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 track counterfeit or substandard medicines. In this work, we will examine spectroscopic signals exhibiting spectral sparsity in some form, although the results can relatively easily 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. Alternatively, 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

\[
y(t_n) = \sum_{k=1}^{N} \theta_k f(t_n, \theta_k) + w(t_n),
\]

where \( f \) is a known complex-valued function, \( \theta_k \in \Theta \subset \mathbb{R}^p \) is an unknown parameter vector, \( \theta_k \in \Theta \) 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 \( \Sigma_w \).

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

\[
f_L(t_n, \theta_k) = e^{-\theta_k, t_n^2 + j \theta_k, t_n}
\]

\[
f_G(t_n, \theta_k) = e^{-\theta_k, t_n^2} + j \theta_k, t_n
\]

\[
f_V(t_n, \theta_k) = e^{-\theta_k, t_n^2} - \theta_k, t_n^2 + j \theta_k, t_n
\]

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, \( \{\theta_k\}_{k=1}^{S} \). It should be noted that if the number of parameters best suited to model each component may vary from 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 \( \theta_k \) to a finite collection of values, \( \Theta_{k,p} = \{ \theta_k^l \}_{l=1}^{N_k,p} \), where \( \theta_k^l \) 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 \( \Theta_{k,p} = \overline{\Theta} \). 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

$$y(t_n) = \sum_{i=1}^D a_i f(t_n, \theta_i) + w(t_n)$$

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

$$y_M = Ax + w_M$$

$$A = [f_1 \ldots f_D]$$

$$f_i = [f(t_1, \theta_i) \ldots f(t_M, \theta_i)]^T$$

$$y_M = [y(t_1) \ldots y(t_M)]^T$$

where $(\cdot)^T$ denotes the transpose, $x \in \mathbb{C}^D$, and $w_M$ is formed similarly to $y_M$. It should be stressed that the $\theta_i$ appearing in (4) thus corresponds to the properly selected element from $\Theta$, such that $(3)$ coincides with (1) and (2). The estimation problem is thus transformed from a nonlinear optimization problem to the familiar set of linear algebra. To estimate $x$ and $[a_0, \theta_0]_{n=1}^s$ now involves finding which dictionary elements, $x_k \in x$, that are non-zero, since each $x_k$ in this way corresponds to values for $a_k$ and $\theta_k$, and as a result $x$ 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 $\theta_0$, the number of potential dictionary elements in $\Theta$ needs to be large, and, in fact, usually larger than the number of observations. This makes (3) an under-determined linear system and, therefore, one may safely assume that $A$ has full rank, it has infinitely many solutions.

Focusing on the $\ell_q$ norm, formed as

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

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, $p(x)$ is the $\ell_1$ 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 $\ell_q$ norm leads to the model

$$p(y_M|x) = \frac{1}{\pi^M |\Sigma|^{1/2}} e^{-\frac{1}{2}(y_M-Ax)^\top \Sigma^{-1}(y_M-Ax)}$$

$$p(x) \propto \prod_{k=1}^n e^{-\lambda|a_k|^q}$$

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

$$\hat{x} = \arg \max_x \ln p(x|y_M)$$

$$= \arg \max_x \ln p(y_M|x) + \ln p(x).$$

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

$$p(x_i) = \int_0^\infty p(x_i|\xi)p(\xi)d\xi$$

$$= \int_0^\infty \frac{\xi}{\pi} e^{\frac{1}{\xi^2}} p(\xi)d\xi$$

where $(\cdot)^\star$ 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 $p_\xi$ and the standard complex normal distribution. However, by instead considering $\xi$ as a latent variable, or, a missing sample, a MAP estimate of $x$ can be obtained by the EM algorithm [20]. Extending the derivation in [21] to complex variables, the E-step yields

$$E(\hat{\xi}|x_i) = \int_0^\infty \hat{\xi} p(\hat{\xi}|x_i)d\hat{\xi}$$

$$= \int_0^\infty \frac{1}{p(x_i|\hat{\xi})} p(\hat{\xi})d\hat{\xi}$$

$$= -\frac{1}{p(x_i|\hat{\xi})} \int_0^\infty \frac{d}{d\hat{\xi}} p(x_i|\hat{\xi}) p(\hat{\xi})d\hat{\xi}$$

$$= -\frac{1}{p(x_i|\hat{\xi})} \int_0^\infty p(\hat{\xi}) p(\hat{\xi})d\hat{\xi}$$

$$= -\frac{p(x_i)}{p(x_i|x_i)} \frac{1}{2} \lambda |x_i|^{q-2},$$

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

$$\hat{x} = \arg \min \Psi(x, \hat{\xi})$$
where
\[ \Psi(x, \xi) = -\ln p(y_M|x)p(x|\xi) \]
\[ = x^H A^H \Sigma_w^{-1} A x - x^H A^H y_M - y_M^H \Sigma_w^{-1} A x + x^H P x + \text{constant}, \]
with \( \xi \) denoting a vector containing the current best estimates of the latent variables from the E-step, as given in (10), \( P = \text{diag}(\xi)^{-1} \), with \( \text{diag}(\xi) \) denoting the diagonal matrix formulated with the vector \( \xi \) along the diagonal, or alternatively, if \( \xi \) is a matrix, as the diagonal matrix resulting from retaining only the diagonal elements of \( \xi \), yielding
\[ \hat{x} = P A^H \left[ \hat{A} P A^H + \Sigma_w \right]^{-1} y_M \]

Thus, in order to use the EM algorithm in this fashion, one needs to know the covariance matrix of the noise, \( \Sigma_w \) as well as the value of the tuning parameter \( \lambda \). In the FOCUSS algorithm, the noise is assumed circularly symmetric Gaussian with known variance, i.e., \( \Sigma_w = \eta I \), where \( \eta \) is a known positive constant. To determine \( \lambda \), several methods have been suggested; re-solving for different fixed \( \lambda \) 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 \( \lambda \) and to handle cases where the noise variance cannot be assumed known, the SLIM algorithm suggests an iterative estimation of the noise variance with \( \lambda \) set to 2/\( q \), with \( q \) being a user defined constant. This is accomplished by assuming \( \Sigma_w = \eta I \), where \( \eta \) has an improper uniform prior distribution, \( p(\eta) = 1 \). This leads to an iterative estimate of \( \eta \), such that it is included in the M-step. 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 A matrix.

### 4. Utilizing Prior Knowledge

In many applications, it is possible to assume some a priori estimate of the noise variance, \( \sigma_{\eta}^2 \). For spectroscopic signals, such an estimate may, for instance, be obtained from measurement when the signal of interest can reasonably be assumed to only affect the measurements marginally. Here, this information is incorporated as an a priori distribution on the variance, which is then included 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
\[ \hat{\sigma}_\eta^2 = \frac{1}{n_0} \sum_{i=1}^{n_0} |\epsilon_i|^2 \]

(11)

which would then imply that \( \hat{\eta} \sim \frac{\sigma_{\eta}^2}{\sigma_{\eta}^2 + \chi^2(2n_0)} \). 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-\( \chi^2 \) family, described by the density [22]
\[ f_\eta(x) = \frac{e^{-n_0 \sigma_{\eta}^2/2}}{\chi^{n_0/2} \prod \Gamma(n_0)} \]

for the parameters \( n_0 > 0, \sigma_{\eta}^2 > 0, \) with \( \Gamma(\cdot) \) denoting the Gamma-function. In particular, for the scaled inv-\( \chi^2 \), the mean and the vari-
The scaled inverse of the covariance matrix can be found as

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

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

As shown in Figure 1(b), the scaled inverse covariance matrix closely resembles the desired \(\chi^2\) 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

$$\max_{\eta, q} f_q(\eta) (\pi^M \eta^{-M} e^{-\frac{1}{2} |y - A^\eta|^2} \prod_{k=1}^N e^{-\frac{1}{2} |y_k|^2})$$

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

$$\{(\hat{x}, \hat{\eta})\} = \arg \max_{x, \eta} \log p(y | M, x; \eta) + \log p(x; \eta) + \log f_q(\eta)$$

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

where

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

with \(c_1 = M + n_0 + 1\). The update formulas for \(x\) will thus be the same as the one for SLIM, as given in Table 1; however, for \(\eta\), one obtains

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

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

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

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 \(\eta\) will cause some bias on the estimate \(\eta_0\) 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 \(x\) 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

$$y(t_k) = e^{j2\pi f_1 t_k} \beta_1 t_k + e^{j2\pi f_2 t_k} \beta_2 t_k + w(t),$$

where \(f_1 = 0.03\text{Hz}, f_2 = 0.12\text{Hz}, \beta_1 = 0.01\text{Hz}, \beta_2 = 0.012\text{Hz}, w(t)\) is a white circularly symmetric Gaussian noise with variance \(\sigma_w^2\), and \(\phi_1, \phi_2\) 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 \(\sigma_s^2\) 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, \(\sigma_w^2\), 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


Table 2: The PK-SLIM algorithm

Initialization:

\[
\mathbf{x}^{(0)} = \left[ \text{diag} \left( \mathbf{A}^H \mathbf{A} \right) \right]^{-1} \mathbf{A}^H \mathbf{y}_M \\
\eta^{(0)} = \sigma_0^2
\]

Then, for iteration \( r = 1, \ldots, m, \)

\[
\mathbf{x}^{(r+1)} = \mathbf{P}^{(r)} \mathbf{A}^H \left[ \mathbf{A} \mathbf{P}^{(r)} \mathbf{A}^H + \eta^{(r)} \mathbf{I} \right]^{-1} \mathbf{y}_M \\
\mathbf{p}^{(r+1)} = \left[ \left| x_1^{(r+1)} \right|^{-2-q} \ldots \left| x_D^{(r+1)} \right|^{-2-q} \right]^T \\
\mathbf{P}^{(r+1)} = \text{diag} \left( \mathbf{p}^{(r+1)} \right) \\
\eta^{(r+1)} = \frac{\left| \mathbf{y}_M - \mathbf{A} \mathbf{x}^{(r)} \right|^2 + \kappa_0 \sigma_0^2}{M + m_0 + 1}
\]


