

# ON THE RELATION BETWEEN SPARSE SAMPLING AND PARAMETRIC ESTIMATION

*Christian D. Austin, Emre Ertin, Joshua N. Ash, Randolph L. Moses*

The Ohio State University  
Department of Electrical and Computer Engineering  
2015 Neil Avenue, Columbus, OH 43210, USA  
Email: {austinc, ertine, ashj, randy}@ece.osu.edu

## ABSTRACT

We consider the relationship between parameter estimation of an additive model and sparse inversion of an under-determined matrix (dictionary) in a linear system. The dictionary is constructed by sampling parameters of the additive model. Parameters and model order are estimated using regularized least-squares inversion. We investigate equi-spaced and Fisher information inspired parameter sampling methods for dictionary construction, and present an example quantifying parameter estimation error performance for the different sampling methods. These results indicate that estimation performance is degraded by sampling the parameter space either too finely or too coarsely.

**Index Terms**— Parameter estimation, Model order estimation, Sparse reconstruction

## 1. INTRODUCTION

In this paper, we examine the relationship between parameter estimation of an additive model and sparse reconstruction. Parametric model estimation problems typically consist of two parts: model order selection and estimation of the value of each parameter in the model. In this context, parameters are estimated for a fixed model order and then a model order selection cost is evaluated using this model and its parameter estimates; the model with the lowest cost is the estimated model [1]. Parameters that span a continuous range can be estimated over that range, or alternatively, over discrete samples within that range. The latter method of estimation is typically combinatoric in the model order and may be computationally intractable. However, sparse reconstruction methods can be used to approximate the solution to some combinatoric problems [2].

In sparse reconstruction we wish to determine the input to an under-determined system under the assumption that the input vector is sparse (has many more zero components than non-zero components). The input vector can be estimated by  $\ell_p$ ,  $0 < p \leq 1$ , regularized least-squares inversion. It has been shown in the compressive sensing literature that if certain intercolumn correlation properties are satisfied by the system

matrix, then  $\ell_p$ ,  $0 < p \leq 1$ , regularized least-squares inversion error on the input vector can be bounded [2, 3].

In this work, we pose joint model order selection and parameter estimation of an additive component model as a sparse reconstruction problem. An additive model is the summation of a continuous parameter component function over different parameters. We sample the parameter space, and at each sample, evaluate the component function, forming an under-determined system (dictionary) matrix. The parametric model estimation problem consists of selecting a small number of columns from this dictionary.

In the next section of this paper, we formulate parametric model estimation of an additive model as a dictionary selection problem. We then discuss the use of sparse reconstruction to both select the model order and estimate parameters, while avoiding a combinatoric search, and we examine ways of sampling the parameter space. We conclude with an example demonstrating estimation performance for different parameter sampling methods.

## 2. PARAMETER SAMPLING AND ESTIMATION

An additive parametric model is defined here as a linear combination of functions of the form  $f(t, \theta) \in \mathbb{C}$ . The parameter  $t \in \mathbb{R}$  denotes the design parameter, an independent variable of the measurement process, for example, time, and  $\theta \in \mathbb{R}^n$  denotes the unobservable parameter, which is the parameter vector that we wish to estimate.

We sample the additive parametric model at  $t_i$ , for  $i = 1, \dots, N$ ; then measurements are modeled by

$$y_i = \sum_{m=1}^M x_m f(t_i, \theta_m) + \epsilon_i, \quad i = 1, \dots, N, \quad (1)$$

where  $M$  is a fixed unobservable model order;  $\theta_m$  and  $x_m$  are the true unobservable parameters and amplitudes, respectively for the  $m^{\text{th}}$  parametric function, ( $m = 1, \dots, M$ ), and where  $\epsilon_i$  is Gaussian white noise.

We wish to estimate the model order  $M$ , and parameters  $\{x_m, \theta_m\}_{m=1}^M$ . In general, this is a non-linear estimation problem in  $\theta_m$ . The approach we take is to sample  $f(t_i, \theta)$

on a fine grid  $\{\bar{\theta}_k\}_{k=1}^K$  of  $\theta$  and to select a subset of these samples, so that a linear combination of the subset fits the measurements well. We represent (1) as the linear system

$$y = Ax + \epsilon,$$

where

$$a(\bar{\theta}_k) = [f(t_1, \bar{\theta}_k), \dots, f(t_N, \bar{\theta}_k)]^T, \quad (2)$$

and

$$A = [a(\bar{\theta}_1), \dots, a(\bar{\theta}_K)], \quad (3)$$

is the system dictionary;  $x = [x_1, \dots, x_K]^T$  is the sparse amplitude signal;  $y = [y_1, \dots, y_N]^T$  is the vector of measurements, and  $\epsilon = [\epsilon_1, \dots, \epsilon_N]^T$  is the additive noise vector. It is assumed that the number of additive components in (1),  $M$ , is much smaller than the number of columns in the dictionary,  $K$ , ( $M \ll K$ ); hence, the selected subset should be sparse in the number of dictionary columns. We also assume that  $A$  is normalized to have unit length columns.

The sparse estimated amplitude signal,  $x$ , is used to estimate the model parameters. Define the ordered signal  $x_k$  as  $|x_{I_1}| \geq \dots \geq |x_{I_K}|$ . The model order estimate,  $\hat{M}$ , amplitude estimates,  $\hat{x}_k$ , and unobservable parameter estimates,  $\hat{\theta}_k$ , are then given by

$$\begin{aligned} \hat{x}_k &= \frac{x_{I_k}}{\|a(\bar{\theta}_{I_k})\|_2}, \\ \hat{\theta}_k &= \bar{\theta}_{I_k}, \\ \hat{M} &= \# \text{ non-zero } \hat{x}_k \quad k = 1, \dots, K, \end{aligned} \quad (4)$$

where the amplitude signal is scaled to adjust for dictionary normalization.

### 3. DICTIONARY SUBSET SELECTION

When estimating parameters by parameter sampling, accurate selection of dictionary columns is important. There are several methods of selecting a subset of dictionary elements. Minimum subset selection is a natural method of column selection that minimizes the cost function

$$J(x) = \left( \frac{1}{\sigma^2} \right) \|y - Ax\|_2^2 + \mu \|x\|_0, \quad (5)$$

where  $\sigma^2$  is noise variance;  $\|\cdot\|_0$  counts the number of non-zero entries in its argument, and  $\mu$  is a constant. For different values of  $\mu$ , minimum subset selection corresponds to different information criteria of model order selection; for example, if the noise is Gaussian, then  $\mu = 2$  is used in Akaike Information Criterion (AIC);  $\mu = \ln(N)$  is used in Bayesian Information Criterion (BIC), and  $\mu$  can be chosen to directly penalize non-zero entries, as in Generalized Information Criterion (GIC).

In general, minimization of (5) is combinatoric and intractable. The use of  $\|x\|_p = \left( \sum_{k=1}^K |x_k|^p \right)^{1/p}$  with  $0 <$

$p \leq 1$ , instead of  $\|x\|_0$ , to measure sparsity, results in a cost function that is tractable to minimize and also enforces solution sparsity. Using this modified cost function, the amplitude parameter vector can be estimated by  $\ell_p$  regularized least-squares inversion

$$\tilde{x}(\lambda) = \underset{x}{\operatorname{argmin}} \|y - Ax\|_2^2 + \lambda \|x\|_p^p, \quad (6)$$

where  $\lambda$  is a user determined sparsity parameter. When  $p = 1$ , this optimization problem is often called basis pursuit denoising [4, 5]. After solving (6), we threshold all components of  $\tilde{x}(\lambda)$  that are more than  $\tau$  dB down from the largest component of  $\tilde{x}(\lambda)$ :

$$x_i(\lambda) = \begin{cases} 0, & \text{if } 20 \log \left( \frac{\max_{k=1, \dots, K} \tilde{x}_k(\lambda)}{\tilde{x}_i(\lambda)} \right) > \tau, \\ \tilde{x}_i(\lambda), & \text{otherwise} \end{cases}, \quad (7)$$

where subscripts denote vector components. The amplitude signal vector,  $x(\lambda)$ , is used in (4) to estimate parameters.

The amplitude signal,  $x$ , and hence, the amplitude parameter estimate,  $\hat{x}$ , are functions of  $\lambda$ ; hence,  $\lambda$  selection is important to model order and parameter estimation performance. One method of choosing  $\lambda$  is cross-validation [6, 7]. Cross-validation methods may be computationally expensive or necessitate the collection of extra training data. We choose  $\lambda$  such that  $J(\hat{x}(\lambda))$  is minimized. This  $\lambda$  selection method can be viewed as minimizing (5) over a subset of amplitude parameters parameterized by  $\lambda$ . The hope is that the subset will include a solution close to the solution of the minimum subset problem.

The envelope of the cost function  $J(\hat{x}(\lambda))$  is convex, but the cost function itself has local minima as a result of the discontinuity of the  $\|\cdot\|_0$  term. We use a tree-like grid search to find an approximate minimum of  $J(\hat{x}(\lambda))$ . At the current stage of the search, the cost function is evaluated on a grid of points within a bounded interval of  $\lambda$  values. The two points with the smallest cost are retained and denoted as  $\lambda_1$  and  $\lambda_2$ . The interval of next search stage is centered at  $\frac{\lambda_1 + \lambda_2}{2}$  and has length  $|\lambda_2 - \lambda_1|$ , and the minimum from a grid of points on this interval is searched for. The search stops after a user-determined number of stages. The point with minimum cost at the final stage is chosen as  $\lambda$ .

### 4. DICTIONARY SAMPLING

When forming a dictionary from  $a(\theta)$ , the unobservable parameter  $\theta$  is sampled to form columns of the dictionary matrix, (3). The choice of parameter spacing affects parameter estimation performance. It is desirable to sample columns as closely as possible in  $\theta$  space to avoid high quantization error. However, for computational purposes, the number of columns in the dictionary matrix must be constrained; so,  $\theta$  cannot be quantized to an arbitrarily small number. Furthermore, as  $\theta$  sampling becomes finer, intercolumn correlation

increases, and when (6) is employed, amplitude parameter estimates become less sparse, spreading around the true value.

If the unobservable parameter  $\theta$  is constrained to a region in  $\mathbb{R}^n$ , this region could be divided into an equi-spaced grid, and dictionary columns could be sampled on this grid. However, when estimating parameters in noise, average parameter estimation error depends on both the component function  $f$  and the location of the unobserved parameter that is estimated.

We propose a grid sampling method based on Fisher Information of the component function. For scalar  $\theta$  (i.e.  $n = 1$ ), this sample spacing is of the form

$$\bar{\theta}_{k+1} = \bar{\theta}_k + \Delta_{\mathcal{I}}, \quad \Delta_{\mathcal{I}} = c \mathcal{I}(a(\bar{\theta}_k))^{-1/2}. \quad (8)$$

Fisher information is denoted as  $\mathcal{I}$ , and  $\mathcal{I}^{-1}$  is the Cramér-Rao lower bound for the variance of an unbiased estimator of  $\theta$ ;  $a(\theta)$  is defined in (2), and  $c$  is a user determined constant that determines the total size  $K$  of the sampled dictionary.

If we assume that  $\epsilon_i$  is complex Gaussian white noise,  $M = 1$  in (1), and  $a(\theta)$  is normalized, it is shown in the Appendix that  $c$  can be interpreted as a function of the local intercolumn correlation of the dictionary matrix. Under this interpretation, samples in (8) are spaced such that, locally, the intercolumn correlation is approximately constant across  $\theta$ . For a fixed number of  $\theta$  samples, this method distributes samples more finely in regions of higher information. For vector  $\theta$  (i.e.  $n > 1$ ), an analogous vector sampling approach to (8) can be used based on an eigendecomposition of the  $n \times n$  matrix,  $\mathcal{I}(a(\bar{\theta}_k))$ , which is also discussed in the Appendix.

## 5. EXAMPLE

In this section, we examine dictionary sampling and parameter estimation of a decaying exponential model, defined as

$$y_i = \sum_{m=1}^M x_m e^{-\alpha_m t_i} + \epsilon_i.$$

The component function is  $f(t_i, \alpha_m)$ ; variables  $\alpha_m > 0$  are the unobservable decay parameters;  $t_i$  are the design parameters in units of time, and  $\epsilon_i$  is additive Gaussian white noise. When the model order is 1, we wish to estimate the additive component function in noise (9), where  $y = [y_1, \dots, y_N]^T \in \mathbb{R}^N$ ,  $x_1 = x \in \mathbb{R}$ , and  $a(\alpha) = [e^{-\alpha t_1}, \dots, e^{-\alpha t_N}]^T$ .

### 5.1. Dictionary Sample Spacing

In this example, we investigate the performance of the Fisher information dictionary sampling method (8). This type of sampling requires calculation of the Fisher information of the additive component function in noise, given by

$$y = xa(\theta) + \epsilon, \quad (9)$$

where  $x$  is an amplitude scalar;  $y$  is the measurement vector, and  $\epsilon$  is Gaussian white noise vector with variance  $\sigma^2$ . The distribution of (9) is

$$p_{\theta}(y) \sim \mathcal{N}(xa(\theta), \sigma^2 I), \quad (10)$$

where  $I$  denotes the identity matrix. The Fisher information of (10) is well-known to be

$$\mathcal{I}(a(\theta)) = \frac{|x|^2}{\sigma^2} J_a(\theta)^H J_a(\theta), \quad (11)$$

where  $J_a(\theta)$  denotes the Jacobian of  $a(\theta)$ , and  $H$  denotes Hermitian transpose.

For the decaying exponential model,  $J_a(\alpha) = [-t_1 e^{-\alpha t_1}, \dots, -t_N e^{-\alpha t_N}]^T$ . Using (11), the Fisher information of the component function model is

$$\mathcal{I}(a(\alpha)) = \frac{|x|^2}{\sigma^2} \sum_{i=1}^N t_i^2 e^{-2\alpha t_i}.$$

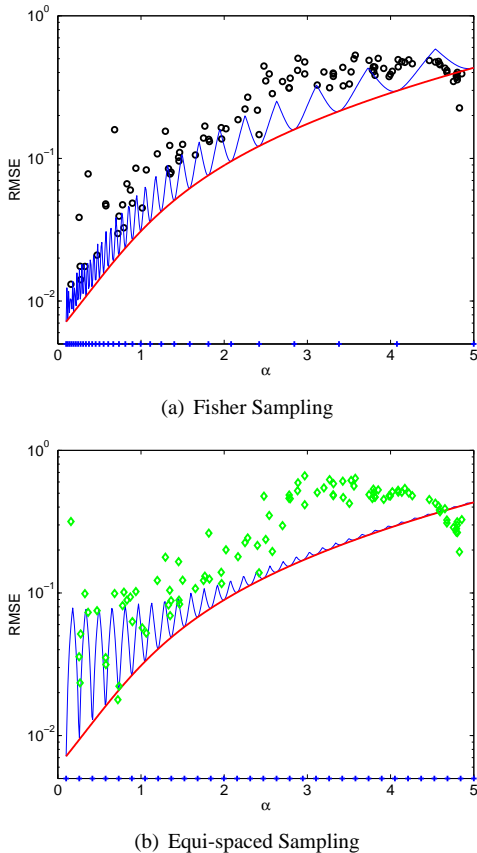
The Fisher information is a function of  $\alpha$ , and Fisher sampling will not be equi-spaced. We note that in contrast to the decaying exponential model, in a complex sinusoids in noise model, where we wish to estimate unobservable frequency parameters, Fisher information is not a function of frequency; so, Fisher sampling is equivalent to equi-spaced sampling.

### 5.2. Estimation Performance

We now quantify parameter sampling estimation performance using (4) and compare equi-spaced dictionary sampling to Fisher information sampling. In the following simulation  $t$  is sampled uniformly so that  $t_i = \Delta_t i$ , where  $\Delta_t = 0.1$ , and the number of time samples is  $N = 32$ . Noise variance is  $\sigma^2 = 10^{-3}$ ; the model order is  $M = 1$ , and for purposes of error analysis, we associate the largest magnitude estimated amplitude parameter with the true parameter. Dictionaries with 32 and 256  $\alpha$  samples distributed over a range of  $[0.1, 5]$  are analyzed. In the  $\lambda$  selection cost, (5),  $\mu = 5$ ; a value of  $p = 0.9$  is used in (6), and a threshold  $\tau = 40$  is used in (7) when estimating amplitude parameters. The optimization algorithm solving (6) is initialized at  $A^T A y$ .

When  $p = 0.9$ , (6) is not a convex optimization problem; however, the dictionaries that we consider have high intercolumn correlation, and using  $p < 1$  has been shown to be beneficial in reducing  $\ell_2$  estimation error in this case [3]. We note that the proposed estimation method also estimates model order; so, in addition to the largest magnitude estimated amplitude component, there may be other non-zero estimated amplitude components spread about the true parameter. Another error metric, such as the squared-difference between true and estimated amplitude parameters, could be used to capture this model order mismatch. For brevity, we do not investigate this error here; however, empirical results from this work suggest that lower amplitude squared-difference spreading error is achieved for  $p = 0.9$ , than for  $p = 1$ .

Figure 1 shows the empirical root mean squared error (RMSE) of the sampled parameter estimate,  $\hat{\alpha}$ , for Fisher and equi-spaced sampled dictionaries of size  $K = 32$  samples. RMSE is calculated as the square-root of the average squared difference between the true parameter and the estimated parameter with largest magnitude amplitude component. RMSE is calculated for a set of 100  $\alpha$  parameters drawn from a uniform distribution over  $[0.1, 5]$ . For each  $\alpha$ , 100 Monte-Carlo realizations of the measurement vector  $y$  are generated.

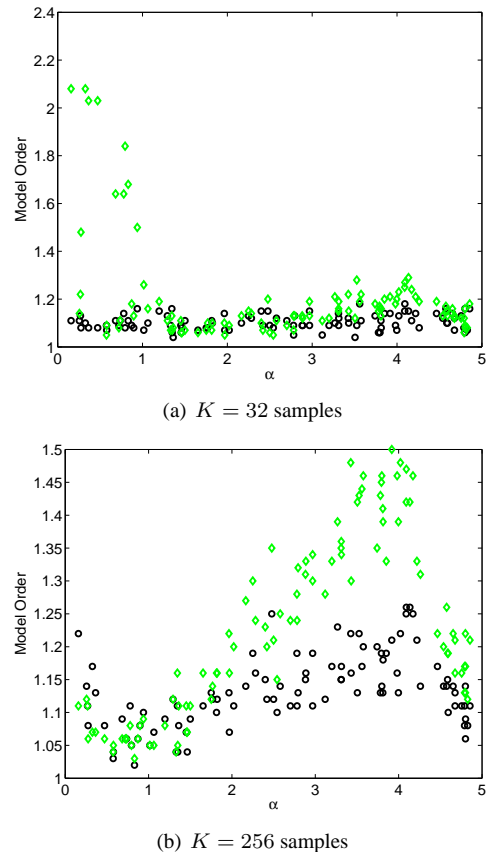


**Fig. 1.** RMSE( $\hat{\alpha}$ ) versus  $\alpha$  for 32 sample dictionaries. Black circles are from a Fisher information sampled dictionary, and green diamonds are from an equi-spaced sampled dictionary.

The solid red line is the square-root of the Cramér-Rao lower bound (CRB) for unbiased estimates of  $\alpha$ , and the solid blue line is  $\sqrt{\text{CRB} + \text{bias}^2}$ ; bias is defined as the distance between the true  $\alpha$  parameter and the closest  $\alpha$  sample in the dictionary. The RMSE approximated by the CRB appears to show how bias quantization affects RMSE across  $\alpha$ . For Fisher sampling, the bias results in a maximum excess relative error that is constant in a log scale across the range of parameter values; whereas, this is not the case with equi-spaced sampling. For both sampling cases, the RMSE of  $\hat{\alpha}$  appears to be mostly above the RMSE approximated by the CRB. Em-

pirical RMSE decreases around  $\alpha = 5$ ; this is a boundary effect, where the estimator is biased toward selecting the dictionary end point. Although not shown in this figure, we also note that as the number of dictionary samples increase, RMSE appears to remain mostly above the RMSE approximated by the CRB, and RMSE decreases in some regions of  $\alpha$ , while it increases in others. We explore the effect of sampling on error in more depth below.

Parameter sample spacing affects RMSE, which appears to be directly related to model order mismatch. Figures 2(a) and 2(b) show the mean estimated model order for each  $\alpha$  in the preceding error plots and for  $K = 256$  sample dictionaries, respectively. At small  $\alpha$  in Figure 1, most of the RMSE samples are greater for the equi-spaced dictionary than for the Fisher sampled dictionary. Points that have higher RMSE in this region also appear to have larger model order mismatch. In general, increases in mean model order appear to correspond with increases in error and vice-versa.



**Fig. 2.** Mean estimated model order versus  $\alpha$ ; the true model order is 1. Black circles are from a Fisher information sampled dictionary, and green diamonds are from an equi-spaced sampled dictionary.

We now demonstrate how RMSE is attributed, at least in part, to model order mismatch by the dictionary section pro-

cedure. Consider the solution to the dictionary selection problem as a two-stage model order selection and parameter estimation problem. First, a model order, which is the number of non-zero entries in the amplitude vector  $x$ , is chosen; if the model order is  $L$ , we call the amplitude vector  $L$ -sparse. An estimate of  $x$  is calculated as

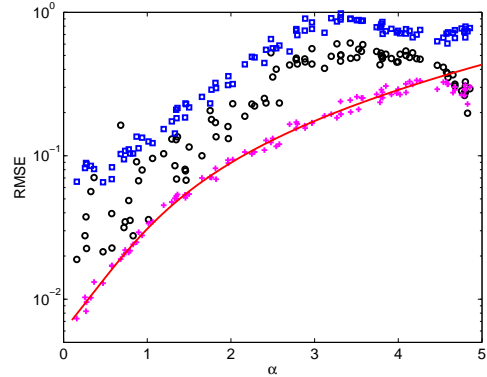
$$\hat{x}^L = \underset{x: L\text{-sparse}}{\operatorname{argmin}} \|Ax - y\|_2^2. \quad (12)$$

The solution of (12) involves brute-force estimation of all  $\binom{K}{L}$   $L$ -sparse  $x$ 's, which is tractable only for small  $L$ . Denote the estimates of  $\alpha$  from  $\hat{x}^1$ ,  $\hat{x}^2$ , and  $\ell_p$  estimates  $\hat{x}$  as  $\hat{\alpha}^1$ ,  $\hat{\alpha}^2$  and  $\hat{\alpha}$ . Figure 3 shows  $\hat{\alpha}^1$ ,  $\hat{\alpha}^2$  and  $\hat{\alpha}$  parameter RMSE for 256 dictionary samples. Again, the location of the largest magnitude estimated amplitude parameter is associated with the true  $\alpha$  parameter. On average, when the model order is overestimated, as in  $\hat{x}^2$ , error in the  $\alpha$  estimate is higher than error when model order is correctly determined, as in  $\hat{x}^1$ . In most cases, RMSE of  $\hat{\alpha}$  falls between the RMSE of  $\hat{\alpha}^1$  and  $\hat{\alpha}^2$  estimates. Figure 3 taken together with the mean model order in Figure 2(b), suggest that  $\hat{\alpha}$  RMSE can be approximated by a mixture of  $\alpha$  RMSE for  $\hat{x}^L$ , with  $L \geq 1$  and mixture weights given by the probability distribution of a  $L$ -sparse  $\hat{x}$  solution. For example, at small  $\alpha$  and for a dictionary size of 256 samples, Fisher information sampling results in mostly larger RMSE points than for equi-spaced sampling, since Fisher sampling has larger mean model order.

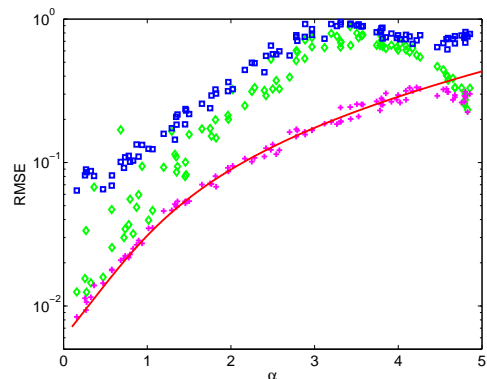
The preceding discussion shows that non equi-spaced sampling on  $\alpha$  can be beneficial. However, model order and parameter estimation performance can suffer if sampling is either “too coarse” or “too fine”. The effect of sample spacing on RMSE can be seen by comparing  $K = 32$  and  $K = 256$  sample dictionaries in Figures 1 and 3. Within some regions of  $\alpha$ , RMSE is improved by increasing the number of dictionary samples; whereas, in other regions, performance is degraded by increasing the number of elements. Figure 4 shows RMSE versus  $\alpha$  sample spacing in an equi-spaced dictionary,  $\Delta\alpha$ , for a model with true  $\alpha = 0.257$ . The black line is the RMSE approximated by the CRB, as defined in Figure 1, and it shows the effect that bias has on RMSE. This figure demonstrates that minimum RMSE is attained between sampling extremes; furthermore, minimum RMSE for different  $\alpha$  may occur for different sample spacings. Ideally, one could estimate the minimum RMSE sampling function from the minima of curves such as Figure 4 for different  $\alpha$ , and a dictionary could be sampled according to this function.

## 6. CONCLUSION

In this paper we have investigated the connection between parameter estimation of additive models and sparse reconstruction methods. The connection is made by sampling the continuous parameter space to obtain an under-determined matrix (dictionary) of a linear system. One cannot, in general,



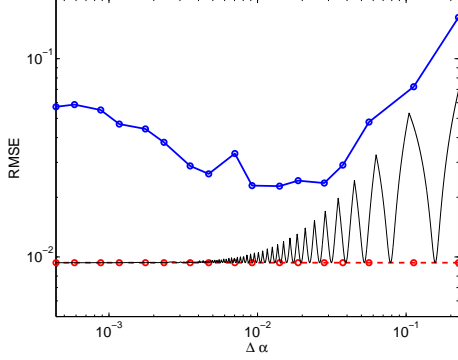
(a) Fisher Sampling



(b) Equi-spaced Sampling

**Fig. 3.** RMSE versus  $\alpha$  for 256 sample dictionaries. Black circles are from  $\hat{\alpha}$  estimates and a Fisher information sampled dictionary, and green diamonds are from  $\hat{\alpha}$  estimates and an equi-spaced sampled dictionary. Magenta pluses are from  $\hat{\alpha}^1$  estimates, and blue squares are from  $\hat{\alpha}^2$  estimates. The red line is the square-root of the Cramér-Rao lower bound for unbiased estimates of  $\alpha$ .

obtain orthogonal or nearly orthogonal dictionary elements, and we proposed a parametric model estimation method that estimates model order and parameters by selecting a subset of columns from the system dictionary. We utilize  $\ell_p$ ,  $0 < p \leq 1$  regularized least-squares inversion to select a subset of parameters and avoid a combinatoric search. Two different methods of sampling the parameter space are explored: equi-spaced sampling, and Fisher information sampling. Parameter estimation performance of the decay parameter in a decaying exponential model demonstrated that neither equi-spaced sampling, nor Fisher information sampling offer superior performance over all parameter values. Parameter sampling that is very fine or coarse results in model order overestimation error, and increased RMSE. Error resulting from coarse sampling is attributed, at least in part, to quantization bias, while we conjecture that that error from fine sampling is an effect of high intercolumn correlation. Simulation results suggest



**Fig. 4.** RMSE versus  $\Delta\alpha$  for  $\alpha = 0.257$ , shown as blue lines. Red broken lines are the square-roots of the Cramér-Rao lower bound for unbiased estimates of  $\alpha$ .

that there is sample spacing that achieves minimum RMSE for each parameter value, and hence, a minimum RMSE parameter spacing function. It may be possible to improve parameter estimation performance by sampling parameter space using this minimum RMSE parameter spacing function.

## 7. APPENDIX

In this section, we show the relation between dictionary intercolumn correlation and Fisher information. For simplicity, we consider dictionaries formed from real component functions  $f$  and amplitude parameters in (1). A discussion for the complex case is similar.

Define dictionary columns  $a(\theta)$  as in (2), and assume they have unit length. The Taylor series linear approximation of  $a(\theta)$  evaluated at  $\theta = \theta_k$  is

$$a(\theta_k) + J_a(\theta_k)(\theta - \theta_k),$$

where  $J_a(\theta)$  is the Jacobian of  $a$ , and the approximation holds for  $\|\theta - \theta_k\|$  small. Normalizing this Taylor series approximation gives the approximation

$$\begin{aligned} a(\theta) &\approx \frac{a(\theta_k) + J_a(\theta_k)(\theta - \theta_k)}{\|a(\theta_k) + J_a(\theta_k)(\theta - \theta_k)\|_2} \\ &= \frac{a(\theta_k) + J_a(\theta_k)(\theta - \theta_k)}{\sqrt{1 + (\theta - \theta_k)^T J_a(\theta_k)^T J_a(\theta_k)(\theta - \theta_k)}}. \end{aligned} \quad (13)$$

The last equality uses the fact that  $\|a(\theta)\|_2 = 1, \forall \theta$ , and hence  $a(\theta)^T J_a(\theta) = (\nabla \|a(\theta)\|_2)^T = 0$ , so  $\langle a(\theta_k), J_a(\theta_k)(\theta - \theta_k) \rangle_2 = 0$ . From (13), it follows that correlation can be approximated locally as

$$\begin{aligned} \langle a(\theta_k), a(\theta_k + \Delta\theta) \rangle_2 &\approx \frac{\|a(\theta_k)\|_2 + \langle a(\theta_k), J_a(\theta_k)\Delta\theta \rangle_2}{\sqrt{1 + \Delta\theta^T J_a(\theta_k)^T J_a(\theta_k)\Delta\theta}} \\ &= \frac{1}{\sqrt{1 + \Delta\theta^T J_a(\theta_k)^T J_a(\theta_k)\Delta\theta}}. \end{aligned} \quad (14)$$

We wish to minimize  $\theta$  sample spacing,  $\|\Delta\theta\|_2$  such that intercolumn correlation satisfies  $\langle a(\theta_k), a(\theta_{k+1}) \rangle_2 = \rho$  locally, where  $0 \leq \rho \leq 1$ . From (14), increasing the quadratic term in the denominator decreases correlation; thus, the sampling criterion is

$$\theta_{j+1} = \theta_k + \Delta\theta, \quad \Delta\theta = \sqrt{\frac{1}{\rho^2} - 1} \frac{v_{\max}}{\sqrt{\lambda_{\max}}}$$

where  $v_{\max}$  and  $\lambda_{\max}$  are the unit norm maximum eigenvector and eigenvalue of  $J_a(\theta_k)^T J_a(\theta_k)$ , respectively. As discussed in Section 5, the Fisher information of the component function  $f$  in complex Gaussian white noise is given by (11); so,  $J_a(\theta_k)^T J_a(\theta_k)$  is a scaled version of this Fisher information matrix evaluated at  $\theta_k$ . When  $\theta$  is one-dimensional, the maximum eigenvalue is a scaled version of the square-root of Fisher information, and step size is inversely proportional to the square-root of the Fisher information.

## 8. REFERENCES

- [1] P. Stoica and R. Moses, *Spectral Analysis of Signals*. New Jersey: Pearson Prentice Hall, 2005.
- [2] E. Candés, J. Romberg, and T. Tao, “Stable signal recovery from incomplete and inaccurate measurements,” *Comm. Pure Appl. Math.*, vol. 59, no. 8, pp. 1207–1223, 2006.
- [3] R. Saab, R. Chartrand, and O. Yilmaz, “Stable sparse approximations via nonconvex optimization,” in *IEEE Int. Conf. on Acoustics, Speech, and Signal Processing (ICASSP)*, Las Vegas, Nevada, April 2008.
- [4] S. Chen, D. Donoho, and M. Saunders, “Atomic decomposition by basis pursuit,” *SIAM Journal on Scientific Computing*, vol. 20, no. 1, pp. 33–61, 1998.
- [5] M. Figueiredo, R. Nowak, and S. Wright, “Gradient projection for sparse reconstruction: Application to compressed sensing and other inverse problems,” *IEEE Journal of Selected Topics in Signal Processing*, vol. 1, no. 4, pp. 586–597, Dec. 2007.
- [6] Ö. Batu and M. Çetin, “Hyper-parameter selection in non-quadratic regularization-based radar image formation,” in *Algorithms for Synthetic Aperture Radar Imagery XV*. Orlando, FL.: SPIE Defense and Security Symposium, March 17–20 2008.
- [7] P. Boufounos, M. F. Duarte, and R. G. Baraniuk, “Sparse signal reconstruction from noisy compressive measurements using cross validation,” in *IEEE 14<sup>th</sup> Workshop on Statistical Signal Processing 2007, (SSP '07)*, August 2007, pp. 299–303.