SPARSE ESTIMATION OF SPECTROSCOPIC SIGNALS

Stefan Ingi Adalbjörnsson and Andreas Jakobsson

Dept. of Mathematical Statistics, Lund University, Sweden
Email: {sia, aj}@maths.lth.se

ABSTRACT

This work considers the semi-parametric estimation of sparse spectroscopic signals, aiming to form a detailed spectral representation of both the frequency content and the spectral line widths of the occurring signals. Extending on the recent FOCUSS-based SLIM algorithm, we propose an alternative prior for a Bayesian formulation of this sparse reconstruction method, exploiting a proposed suitable prior for the noise variance. Examining three common models for spectroscopic signals, the introduced technique allows for reliable estimation of the characteristics of these models. Numerical simulations illustrate the improved performance of the proposed technique.

1. INTRODUCTION

The problem of finding an appropriate representation for signals exhibiting various forms of spectral sparsity is commonly occurring in a variety of fields, such as, for example, in characterizing and analyzing several forms of spectroscopic signals, or in the detailing of broadband sources impinging on an array of sensors. Such signals can often be viewed as consisting of a collection of broad spectral lines, where the line widths are typically of notable interest. For instance, in a communication scenario, the line width may carry information of the angular or Doppler spread of the sources (see, e.g., [1, 2]). Similarly, in the analysis of nuclear quadrupole resonance (NQR) signals from various forms of solid-state explosives, narcotics, and drugs, the line width detail important features of the decay of the spectral lines, allowing for a more accurate identification of such substances (see, e.g., [3]). In some applications, the shape of the spectral line may in itself also be of notable interest. For example, in the above noted spectroscopic application, it is reasonable to assume that the line shape may allow for the identification of the manufacturing plant of an explosive or drug containing a given active substance but being produced in different factories, a problem of significant interest, e.g., in trying to detect and trace counterfeit or substandard medicines. In this work, we will examine spectroscopic signals exhibiting spectral sparsity in some form, although the results can relatively easy be recast to a variety of related problems. In particular, the work examines the possibility to treat a collection of decaying spectral lines as sparse components, with a set of parameters detailing both the frequency and the decay of the spectral line. The work may in this sense be viewed as an extension of our recent contribution detailing spectroscopic signals using the damped IAA-style algorithm presented in [4]. Herein, we will take an approach based on the sparse reconstruction of the signal components, basing our algorithm on the framework developed in [5,6]. Rewriting the FOCUSS-based SLIM algorithm using a Bayesian framework clarifies the inference of the assumed prior on the noise variance. Here, exploiting this insight, we propose an alternative, more reasonable, prior, which we then show only imply a minor, but significant, modification, of the SLIM algorithm. The proposed prior exploits that spectroscopic signal generally decay quite rapidly, thereby allowing for a reasonable estimate of the noise variance by simply examining measurements from when the signal of interest can be assumed to have decayed sufficiently. Al-

ternatively, such an estimate can be obtained from prior measurements or calibrations, or by exploiting knowledge of the measurement setup. Using the so-obtained *a priori* estimate, we propose the use of a prior behaving close to what can be expected of that of the variance estimate, but having the benefit of dramatically simplifying the resulting calculations.

The rest of the paper is organized as follows: in the following section, we present the problem formulation and its connection to sparse reconstruction. In Section 3, the related FOCUSS [5] and SLIM [6] algorithms for solving the sparse reconstruction problem are presented in a Bayesian framework. Then, in Section 4, we propose an extension of the SLIM algorithm by utilizing the assumed availability of a prior estimate of the noise variance and its reliability. Finally, Section 5 presents numerical evaluations of the performance of the proposed algorithm.

2. PROBLEM FORMULATION

Let $y(t_n) \in \mathbb{C}$, for $n = 0, 1, \ldots, M$, represent a data sequence under consideration, where t_n are consecutive but not necessarily evenly spaced sample times, and where the data may be considered to be formed as

$$y(t_n) = \sum_{k=1}^{s} a_k f(t_n, \theta_k) + w(t_n), \qquad (1)$$

where f is a *known* complex-valued function, $\theta_k \in \Theta \subset \mathbb{R}^{m_k}$ is an unknown parameter vector, $a_k \in \mathbb{C}$ the amplitude associated with each component in the sum, and $w(t_n)$ is a (possibly non-white) circularly symmetric Gaussian noise with covariance matrix Σ_w. Common spectroscopic examples of functions detailing some or all of the signal components of a typical signal include the Lorentzian, Gaussian, and Voigt line shapes, where

$$f_L(t_n, \theta_k) = e^{-\theta_{k,1}t_n + j\theta_{k,2}t_n}$$
$$f_G(t_n, \theta_k) = e^{-\theta_{k,1}t_n^2 + j\theta_{k,2}t_n}$$
$$f_V(t_n, \theta_k) = e^{-\theta_{k,1}t_n - \theta_{k,2}t_n^2 + j\theta_{k,3}t_n}$$

respectively, with, possibly, the set of parameters themselves depending on some common parameters (see, e.g., [7,8]). Commonly, the number of spectral components, s, may be unknown due to the presence of interference or different kinds of spurious signals, thereby making parametric techniques, like non-linear least squares, unsuitable. Here, we are therefore primarily interested in estimating both the model order, s, and the unknown parameter detailing the components, $\{a_k, \theta_k\}_{k=1}^{s}$; it should be noted that number of parameters best suited to model each component may vary among the components, and is generally unknown. One option for solving such a problem is to transform the problem into a sparse reconstruction setting (see, e.g., [9]). This is achieved by restricting the parameter space for θ_k to a finite collection of values, $\tilde{\Theta}_{k,p} = \{\theta_{k,p}^l\}_{l=1}^{N_{k,p}}$, where $\theta_{k,p}^l$ denotes the l:th grid point for the p:th coefficient and the k:th element, with $N_{k,p}$ denoting the number of elements of the allowed parameter space. For notational simplicity, but without loss of generality, we here set $N = N_{k,p}, \forall k, p$, such that $\tilde{\Theta}_{k,p} = \tilde{\Theta}$. The continuous parameter space of each function f has thereby been

approximated to be restricted to a set of N discrete elements. By further expanding the sum in (1) to be formed over $D \gg s$ dictionary functions, so that

$$y(t_n) = \sum_{l=1}^{D} a_l f(t_n, \theta_l) + w(t_n) \qquad (2)$$

where thus only s of the D amplitudes a_l are non-zero, one may express the problem succinctly in matrix notation as

$$\begin{aligned}
\mathbf{y}_M &= \mathbf{A}\mathbf{x} + \mathbf{w}_M \qquad (3) \\
\mathbf{A} &= [\ \mathbf{f}_1 \ \ \dots \ \ \mathbf{f}_D \ ] \\
\mathbf{f}_i &= [\ f(t_1, \theta_k^i) \ \ \dots \ \ f(t_M, \theta_k^i) \ ]^T \qquad (4) \\
\mathbf{y}_M &= [\ y(t_1) \ \ \dots \ \ y(t_M) \ ]^T
\end{aligned}$$

where $(\cdot)^T$ denotes the transpose, $\mathbf{x} \in \mathbb{C}^D$, and $\mathbf{w}_M$ is formed similarly to $\mathbf{y}_M$. It should be stressed that the θ_k^i appearing in (4) thus corresponds to the properly selected element from $\tilde{\Theta}$, such that (3) coincides with (1) and (2). The estimation problem is thus transformed from a nonlinear optimization problem to the familiar setting of linear algebra. To estimate s and $\{a_k, \theta_k\}_{k=1}^s$ now involves finding which dictionary elements, $x_k \in \mathbf{x}$, that are non-zero, since each x_k in this way corresponds to specific values for a_k and θ_k, and as a result s may be found as being equal to the number of non-zero x_k values. Thus, the parameter estimate will be limited by the grid on which the parameter space is discretized, meaning that in order to get a high resolution estimate of θ_k, the number of potential dictionary elements in $\tilde{\Theta}$ needs to be large, and, in fact, usually larger than the number of observations. This makes (3) an underdetermined linear systems and, since one may safely assume that $\mathbf{A}$ has full rank, it has infinitely many solutions. A commonly used approach to solve this form of under-determined system of equations is by choosing the solution with the minimum ℓ_2 norm. The resulting solution is unfortunately not of interest here since it tends to give very few zero elements. However, if one assumes that it is known that the signal in (1) allows for a sparse representation, then for the transformed problem, one can assume that only a few elements in $\mathbf{x}$ are non-zero, which means that the problem is instead to find an appropriate sparse solution among the infinitely many possible solutions. This can be achieved by adding a penalty term, $\rho(\cdot)$, such that small values of $\mathbf{x}$ are shrunken to zero, i.e., the solution is reformulated to finding $\mathbf{x}$ such that

$$\min_{\mathbf{x}} ||\mathbf{y}_M - \mathbf{A}\mathbf{x}||_2^2 + \rho(\mathbf{x}), \qquad (5)$$

where often the penalty term is formed as a scaling of the ℓ_1 norm, $\sum_{r=1}^{M} |x_r|$, see e.g., [10, 11]. The literature on finding such sparse solutions is rich, being of interest in a wide range of fields, such as, for instance, machine learning [12, 13], compressive sensing [14], statistics [11], and signal processing [5, 6].

3. FOCUSS/SLIM

The above formulation results in a convex optimization problem, allowing for a unique but not necessarily optimal solution in the sense that it may not be the sparsest possible. In fact, in many cases the ℓ_1 solution is not sparse enough, which has lead to investigations that replace the ℓ_1 penalty with another function that will give even sparser results, see e.g., [15] and the references therein. Often, the so-called ℓ_q norm, formed as

$$\rho(x) = ||x||_q^q = \sum_{i=1}^{M} |x_i|^q, \quad 0 < q \le 1, \qquad (6)$$

although not a true norm, has been suggested, both with theoretical [16, 17], and empirical [5, 6, 9, 18] support. Here, we will restrict

our attention to these kind of approaches, and in particular focus on FOCUSS-based techniques such as the SLIM algorithm. Further insight of these algorithms may be gained by re-deriving them in a probabilistic setting based on an EM formulation of the algorithms. A probabilistic viewpoint of (5) is possible, as pointed out in [11], by interpreting it as an estimate of the mode of the posterior distribution, or the maximum *a posteriori* (MAP) estimate, e.g., if, for example, $\rho(\mathbf{x})$ is the ℓ_1 penalty, the optimization criterion in (5) can be seen as a log-likelihood that is consistent with a MAP estimate of a model with Gaussian observations and independent identical Laplace priors on the amplitudes of x_i. In the same way, the ℓ_q norm leads to the model

$$\begin{aligned}
p(\mathbf{y}_M|\mathbf{x}) &= \frac{1}{\pi^M |\boldsymbol{\Sigma}_w|} e^{-(\mathbf{y}_M - \mathbf{A}\mathbf{x})^H \boldsymbol{\Sigma}_w^{-1}(\mathbf{y}_M - \mathbf{A}\mathbf{x})} \\
p(\mathbf{x}) &\propto \prod_{k=1}^{n} e^{-\lambda |x_k|^q} \qquad (7)
\end{aligned}$$

where $(\cdot)^H$ denotes the conjugate transpose, $|\cdot|$ the determinant, λ is a tuning parameter that affects the sparsity of the solution, and $p(\mathbf{y}_M|\mathbf{x})$ is the probability density function of $\mathbf{y_M}$ given $\mathbf{x}$. The MAP estimate of $\mathbf{x}$, given the observed measurements, can be found as

$$\begin{aligned}
\hat{\mathbf{x}} &= \arg\max_{\mathbf{x}} \ln p(\mathbf{x}|\mathbf{y}_M) \qquad (8) \\
&= \arg\max_{\mathbf{x}} \ln p(\mathbf{y}_M|\mathbf{x}) + \ln p(\mathbf{x}). \qquad (9)
\end{aligned}$$

This estimate can be formed by observing that the penalty, i.e., the probability density $p(x_i) \propto \exp(-\lambda |x_i|^q)$ can be represented as a complex scale mixture of normal distributed variables [19], i.e., there exists a positive random variable ξ with probability distribution p_ξ such that $X = Z/\xi^{1/2}$, where Z is standard complex normal distributed, which is equivalent with

$$\begin{aligned}
p(x_i) &= \int_0^{\infty} p(x_i|\xi)p(\xi)d\xi \\
&= \int_0^{\infty} \frac{\xi}{\pi} e^{\xi x_i^* x_i} p(\xi)d\xi
\end{aligned}$$

where $(\cdot)^*$ denotes the complex conjugate. This can prove useful in cases where, e.g., it is difficult to sample from the distribution $p(x_i)$ but simpler to draw samples from both p_ξ and the standard complex normal distribution. However, by instead considering ξ as a latent variable, or, a missing sample, a MAP estimate of $\mathbf{x}$ can be obtained by the EM algorithm [20]. Extending the derivation in [21] to complex variables, the E-step yields

$$\begin{aligned}
E(\xi_i|x_i) &= \int_0^{\infty} \xi_i p(\xi_i|x_i)d\xi_i \\
&= \int_0^{\infty} \xi_i \frac{p(x_i|\xi_i)p(\xi_i)}{p(x_i)} d\xi_i \\
&= \frac{-1}{p(x_i)x_i} \int_0^{\infty} \frac{d}{dx_i^*} p(x_i|\xi_i)p(\xi)d\xi_i \\
&= \frac{-1}{p(x_i)x_i} \frac{d}{dx_i^*} \int_0^{\infty} p(x_i|\xi_i)p(\xi)d\xi_i \\
&= -\frac{p'(x_i)}{p(x_i)x_i} = \frac{q}{2}\lambda |x_i|^{q-2}, \qquad (10)
\end{aligned}$$

The following M-step then finds the $\mathbf{x}$ that minimizes the negative logarithm of the posterior distribution conditional upon the latent variables as they were calculated in the E-step, i.e.,

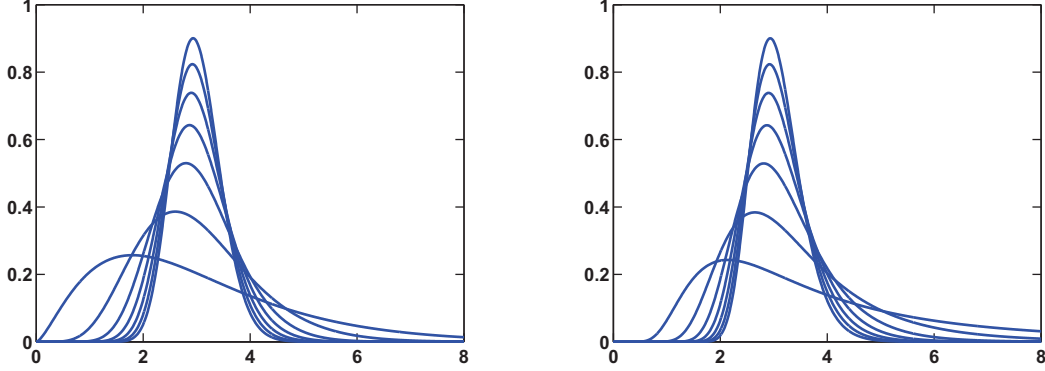$$\hat{\mathbf{x}} = \arg\min \Psi(\mathbf{x}, \hat{\xi})$$

Figure 1: (a): Density function of the $\frac{\sigma^2}{2n_0}\chi^2(2n_0)$ distribution, (b): Density functions for the scaled inv-$\chi^2(n_0, \sigma^2)$ distribution, where $2n_0$=5, 15, 30, 45, 60, 75, 90, and $\sigma^2 = 3$. For higher values of n_0, one gets a higher peak in the density function.

where

$$
\begin{aligned}
\Psi(\mathbf{x}, \hat{\xi}) & = -\ln p(\mathbf{y}_M|\mathbf{x})p(\mathbf{x}|\hat{\xi}) \\
& = \mathbf{x}^H\mathbf{A}^H\boldsymbol{\Sigma}_w^{-1}\mathbf{A}\mathbf{x} - \mathbf{x}^H\mathbf{A}^H\boldsymbol{\Sigma}_w^{-1}\mathbf{y}_m - \mathbf{y}_M^H\boldsymbol{\Sigma}_w^{-1}\mathbf{A}\mathbf{x} + \\
& \quad \mathbf{x}^H\mathbf{P}\mathbf{x} + \text{constant},
\end{aligned}
$$

with $\hat{\xi}$ denoting a vector containing the current best estimates of the latent variables from the E-step, as given in (10), $\mathbf{P} = diag(\hat{\xi})^{-1}$, with $diag(\mathbf{z})$ denoting the diagonal matrix formulated with the vector $\mathbf{z}$ along the diagonal, or alternatively, if $\mathbf{z}$ is a matrix, as the diagonal matrix resulting from retaining only the diagonal elements of $\mathbf{z}$, yielding

$$
\hat{\mathbf{x}} = \mathbf{P}\mathbf{A}^H\left[\mathbf{A}\mathbf{P}\mathbf{A}^H + \boldsymbol{\Sigma}_w\right]^{-1}\mathbf{y}_M
$$

Thus, in order to use the EM algorithm in this fashion, one needs to know the covariance matrix of the noise, $\boldsymbol{\Sigma}_w$ as well as the value of the tuning parameter λ. In the FOCUSS algorithm, the noise is assumed circularly symmetric Gaussian with known variance, i.e., $\boldsymbol{\Sigma}_w = \eta\mathbf{I}$, where η is a known positive constant. To determine λ, several methods have been suggested; re-solving for different fixed λ values to reach the assumed quality of fit or until assumed sparsity level is reached, or using an L-curve method. To resolve the issue of specifying λ, and to handle cases where the noise variance cannot be assumed known, the SLIM algorithm suggests an iterative estimation of the noise variance with λ set to $2/q$, with q being a user defined constant. This is accomplished by assuming $\boldsymbol{\Sigma}_w = \eta\mathbf{I}$, where η has a improper uniform prior distribution, $p(\eta) = 1$. This leads to a iterative estimate of η, such that it is included in the M-step. The resulting SLIM algorithm is summarized in Table 1. Convergence is reasonably fast, with 10-15 iterations, depending on the problem, often being sufficient. It should be kept in mind that each iteration becomes faster since in every iteration some x_i are set to zero, which is equivalent with the appropriate column vector being removed from the $\mathbf{A}$ matrix.

4. UTILIZING PRIOR KNOWLEDGE

In many applications, it is possible to assume some *a priori* estimate of the noise variance, σ_0^2. For spectroscopic signals, such an estimate may, for instance, be obtained from measurement when the signal of interest can reasonably be assume to only affect the measurements marginally. Here, this information is incorporated as an *a priori* distribution on the variance, which is then included

Table 1: The SLIM algorithm

Initialization:

$$
\begin{aligned}
\mathbf{x}^{(0)} & = \left[diag\left(\mathbf{A}^H\mathbf{A}\right)\right]^{-1}\mathbf{A}^H\mathbf{y}_M \\
\eta^{(0)} & = \frac{1}{M}\left\|\mathbf{y}_M - \mathbf{A}\mathbf{x}^{(0)}\right\|_2^2
\end{aligned}
$$

Then, for iteration $t = 1, \ldots, m$,

$$
\begin{aligned}
\mathbf{x}^{(t+1)} & = \mathbf{P}^{(t)}\mathbf{A}^H\left[\mathbf{A}\mathbf{P}^{(t)}\mathbf{A}^H + \eta^{(t)}\mathbf{I}\right]^{-1}\mathbf{y}_M \\
\mathbf{p}^{(t+1)} & = \left[\left|x_1^{(t+1)}\right|^{2-q} \quad \cdots \quad \left|x_D^{(t+1)}\right|^{2-q}\right]^T \\
\mathbf{P}^{(t+1)} & = diag(\mathbf{p}^{(t+1)}) \\
\eta^{(t+1)} & = \frac{1}{M}\left\|\mathbf{y}_M - \mathbf{A}\mathbf{x}^{(t+1)}\right\|_2^2
\end{aligned}
$$

in the framework presented above. This prior should preferably be selected to reflect the uncertainty in the estimated variance. For example, under the assumption that the measurements consist of n_0 independent and circularly symmetric zero mean Gaussian variables, the variance can be estimated as

$$
\hat{\eta} = \frac{1}{n_0}\sum_{i=1}^{n_0}|e(t)|^2 \tag{11}
$$

which would then imply that $\hat{\eta} \sim \frac{\sigma_0^2}{2n_0}\chi^2(2n_0)$. This distribution is shown in Figure 1(a) for different value of n_0. However, use of this distribution as a prior is somewhat problematic as it is not from a conjugate family for the variance of the multivariate normal distribution, and will thus lead to a more difficult inference problem [22]. A more convenient prior may be selected from the so-called inv-χ^2 family, described by the density [22]

$$
f_\eta(x) = \frac{e^{-n_0\sigma_0^2/x}}{x^{1+n_0}}\frac{(\sigma_0^2 n_0)^{n_0}}{\Gamma(n_0)},
$$

for the parameters $n_0 > 0$, $\sigma_0^2 > 0$, with $\Gamma(\cdot)$ denoting the Gamma-function. In particular, for the scaled inv-χ^2, the mean and the vari-
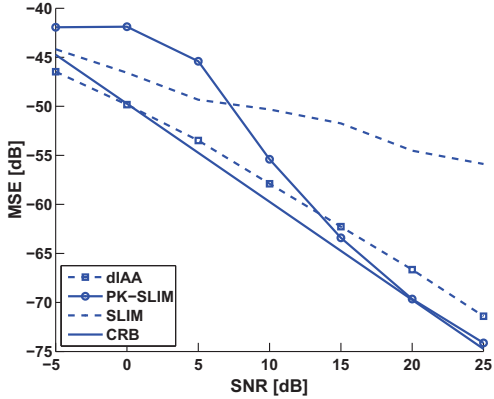
Figure 2: The MSE of the estimated damping parameter for the first component, as compared with the corresponding CRB.

Figure 3: the MSE of the estimated amplitude for the first component, as compared with the corresponding CRB.

ance can be found as

$$E(X) = \frac{n_0 \sigma_0^2}{n_0 - 2}$$

$$V(X) = \frac{2n_0^2 \sigma_0^4}{(n_0 - 2)^2 (n_0 - 4)}.$$

As shown in Figure 1(b), the scaled inv-χ^2 distribution closely resembles the desired χ^2 distribution, especially for larger values of n_0, but has the added benefit, as shown below, that it can incorporate the prior knowledge about the variance without increasing the complexity of the algorithm. Using the above presentation of the SLIM algorithm, the prior can be incorporated in the likelihood as

$$\max_{\mathbf{x}, \eta} f_\eta(\eta) (\pi \eta)^{-M} e^{-||\mathbf{y}_M - \mathbf{Ax}||^2/\eta} \prod_{k=1}^{N} e^{-2|x_k|^q/q}$$

Just as for the SLIM algorithm, we are interested in the MAP estimate of η and x, which is equivalent with finding

$$\{\hat{\mathbf{x}}, \hat{\eta}\} = \arg\max_{\mathbf{x}, \eta} \ln p(\mathbf{y}_M | \mathbf{x}, \eta) + \ln p(\mathbf{x}) + \ln f_\eta(\eta)$$

$$= \arg\min_{\mathbf{x}, \eta} g_q(\mathbf{x}, \eta)$$

where

$$g_q(\mathbf{x}, \eta) \triangleq c_1 \log \eta + \frac{n_0 \sigma_0^2}{\eta} + \frac{1}{\eta} ||\mathbf{y}_M - \mathbf{Ax}||_2^2 + \sum_{k=1}^{n} \frac{2}{q} |x_k|^q,$$

with $c_1 = M + n_0 + 1$. The update formulas for $\mathbf{x}$ will thus be the same as the one for SLIM, as given in Table 1; however, for η, one obtains

$$\eta^{(t+1)} = \arg\min_{\eta} g_q(\mathbf{x}, \eta).$$

If we consider $\mathbf{x}$ to be fixed at the current best estimate, one may find the minimum of $g_q(\mathbf{x}, \eta)$ by setting the derivative with respect to η to zero, where

$$\frac{d}{d\eta} g_q(\mathbf{x}^{(t+1)}, \eta) = \frac{(M + n_0 + 1)}{\eta} - \frac{1}{\eta^2} \left(||\mathbf{y}_M - \mathbf{Ax}||_2^2 + \frac{n_0 \sigma_0^2}{\eta} \right)$$

which lead to the new updating formulas given in Table 2. It is worth noting that, when $n_0 \to 0$, the new algorithm will coincide
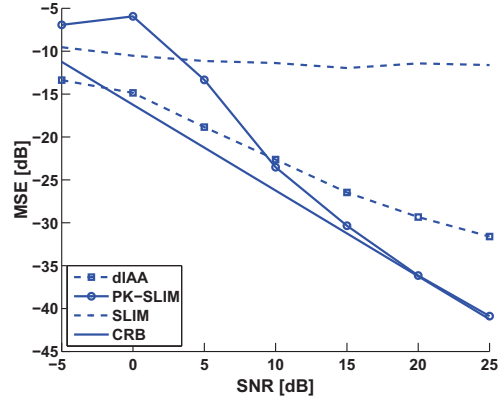
with the SLIM algorithm, and as $n_0 \to \infty$, one obtains the FOCUSS algorithm. Since the prior distribution on $|x_i|$ will cause some bias on the estimate a_k in (1), it is suggested that one only uses the results from the above proposed prior knowledge SLIM (PK-SLIM) to pick the support of the $\mathbf{x}$ vector, and then re-estimating the amplitudes by least squares or a similar robust technique.

5. NUMERICAL EXAMPLES

To examine the performance of the proposed algorithm, we examine the estimation of an irregularly sampled signal simulating a stochastic NQR measurement of TNT [23, 24]. The signal consists of two damped sinusoids that may be expressed as

$$y(t_k) = e^{j2\pi f_1 t_k - \beta_1 t_k + j\phi_1} + e^{j2\pi f_2 t_k - \beta_2 t_k + j\phi_2} + w(t_k), \quad (12)$$

where $f_1 = 0.03$Hz, $f_2 = 0.12$Hz, $\beta_1 = 0.01$Hz, $\beta_2 = 0.012$Hz, $w(t_k)$ is a white circularly symmetric Gaussian noise with variance σ_w^2, and ϕ_1, ϕ_2 are uniformly distributed on $[0, 2\pi]$. The sampling rate is normalized to 1Hz, with 6 blocks of 16 samples collected with 10.5s between each block. For each signal to noise ratio (SNR), defined here as $\sigma_w^{-2} \sigma_s^2$, where σ_s^2 denotes the signal power, 250 Monte Carlo simulations are run to estimate the mean square error (MSE) of the parameter estimates. Figure 2 and 3 illustrate the performance of the discussed estimators for the first component (the second behave similarly), as compared to the corresponding Cramér-Rao bound (CRB) [25]. In the figures, the recent dIAA algorithm [4] is also included for comparison. As can be seen from the figures, the presented PK-SLIM algorithms outperform the other algorithms noticeably for higher SNRs. Here, the assumed prior estimate of the noise variance, σ_0^2, has been estimated from merely $n_0 = 25$ measurements. Furthermore, we here use $q = 0.01$, noting that the results generally improved as q got smaller.

REFERENCES

[1] D. Astely and B. Ottersten, "The Effect of Local Scattering on Direction of Arrival Estimation with MUSIC," *IEEE transactions on Signal Processing*, vol. 45, pp. 3220–3234, 1999.

[2] A. Jakobsson, A. L. Swindlehurst, and P. Stoica, "Subspace-Based Estimation of Time-Delays and Doppler Shifts," *IEEE Trans. on Signal Processing*, vol. 46, no. 9, pp. 2472–2483, September 1998.

[3] A. Jakobsson, M. Mossberg, M. Rowe, and J. A. S. Smith, "Exploiting Temperature Dependency in the Detection of NQR Signals," *IEEE Transactions on Signal Processing*, vol. 54, no. 5, pp. 1610–1616, May 2006.

Table 2: The PK-SLIM algorithm

Initialization:

$$\mathbf{x}^{(0)} = \left[diag\left(\mathbf{A}^H \mathbf{A} \right) \right]^{-1} \mathbf{A}^H \mathbf{y}_M$$

$$\eta^{(0)} = \sigma_0^2$$

Then, for iteration $t = 1, \dots, m$,

$$\mathbf{x}^{(t+1)} = \mathbf{P}^{(t)} \mathbf{A}^H \left[\mathbf{A} \mathbf{P}^{(t)} \mathbf{A}^H + \eta^{(t)} \mathbf{I} \right]^{-1} \mathbf{y}_M$$

$$\mathbf{p}^{(t+1)} = \left[\begin{array}{ccc} \left| x_1^{(t+1)} \right|^{2-q} & \cdots & \left| x_D^{(t+1)} \right|^{2-q} \end{array} \right]^T$$

$$\mathbf{P}^{(t+1)} = diag(\mathbf{p}^{(t+1)})$$

$$\eta^{(t+1)} = \frac{||\mathbf{y}_M - \mathbf{A}\mathbf{x}||_2^2 + n_0 \sigma_0^2}{M + n_0 + 1}$$

[4] E. Gudmundson, P. Stoica, J. Li, A. Jakobsson, M. D. Rowe, J. A. S. Smith, and J. Ling, "Spectral Estimation of Irregularly Sampled Exponentially Decaying Signals with Applications to RF Spectroscopy," *Journal of Magnetic Resonance*, vol. 203, no. 1, pp. 167–176, March 2010.

[5] B. D. Rao, K. Engan, S. F. Cotter, J. Palmer, and K. Kreutz-delgado, "Subset Selection in Noise Based on Diversity Measure Minimization," *IEEE Trans. Signal Processing*, pp. 760–770, 2003.

[6] X. Tan, W. Roberts, J. Li, and P. Stoica, "Sparse Learning via Iterative Minimization With Application to MIMO Radar Imaging," *IEEE Trans. Signal Processing*, vol. 59, no. 3, pp. 1088–1101, March 2011.

[7] S. D. Somasundaram, A. Jakobsson, J. A. S. Smith, and K. Althoefer, "Exploiting Spin Echo Decay in the Detection of Nuclear Quadrupole Resonance Signals," *IEEE Trans. Geoscience and Remote Sensing*, vol. 45, no. 4, pp. 925–933, April 2007.

[8] S. I. Adalbjörnsson and A. Jakobsson, "Relax-Based Estimation of Voigt Lineshapes," in *18th European Signal Processing Conference, EUSIPCO 2010*, Aalborg, Denmark, 2010, pp. 1053–1057.

[9] C. D. Austin, R. L. Moses, J. N. Ash, and E. Ertin, "On the Relation Between Sparse Reconstruction and Parameter Estimation With Model Order Selection," *IEEE Journal of Selected Topics in Signal Processing*, vol. 4, pp. 560–570, 2010.

[10] S. S. Chen, D. L. Donoho, , and M. A. Saunders, "Atomic Decomposition by Basis Pursuit," *SIAM Review*, vol. 43, pp. 129–159, 2001.

[11] R. Tibshirani, "Regression shrinkage and selection via the Lasso," *Journal of the Royal Statistical Society B*, vol. 58, no. 1, pp. 267–288, 1996.

[12] M. Figueiredo, "Adaptive Sparseness for Supervised Learning," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 25, pp. 1150–1159, September 2003.

[13] M. Tipping, "Sparse Bayesian Learning and the Relevance Vector Machine," *J. Mach. Learn. Res.*, vol. 1, pp. 211–244, September 2001.

[14] D.L. Donoho, "Compressed Sensing," *Information Theory, IEEE Transactions on*, vol. 52, pp. 1289 –1306, 2006.

[15] G. Gasso, A. Rakotomamonjy, and S. Canu, "Recovering Sparse Signals with a Certain Family of Nonconvex Penalties and DC programming," *IEEE Trans. on Signal Processing*, vol. 57, pp. 4686–4698, December 2009.

[16] R. Chartrand and V. Staneva, "Restricted Isometry Properties and Nonconvex Compressive Sensing," *Inverse Problems*, vol. 24, no. 035020, pp. 1–14, 2008.

[17] S. Foucart and M. Lai, "Sparsest Solutions of Underdetermined Linear Systems via lq-minimization for $0 < q \leq 1$," *Applied and Computational Harmonic Analysis*, vol. 26, no. 3, pp. 395 – 407, 2009.

[18] R. Chartrand, "Exact reconstruction of sparse signals via nonconvex minimization," *IEEE Signal Processing Letters*, vol. 14, no. 10, pp. 707–710, Oct. 2007.

[19] T. Park and G. Casella, "The Bayesian Lasso," *Journal of the American Statistical Association*, vol. 103, pp. 681–686, June 2008.

[20] A.P. Dempster, N.M. Laird, and D.B. Rubin, "Maximum likelihood from incomplete data via the EM algorithm," *J. Roy. Stat. Soc. Ser.*, vol. 39, pp. 1–38, 1977.

[21] J. A. Palmer, D. P. Wipf, K. Kreutz-Delgado, and B. D. Rao, "Variational EM algorithms for Non-Gaussian Latent Variable Models," in *Advances in Neural Information Processing Systems 18*, 2006.

[22] A. Gelman, J. B. Carlin, H. S. Stern, and D. B. Rubin, *Bayesian Data Analysis, Second Edition (Chapman & Hall/CRC Texts in Statistical Science)*, Chapman and Hall/CRC, 2nd edition, 2003.

[23] M.-Y. Liao and D. B. Zax, "Analysis of Signal-to-Noise Ratios for Noise Excitation of Quadrupolar Nuclear Spins in Zero Field," *Journal of Physical Chemistry*, vol. 100, pp. 1483–1487, 1996.

[24] S. D. Somasundaram, A. Jakobsson, M. D. Rowe, J. A. S. Smith, N. R. Butt, and K. Althoefer, "Robust Detection of Stochastic Nuclear Quadrupole Resonance Signals," *IEEE Transactions on Signal Processing*, vol. 56, no. 9, pp. 4221–4229, September 2008.

[25] Y. Yao and S. P. Pandit, "Cramér-Rao Lower Bound for a Damped Sinusoidal Process," *IEEE Trans. Signal Processing*, vol. 43, no. 4, pp. 878–885, April 1995.