A DATA-ADAPTIVE REGULARIZATION METHOD FOR LINE SPECTRUM ESTIMATION

Ignacio Santamaría, Antonio Artés, Rafael González, Carlos Pantaleón*
DICOM, ETSII y Telecom. University of Cantabria, Santander, SPAIN
* Dpto. TSC, University of Alcala, Madrid, SPAIN
Tel: +34-942-201552; fax: +34-942-201488
e-mail: nacho@gtas.dicom.unican.es

ABSTRACT

We present a data-adaptive regularization method for estimation of line spectra. A Gaussian mixture is used as a suitable prior distribution and a different regularization term is associated to each component of the mixture. The regularized functional is minimized by applying an iterative procedure. The parameters of the mixture as well as the noise variance are updated at each iteration. In this way, the method can be applied even if accurate estimates of these parameters are not available. We apply the method to 1-D and 2-D problems showing its high-resolution ability and good performance.

1 INTRODUCTION

The problem of finding a sparse solution to a linear inverse problem appears in applications such as seismic deconvolution, deblurring of astronomical star field images or estimation of line spectra. Some existing techniques to obtain sparse solutions include exhaustive search techniques and greedy algorithms [1]. These approaches, however, are only suitable for small size problems. When the size of the problem increases, some alternative approaches formulate the restoration of a sparse signal as a constrained nonlinear minimization problem and use Linear Programming techniques to solve it [2, 3]. Recently, it has been demonstrated that a sparse solution can be obtained by solving iteratively a weighted minimum norm problem [4, 5].

An alternative group of methods use different measures of sparsity to derive penalty terms which are used to regularize the typical l_2-norm of the error [6, 7]. In particular, in [7] it has been proposed a penalty term which has proven quite effective for obtaining sparse spectra. However, its ability to resolve close peaks reduces noticeably if we do not have accurate estimates of the hyperparameters defining the distributions of the noise and the signal.

In this paper we use a more general regularization term which considers a mixture of two Gaussian distributions as a prior for the solution. Each component of the mixture is associated to a different penalty term. In this way, a sample can be penalized in a different way depending on whether it belongs to the solution or not. We use an updating procedure for the hyperparameters, which avoids the requirement of having accurate estimates in advance. At low SNR's we show that the proposed data-adaptive regularization method improves its resolution ability with respect to the non-adaptive version.

This work has been partially supported by the European Commission and the Spanish Government under Grant 1FD97-1863-C02-01.

2 PROBLEM FORMULATION

We consider herein the problem of estimating M spectral samples of a discrete-time 1-D data sequence $\mathbf{x} = (x_0, \cdots, x_{N-1})^T$ with $M > N$. If we know in advance that the data consists of a few dominant spectral lines, the harmonic retrieval problem can be cast as the following constrained nonlinear minimization

$$\min P(\mathbf{X}), \quad \text{subject to} \quad ||\mathbf{x} - \mathbf{FX}|| \leq \mathbf{e} \tag{1}$$

where $\mathbf{F}$ is a $N \times M$ matrix with elements $f_{nk} = (1/M)\exp(j2\pi nk/M)$, $P(\mathbf{X})$ is a nonlinear objective function which incorporates our knowledge about the solution, and the constraint vector $\mathbf{e}$ represents the noise and other sources of uncertainty.

The selection of an adequate objective function, $P(\mathbf{X})$, is a key step in order to control the shape of the solution. For instance, some methods use l_p-norms or quasi-norms ($p \neq 2$), which are known to favor sparse solutions. The sparse solution can be obtained by applying Linear Programming techniques [2, 3] or weighted norm minimization algorithms such as FOCUSS (FOCal Undetermined System Solver) [4, 5].

Alternatively, instead of solving (1) directly, we can minimize the following regularized functional

$$J(\mathbf{X}) = \frac{1}{\sigma^2}||\mathbf{x} - \mathbf{FX}||_2^2 + P(\mathbf{X}) \tag{2}$$

where σ^2 is the noise variance. This regularized approach admits the following Bayesian interpretation: the minimization of (2) gives the maximum *a posteriori* (MAP) estimate of $\mathbf{X}$ if we assume that the noise is Gaussian and that $P(\mathbf{X})$ may be regarded as a prior distribution for the complex vector $\mathbf{X}$.

In this context, several regularization terms, $P(\mathbf{X})$, have been recently applied to different problems [6, 7]. In particular, the following regularizer has been proposed in [7] as a measure of sparsity

$$P(\mathbf{X}) = \sum_{k=0}^{M-1} \ln\left(1 + \frac{X_k X_k^*}{\sigma_x^2}\right). \tag{3}$$

This regularizer has proven quite effective in finding sparse spectra; however, it suffers from two drawbacks. The first one is that, from our experience the estimation of the hyperparameters (σ_x^2 and σ^2) is a critical step. The second drawback is that all the samples are treated in the same way, i.e., no matter whether they belong to the solution or not, they are penalized as $\ln\left(1 + X_k X_k^*/\sigma_x^2\right)$. Intuitively, it seems more reasonable that the samples belonging to the solution would be less penalized (ideally, not penalized at all) than the samples that must be pulled to zero.

3 ADAPTIVE REGULARIZATION

3.1 Regularization by a mixture model

In this section we propose a more general adaptive regularization method, which avoids the drawbacks of using (3). The main idea consists of using a mixture of models: G_1 (modeling the nonzero samples of the sparse solution) and G_2 (for the samples that must be zeroed). Using this approach, the samples can be penalized in a different manner depending on whether they are better modeled by G_1 or G_2

$$X_k = \begin{cases} \text{modeled by } G_1 & \Rightarrow \text{ penalty term } P_1(X_k), \\ \text{modeled by } G_2 & \Rightarrow \text{ penalty term } P_2(X_k); \end{cases} \tag{4}$$

therefore, the proposed regularizer can be written as

$$P(\mathbf{X}) = \sum_{k=0}^{M-1} \sum_{j=1}^{2} I_j(X_k) \ln\left(1 + \frac{X_k X_k^*}{\sigma_j^2}\right) \tag{5}$$

where

$$I_j(X_k) = \begin{cases} 1, & \text{if } X_k \text{ is modeled by } G_j \\ 0, & \text{otherwise;} \end{cases} \tag{6}$$

is a selection variable, which counts the number of samples modeled by each distribution.

We now describe the iterative procedure to minimize the cost function. Considering a fixed set of parameters (σ^2, σ_1^2 and σ_2^2), taking derivatives of (2) with respect to $\mathbf{X}^*$ and equating to zero, we obtain that the solution to our problem must fulfill

$$\mathbf{X} = \left(\sigma^2 \mathbf{R} + \mathbf{F}^H \mathbf{F}\right)^{-1} \mathbf{F}^H \mathbf{x} \tag{7}$$

where $\mathbf{R}$ is a $M \times M$ diagonal matrix with elements

$$R_{kk} = \sum_{j=1}^{2} \frac{1}{\sigma_j^2} \frac{I_j(X_k)}{1 + X_k X_k^*/(\sigma_j^2)}. \tag{8}$$

Since the elements of the regularization matrix $\mathbf{R}$ depend on the solution $\mathbf{X}$, an iterative procedure is needed to solve (7). An efficient method is suggested in [7]: starting from an initial solution $\mathbf{X}_0$ (given by the DFT, for instance) we calculate the matrix $\mathbf{R}_0$; then, at each iteration the method obtains

$$\mathbf{b}_n = \left(\sigma^2 \mathbf{I}_N + \mathbf{F}\mathbf{R}_n^{-1}\mathbf{F}^H\right)^{-1} \mathbf{x} \tag{9}$$

and the solution is updated as

$$\mathbf{X}_n = \mathbf{R}_n^{-1} \mathbf{F}^H \mathbf{b}_n. \tag{10}$$

3.2 Updating the hyperparameters

The improvement and novelty of the proposed method with respect to [7] come from the fact that the selection variables $I_j(X_k)$ as well as the parameters of the penalty terms can be updated at each iteration. Specifically, if $p_j(X_k)$ denotes the probability density of X_k under model G_j; then, applying Bayes, $I_j(X_k)$ can be estimated as a posterior probability

$$\hat{I}_j(X_k) = Prob\left(G_j | X_k\right) = \frac{\pi_j p_j(X_k)}{\sum_i \pi_i p_i(X_k)} \tag{11}$$

where π_j ($j = 1, 2$), are the prior probabilities for each distribution and they are therefore constrained to sum 1.

Unlike standard Bayesian regularization procedures for which the penalty term may be regarded as a prior distribution for the solution, here the prior is the mixture of distributions G_1 and G_2, while each penalty terms is considered as a deterministic cost functions associated to each distribution of the mixture.

Therefore, keeping the regularizer term as (5), we could use several distributions as components of the mixture. A reasonable choice, which leads to easy and compact formulas for updating its parameters, is to use zero-mean Gaussians with different variances σ_1^2 and σ_2^2. Note that these variances coincide with those used in the penalty terms. Using this choice, the mixture parameters at the nth iteration can be updated according to

$$\sigma_{j,n}^2 = \gamma \sigma_{j,n-1}^2 + (1-\gamma) \frac{\sum_k \hat{I}_j(X_k) X_k X_k^*}{M \sum_k \hat{I}_j(X_k)}, \tag{12}$$

$$\pi_{j,n} = \gamma \pi_{j,n-1} + (1-\gamma) \frac{\sum_k \hat{I}_j(X_k)}{M} \tag{13}$$

γ being a constant near to 1, which forces a smooth evolution of the mixture.

In a typical situation the Gaussians of the mixture are initialized with slightly different variances. As long as the algorithm proceeds, one of the Gaussians shrinks its variance and then, its penalty term increases and the samples modeled by that Gaussian are strongly pushed to zero. Conversely, the other Gaussian tends to increase its variance and then, since $X_k X_k^*/\sigma_j^2 < 1$, the penalty for these samples can be made arbitrarily low even if they have large values.

A similar approach can be used to update the noise variance at each iteration; specifically, we have applied the following iteration

$$\sigma_n^2 = \gamma\sigma_{n-1}^2 + \frac{(1-\gamma)}{N}||\mathbf{x} - \mathbf{F}\mathbf{X}_{n-1}||_2^2 \tag{14}$$

where $\mathbf{X}_{n-1}$ is the solution after $n-1$ iterations, which is given by (10)

3.3 Relationship to FOCUSS algorithms

The proposed data-adaptive regularization method can be related to weighted minimum norm algorithms such as FOCUSS [4, 5]. To show this connection, let us consider a situation without noise; in this case, our reconstruction problem can be rewritten as

$$\min P(\mathbf{X}), \qquad \text{subject to} \qquad \mathbf{x} = \mathbf{F}\mathbf{X}. \tag{15}$$

To solve this problem we introduce a vector of Lagrange multipliers $\lambda = (\lambda_1, \cdots, \lambda_N)^T$ and minimize

$$J(\mathbf{X}) = P(\mathbf{X}) + \lambda^T(\mathbf{x} - \mathbf{F}\mathbf{X}). \tag{16}$$

Its solution is given by

$$\mathbf{X} = \mathbf{R}^{-1}\mathbf{F}^H\left(\mathbf{F}\mathbf{R}^{-1}\mathbf{F}^H\right)^{-1}\mathbf{x} \tag{17}$$

where $\mathbf{R}$ is given by (8). Obviously, this solution coincides with (10) substituting $\sigma^2 = 0$ in (9).

As it is shown in [5], the same result can be alternatively obtained by minimizing the following weighted norm problem

$$\min ||\mathbf{W}^{-1}\mathbf{X}||, \qquad \text{subject to} \qquad \mathbf{x} = \mathbf{F}\mathbf{X}, \tag{18}$$

in this case, the weighting matrix $\mathbf{W}$ must be chosen as

$$\mathbf{W}^2 = \mathbf{R}^{-1}. \tag{19}$$

Note that the weighting matrix is positive definite.

In this context, the proposed approach can be regarded as a FOCUSS-like method in which we use a data-adaptive weighting matrix. In fact, at each iteration we solve a weighted constrained minimum norm problem as

$$\mathbf{X}_n = \mathbf{W}_n(\mathbf{F}\mathbf{W}_n)^\sharp\mathbf{x} \tag{20}$$

where the superscript $\sharp$ denotes the Moore-Penrose pseudoinverse. The weighting matrix is obtained from (19) and (8) using the solution from the previous iteration.

This relationship between the proposed approach and FOCUSS can be extended to a noisy situation by considering at each step a regularized weighted minimum norm problem.

4 SIMULATION RESULTS

4.1 1-D Line Spectrum

To show the performance of the method we first consider a 1-D data sequence with $N = 32$ data samples. The register contains two spectral lines at frequencies 0.33 and 0.33+0.4/32 (i.e., the frequency spacing is clearly below the Fourier resolution limit given by $1/N$). Figure 1 shows the original spectrum and the final sparse solution obtained with the proposed method using $M = 256$; the SNR is 5 dB. The mixture parameters are initialized as $\pi_1 = \pi_2 = 0.5$, $\sigma_1^2 = \sigma_x^2/8$ and $\sigma_2^2 = \sigma_x^2/16$, with σ_x^2 being the variance of the data. On the other hand, the noise variance is initialized as $\sigma^2 = \sigma_1^2$. A maximum of 15 iterations are carried out. Figure 2 shows the evolution of the variances for the mixture and the noise.

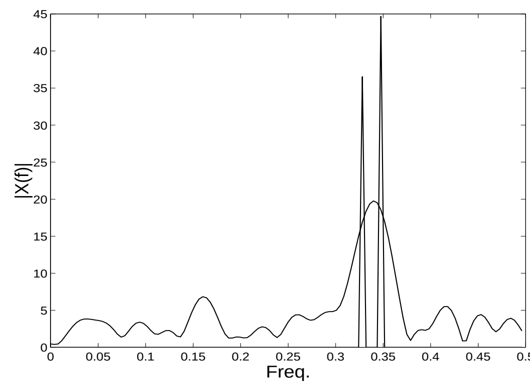


Figure 1: Original spectrum and sparse solution.

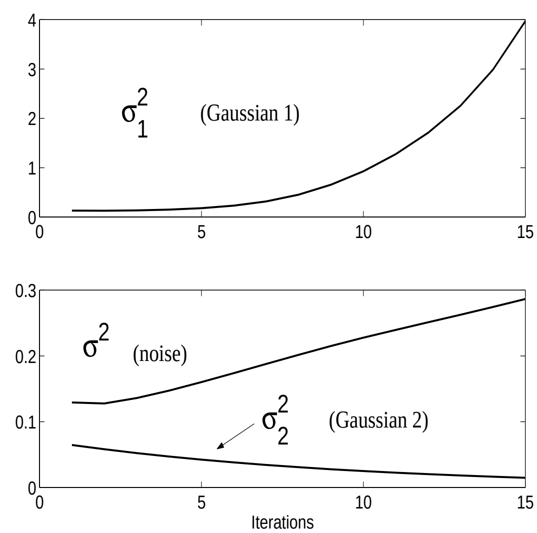


Figure 2: Evolution of the mixture and noise variances.

Finally, Figure 3 shows the resolution probabilities for the proposed adaptive regularization method, the non-adaptive regularization method of [7] and a high-resolution method such as ESPRIT. For the non-adaptive method of [7] we have selected as fixed hyperparameters the true noise variance and the best performing signal variance. For each SNR we average the results of 500 trials; in each trial the two sources were considered resolved if the spectrum had two peaks and each peak was within an interval $[0.9f_{true}, 1.1f_{true}]$.

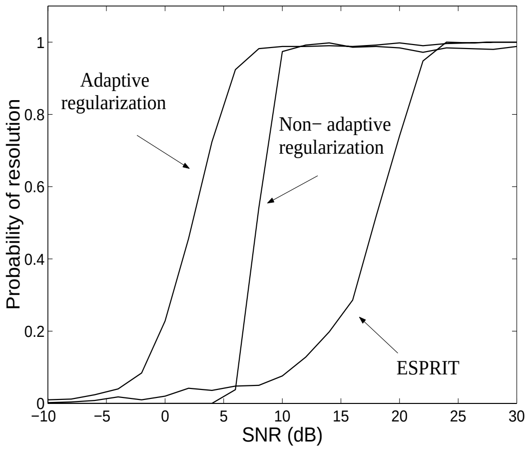


Figure 3: Probability of resolution.

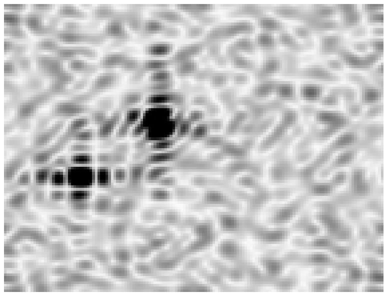
4.2 2-D Line Spectrum

For 2-D data, the line spectrum estimate can be obtained by applying the proposed 1-D approach to each row of the data matrix to obtain an intermediate $N \times M$ matrix and then to the columns of this matrix to obtain the final $M \times M$ estimate. A similar result is obtained if the columns are processed first.

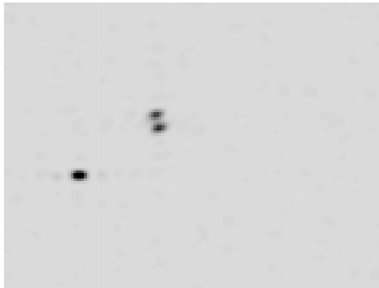
In particular, we consider its application to Synthetic Aperture Radar (SAR) imaging. As an example, Figure 4 shows a SAR image (24 × 24) obtained using a 2D-FFT (M=256) and the result after applying the proposed adaptive regularization method. The original data contained three spectral lines, which simulate corner reflectors and the SNR was 5 dB.

5 CONCLUSIONS

In this paper we have presented a new adaptive regularization method suitable for estimation of line spectra. The method incorporates an updating procedure for the hyperparameters; therefore, it can be applied without any knowledge about the signal or the noise. We have shown that the method can be reformulated as a weighted minimun norm algorithm. Finally, some computer simulations suggest that the proposed method outperforms its non-adaptive version mainly at low SNRs.



a)



b)

Figure 4: Spectrum of a 2-D sequence. a) FFT, b) Proposed method.

References

[1] B. K. Natarajan, “Sparse approximate solutions to linear systems”, *SIAM J. Comput.*,vol. 24, pp. 227-234, April 1995.

[2] B. D. Jeffs, M. Gunsay, “Restoration of blurred star field images by maximally sparse optimization”, *IEEE Trans. Image Processing*, vol. 2, pp. 202-211, 1993.

[3] M. S. O’Brien, A. N. Sinclair, S. M. Kramer, “Recovery of a sparse spike time series by l_1 norm deconvolution”, *IEEE Trans. Signal Processing*, vol. 42, pp. 3353-3365, 1994.

[4] I. F. Gorodnitsky, B. D. Rao, “Sparse signal reconstruction from limited data using FOCUSS: A reweighted minimum norm algorithm”, *IEEE Trans. Signal Processing*, vol. 45, pp. 600-615, 1997.

[5] B. D. Rao, K. Kreutz-Delgado, “An affine scale methodology for best basis selection”, *IEEE Trans. Signal Processing*, vol. 47, pp. 187-199, 1999.

[6] I. Santamaría, C. Pantaleón, A. Artés, “Sparse deconvolution using adaptive mixed-Gaussian models”, *Signal Processing*, vol. 54, pp. 161-171, 1996.

[7] M. D. Sacchi, T. J. Ulrych, C. J. Walker, “Interpolation and extrapolation using high-resolution discrete Fourier transform”, *IEEE Trans. Signal Processing*, vol. 46, pp. 31-37, 1998.