispersion extraction in a band. To this end we make the following assumption regarding the attenuation.

Assumption 2.1: Assume that for the given band the modal attenuation for each mode is negligible, i.e. $a_m(f_0) \approx 0$ for all $m = 1, 2, .., M$. This assumption is not very restricting as in many cases the attenuation is mild. Moreover ignoring the attenuation does not bias the wavenumber estimates. Extensions to handle attenuation is a subject of future work and is beyond the scope of this paper. Therefore the results and methods presented in this paper only extend to the cases with mild or little attenuation.

Problem Statement - Under the assumption 2.1 and given the data $\mathbf{Y}_F$ in the frequency band F we want to estimate the model order M, the slowness dispersion curves of the modes as modeled by the wavenumber response $\mathbf{P}_F(m)$ in the band F, and the mode spectrum $\mathbf{S}_F(m)$ for all $m = 1, 2, ..., M$.

At this point it is worth to point out two relevant pieces of work based for dispersion extraction based on local linear approximation to the dispersion curves over disjoint frequency bands. In [16], for each band a variation of minimum variance spectral estimator of [12] was proposed to jointly estimate the group and the phase slowness. Although this was shown to be superior to the conventional beamforming estimator in achieving a higher resolution, however it remains very sensitive to noise and model mismatch. To overcome this a Maximum-Likelihood (ML) method based on the same linear approximation of the dispersion curves in disjoint frequency bands was proposed in [23]. As pointed out before the resulting algorithm is combinatorial and non-convex in nature and therefore suffers from a huge computational complexity. Thus there was a lack of balance of computational complexity vs the estimation quality. The method proposed in this paper can be thought of as providing such a balance by suitable convex relaxation of the ML type estimator.

III. A SPARSE SIGNAL RECONSTRUCTION FRAMEWORK FOR DISPERSION EXTRACTION IN F-K DOMAIN

In this section we will introduce the framework used for dispersion extraction in a band for the broadband set-up proposed in the previous section.

We begin by introducing the following notion of a broadband basis element in a band.

Definition 3.1: A broadband basis element $\mathbf{P}_F(k(f_0), k'(f_0))$ (say in band F) corresponding to a given phase slowness $\frac{k(f_0)}{f_0}$ and group slowness $k'(f_0)$ is a block diagonal matrix $\in \mathbb{C}^{L \cdot N_f \times N_f}$ given by

$$\mathbf{P}_F(k(f_0), k'(f_0)) = \begin{bmatrix} \mathbf{\Phi}(f_1) & & & \\ & \mathbf{\Phi}(f_2) & & \\ & & \ddots & \\ & & & \mathbf{\Phi}(f_{N_f}) \end{bmatrix} \tag{14}$$

of sampled exponential vectors $\mathbf{\Phi}(f)$ given by

$$\mathbf{\Phi}(f) = \begin{bmatrix} e^{-i2\pi(k(f_0)+k'(f_0)(f-f_0))z_1} \\ e^{-i2\pi(k(f_0)+k'(f_0)(f-f_0))z_2} \\ \vdots \\ e^{-i2\pi(k(f_0)+k'(f_0)(f-f_0))z_L} \end{bmatrix} \tag{15}$$

In the following we will also refer to a broadband basis element as a broadband propagator.

Under the Assumption 2.1 and from Equation 13 one can easily see that the data in a given band is a *superposition of broadband basis elements* defined by 3.1. Under this set-up our methodology for dispersion extraction in a band consists of the following.

1. Given the band F around a center frequency f_0, form an *over-complete basis* of broadband propagators $\mathbf{P}_F(k(f_0), k'(f_0))$ in the $(f-k)$ domain spanning a range of group and phase slowness. This is described in the next section.
2. Assuming that the broadband signal is in the span of the broadband basis elements from the over-complete dictionary, the presence of a few significant modes in the band implies that the signal representation in the over-complete basis is *sparse*. In other words that the signal is composed of a superposition of few broadband propagators in the over-complete basis.
3. The problem of slowness dispersion extraction in the band can then be mapped to that of finding the sparsest signal representation in the over-complete basis of broadband propagators.

To put simply we map the problem of dispersion estimation to that of finding the sparsest signal representation in an over-complete basis of broadband propagators spanning a range of group and phase slowness. Before we go into the precise mathematical development we would like to introduce some notation used in the rest of this paper.

A. Notation

In the rest of the paper we will use the following notation. The vectors are denoted by small boldface letters, e.g. $\mathbf{x}$ and the matrices are denoted by boldface capital letters, e.g. $\mathbf{A}$ with the i,j-th element denoted by $\mathbf{A}_{ij}$. The same definition extends to higher dimensional arrays. For quantities related to dispersion we will use $k'(f)$ and $s^g(f)$ to denote group slowness interchangeably and $\frac{k(f)}{f}$ and s^ϕ to denote phase slowness interchangeably. We will often use the following matrix operations.

- Reshape$(\mathbf{x}, n_1, n_2)$ for a vector $\mathbf{x} \in \mathbb{C}^{n_1.n_2 \times 1}$ is a $n_1 \times n_2$ matrix with elements taken column by column from $\mathbf{x}$. Similarly Reshape$(\mathbf{x}, n_1, n_2, n_3)$ for a vector $\mathbf{x} \in \mathbb{C}^{n_1.n_2.n_3 \times 1}$ is a 3-dimensional $n_1 \times n_2 \times n_3$ array with elements taken 3rd dimension wise first, and then taken 2nd dimension wise, i.e. the (i_1, i_2, i_3)-th element of the matrix is populated according to $i \to (i_1, i_2, i_3) : i = i_1 + n_2(i_2 - 1 + n_3(i_3 - 1)), i_1 = 1, 2, ..., n_1$.
- norms$(\mathbf{A}, p, \text{dim})$ is the matrix (vector) of p-norms along the dimension dim. For example if $\mathbf{A} \in \mathbb{C}^{n_1 \times n_2}$ then norms$(\mathbf{A}, 2, 1)$ is a vector of length n_2 with i-th entry corresponding to the ℓ_2 norm of i-th column.

B. Over-complete basis of broadband propagators in a band

In order to construct the over-complete basis we pick a range of wavenumbers $k_i, i = 1, 2, ..., n_1$ at the center frequency f_0 and a range of group slowness $k_j^{'}, j = 1, 2, ..., n_2$ for the given band. Note that the range of wavenumbers in turn can be picked from the range of phase slowness at the center frequency using the relation $s_i^\phi(f_0) = \frac{k_i(f_0)}{f_0}$. In line with the broadband system given by Equation 11, the over-complete basis of broadband propagators for acoustic signal representation in a band can be written as

$$\mathbf{\Phi}_F = \Big[\Psi_1(F) \mid \Psi_2(F) \mid ... \mid \Psi_N(F)\Big] \in \mathbb{C}^{L.N_f \times N.N_f} \tag{16}$$

where $\Psi_{i+n_1\cdot(j-1)}(F) = \mathbf{P}_F(k_i(f_0), k_j^{'}(f_0))$, $N = n_1 \times n_2$ is the number of broadband basis elements in the over-complete dictionary.

A pictorial representation of an over-complete basis of broadband propagators in the given band in the f-k domain is shown in Figure 5. To this end we note the following trivial result.

Lemma 3.1: Any set of L broadband basis elements Ψ_Ω, $\Omega \subset \{1, 2, .., N\}, |\Omega| = L$ contained in $\mathbf{\Phi}$ are linearly independent.

Proof: The proof follows by construction. ∎

C. Problem set-up in the over-complete basis

In order to bring out the salient features of the problem formulation and dispersion extraction methodology we make the following assumption for now.

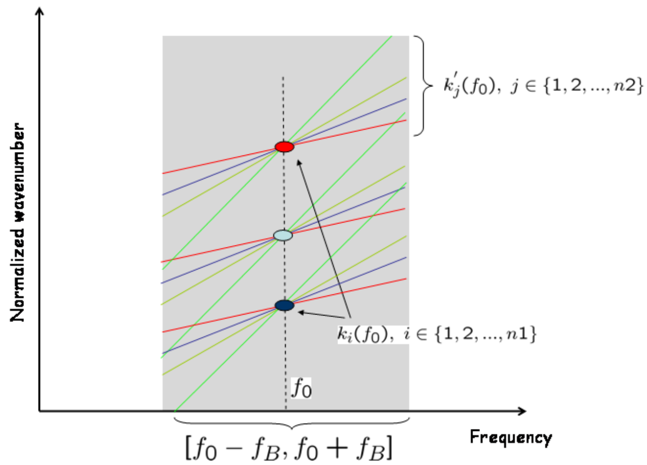


Fig. 5. Figure showing a collection of broadband basis in a given band around the center frequency f_0

Assumption 3.1: The broadband basis elements $\mathbf{P}_m(F)$ corresponding to the M modes are contained in the constructed over-complete basis $\mathbf{\Phi}$.

Note that this assumption is not critical for our method and we will indeed relax this assumption later on. Recall that the signal in band F is a superposition of M modes and can be written as

$$\mathbf{S}_F = \sum_{m=1}^{M} \mathbf{P}_F(m)\mathbf{S}_F(m) \tag{17}$$

Under the assumption 3.1, let the set of broadband propagators $\Psi_\Omega(F)$, for some $\Omega \subset \{1, 2, ..., N\}$ contained in the dictionary $\mathbf{\Phi}$ denote the basis elements corresponding to the true modes in $\mathbf{S}_F$. Let $\mathbf{x}_F \in \mathbb{C}^{N.N_f \times 1}$ denote the coefficient vector such that $\mathbf{x}_F(\Omega) = [\mathbf{S}_F(1), ..., \mathbf{S}_F(M)]^T$, i.e., the coefficient vector restricted to the true support in the broadband basis is equal to the spectrum of the modes. Note that $\mathbf{x}_F(\Omega) \in \mathbb{C}^{M.N_f \times 1}$. Then clearly the coefficient vector $\mathbf{x}_F$ synthesizes the signal $\mathbf{S}_F$, i.e., $\mathbf{\Phi}_F\mathbf{x}_F = \mathbf{S}_F$. We will now point out some special properties of this coefficient vector $\mathbf{x}_F$ and relate them to modes and the corresponding dispersion parameters in the given band. To this end let the support of a vector (matrix) be denoted by

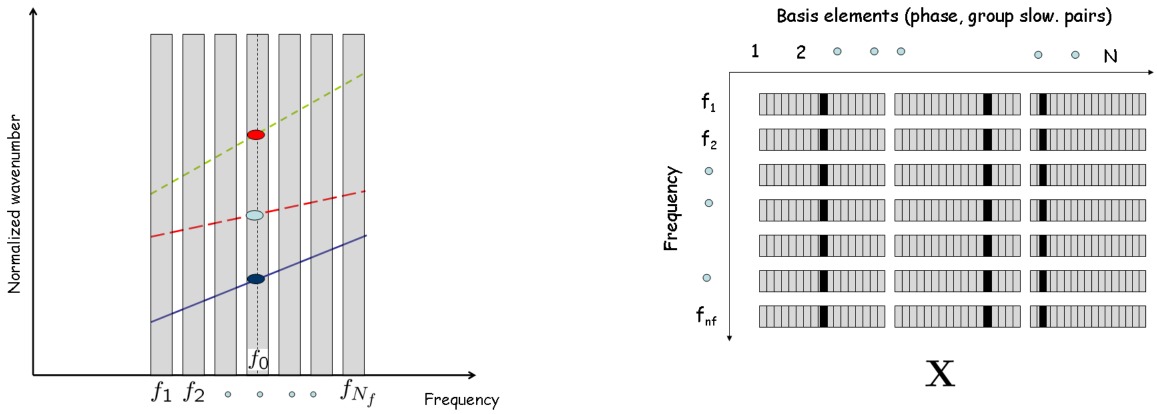
$$\text{Supp}(\mathbf{x}) = \mathbf{1}_{\mathbf{x} \neq \mathbf{0}} \tag{18}$$

where $\mathbf{1}_.$ denotes the indicator function the indicator function of the set in the subscript. In other words the vector Supp($\mathbf{x}$) has a value 1 at places where $\mathbf{x}$ has a non-zero entry and zero otherwise. Then we make the following observation regarding the structure of the support of the coefficient vector $\mathbf{x}_F$.

Remark 3.1: For the given band F if

A1. The number of modes $M << N$.

A2. The frequency support Supp$(S_F(m))$ of each mode m is equal to N_f.



(a) (b)

Fig. 6. Schematic showing the column sparsity of the signal support in $\mathbf{\Phi}$. Sparsity of the number of modes in the band (a) implies a column sparsity in mode representation in the broadband basis in the band (b).

then the reshaped vector - $\mathbf{X}_F = \text{Reshape}(\mathbf{x}_F, N_f, N)$ exhibits *column sparsity*. In other words the cardinality of the *column support*- the number of columns with a non-zero entry - of $\mathbf{X}_F$ is small.

A schematic representation of the structure of the sparsity of the column support of $\mathbf{X}_F$ in the over-determined basis of broadband propagators in a band is shown in Figure 6. The essence of this observation is that the broadband basis elements corresponding to the column support of $\mathbf{X}_F$ in the basis $\mathbf{\Phi}_F$ correspond to the slowness dispersion parameters of the modes present in the data and the cardinality of the column support is related to the model order in the band.

Under this observation the problem of dispersion extraction in the band F can be stated as follows. Given the set of observations, $\mathbf{Y} = [\mathbf{Y}(f_1)\mathbf{Y}(f_2)\ldots\mathbf{Y}(f_{N_f})]^T \in \mathbb{C}^{L\times N_f}$, at L receivers in a band $F = \{f_1, f_2, ..., f_{N_f}\}$ obeying the relation

$$\mathbf{Y}_F = \mathbf{\Phi}_F\mathbf{x}_F + \mathbf{W}_F,\ \mathbf{x}_F \in \mathbb{C}^{N.N_f\times 1} \tag{19}$$

with respect to an over-complete basis $\mathbf{\Phi}_F \in \mathbb{C}^{M.N_f\times n_1.n_2.N_f}$ of broadband propagators; *find* $\mathbf{x}_F$ *such that the corresponding* $\mathbf{X}_F$ *exhibits column sparsity.*

We now define measures or cost functions on column sparsity. The column sparsity is also known as simultaneous sparsity in the literature, see [24] and references therein. As clear from its meaning, column sparsity is measured in terms of the column support and one can measure the column support through the following,

$$J_{0,p}(\mathbf{X}) = ||\text{norms}(\mathbf{X}, 1, p)||_0,\ p > 0 \tag{20}$$

which is essentially the ℓ_0 norm of the vector composed of p norms of the columns of $\mathbf{X}$. In this work we will focus on the case when $p = 2$. The main reason for selecting $p = 2$ is because later on we will relax the $J_{0,2}$ penalty to $J_{1,2}$ penalty and heuristically speaking $p = 2$ measures the energy of the mode spectrum and thus penalizes each mode in proportion to the square root of the energy. Then the problem of dispersion extraction in the above framework can be stated as the following optimization problem

$$\text{OPT}_0 : \text{minimize} \;\; J_{0,2}(\mathbf{X}_F) \tag{21}$$

$$\text{subject to :} \;\; \mathbf{Y}_F = \mathbf{\Phi}_F \mathbf{x}_F \tag{22}$$

We will now outline some sufficient conditions under which the solution to the above optimization problem is unique and results in correct support recovery and hence correct dispersion extraction. To this end we have the following Lemma.

Lemma 3.2: If the frequency support of each mode is N_f and the number of modes $M \leq L/2$ then

1. Solution to OPT_0 is unique
2. $\mathbf{x}_F$ is supported on the right set of broadband basis elements.

Proof: The proof follows from Lemma 3.1 and along lines of Theorem 3 in [25]. See also [26]. ■

In the presence of noise one typically solves the following,

$$\text{minimize} \;\; J_{0,2}(\mathbf{X}_F) \tag{23}$$

$$\text{subject to :} \;\; ||\mathbf{Y}_F - \mathbf{\Phi}_F \mathbf{x}_F||_2^2 \leq \epsilon \tag{24}$$

where ϵ is a parameter is chosen based on some prior knowledge of the noise variance. It is well known that the above problem is equivalent to solving the following problem,

$$arg \min_{\mathbf{x}_F} ||\mathbf{Y}_F - \mathbf{\Phi}_F \mathbf{x}_F||_2^2 + \lambda J_{0,2}(\mathbf{X}_F) \tag{25}$$

for some appropriately chosen value of λ. However note that imposing the $J_{0,2}$ penalty poses a combinatorial problem, which in general is very hard to solve. It was proposed in [27], [26] to relax the $J_{0,2}$-penalty to $J_{1,2}$-penalty which is the closest convex relaxation to $J_{0,2}$-penalty. Motivated by the developments there, in this work we will use the following convex relaxation to solve the sparse reconstruction problem,

$$\text{OPT}_1 : arg \min_{\mathbf{x}_F} \|\mathbf{Y}_F - \mathbf{\Phi}_F \mathbf{x}_F\|^2 + \lambda J_{1,2}(\mathbf{X}_F) \tag{26}$$

Our proposed methodology is similar in spirit to that of [28] in that the sparsity penalty is the same. However there are several key differences from the formulation there and the formulations analyzed in [24],[27], [26]. First note that under a suitable permutation of the columns of the dictionary and corresponding rearrangement of the coefficient vector $\mathbf{x}_F$ one can convert column sparsity to row sparsity which is what is considered in all these papers.

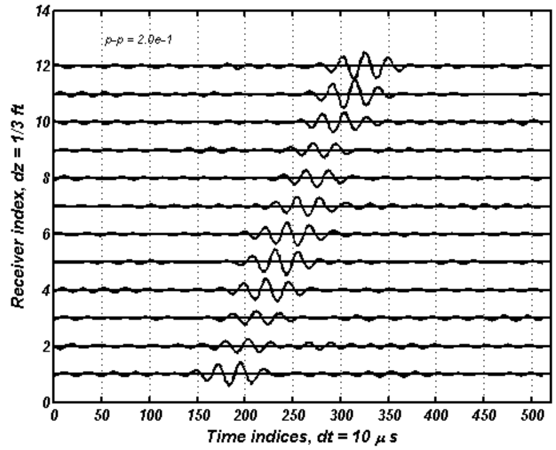


Fig. 7. Noisy data in a band containing two time overlapping modes. The in-band SNR is around 7 dB for the stronger mode and around -3 dB for the weaker mode.

Now note that under this common setting, i.e. common simultaneous sparsity set-up, the over-complete dictionary for each sparse column of row-sparse matrix is the same in [26] [24],[28] while in our setting it is different. Therefore, fundamentally in order to analyze our problem the notion of coherence between the dictionary atoms needs to be extended to the notion of the coherence between the broadband basis elements. This characterization and subsequent analysis is beyond the scope of this work and will be dealt with in future.

We will now address two issues with the framework and the proposed methodology. These issues are,

1. Selection of regularization parameter λ in the optimization problem OPT_1. With respect to the optimization problem OPT_0 this parameter governs the *sparsity* of the solution and thus is critical for model order selection. In the relaxed set up of OPT_1 apart from governing the sparsity of the solution this parameter also affects the spectrum estimates. This is due to solution *shrinkage* resulting from the ℓ_1 part of the $J_{1,2}$ penalty.
2. Assumption 3.1 is not true in general. This has consequences for model order selection as well as for the estimates of dispersion parameters.

In the following we will address the above issues and propose methods to handle them. We will start with the selection of the regularization parameter. We will postpone the details of the implementation for solving the convex optimization problem OPT_1 to section IV. The interested reader can go to that section and come back for further development of the ideas.

D. Selection of regularization parameter λ

In this work we propose a novel approach to select the regularization parameter λ in the optimization problem OPT_1 using tests between distribution of residuals. In this context we recognize that the $J_{1,2}$ penalty plays two roles. The first role is that of general *regularization* of the solution where it prevents the amplification of noise due to *ill-conditioning* of the matrix $\mathbf{\Phi}$. The second role is that of *model order selection* which is essentially related

STEP 1. For a given band F, pick an increasing sequence of the regularization parameter $\Lambda = \lambda_1 < \lambda_2 < ... < \lambda_n$. In practice it is sufficient to choose this sequence as $\Lambda =$ logspace$(a_0, a_1, n), a_0, a_1 \in \mathbb{R}$. This choice implies $\lambda_1 = 10^{a_0}$ and $\lambda_n = 10^{a_1}$. Both a_0 and a_1 are chosen so as to cover the useful range of λ.

STEP 2. Find the sequence of solutions $\mathbf{x}_F^\lambda, \lambda \in \Lambda$ by solving OPT$_1$ and the corresponding sequence of residual vectors $\mathbf{R}_\lambda = \mathbf{Y}_F - \mathbf{\Phi}_F \mathbf{x}_F^\lambda$. Denote the empirical distribution of residuals w.r.t. the solution $\mathbf{x}_F^\lambda$ by $CDF_\lambda(r)$ where r denotes the variable for residual.

STEP 3. Conduct KS tests between the residuals -

3a. Choose two reference distributions, viz., $CDF_1(r)$ corresponding to the minimum value λ_1 of regularization parameter and $CDF_n(r)$ corresponding to the maximum value λ_n of the regularization parameter.

3b. Find the sequence of KS-test statistics $D_{i,n} = \sup_r |CDF_n(r) - CDF_i(r)|$, $i = 1, 2, ..., n$ and the corresponding p-value sequence $P_{i,n}$. Similarly find the sequence of KS-test statistics $D_{i,1} = \sup_r |CDF_1(r) - CDF_i(r)|$, $i = 1, 2, ..., n$ and the corresponding p-value sequence $P_{i,1}$.

STEP 4. Plot the sequences $D_{i,n}$ and $D_{i,1}$ as a function of λ_i. The operating λ say λ^* is then taken as the point of intersection of these two curves. A similar operating point can also be obtained by choosing the intersection point of the p-value sequences $P_{i,n}$ and $P_{1,i}$.

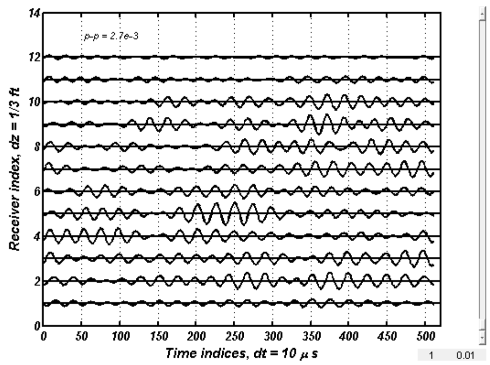
TABLE I

TABLE ILLUSTRATING THE STEPS FOR SELECTION OF THE REGULARIZATION PARAMETER USING KOLMOGOROV-SMIRNOV (KS) TEST BETWEEN RESIDUALS.

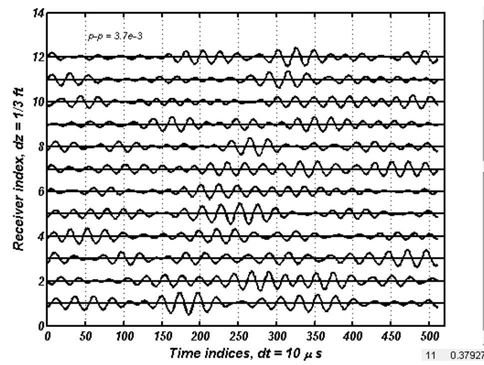
to selection of the sparsest (and correct) basis for signal representation. In the context of finding sparse solutions these two aspects go hand in hand. To this end we note the following points.

1. At a very low value of λ, the solution to OPT$_1$ comes close to the Least Squares (LS)solution. This results in *over-fitting* of the data and the residual is close to zero. Due to ill-conditioning of the matrix $\mathbf{\Phi}$ the noise contribution to the solution gets amplified and one observes many spurious peaks in the solution support.
2. As λ is increased the $J_{1,2}$ penalty kicks in allowing less degrees of freedom for data fitting, thereby combating noise and reducing the spurious peaks in the solution. Due to the sparsity imposed by the $J_{1,2}$ penalty, at a certain λ one hits the sparse signal subspace in the dictionary. However due to shrinkage in the solution implied by the $J_{1,2}$ penalty *signal leakage* occurs into the residual. This implies that *as λ is increased (varied) the distribution of the residual changes*.
3. At very high values of λ we have no data fitting and the solution is driven to zero. Therefore at high values of λ the distribution of the residual *converges* to the distribution of the data.

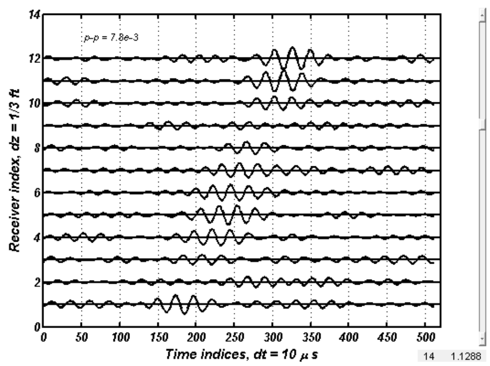
These observations are illustrated in Figure 8 for a synthetic data set in a band with two time-overlapping modes as shown in Figure 7. These observations suggest the following strategy for selecting the regularization parameter.



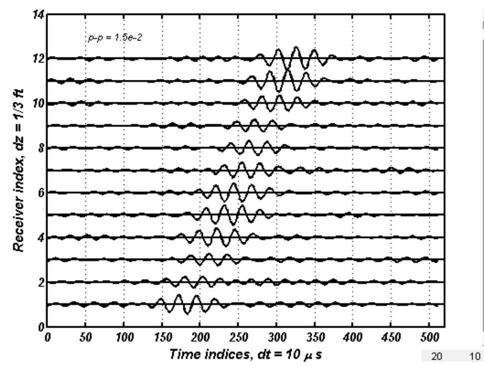
(a) Residual at λ_{min}



(b) Residual at an intermediate λ



(c) Residual at high λ



(d) Residual at λ_{max}

Fig. 8. An example showing the behavior of the residual as a function of λ. In (c) Note the mode leakage, i.e. the leakage of modal energy into the residual due to shrinkage of the coefficients. In (d) note more modal leakage into the residual and residual appears to look like the data - See Figure 7. The Kolmogorov-Smirnov (KS) test is used to detect changes in the distribution of the residuals as a function of λ.

Vary λ over a range and detect changes in distribution of the residuals and select an operating λ that mitigates noise and minimizes the signal leakage into the residual while still finding the right signal subspace. In this work we propose to use the Kolmogorov-Smirnov (KS) test between empirical distributions of the residuals to detect these changes and find an operating range of λ. Given a data sample of the residual $R_1, ..., R_i, ..., R_{N_s}$, where $N_s = N_f.L$ is the number of samples, define the empirical distribution function as follows,

$$CDF(r) = \frac{1}{N_s} \sum_{i=1}^{N_s} \mathbf{1}_{R_i \leq r} \tag{27}$$

Then the KS test statistic for $CDF(r)$ with respect to a *reference distribution* $CDF_{ref}(r)$ is given by

$$D = \sup_r |CDF(r) - CDF_{ref}(r)| \tag{28}$$

The properties of the KS test statistic D, its limiting distribution and the asymptotic properties can be found in [29],[30]. Essentially the KS statistic is a measure of how similar two distributions are. In order to apply the KS-test to our problem we perform the steps shown in table I. The intersection point of the test curves qualitatively signifies the tradeoff between over-fitting and under-fitting of the data. In addition note that as the solution sequence

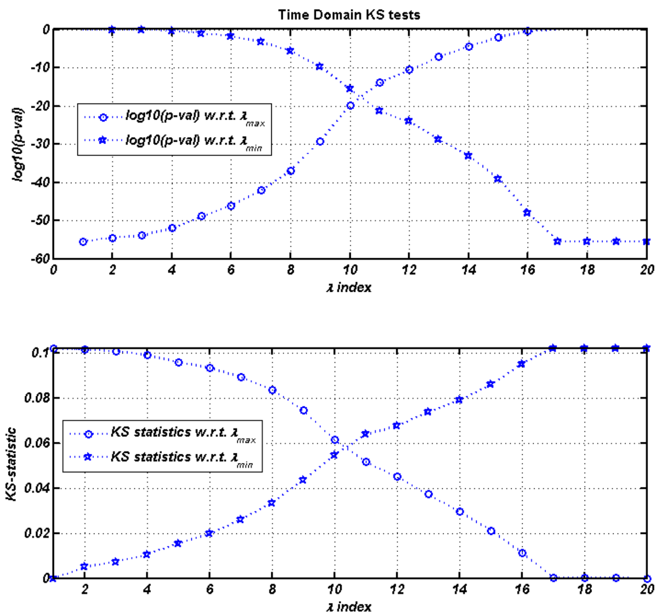


Fig. 9. An example of the KS test statistic sequence and the p-value sequence for determining the operating λ.

$\mathbf{X}_\lambda$ starts hitting the right signal subspace in the over-determined basis the *rate of change* of distribution of the residual as implied by the change in the test statistic, is more rapid due to mode leakage into the residual. In particular one can choose an entire range of λ around the point of intersection. The solutions for this range of λ exhibit stability in terms of getting to the right signal support. For the problem at hand one can perform the KS tests between residuals either in the frequency domain or in the time domain. Since the modes are time compact it is useful to compare the distributions of the residuals in the space time domain. An example of the KS test-curves for distribution of residuals in time domain and the corresponding operating range of λ is shown in Figure 9.

E. Issue of dictionary mismatch with the underlying true parameter values

The Assumption 3.1 is not true in general and in this section we will relax this assumption and address the consequences. A schematic depiction of the dictionary mismatch with respect to the true modal dispersion is shown in Figure 10. Recall that the signal in the band F denoted by $\mathbf{S}_F$ is a superposition of M broadband propagators in the band,

$$\mathbf{S}_F = \sum_{m=1}^{M} \mathbf{P}_F(m)\mathbf{S}_F(m) \tag{29}$$

Let $\hat{\mathbf{S}}_F$ be the best and unique ℓ_2 M'-term approximation to $\mathbf{S}_F$ in the constructed over-complete dictionary Φ. Note that in general the value of M' depends on the quantization of the parameter space and $M' \geq M$. To this end let $\mathbf{x}_F \in \mathbb{C}^{N.N_f \times 1}$ denote the coefficient vector that synthesizes the signal $\hat{\mathbf{S}}_F$, i.e. $\mathbf{\Phi}_F \mathbf{x}_F = \hat{\mathbf{S}}_F$. It is easy to show the following,

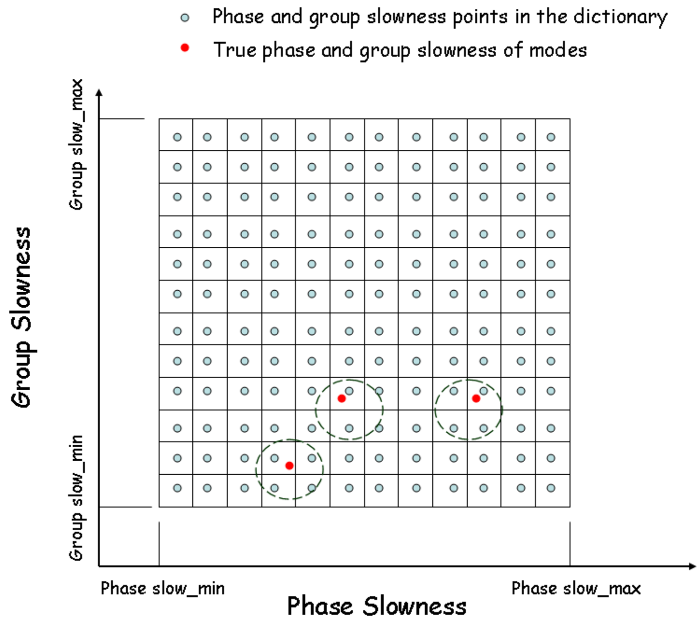


Fig. 10. Schematic showing the discretization of the parameter space of phase and group slowness in a band and the mismatch of the dictionary with respect to the true modes in the data.

Lemma 3.3: Given $\epsilon > 0$ there exists an N_ϵ - the number of quantization points, and a quantization of the slowness parameter domain (phase and group slowness) such that

$$\|\hat{\mathbf{S}}_F(N_\epsilon) - \mathbf{S}_F\| \leq \epsilon \tag{30}$$

,i.e. under appropriate quantization the signal can be well represented in the over-complete basis.

One can then model this mismatch as an additive error in the equation without model mismatch,

$$\mathbf{Y}_F = \mathbf{\Phi}_F \mathbf{x}_F + \underbrace{(\mathbf{S}_F - \hat{\mathbf{S}}_F(N)) + \mathbf{W}_F}_{\tilde{\mathbf{W}}_F},\ \mathbf{x}_F \in \mathbb{C}^{N_f.N \times 1} \tag{31}$$

where now the coefficient vector $\mathbf{x}_F \in \mathbb{C}^{N.N_f \times 1}$ corresponds to the representation for the approximate signal $\hat{\mathbf{S}}_F$. Thus in the presence of dictionary mismatch the problem of dispersion extraction maps to that of finding the dispersion parameters of the best ℓ_2 approximation to the signal in the over-complete basis of broadband propagators. Sparsity in the number of modes M in $\mathbf{S}_F$ still translate to sparsity in the representation M' of the approximation $\hat{\mathbf{S}}_F$ in the over-complete basis $\mathbf{\Phi}_F$ and we still employ the sparse reconstruction algorithm OPT$_1$ for dispersion extraction. However this affects the model order selection and dispersion estimates. In order to understand this in our context we introduce the following notation.

$$\mathbf{X}_{colnorm} = \text{norms}(\mathbf{X}_F, 2, 1) \tag{32}$$

STEP 1. Pick a certain number of peaks say N_p corresponding to the largest values in ModIm$(\mathbf{X}_F)$. Step 2. Perform a k-means clustering of the peak points in the phase and the group slowness domain. STEP 3. Declare the resulting number of clusters as the model order. STEP 4. Mode consolidation - 4a. The mode spectrum corresponding to each cluster is obtained by summing up the estimated coefficients corresponding to the points in the cluster. 4b. The corresponding dispersion estimates are taken to be the average over the dispersion parameters corresponding to the cluster points.

TABLE II
TABLE ILLUSTRATING THE STEPS FOR MODEL ORDER SELECTION AND MODE CONSOLIDATION IN THE PHASE AND GROUP SLOWNESS DOMAIN ($(f-k)$ PROCESSING).

i.e., $\mathbf{X}_{colnorm} \in \mathbb{R}^{N\times 1}$ is the column norm of the reshaped vector $\mathbf{X}_F$. We further reshape $\mathbf{X}_{colnorm}$ as Reshape$(\mathbf{X}_{colnorm}, n_1, n_2)$ and obtain a modulus image of $\mathbf{X}_F$ in phase and group slowness domain over the range of values chosen. Note that the support of the modulus image corresponds to the broadband signal support and the corresponding dispersion parameters can be read from the image, e.g. see Figure 11. For sake of exposition we will denote this modulus image by ModIm$(\mathbf{X}_F)$.

The result of the dictionary mismatch with respect to the true dispersion parameters of the modes is that due to approximate signal representation of the true signal in the over-complete basis, in the solution to the optimization problem OPT$_1$, we obtain *clustered* set of peaks in the modulus image of the column norm of the solution $\mathbf{X}_F$. These clustered set of points correspond to the set of phase and group slowness points that are closest to the true phase and group slowness points. The mode clusters can also form due to very high level of *coherence* between the broadband basis elements. Therefore we need to do clustering for model order selection. For this we choose the k-means clustering algorithm as implemented in MATLAB function [31] to do the clustering and follow the steps as outlined in table II.

In summary so far we have provided a framework for dispersion extraction in terms of reconstruction of a sparse set of features in an over-complete basis of broadband propagators in the f-k domain. We have developed the algorithmic tools for solving the problem posed in this framework, viz., (a) the $J_{1,2}$ penalized reconstruction algorithm OPT$_1$; (b) the KS tests between sequences of residuals in the time domain and; (c) mode clustering for model order selection in the quantized parameter space of phase and group slowness. A flowchart showing various steps of the proposed dispersion extraction methodology in the $(f-k)$ domain is shown in Figure 12.

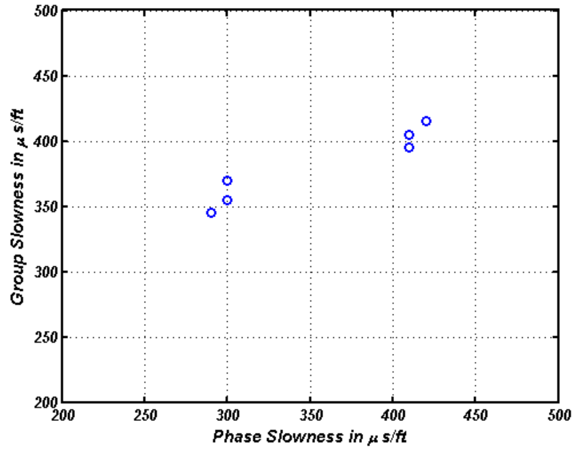


Fig. 11. An example showing the mode clusters in the phase and group slowness domain due to dictionary mismatch corresponding to 6 largest peaks in the modulus image .

F. Using Continuous Wavelet Transform (CWT)

In this work instead of finding a set of disjoint frequency bands for dispersion extraction we make use of continuous wavelet transform (CWT) and extend the broadband dispersion extraction methodology as proposed above in the $(f-k)$ domain in a straightforward manner to the CWT domain where the linear approximation is assumed to be valid in the effective band $\mathcal{F}_a$ around the center frequency f_a of the analyzing wavelet at given scale, see Figure 13. Then one can do the processing at appropriately chosen scales so that their effective bandwidths don't overlap significantly and then combine the estimated curves across scales to obtain the entire dispersion curve.

In particular in the CWT domain at scale a the signal is composed of superposition of modes with spectrum shaped according to the spectrum of the analyzing wavelet at scale a, i.e., in the effective band F_a at scale a the overall signal can be written as

$$\mathbf{S}_F^a = \sum_{m=1}^{M} \mathbf{P}_F^a(m)\mathbf{S}_F^a(m) \tag{33}$$

where $\mathbf{P}_F^a(m)$ is the broadband propagator corresponding to mode m in band F_a and $\mathbf{S}_F^a(m)$ is the mode spectrum of the CWT coefficients of mode m at scale a.

Let $Y_l(a,b)$ denote the array of CWT coefficients of the array data $y(l,t)$ at scale a. Let

$$\mathbf{Y}_a(f) = [Y_1(a,f),..,Y_l(a,f),..,Y_L(a,f)]^T$$

denote the Fourier transform the array of CWT coefficients for $f \in F_a$. Then one can write,

$$\mathbf{Y}_F^a = \mathbf{S}_F^a + \mathbf{W}_F^a \tag{34}$$

where $\mathbf{Y}_F^a = [\mathbf{Y}_a(f_1), \mathbf{Y}_a(f_2), \ldots, \mathbf{Y}_a(f_{N_f})]^T$. Let $\mathbf{\Phi}_F^a$ denote the over-complete basis of broadband propagators

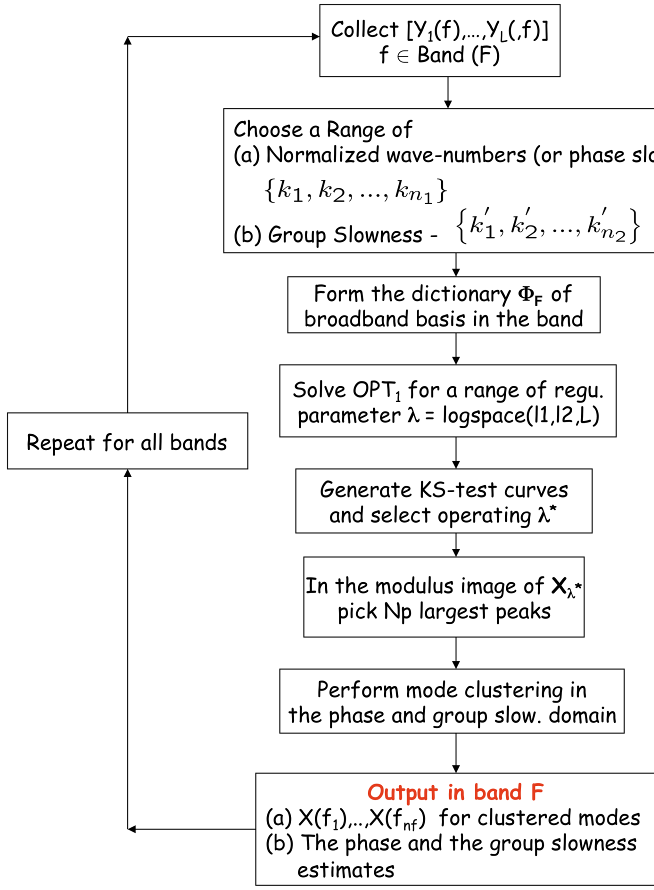


Fig. 12. The flow of the processing in a band for dispersion extraction in the $(f-k)$ domain.

in the $(f-k)$ domain in the effective band F_a at scale a. In this case the complex exponentials composing the basis are constructed with respect to the reference receiver, i.e.

$$\mathbf{\Phi}^a(f) = \begin{bmatrix} e^{-i2\pi(k(f_a)+k(f_a)'(f-f_a))(z_1-z_0)} \\ e^{-i2\pi(k(f_a)+k'(f_a)(f-f_a))(z_2-z_0)} \\ \vdots \\ e^{-i2\pi(k(f_a)+k'(f_a)(f-f_a))(z_L-z_0)} \end{bmatrix} \tag{35}$$

Then similarly to the broadband processing framework proposed earlier we have,

$$\mathbf{Y}_F^a = \mathbf{\Phi}_F^a \mathbf{x}_F^a + \mathbf{W}_F^a \tag{36}$$

where $\mathbf{x}_F^a$ is the vector of coefficients to be estimated corresponding to the modes contained in the CWT of the data at scale a. By construction the estimate $\hat{\mathbf{x}}_F^a$ of the coefficients corresponds to the Fourier transform of the CWT of mode waveforms at scale a at the reference receiver. This fact will be used later on for space time processing and

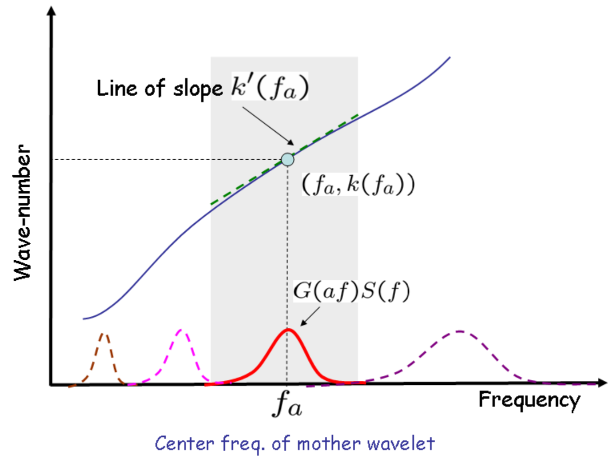


Fig. 13. Schematic showing the first order Taylor series approximation to the dispersion curve in a band around the center frequency f_a of the mother wavelet at scale a.

to reduce interference. In order to solve for this problem we solve the following optimization problem,

$$arg \min_{\mathbf{x}_F^a} ||\mathbf{Y}_F^a - \mathbf{\Phi}_F^a \mathbf{x}_F^a||^2 + \lambda_a J_{1,2}(\mathbf{X}_F^a) \tag{37}$$

For selection of the regularization parameter λ_a we repeat the steps proposed earlier for the broadband processing of the data in the $(f-k)$ domain. This step is followed by a model order selection and a mode consolidation step.

IV. Evaluation on synthetic data

We will now present the dispersion extraction results using the above approach for a two mode synthetic problem with considerable time overlap. For the continuous wavelet transform we use Morlet wavelet [32] which has shown to be promising for geophysical applications. The mother wavelet is given by

$$g(t) = e^{\frac{-t^2}{2\sigma^2}} e^{j\omega_0 t} \tag{38}$$

Although not strictly admissible this wavelet is practically admissible for $\omega_0 \geq 5$ [33]. Since the envelope is Gaussian this wavelet has good time frequency localization properties. For the simulations we use $\omega_0 = 2\pi$ and $\sigma^2 = 1$. The algorithm that is used to implement the Continuous Wavelet Transform using the Morlet wavelet is detailed in [34] that uses symmetry properties of the transform for a faster implementation.

We now come to the details of the implementation for the broadband f-k processing of the CWT data at a given scale a. Recall that the f-k processing at a given scale consists of solving the optimization problem.

$$\text{OPT}_1 : arg \min_{\mathbf{x}_F^a} ||\mathbf{Y}_F^a - \mathbf{\Phi}_F^a \mathbf{x}_F^a||^2 + \lambda_a J_{1,2}(\mathbf{X}_F^a) \tag{39}$$

in a given band F_a at scale a in the $(f-k)$ domain. For a fixed value of λ_a in order to numerically find the solution to OPT_1 we use CVX which is a general purpose convex optimization toolbox for MATLAB developed at Stanford University. We will not go into the details of the implementation and the reader is referred to [35],[36]. The effective band F_a at scale a is chosen based on the effective spread of the Gaussian envelope of the analyzing wavelet at the particular scale. Note that at higher frequencies, i.e. lower scales the effective band is larger and vice-versa. For our experiments we choose the center frequency as $f_a = \frac{2\pi}{a}$ Hz and effective band F_a is taken to be the set of discrete frequencies (DFT) contained in the set

$$F_a \subset [0.67 f_a \;\; 1.33 f_a] \tag{40}$$

Note that at low frequencies the frequency resolution of the CWT is good but the time resolution is bad and vice-versa. Therefore the effective time support in the space-time processing is taken as the effective time support of the analyzing wavelet at that scale. In this work we will compare the results with the existing state of the art method for multi-mode dispersion extraction which essentially employs the modified matrix pencil (MP) method of [5] followed by a post processing for mode association across the frequencies.

a) Computational complexity of the algorithm: The optimization program OPT_1 solved as a Semi Definite Program using well known interior point method, see [37], is known to have a worst-case computational complexity that is polynomial in the number of constraints, $2L \times N_f$ (the factor of 2 comes due to the variables being complex) with the number of iterations scaling with square root of the variable size $2N$ and logarithmic with respect to the inverse of the error accuracy required. This implies that practically for a fixed desired accuracy error and the fixed parameters L and N_f the computational complexity is polynomial in N. This is in stark contrast to the combinatorial complexity (in the problem size N) for the ML broadband formulation of [23].

A. Synthetic Data

In order to test our method we use the synthetic data set with two modes whose $(f-k)$ response (2-D FFT) is shown in Figure 14(a)-(b). Shown in Figure 14(c)-(d) are the CWT maps of the modes at the first receiver. Note that the modes are close to each other in time as evidenced by the CWT maps. Note the difference in the effective spectral support of the two modes. This is usually the case in practice for real field data. In Figure 14(e) we show the array waveforms for the superposition of two modes in additive noise. The SNR is -10 dB with respect to the total energy of both the modes across the array and for the entire time window of acquisition. In Figure 14(f) we show the noisy CWT map of the superposed waveforms at receiver 1. Note that there is no time frequency separation here and the modes cannot be separated by methods proposed in [19], [22].

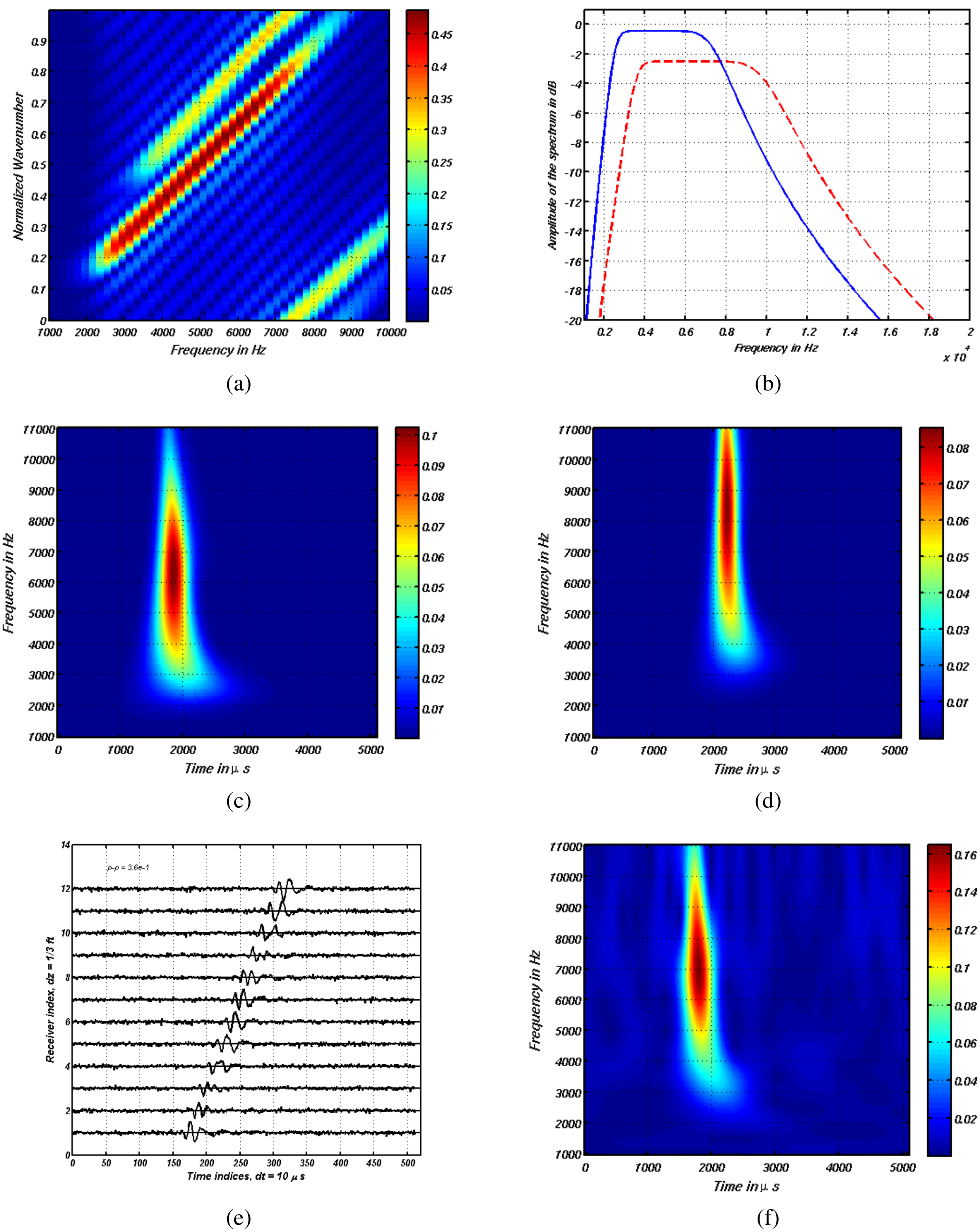
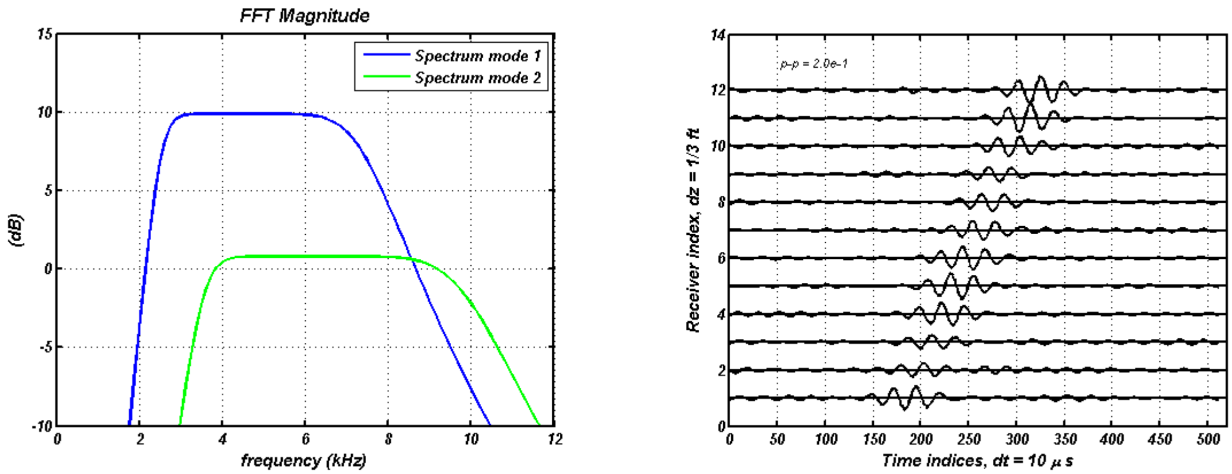


Fig. 14. (a) The $(f-k)$ dispersion of the modes. (b) Individual mode spectrum. (c) CWT map of mode 1 at receiver 1. (d) CWT map of mode 2 at receiver 1. (e) Array waveforms with two modes superposed in space time plus AWGN noise. (f) CWT map of the array data at receiver 1.



(a) (b)

Fig. 15. (a) Figure showing the frequency spectrum of the modes. Note the frequency overlap. (b) Figure showing the data in the band 3.7 kHz - 5.2 kHz. Note the significant time overlap in the modes.

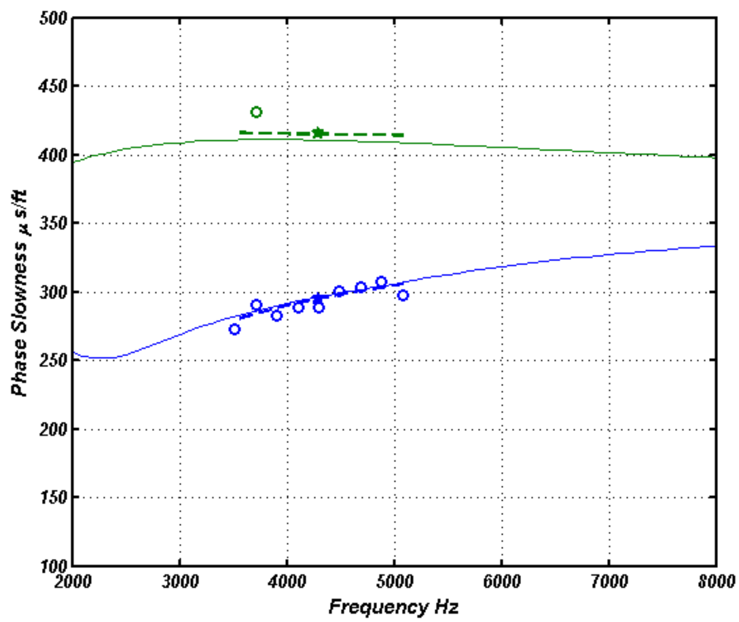
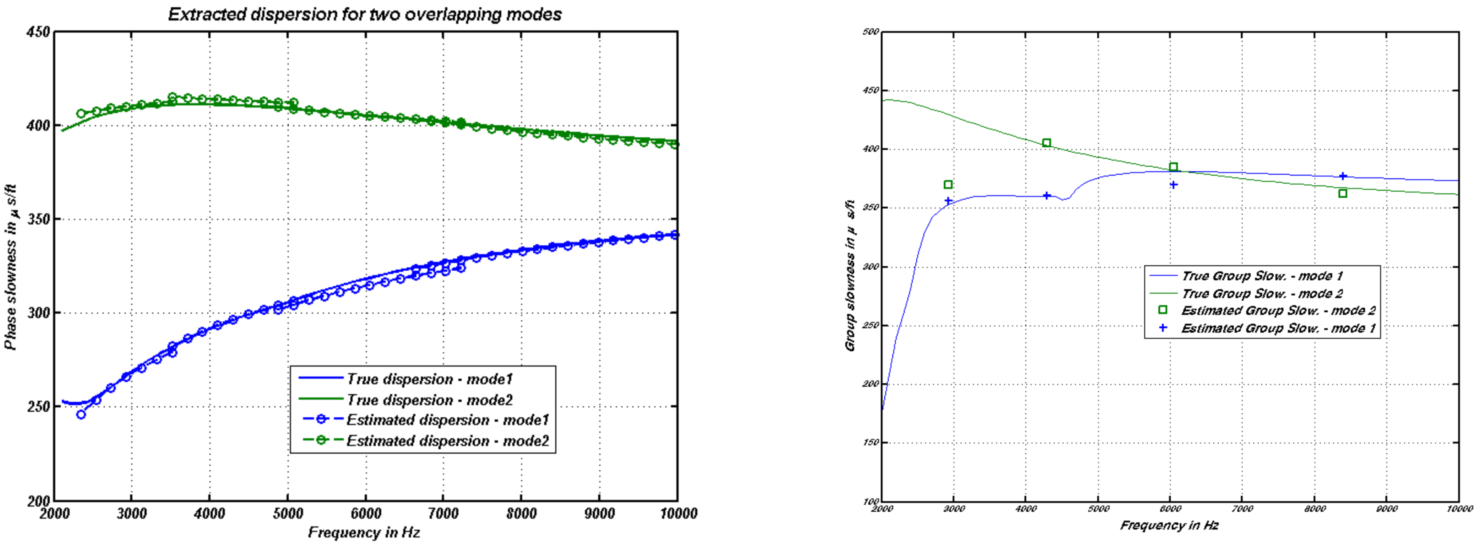


Fig. 16. Dispersion Extraction results in the given band. The thin solid lines are the true dispersion curves for the two modes. The pentagrams are the estimates of the phase slowness at the center frequency of 4.5 kHz as obtained using the proposed method. The thick dashed lines are the corresponding estimated dispersion curves in the band. The blue and green circles are the dispersion curves obtained using the narrowband Matrix Pencil method of [5]. Note the superior performance of the proposed method over the Matrix Pencil based method.

B. Comparison of the proposed method with Matrix Pencil (MP)

We will now compare the broadband dispersion extraction methodology in the $(f-k)$ domain as proposed in section III with the narrow band matrix pencil (MP) approach proposed in [5] for dispersion extraction. In order to compare we consider the noisy synthetic data described above and consider a band $[3.7\, 5.2]$ kHz. We further artificially scale one mode to be 10-dB below the other one, see Figure 15 (a). This results in an *in band* SNR of



(a) (b)

Fig. 17. (a) Extracted dispersion curves for phase slowness in frequency bands. The solid lines denote the true underlying dispersion curves for the two modes. The extracted curves are shown by solid lines marked with circles. (b) Extracted group slowness in frequency bands.

7 dB for the stronger mode and an SNR of -3 dB for the weaker mode. The noisy CWT data in the given band is shown in Figure 15(b). The results of the two processing schemes , viz., narrowband matrix pencil (MP) method and the sparsity penalized broadband dispersion extraction method are shown in Figure 16. Note that although MP based method has reasonable estimates for the stronger mode it completely looses the weaker mode. Also note that the dispersion curves extracted using the MP method are less smooth compared to that obtained from the broadband f-k processing.

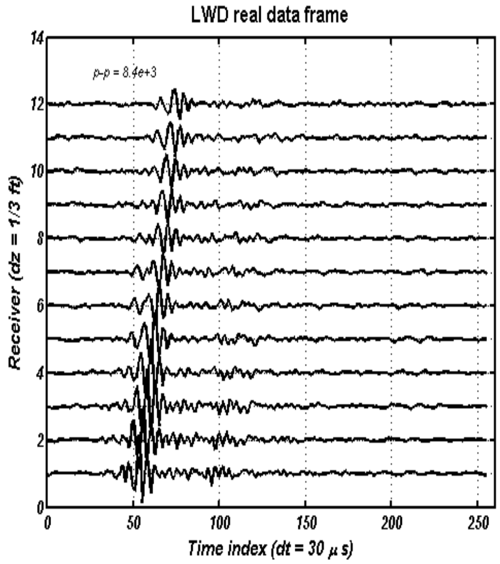
C. Overall dispersion extraction: Synthetic data

We now show the performance of the overall methodology using both the hybrid methodology of section III-F using both the $(f-k)$ processing and t-z processing. For sake of brevity we will not show the mode consolidation and extracted CWT coefficients at each scale. Shown in Figure 17(a) are the phase slowness dispersion curves of two modes with the corresponding group slowness estimates in each band shown in Figure 17(b) Note that the group estimates are less robust than the phase slowness estimates especially at low frequencies where the curvature in the dispersion curve is more.

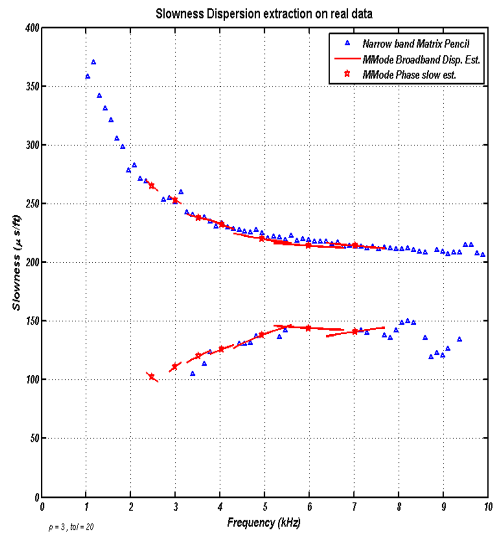
V. EVALUATION ON REAL DATA

Next we present illustration of the performance of the proposed methodology on a real data frame. The data considered is a typical dipole sonic logging while drilling (LWD) data where there is a weaker formation flexural mode significantly time-frequency overlapped with a much stronger interfering tool mode. In Figure 18(b) we show the performance of the proposed methodology on slowness dispersion extraction on one such frame shown in Figure

18(a). Again note the superior performance of the method compared to the current industry standard multi-mode processing method based on the matrix pencil method proposed in [5].



(a)



(b)

Fig. 18. (a) Noisy dipole data acquired at a given depth. (b) Dispersion extraction results using the proposed broadband dispersion extraction and the currently commercially used matrix pencil approach . Note the superior performance of the method especially in band around 4 kHz and 6 kHz where the matrix pencil based approach fails to even detect the weaker mode.

VI. Conclusion and future directions

In this work we presented and evaluated a broadband dispersion extraction methodology for automatic dispersion extraction. Synthetic examples show its superior performance over other methods in presence of heavy noise and interference. However there are several theoretical and computational challenges that need to be addressed for the proposed broadband methodology. On the theoretical side, signal reconstruction using simultaneous sparse penalization has been considered and analyzed in [24],[26]. In particular they consider *row sparsity* and use the notion of coherence between dictionary atoms as a primary tool to analyze the simultaneous sparse set-up. Under a suitable permutation of the columns of the dictionary and corresponding rearrangement of the coefficient vector $\mathbf{x}_F$ we can convert column sparsity to row sparsity considered in all these papers. Now note that under this common simultaneous sparsity set-up, the over-complete dictionary for each sparse column of row-sparse matrix is the same in [26],[24], while in our setting it is different. Therefore, fundamental to analyzing our problem is the need to extend the notion of coherence between the dictionary atoms to that of coherence between the broadband basis elements. On the computation side even though the optimization problem OPT_1 is convex, in practice the computational requirements become severe with the number of broadband basis elements chosen. It can be clearly seen that this number also affects the reliability of the dispersion estimates. Therefore there is a tradeoff between computational resources and estimation quality. This characterization and subsequent analysis will be dealt with in future.

References

[1] F. L. Paillet and C. H. Cheng, *Acoustic Waves in Boreholes*. CRC Press, 1991.

[2] B. Sinha and S. Zeroug, *Geophysical Prospecting Using Sonics and Ultrasonics*. Wiley Encyclopedia of Electrical and Electronic Engineering, 1999.

[3] S. Lang, A. Kurkjian, J. McClellan, C. Morris, and T. Parks, “Estimating slowness dispersion from arrays of sonic logging waveforms,” *Geophysics*, vol. 52, no. 4, pp. 530–544, April 1987.

[4] A. Kurkjian, “Numerical computation of the individual far-field arrivals excited by an acoustic source in a borehole,” *Geophysics*, vol. 50, pp. 852–866, 1985.

[5] M. P. Ekstrom, “Dispersion estimation from borehole acoustic arrays using a modified matrix pencil algorithm,” in *29th Asilomar Conference on Signals, Systems and Computers*, vol. 2, 1995, pp. 449–453.

[6] Y. Hua and T. K. Sarkar, “On SVD for estimating generalized eigenvalues of singular matrix pencil in noise,” *IEEE Transactions on Signal Processing*, vol. 39, no. 4, pp. 892–899, April 1991.

[7] R. Roy and T. Kailath, “ESPRIT - estimation of signal parameters via rotational invariance techniques,” *IEEE Transactions on Acoustics Speech and Signal Processing*, vol. 43, no. 7, pp. 984–995, July 1989.

[8] M. Viberg, “Subspace fitting concepts in sensor array processing,” PhD Thesis, Linkoping University, 1989.

[9] M. Wax and T. Kailath, “Detection of signals by information theoretic criteria,” *IEEE Transactions on Acoustics Speech and Signal Processing*, vol. 33, no. 2, pp. 387–392, April 1985.

[10] J. Rissanen, “Modeling by shortest data description,” *Automatica*, vol. 14, pp. 465–471, 1978.

[11] C. Kimball and T. Marzetta, “Semblance processing of borehole acoustic array data,” *Geophysics*, vol. 49, no. 3, pp. 264–281, March 1984.

[12] J. Capon, “High resolution frequency-wavenumber analysis,” *Proceedings of the IEEE*, vol. 57, no. 8, pp. 1408–1418, August 1969.

[13] K. Hsu and A. B. Baggeroer, “Application of the maximum likelihood method (MLM) for sonic velocity logging,” *Geophysics*, vol. 51, no. 3, pp. 780–787, March 1986.

[14] S. Umesh and D. W. Tufts, “Estimation of parameters of exponentially damped sinusoids using fast maximum likelihood estimation with application to nmr spectroscopy data,” *IEEE Transactions on Signal Processing*, vol. 44, no. 9, pp. 2745–2259, Spetember 1996.

[15] K. Ellefsen, D. Burns, and C. Chena, “Homomorphic processing of the tube wave generated during acoustic logging,” *Geophysics*, vol. 58, no. 10, pp. 1400–1407, Oct 1993.

[16] K. Hsu, C. Esmersoy, and T. L. Marzetta, “Simultaneous phase and group slowness estimation of dispersive wavefields,” in *Proc. th Annual International Geoscience and Remote Sensing Symposium (IGARSS). ‘Remote Sensing Science for the Nineties’*, 20–24 May 1990, pp. 367–370.

[17] C. Wang, “Sonic well logging methods and apparatus utilizing parametric inversion dispersive wave processing,” *US Patent, US7120541 B2*, 2004.

[18] S. Aeron, S. Bose, and H. P. Valero, “Automatic dispersion extraction using continuous wavelet transform,” in *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Las Vegas, NV, 2008.

[19] A. Roueff, J. I. Mars, J. Chanussot, and H. Pedersen, “Dispersion estimation from linear array data in the time-frequency plane,” *IEEE Transactions on Signal Processing*, vol. 53, no. 10, pp. 3738–378, October 2005.

[20] F. Auger and P. Flandrin, “Improving the reliability of time frequency time-scale representation by the reassignment method,” *IEEE Transactions on Signal Processing*, vol. 43, no. 5, pp. 1068–1089, 1995.

[21] W. H. Prossner and M. D. Seale, “Time frequency analysis of the dispersion of lamb modes,” *Journal of Acoustical Society of America*, vol. 105, no. 5, pp. 2669–2676, May 1999.

[22] H. A. Pedersen, J. Mars, and P.-O. Amblard, “Improving group velocity measurements by energy reassignment,” *Geophysics*, vol. 68, pp. 677–684, 2003.

[23] K. Hsu and C. Esmersoy, “Parametric estimation of phase and group slowness from sonic logging waveforms,” *Geophysics*, vol. 57, pp. 978–985, 1992.

[24] J. A. Tropp, A. C. Gilbert, and M. J. Strauss, “Algorithms for simultaneous sparse approximation. part II: Convex relaxation,” *Signal Processing, special issue on Sparse approximations in signal and image processing*, vol. 86, pp. 572–588, April 2006.

[25] D. L. Donoho and M. Elad, “Optimally sparse representation in general (nonorthogonal) dictionaries via ℓ_1 minimization,” *Proceedings of the National Academy of Sciences of the United States of America*, vol. 100, no. 5, p. 21972202, March 2003.

[26] J. Chen and X. Huo, “Theoretical results on sparse representations of multiple-measurement vectors,” *IEEE Transactions on Signal Processing*, vol. 54, pp. 4634–4643, December 2006.

[27] J. A. Tropp, “Algorithms for simultaneous sparse approximation. Part I: Greedy pursuit,” *Signal Processing*, vol. 86, no. 3, pp. 572–588, March 2006.

[28] D. M. Malioutov, M. Cetin, and A. S. Willsky, “Source localization by enforcing sparsity through a Laplacian prior: an SVD-based approach,” in *IEEE Workshop on Statistical Signal Processing (SSP)*, St. Louis, MO, September 2003.

[29] W. Feller, “On the Kolmogorov-Smirnov limit theorems for empirical distributions,” *Annals of Mathematical Statistics*, vol. 19, p. 177, 1948.

[30] A. N. Kolmogorov, “Sulla determinazione empirica di una legge di distribuzione,” *Giornale dell Istituto Italiano degli Attuari*, vol. 4, p. 83, 1933.

[31] H. Spath, *Cluster Dissection and Analysis: Theory, FORTRAN Programs, Examples*, T. by J. Goldschmidt, Ed. New York: Halsted Press, 1985.

[32] P. Goupillaud, A. Grossman, and J. Morlet, “Cycle-octave and related transforms in seismic signal analysis,” *Geoexploration*, vol. 23, pp. 85–102, 1985.

[33] N. Delprat, B. Escudii, P. Guillemain, R. Kronland-Martinet, P. Tchamitchian, and B. Torrésani, “Asymptotic wavelet and Gabor analysis: Extraction of instantaneous frequencies,” *IEEE Transactions on Information Theory*, vol. 38, no. 2, pp. 644–664, March 1992.

[34] H. P. Valero, “Seismic endoscopy,” Ph.D. dissertation, Institut de Physique du Globe de Paris, Paris, France, 1997.

[35] M. Grant and S. Boyd, “CVX: Matlab software for disciplined convex programming,” http://stanford.edu/ boyd/cvx, October 2008.

[36] ——, *Graph implementations for nonsmooth convex programs, Recent Advances in Learning and Control (a tribute to M. Vidyasagar)*, ser. Lecture Notes in Control and Information Sciences, S. B. V. Blondel and H. Kimura, Eds. Springer, 2008.

[37] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge University Press, 2006.