Markov Chain Monte Carlo Inference of Parametric Dictionaries for Sparse Bayesian Approximations

Theodora Chaspari, Andreas Tsiartas, Panagiotis Tsilifis, and Shrikanth S. Narayanan

University of Southern California, Los Angeles, CA

Sparse representations model a signal as a combination of small number of components from a large overcomplete set of exemplar atoms, called dictionary. Let $\mathbf{D} \in \Re^{P \times K}$ be a dictionary with K atoms $\mathbf{d_1}, \ldots, \mathbf{d_K} \in \Re^P$ and $\mathbf{x} \in \Re^P$ be a P-dimensional signal, such that $\mathbf{x} = \mathbf{Dc} + \boldsymbol{\epsilon}$, where $\mathbf{c} \in \Re^K$ are the corresponding atom coefficients. Dictionary learning (DL) is a widely used technique towards compact and reliable representations focusing on learning atoms from the available training data. It includes several well-known algorithms, such as the K-SVD (Aharon *et al.*, 2006) and the method of optimal directions (Engan *et al.*, 1999).

Parametric dictionaries contain atoms expressed as a function of a parameter set, say $\mathbf{d_k} = \phi(\boldsymbol{\theta_k})$, $\boldsymbol{\theta_k} \in \Re^Q$, $Q < P$, where the atom parameters $\boldsymbol{\theta_k}$ are optimized with respect to criteria involving desirable properties. Parametric DL is more likely to converge faster, have more efficient implementations compared to the non-parametric problem, and further provide higher interpretability yielding important meta-information about the signal structure. Previous studies on parametric DL have attempted fitting the dictionary parameters with least-squares based techniques (Ataee *et al.*, 2010), or in the light of designing dictionaries with minimum coherence (Yanghoobi *et al.*, 2009).

We propose a Bayesian framework for learning the parameters of dictionary atoms because of its ability to perform full estimates of the problem variables and generalize well on unseen data. Although Bayesian approaches have been proposed in non-parametric DL (Olshausen and Field, 1996), to the best of our knowledge this is the first study examining this framework on parametric dictionaries.

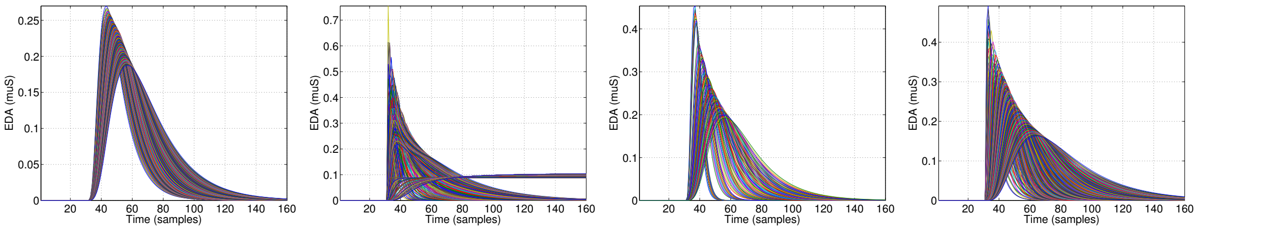
Sparseness is ensured through the use of indicator variables that select the atoms of the dictionary for representing each signal, such that:

$$\mathbf{x} = \mathbf{D}\left(\mathbf{s} \circ \mathbf{z}\right) + \boldsymbol{\epsilon} \tag{1}$$

where $\mathbf{z} \in \Re^K$, $\mathbf{s} \in \Re^K$, $||\mathbf{s}||_0 = ||\mathbf{z}||_0 = L \ll K$, and "$\circ$" represents the Hadamard or entrywise product. We assume that $\mathbf{z}$ follows the Wallenius' non-central hypergeometric distribution, i.e. $\mathbf{z} \sim \text{Wallenius}\left(\mathbf{1_K}, L, \boldsymbol{\pi}\right)$, that models the sampling of L dictionary atoms out of the possible K without replacement, where $\boldsymbol{\pi} \in \Re^K$ are the weights for each atom.

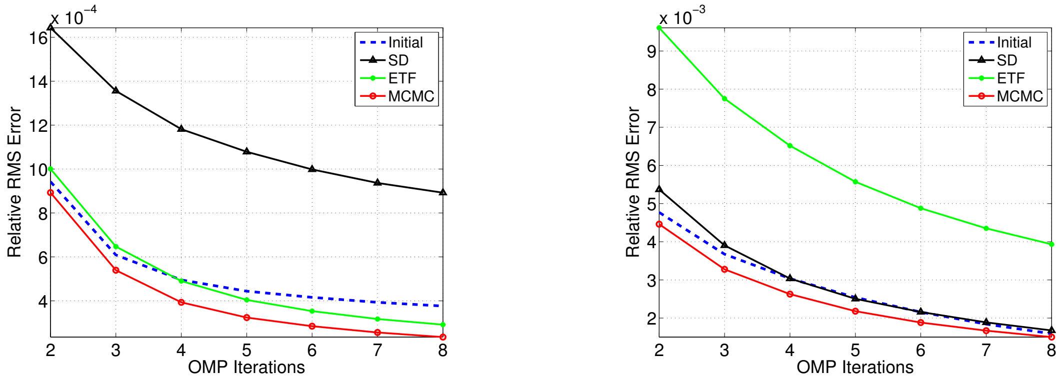
We further assume independent atom coefficients following the normal distribution with mean $\boldsymbol{\mu_s} \in \Re^K$ and precision γ_s, i.e. $\mathbf{s} \sim \text{Normal}\left(\boldsymbol{\mu_s}, \gamma_s^{-1}\mathbf{I_L}\right)$. We also hypothesize dictionary parameters of normal distribution $\boldsymbol{\theta_k} \sim \text{Normal}\left(\mathbf{g_k}, \mathbf{G}^{-1}\right)$ with mean $\mathbf{g_k} \in \Re^Q$ and precision $\mathbf{G_n} \in \Re^{Q \times Q}$. Finally, we assume zero mean Gaussian noise with variance γ_ϵ^{-1}, i.e. $\boldsymbol{\epsilon} \sim \text{Normal}\left(\mathbf{0}, \gamma_\epsilon^{-1}\mathbf{I_P}\right)$. Appropriately selected conjugate priors are imposed to the parameters of the aforementioned distributions in order to further handle uncertainty.

Inference is performed with a Markov Chain Monte Carlo approach, that uses block sampling to generate the variables of the Bayesian problem. Since the parameterization of dictionary atoms results in posteriors that cannot be analytically computed, we use a



(a) Initial (b) SD (c) ETF (d) MCMC

Figure 1: Example of initial dictionary and dictionaries learnt Steepest Descent (SD), Equiangular Tight Frame (ETF), and Markov Chain Monte Carlo Bayesian inference (MCMC). An instance of phasic atoms shifted with $t_0 = 30$ is shown.



(a) Synthetic data (b) Real data

Figure 2: Relative root mean square (RMS) error between original and reconstructed signal with respect to the number of orthogonal matching pursuit OMP iterations. Dictionaries are learnt with Steepest Descent (SD), Equiangular Tight Frame (ETF), and Markov Chain Monte Carlo Bayesian inference (MCMC).

Metropolis-Hastings-within-Gibbs framework, according to which variables with closed-form posteriors are generated with the Gibbs sampler, while the remaining ones with the Metropolis Hastings from appropriate candidate-generating densities.

We demonstrate the ability of our algorithm to represent synthetic data and real biomedical signals from the publicly available database of emotion analysis using physiological signals (DEAP) (Koelstra *et al.*, 2012). Our proposed approach is compared to previous Steepest Descent (SD) (Ataee *et al.*, 2010) and Equiangular Tight Frame (ETF) (Yanghoobi *et al.*, 2009) methods. Visual inspection of the final dictionaries indicates that SD does not always preserve the initial dictionary structure (Figs. 1a-b), while ETF yields less variable dictionaries (Fig. 1c). Our Bayesian approach results in a variety of dictionary atoms preserving the original structure (Fig. 1d). Signal reconstruction is further improved with our proposed Bayesian DL method (Fig. 2). Dictionaries trained using SD perform quite poorly on unseen synthetic data (Fig. 2a). This might occur because the simple structure of synthetic data causes significant overfitting to least-squares-based methods. Despite the fact that ETF DL is not prone to overfitting, since it does not take into account exemplar data during training, it lacks adaptation to more complex real data (Fig. 2b).

The authors would like to thank Prof. Urbashi Mitra and Mr. Sajjad Beygi at USC for their valuable feedback during the course of this research.