PARAMETER ESTIMATION USING SPARSE RECONSTRUCTION WITH DYNAMIC DICTIONARIES

Christian D. Austin, 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, ashj, randy}@ece.osu.edu

ABSTRACT
We consider the problem of parameter estimation for signals characterized by sums of parameterized functions. We present a dynamic dictionary subset selection approach to parameter estimation where we iteratively select a small number of dictionary elements and then alter the parameters of these dictionary elements to achieve better signal model fit. The proposed approach avoids the use of highly oversampled (and highly correlated) dictionary elements, which are needed in fixed dictionary approaches to reduce parameter bias associated with dictionary quantization. We demonstrate estimation performance on a sinusoidal signal estimation example.

Index Terms—Sparse reconstruction, Parameter estimation, Compressive Sensing, Dictionary selection, Model order selection

1. INTRODUCTION
We consider the problem of parameter estimation from noisy measurements of signals, where the signal component of the measurement is a sum of parametric signal components (such as a sum of sinusoids). Traditional parameter estimation approaches to this problem typically entail two steps: model order (MO) selection and parameter estimation [1]. Model order may be selected using metrics, such as information criteria metrics, that trade off goodness of signal fit with the number of components in the model. Parameter estimates are typically obtained using methods such as least-squares (LS), maximum-likelihood (ML), or maximum-a-posteriori (MAP) estimation [1].

We present an alternative approach, based on dictionary subset selection and refinement, to the joint model order and parameter estimation problem. In this approach, the continuous parameter space is sampled, and each sample generates a dictionary entry vector; the measured signal is then modeled as weighted sum of a small subset of these dictionary elements. The approach applies to problems described by an additive component model in which the signal is described as linear combinations of a common parametric function evaluated at different parameter values, one for each component.

An example is the ubiquitous sinusoids-in-noise model.

One limitation of the fixed dictionary approach is that parameter estimates are quantized to the level of the parameter sampling and may therefore exhibit quantization error [2]. One way to reduce this parameter quantization error is to sample the parameter space more finely—at the cost of increasing the size of the dictionary and resulting in increased computational complexity. In addition, finely-sampled dictionaries have high intercolumn correlation, so the dictionaries almost always violate the Restricted Isometry Property (RIP) needed for provable convergence to ‘good’ solutions in most sparse reconstruction methods [3, 4]. In this paper, we propose an estimation method based on dynamic dictionaries. The approach is similar to the fixed dictionary case but includes iterative refinement of the dictionary elements as the algorithm progresses. Since the dictionary elements are adapted to the data, low quantization error is obtained while retaining small dictionary sizes, avoiding the problems with a large, finely-sampled fixed dictionary. The proposed method is similar to dictionary learning in sparse reconstruction problems [5, 6]. In dictionary learning methods, the dictionary is learned before sparse reconstruction by optimizing a cost function with respect to training data, whereas in the dynamic dictionary method considered here, the dictionary and multiplicative coefficients are estimated jointly using the measured data.

2. ADDITIVE COMPONENT MODEL
The additive component model

\[ y(t_i) = \sum_{m=1}^{M} x_m f(t_i, \theta_m) + \epsilon_i, \quad i = 1, \ldots, N \]  

is defined as a linear combination of component functions \( f(t, \theta) \in \mathbb{C} \) along with additive noise \( \epsilon_i \). The form of \( f(t, \theta) \) is assumed to be known, but the parameters \( \theta_m \in \mathbb{R}^n \), and the amplitude \( x_m \in \mathbb{C} \), for each component \( m \) are unknown. The fixed, but unknown, model order is \( M \). The problem of interest is to estimate \( M \) and the component parameters \( \{\theta_m, x_m\}_{i=1}^{M} \), from a set of measurements \( \{y(t_i)\}_{i=1}^{N} \).
The model (1) can be represented by the linear system
\[ y = A(\bar{\theta})x + \epsilon, \]
where \( y = [y_1, \ldots, y_N]^T \) is the vector of measurements; \( x = [x_1, \ldots, x_K]^T \) is the sparse amplitude signal; \( \epsilon = [\epsilon_1, \ldots, \epsilon_N]^T \) is the additive noise vector,
\[ A(\bar{\theta}) = [a(\bar{\theta}_1), \ldots, a(\bar{\theta}_K)], \quad N \times K \]
and
\[ a(\bar{\theta}_k) = [f(t_1, \bar{\theta}_k), \ldots, f(t_N, \bar{\theta}_k)]^T. \]
The vector \( \bar{\theta}_k \) denotes the \( k \)th sample drawn from the parameter space and \( a(\bar{\theta}_k) \) is the associated \( k \)th dictionary element. A collection of \( K \) samples \( \theta = \{\theta_1 \ldots \theta_K\} \) establishes the entire parameterized dictionary \( A(\bar{\theta}) \). It is assumed in constructing \( \bar{\theta} \) that the candidate parameter space is bounded and that \( K > M \), so that the amplitude vector \( x \) is sparse in the parameterized dictionary. Often \( \bar{\theta} \), and hence, \( A(\bar{\theta}) \) are fixed \( a-priori \) in sparse reconstruction problems, but in this paper, we consider dynamic, data-driven selection of \( \bar{\theta} \).

3. MODEL ESTIMATION

Parameters and model order of the additive component model are estimated by solving the regularized LS cost function
\[ \begin{align*}
\hat{x}, \hat{\theta} & = \underset{x, \hat{\theta}}{\text{argmin}} \| y - A(\bar{\theta})x \|^2_2 + \lambda \|x\|^p_\nu + \mu g(\bar{\theta}),
\end{align*} \]
where \( \| \cdot \|_p \) is the \( \ell_p \) quasi-norm and \( 0 < p \leq 1 \). The function \( g(\cdot) \) is a penalty for particular parameter samplings \( \bar{\theta} \). For example, \( g(\cdot) \) might penalize for \( \bar{\theta} \) that result in highly correlated \( A(\bar{\theta}) \) columns. The optimization (5) contains two user-selected settings: \( \lambda \), which controls the sparsity in dictionary element selection, and \( \mu \), which controls the weight of the penalty function. Once (5) is minimized, model order selection and parameter estimates for the model in (1) are obtained by thresholding \( \hat{x} \) with \( \eta \) [4]
\[ \hat{x} = \begin{cases} 
\hat{x} & |\hat{x}| > \eta \\
0 & \text{otherwise}
\end{cases}. \]

For the model in (1), the model order estimate, \( \hat{M} \), is the number of nonzero elements of \( \hat{x} \), and the components of the parameter estimate, \( \hat{\theta} \), are the values of \( \hat{\theta} \) at non-zero locations of \( \hat{x} \). The amplitude estimate, \( \hat{x} \), is \( A^\dagger(\bar{\theta})y \), where \( A^\dagger(\bar{\theta}) \) is the pseudoinverse of the \( N \times \hat{M} \) matrix \( A(\bar{\theta}) \) defined as in (3). In general, (5) is a nonconvex optimization problem, even for \( p = 1 \), because of the nonlinear dependence of \( \bar{\theta} \) in \( A(\bar{\theta}) \).

When implementing an algorithm to solve (5), the number of columns \( K \) in \( A(\bar{\theta}) \) must first be determined. Whereas in a fixed dictionary problem, \( K \) is chosen based on parameter quantization considerations, here \( K \) is chosen based on the maximum hypothesized model order as well as region-of-attraction considerations to avoid convergence of (5) to local minima. For many problems, \( K \) will be much smaller for the dynamic dictionary than for a fixed dictionary.

To solve (5), we use a block coordinate descent approach. The first block of the algorithm optimizes over \( x \) while keeping \( \bar{\theta} \) fixed
\[ \hat{x}_{n+1} = \underset{x}{\text{argmin}} \| y - A(\bar{\theta}_n)x \|^2_2 + \lambda \|x\|^p_\nu. \]
This block is the standard basis pursuit problem and is solved using the majorization-minimization algorithm in [7]. The second block of the algorithm optimizes over \( \bar{\theta} \) while fixing \( x \)
\[ \hat{\theta}_{n+1} = \underset{\bar{\theta}}{\text{argmin}} \| y - A(\bar{\theta})\hat{x}_{n+1}\|^2_2 + \mu g(\bar{\theta}). \]
This optimization is implemented using a trust-region subspace algorithm [8]. In both cases, other algorithms could be used.

The cost functions (5) or (8) are non-convex optimization problems, and initial parameter settings will affect estimation results. One possible choice for initial parameters, \( \theta_0 \), is to sample the parameters uniformly over a bounded region of interest. Alternatively, non-uniform sampling strategies based on Fisher information have also been proposed [2].

4. SIMULATION EXAMPLES

As an initial illustration of dynamic dictionary algorithm properties, we consider the problem of estimating complex sinusoids in additive white Gaussian noise with variance \( \sigma^2 \). We compare performance with two fixed dictionary approaches and the Cramèr-Rao lower bound (CRB); comparison of fixed dictionary approaches with traditional parameter estimation methods are presented in [4] and due to space limitations are not repeated here. Algorithm settings, \( \lambda \) and \( \mu \), are set manually, although automatic methods for selecting \( \lambda \) have been proposed [4, 9].

We first examine the estimation of \( M = 2 \) well-separated sinusoids in noise. The sinusoids-in-noise model is defined by (1) with component function \( f(t, \gamma_m) = e^{j2\pi \gamma_m t} \), where \( \gamma_m \) is the \( m \)th frequency in Hertz and \( t \) is the time in seconds. Each sinusoid has a magnitude \( |x_m| = 1 \), and the frequencies are 3.13 and 7.19 Hz. One frequency aligns with the coarse, fixed dictionary sample, while the other lies halfway between two dictionary samples. An SNR, defined as \( 10 \log(|x_m|^2/\sigma^2) \), of 10 dB is used here.

We use \( N = 16 \) time measurements, and \( K = 16 \) initial frequency samples for estimation using the dynamic dictionary approach. The function \( g(\gamma) = \sum_{k=1}^{K-1} \frac{1}{(\gamma_k - \gamma_{k+1})^2} \) is used in (5) to penalize frequency estimates that grow close together and increase dictionary correlation. We compare these results with performance from a fixed dictionary approach using a coarsely-sampled dictionary with 16 elements, and with a finely-sampled dictionary of 4096 elements. The fixed dictionary method is implemented via the basis pursuit algorithm as in [4]. Dictionary elements in both methods are uniformly
Fig. 1. Estimation of two well-separated sinusoids in noise. Left column: proposed dynamic dictionary ($\lambda = 2$, $\mu = 2000$). Middle column: coarse fixed dictionary ($\lambda = 5$). Right column: finely sampled fixed dictionary ($\lambda = 5$). Top row: estimated model order histograms. Second row: a representative sample realization. Bottom row: overlayed ensemble of 200 Monte-Carlo estimates when selected model order is 2; the bottom-right figure is empty because the finely sampled fixed dictionary estimator never selects the correct model order. Red vertical lines show the locations of true sinusoid frequencies which have magnitude 1, shown by a horizontal red dotted line. The other solid vertical lines indicate estimated frequencies and amplitudes.

spaced in frequency, and the initial dynamic dictionary sampling is the same as the coarse, fixed dictionary. All algorithms use $p = 1$, and $\eta$ 35 dB below $\|\hat{x}\|_\infty$; the remaining algorithm settings are chosen manually to provide good results. Performance results are based on 200 Monte-Carlo realizations for each estimator.

Figure 1 shows estimation results. For the dynamic dictionary approach (left column), the correct model order of 2 is obtained in all 200 Monte-Carlo estimates, and the resulting parameter estimates cluster tightly around the true parameters, as shown in the ensemble plot. The realization plot demonstrates that the algorithm converges to an estimate in which all but two magnitudes are zero, even though 16 dictionary elements are present; furthermore, dictionary elements adapt their frequencies so that two of them (nearly) align to the true frequencies.

Estimates for both a coarse and fine fixed dictionary are shown in Fig. 1. For the coarse dictionary (middle column), the model order is overestimated as 3 in most cases, due to selection of two nonzero dictionary elements that surround the second frequency; this is illustrated in the single realization plot Fig. 1(e). For the finely sampled dictionary (right column), model order is significantly overestimated, and both frequencies are modeled by numerous dictionary elements that surround the true frequencies (Fig. 1(f)). This is a result of many highly correlated columns being selected by the al-
algorithm. Sparser solutions are possible by using \( p < 1 \) \cite{4}, but comparable results to the dynamic dictionary approach are not possible using \( p = 1 \).

Table 1 shows frequency estimation performance for the dynamic and coarse fixed dictionary methods, conditioned on the estimators determining the correct model order of 2. For the dynamic dictionary, both frequencies are estimated with RMSE values close to the CRB (specifically, about 5% and 12% higher). This is competitive with traditional parameter estimation techniques, such as ML or ESPRIT; see also \cite{4}. In contrast, the fixed dictionary method gives an RMSE value of 0 for the sinusoid that is exactly aligned with a dictionary element, but a much higher RMSE of 0.3125 Hz for the second frequency. The second RMSE is due entirely to parameter quantization error, as one-half the dictionary spacing is 0.3125 Hz in this example.

Average model order and RMSE performance as a function of SNR for the dynamic and coarse fixed dictionary methods are shown in Fig. 2. As SNR increases, the expected model order approaches 2 for the dynamic dictionary and 3 in the fixed dictionary. Dynamic dictionary frequency estimation approaches the CRB at around 0 dB; whereas the fixed dictionary method approaches the quantization error solution with RMSE of 0 and 0.3125 Hz.

![Average model order and RMSE versus SNR with \( K = 16 \). Solid lines: dynamic dictionary method; dash-dotted lines: fixed dictionary method. RMSE curves are conditioned on estimated model order of 2. RMSE scale is in dB. Green and black colors correspond to sinusoid estimates in Fig. 1 (g) and (h).](image)

5. CONCLUSION

We have presented a dictionary subset selection method for model order and parameter estimation in additive component models. This method uses a data-adaptive dictionary to address problems of RIP violation and large computation times that result from finely-sampled and highly correlated dictionaries. By allowing the dictionary columns to dynamically change, parameter estimation quantization error is reduced and dictionary size is decreased as compared to fixed dictionary methods. Initial results on sinusoidal frequency estimation show effective model order selection and parameter estimation mean-squared error that is close to the Cramér-Rao bound. In contrast, fixed dictionary estimators suffer from either high quantization error, when dictionary spacing is coarse, or model order overestimation when dictionary spacing is fine.

Future work will compare the dynamic dictionary method presented in this paper with the fixed, finely sampled dictionary method using an \( \ell_p \) quasi-norm penalty, with \( p < 1 \). In addition, other methods for solving the non-convex optimization dynamic dictionary cost function will be explored.

6. REFERENCES