APPLICATION OF ℓ_P-REGULARIZED LEAST SQUARES FOR $0 \leq P \leq 1$ IN ESTIMATING DISCRETE SPECTRUM MODELS FROM SPARSE FREQUENCY MEASUREMENTS

Mu-Hsin Wei, James H. McClellan, and Waymond R. Scott, Jr.

Georgia Institute of Technology
School of Electrical and Computer Engineering
777 Atlantic Drive NW, Atlanta, GA 30332-0250

ABSTRACT

It is difficult to robustly estimate the parameters of an additive exponential model from a small number of frequency-domain measurements, especially when the model order is unknown and the parameters must be constrained to be real. Recent work in sparse sampling and sparse reconstruction casts this problem as a linear dictionary selection problem by densely sampling the parameter space. We present a modified ℓ_p-regularized least squares algorithm, for $0 \leq p \leq 1$, and show that it is effective when the frequency sampling is sparse over a couple of decades and the parameters must be estimated over more than four decades. An empirical method for choosing the regularization parameter is also studied. Using tests on synthetic data and laboratory measurements for an EMI application, the proposed method is shown to provide robust estimates of the model parameters up to eighth order.

***Index Terms*—** Parameter estimation, ℓ_1 minimization, sum of exponentials, basis pursuit.

1. INTRODUCTION

Additive exponential models are commonly used in science and engineering to model a wide range of physical phenomena such as the eddy currents for electromagnetic induction (EMI), dielectric material properties in polymer science, and many others in the fields of chemistry, biology, and speech, to name a few. It can be difficult to extract the model parameters from measurements when the number of parameters is more than a few, when the model order is unknown, or when the number of measurements is very small. Furthermore, when the measurements are made in the frequency domain and the exponential parameters must be real, very few techniques exist to solve the resulting constrained estimation problem. In [1] we developed a constrained linear least-squares method to estimate these models when the parameters are nonnegative, and demonstrated the utility and robustness of this algorithm for wideband EMI systems. The method uses 21 measurements, equi-spaced in the *logarithmic* frequency domain, taken over a range of 2.5 decades, and estimates parameter values over 4.17 decades. We have tested this method on synthetic data and lab data, and have shown that it can reliably extract models. Recently, Austin *et al.* [2] studied parameter estimation for additive models through sparse sampling and reconstruction. They formulated a linear problem by enumerating an overcomplete dictionary of possible models. Then they proposed a sparse nonuniform sampling strategy based on the Fisher information, and demonstrated their method on a time-domain sum-of-exponentials model.

In this paper, we extend the frequency-domain technique [1] to remove the nonnegative constraint. We linearize the estimation problem with a dictionary as in [1, 2], and solve it with ℓ_p-regularized least squares for $0 \leq p \leq 1$. We exploit the fact that after linearizing the problem the solution vector is most likely sparse, and we show that a log-frequency sampling scheme performs nearly the same as one based on Fisher information. The proposed method requires no prior knowledge of the model order K and always returns real parameters. We demonstrate its robustness with results on synthetic and laboratory data, even when using high model orders and frequencies measured over several decades.

This work is supported in part by the US Army REDCOM CERDEC Night Vision and Electronic Sensors Directorate, Science and Technology Division, Countermine Branch and in part by the U. S. Army Research Office under Contract Number W911NF-05-1-0257.

2. ESTIMATION METHOD

The EMI frequency response $H(\omega)$ of a metallic target, which is proportional to a projection of the magnetic polarizability tensor of the target,can be expressed as [3]:

$$H(\omega) = c_0 + \sum_{k=1}^{K} \frac{c_k}{1 + j\omega/\zeta_k} \quad (1)$$

where c_0 is the shift, K the model order, c_k the real spectral amplitudes, and ζ_k the relaxation frequencies. The parameter set $S = \{(\zeta_k, c_k) : k = 1 \dots K\}$ is called the Discrete Spectrum of Relaxation Frequencies (DSRF); each pair (ζ_k, c_k) is one relaxation. The term DSRF and spectrum will be used interchangeably.

It is advantageous to model the EMI signal with (1) because the relaxation frequencies are invariant to target orientation, which is valuable in target detection. However, it is difficult in practice to obtain the model parameters in (1) from a small number of measurements. For most existing estimation methods, a good guess of the model order K is required for the fitting process to converge. Prior knowledge of K, however, is usually unavailable. The highly correlated summands in (1) and the nonlinear relation between $H(\omega)$ and ζ_k also make estimation difficult. Most existing methods often give sub-optimal solutions that are far from the truth, or return complex parameters that do not have physical meaning [4].

When the target response is measured at N distinct frequencies, (1) can be written in matrix form:

$$\begin{bmatrix} H(\omega_1) \\ H(\omega_2) \\ \vdots \\ H(\omega_N) \end{bmatrix} = \underbrace{\begin{bmatrix} 1 & \frac{1}{1+j\omega_1/\zeta_1} & \frac{1}{1+j\omega_1/\zeta_2} & \cdots & \frac{1}{1+j\omega_1/\zeta_K} \\ 1 & \frac{1}{1+j\omega_2/\zeta_1} & \frac{1}{1+j\omega_2/\zeta_2} & \cdots & \frac{1}{1+j\omega_2/\zeta_K} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & \frac{1}{1+j\omega_N/\zeta_1} & \frac{1}{1+j\omega_N/\zeta_2} & \cdots & \frac{1}{1+j\omega_N/\zeta_K} \end{bmatrix}}_{\boldsymbol{Z}} \begin{bmatrix} c_0 \\ c_1 \\ c_2 \\ \vdots \\ c_K \end{bmatrix}$$

$$\boldsymbol{h} = \boldsymbol{Z}\boldsymbol{c} \tag{2}$$

where $\omega_{\min} = \omega_1 < \omega_2 < \cdots < \omega_N = \omega_{\max}$, $\boldsymbol{h}$ is the observation vector, $\boldsymbol{c}$ the spectral amplitude vector augmented by the shift c_0, and $\boldsymbol{Z}$ a matrix containing information about the relaxation frequencies $\boldsymbol{\zeta}$.

To estimate the DSRF (i.e., ζ_k and c_k) from a given observation $\boldsymbol{h}$, the usual approach is to minimize the norm of the error, but this leads to a *nonlinear* optimization problem. Instead, we follow the strategy of basis pursuit to linearize the problem with an overcomplete dictionary. The overcomplete dictionary is a matrix $\tilde{\boldsymbol{Z}}$ that has the same form as $\boldsymbol{Z}$ in (2), but with many more columns. To generate the columns, we enumerate a large set of M possible relaxation frequencies uniformly distributed in the log-ζ space ($M \gg K$), and create one column for each enumerated ζ [1]. Compared to the non-uniform sampling [2] based on the Fisher information where the step size is given by

$$\Delta(\tilde{\zeta}_m) = \alpha \left[\sum_{n=1}^{N} \left(\frac{\omega_n}{\omega_n^2 + \tilde{\zeta}_m^2} \right)^2 \right]^{-\frac{1}{2}}$$

the uniform log-ζ sparse sampling gives similar sample points (Fig. 1).

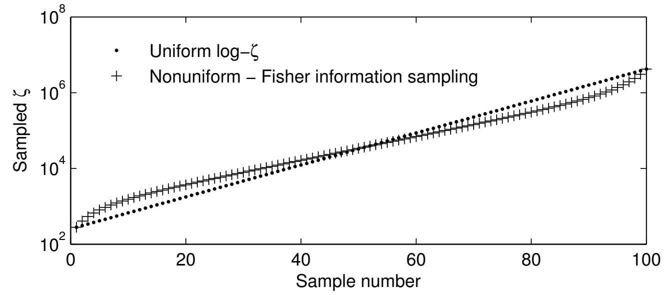


Fig. 1. Samples generated from uniform log-ζ sampling and non-uniform sampling based on the Fisher information.

Since the dictionary matrix $\tilde{\boldsymbol{Z}}$ has (M+1) columns, we redefine the unknown as an (M+1)-element *weighted selector vector* $\tilde{\boldsymbol{c}}$ and rewrite the problem as:

$$\boldsymbol{h} = \tilde{\boldsymbol{Z}}\tilde{\boldsymbol{c}} + error \tag{3}$$

The vector $\tilde{\boldsymbol{c}}$ contains the shift estimator $\tilde{c}_0$ followed by the spectral amplitude estimators $\tilde{c}_m$. We expect the solution for $\tilde{\boldsymbol{c}}$ to have many zero elements because $M \gg K$, i.e., $\tilde{\boldsymbol{c}}$ will be sparse. We utilize the ℓ_p-regularized least squares technique, for $0 \le p \le 1$, because it promotes sparse solutions.

$$\arg\min_{\tilde{\boldsymbol{c}}} ||\tilde{\boldsymbol{Z}}'\tilde{\boldsymbol{c}} - \boldsymbol{h}'||_2^2 + \lambda||\tilde{\boldsymbol{c}}||_p^p \quad ,0 \le p \le 1 \tag{4}$$

$$\text{where } \tilde{\boldsymbol{Z}}' = \begin{bmatrix} \mathfrak{Re}(\tilde{\boldsymbol{Z}}) \\ \mathfrak{Im}(\tilde{\boldsymbol{Z}}) \end{bmatrix} \text{ and } \boldsymbol{h}' = \begin{bmatrix} \mathfrak{Re}(\boldsymbol{h}) \\ \mathfrak{Im}(\boldsymbol{h}) \end{bmatrix}$$

where λ is the regularization parameter. Separating the real and imaginary parts in $\tilde{\boldsymbol{Z}}$ makes the whole system real. Ideally, in the best selected $\tilde{\boldsymbol{c}}$, only those $\tilde{c}_m$ with corresponding $\tilde{\zeta}_m$ that are near a true ζ_k will be nonzero, and they will take on the correct spectral amplitudes c_k. It follows that a DSRF can then be deduced from the nonzero estimated $\tilde{c}_m$ and their corresponding $\tilde{\zeta}_m$.

The ℓ_p-regularized least squares for $p < 1$ can be approximated by the iteratively reweighted ℓ_1 algorithm proposed by Candès *et al.* [5]. The weights are updated as suggested in [6]. We also adopt the ϵ-regularization technique used in the same paper. In summary, (4) is approximated by (see also [7]):

Algorithm 1: Approximated ℓ_p-regularized least square

Input: $\tilde{\boldsymbol{Z}}'$, $\boldsymbol{h}'$, p, λ, $\tilde{\boldsymbol{c}}^0$

$\tilde{\boldsymbol{c}}^n \leftarrow \tilde{\boldsymbol{c}}^0$

for $k \leftarrow 0$ **to** -8 **step** -1 **do**

  $\epsilon \leftarrow 10^k$

  repeat

    $\tilde{\boldsymbol{c}}^{n-1} \leftarrow \tilde{\boldsymbol{c}}^n$

    $w_i^n \leftarrow (|\tilde{c}_i^{n-1}| + \epsilon)^{p-1}$

    $\tilde{\boldsymbol{c}}^n \leftarrow \arg\min ||\tilde{\boldsymbol{Z}}'\tilde{\boldsymbol{c}} - \boldsymbol{h}'||_2^2 + \lambda \sum_{i=1}^{M+1} w_i^n |\tilde{c}_i|$

  until $||\tilde{\boldsymbol{c}}^n - \tilde{\boldsymbol{c}}^{n-1}||_2 < \sqrt{\epsilon}/100$

return $\tilde{\boldsymbol{c}}^n$

The ℓ_1 minimization problem is solved by **l1_ls**, a MATLAB optimizer proposed by Kim *et al.* [8]. We have also found that normalizing the input data $\boldsymbol{h}$, as well as the columns of $\tilde{\boldsymbol{Z}}'$ to have unit ℓ_2 norm increases the accuracy of estimation. Setting entries of $\tilde{\boldsymbol{c}}^0$ to all ones also seems to be effective. The nonzero entries of $\tilde{\boldsymbol{c}}$ selected by (4) are the relaxations need in the estimated DSRF, $\hat{S} = \{(\hat{\zeta}_l, \hat{c}_l) : l = 1 \ldots L\}$.

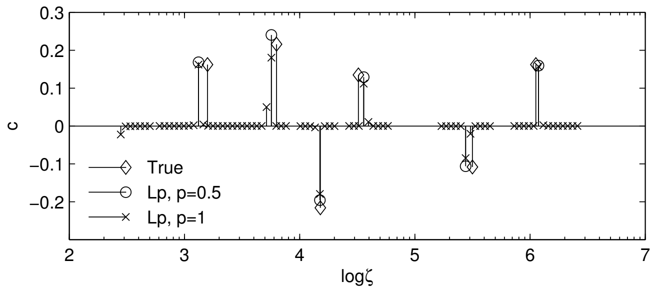


Fig. 2. Estimation of a synthetic six-relaxation DSRF. $p = 0.5$.

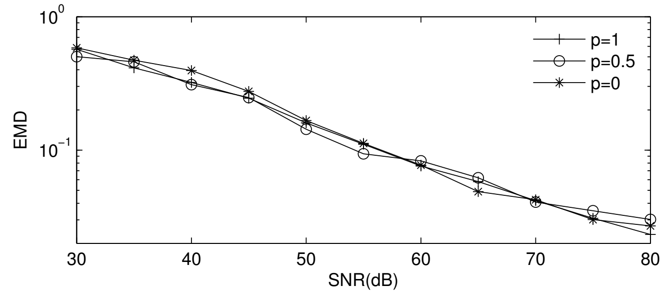


Fig. 3. Monte Carlo simulation on goodness of estimation vs. SNR performed on a four-relaxation DSRF. Sample size is 100 per SNR.

3. ESTIMATION RESULTS

The proposed estimation method is tested against synthetic and laboratory data to show its functionality, accuracy, and stability. The hardware system used is a wideband EMI sensor operating at 21 frequencies approximately logarithmically distributed over the range 300 Hz–90 kHz (2.5 decades) [9]. The synthetic data is generated in accordance with the hardware specification. The range of ζ for estimation is chosen such that $\log(\tilde{\zeta}_{\min})$ and $\log(\tilde{\zeta}_{\max})$ are 2.45 and 6.62, respectively, i.e., 4.17 decades. All estimations are performed with $M = 100$, and all presented spectra are normalized such that $\sum_{i=1} |c_i| = 1$. Spectral amplitudes less than 10^{-5} are not displayed. Unless specified, $p = 0.5$ is chosen as a representative case. The choice of λ is discussed in Section 4.
Notation: ζ and c are the true/theoretical relaxation frequencies and spectral amplitudes; $\hat{\zeta}$ and $\hat{c}$ are the estimates.

3.1. Dissimilarity Measure Between Two DSRFs

In order to evaluate the goodness of the estimate, we need to define a measure of dissimilarity that is appropriate for sparse spectra with multiple peaks. We use the Earth Mover's Distance (EMD) [10] which quantifies the "amount of work" to morph one spectrum into the other. Strictly speaking the EMD is only defined for positive spectra, but we can account for negative spectral amplitudes by defining the distance function between two relaxations (ζ_i, c_i) and ($\hat{\zeta}_j$, $\hat{c}_j$) to be:

$$d_{ij} = \begin{cases} |\log \zeta_i - \log \hat{\zeta}_j| & , c_i \hat{c}_j \geq 0 \\ 1 + |\log \zeta_i - \log \hat{\zeta}_j| & , c_i \hat{c}_j < 0 \end{cases}$$

which penalizes relaxations with different signs. Spectra are made nonnegative and normalized prior to the EMD computation. Notice that the EMD has units of decades.

3.2. Synthetic Six-relaxation DSRF

We test our method on a six-relaxation DSRF synthesized at 65 dB SNR with AWGN (Fig. 2). This is a case that cannot be handled by traditional nonlinear parameter optimization which tends to return complex-valued estimates [4]. Using $p = 0.5$, all six relaxation frequencies are recovered by solving (4). Though the estimation is not perfect, it is nevertheless satisfactory. The estimated model parameters are real, and the deviation from truth is small. The EMD between the estimated and the true DSRF is 0.09 decades. The estimate using $p = 1$ is also shown in Fig. 2. It is less sparse, but its EMD is still small, 0.10 decades. Satisfactory estimates are also observed for model orders up to eight.

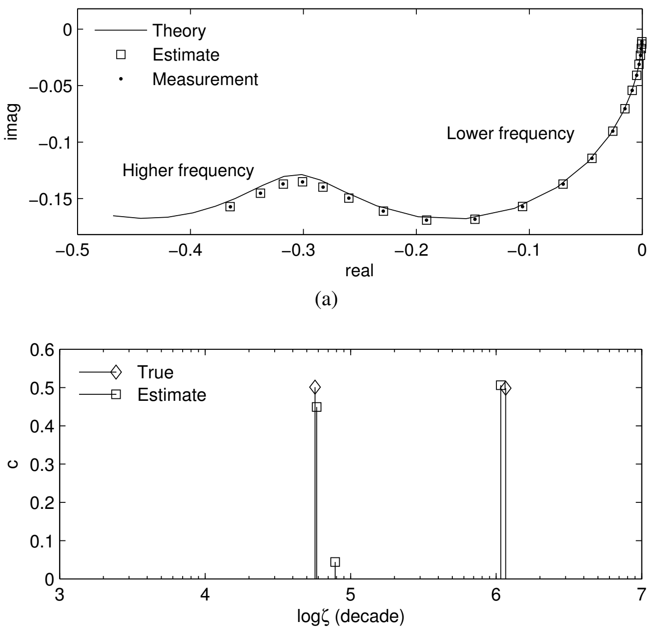
3.3. Signal to Noise Ratio

To see how the proposed method performs in noise, a Monte Carlo simulation versus SNR is run for several p's. The true spectrum is a target with a four-relaxation DSRF including negative relaxations. The simulation result, shown in Fig. 3, shows the robustness of the estimation method at different signal-to-noise ratios. For all p's, the EMD between the estimate and the truth increases as the SNR decreases. This suggests that the proposed method is functional in a range of SNR where the EMD is below some threshold. This threshold, however, depends on the application of the estimated spectrum. For example, if the DSRF produces features for classification, a more robust classifier may tolerate worse estimations and, therefore, allow lower SNR. It seems $p < 1$ offers performance similar to that of $p = 1$, but $p < 1$ does give sparser estimates as demonstrated above. The sparsity, however, is not reflected in the EMD measure.

3.4. Two Coplanar Coaxial Circular Loops

For laboratory data, a target with two coplanar coaxial circular loops of copper wire was constructed. The circumferences of the two loops were chosen to be 200 mm and 150 mm. The larger loop has a wire radius of 0.06 mm, and the smaller one of 0.32 mm. The EMI response of this target was measured in the laboratory and is shown in Fig. 4(a), and the estimated and theoretical DSRF are displayed in Fig. 4(b).

The estimated DSRF deviates from the theory with an EMD of 0.04 decades. We believe the extra estimated relaxation and the deviation is mostly due to the thin-wire approximation used in the theory [1]. The measured frequency response itself deviates from the theory slightly, but the devia-



(a)

(b)

Fig. 4. (a) Laboratory measured frequency response of two coplanar coaxial circular loops. Responses are normalized such that $\|\boldsymbol{h}\|_2 = 1$. (b) Theoretical and estimated DSRF. $\log\zeta_k$ and c_k are (4.76 6.07) and (0.50 0.50), respectively. $\log\hat{\zeta}_l$ and $\hat{c}_l$ are (4.77 4.89 6.03) and (0.45 0.04 0.51), respectively.

tion is small. We conclude that this estimated DSRF correctly represents the physical DSRF of the target.

4. CHOOSING λ

We propose an empirical method to find the best λ in (4) as a function of SNR. First, we build a collection of synthetic spectra with different model orders and a wide variety of distributions of relaxations. For each spectrum at a fixed SNR, the spectrum is estimated with different λ's, and then the λ that gives the smallest error (EMD) is recorded. This is done for a range of SNR and is repeated 100 times to obtain an average. The result shown in Fig. 5 is for the case $p = 0.5$. The EMD plays an important role in finding the best λ when there are many peaks in the spectrum because it combines all the spectrum deviations into one number. Some authors only count the number of perfect reconstructions when evaluating the goodness of fit, but imperfect estimates are often acceptable and some level of error always occurs, especially for ill-conditioned dictionaries.

Surprisingly, for each spectrum, the best λ has a simple relationship with the SNR: $\log\lambda$ is linear vs. SNR. We can, therefore, choose the best λ based on the SNR through a simple linear equation such as (5). Although targets of different model orders have different lines of best λ, we have observed that the minimum EMD is not highly sensitive to the exact choice of λ; changing by one or two decades gives an error

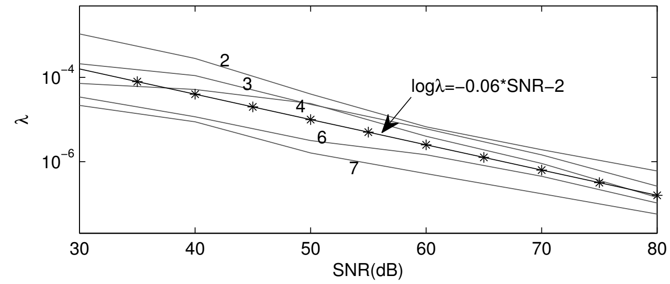


Fig. 5. Best λ vs. SNR for spectra of various model order. The line with markers is chosen to represent the best λ for all model orders.

very close to the minimum. As a result, there is some freedom in choosing the best λ in a practical application. For our problem setup, the λ is chosen by

$$\log\lambda = -0.06 \cdot \text{SNR} - 2 \tag{5}$$

The same empirical method can be repeated for other p's, and the result is also a linear relationship between $\log\lambda$ and SNR. With field data we use (5) along with an estimate of the SNR to determine λ for use in **Algorithm 1**.

5. REFERENCES

[1] M. Wei, W. R. Scott, Jr., and J. H. McClellan, "Robust estimation of the discrete spectrum of relaxations for electromagnetic induction responses," *IEEE Trans. Geosci. Remote Sens.*, to be published, Feb. 2010.

[2] C. D. Austin, E. Ertin, J. N. Ash, and R. L. Moses, "On the relation between sparse sampling and parametric estimation," in *Proc. DSP Workshop*, 2009.

[3] C. E. Baum, "On the singularity expansion method for the solution of electromagnetic interaction problems," Interaction Notes 88, Air Force Weapons Laboratory, 1971.

[4] Y. Das and J. E. McFee, "Limitations in identifying objects from their time-domain electromagnetic induction response," in *Proc. SPIE*, 2002.

[5] E. J. Candès, M. B. Wakin, and S. P. Boyd, "Enhancing sparsity by reweighted ℓ_1 minimization," *J. Fourier Anal. Appl.*, vol. 14, no. 5, pp. 877–905, 2008.

[6] R. Chartrand and W. Yin, "Iteratively reweighted algorithms for compressive sensing," in *ICASSP*, Las Vegas, NV, Mar. 2008, pp. 3869–3872.

[7] M. A. T. Figueiredo and R. D. Nowak, "A bound optimization approach to wavelet-based image deconvolution," in *ICIP*, Genoa, Italy, 2005, vol. 2, pp. 782–785.

[8] S. J. Kim, K. Koh, M. Lustig, and S. Boyd, "An efficient method for compressed sensing," in *ICIP*, San Antonio, TX, 2007, vol. 3, pp. 117–120.

[9] W. R. Scott, Jr., "Broadband array of electromagnetic induction sensors for detecting buried landmines," in *Proc. IGARSS*, Boston, MA, July 2008.

[10] Y. Rubner, C. Tomasi, and L. J. Guibas, "A metric for distributions with applications to image databases," in *Proc. ICCV*, Bombay, India, Jan. 1998, pp. 59–66.