EXPLICIT VERSUS IMPLICIT SOURCE ESTIMATION FOR BLIND MULTIPLE INPUT SINGLE OUTPUT SYSTEM IDENTIFICATION

Austin J. Brockmeier and Jose C. Principe

University of Florida
Electrical and Computer Engineering
Gainesville, Florida

ABSTRACT

Sparsely-activated time series are found in many physical systems. In these cases, the signals can be approximated by convolution of sparse sources with a set of shift-invariant filters. When there is access to only one sensor, such that there is a single observation signal, identifying the source signals appears to be an ill-posed problem, but for very sparse sources it is still possible to learn the system. We discuss analysis techniques for sparsely activated signals, which retrieve sparse sources given the filters, and identify conditions when algorithms based on independent component analysis (ICA) and sparse coding can blindly estimate filters from a single noisy time-series. Many qualitative results have been made for learning shift-invariant bases on natural signals, but for a thorough understanding of the effect of sparsity, we quantitatively analyze results on synthetic examples, comparing how ICA and shift-invariant sparse coding approaches perform for multiple-source blind system identification.

Index Terms— blind channel estimation, dictionary learning, ICA, sparse coding

1. INTRODUCTION

Sparsely-activated signals are evident in a number of natural signals: neural spike trains [1, 2, 3], electrocardiograms [4], and seismic recordings. Signals such as these have two key distinctions: they consist of the same waveforms appearing repeatedly—but only occasionally—throughout the signal. These time-series signals can be modeled as the sparse excitation of a single filter. A signal of this form is known as shot noise. Mathematically, it corresponds to the convolution of a train of Dirac delta functions, with possibly varying amplitude, with a filter representing the system's impulse response [5].

Even more real-world signals are approximated by a combination of shot-noise processes—for example, recordings from an electrode in the brain near multiple neurons. This summation of single-source signals forms a multiple-input single-output (MISO) linear system. In this system, each component has a unique filter, and each source is assumed to be independently and sparsely excited. This model can explain signals such as sound or electromagnetic waves that are emitted in a passive physical environment with a linear medium.

The shot-noise model can be seen as a form of time-series analysis where the energy is not only temporally localized and sparsely distributed, but also where the same waveform is recurrent throughout the signal. By only recording the waveform index, amplitude, and timing—a so-called atomic decomposition [6]—this approach provides significant compression. The atomic decomposition is only useful if selected waveforms match the signal such that a good approximation of the whole signal, or a relevant component, can be achieved with a limited number of atoms. Instead of using predefined waveforms the filters of the MISO model can be learned directly from the data.

There are two problems entwined in this learning problem: learning the dictionary—i.e., the set of filters—to represent the signal, and inferring the index, amplitude, and timing for the train of Dirac deltas, corresponding to the unobserved sources. Together these problems are known as blind deconvolution or blind system identification.

This study's primary contribution is the comparison of two distinct approaches for blind system identification with sparse sources. The first approach is to assume a generative model for the signal, with constraints in the form of sparse priors for the sources. Using a normal distribution for the noise, the optimal model is the one that minimizes the least-squares reconstruction cost while complying with the sparsity constraint, or jointly maximizing the sparsity of the source. Based on prior work in computational neuroscience, this approach is often referred to as sparse coding [7]. A number of researchers have shown how filters which describe natural signals can be efficiently estimated directly from data [8, 9, 10, 11, 12]. In particular, using matching-pursuit as a proxy for a maximum a posterior estimate, an efficient algorithm for learning sparse bases [13] has been extended to the time-series case [14, 15, 16]. In these cases the sources must be explicitly estimated.

The second approach avoids the explicit estimation of the sources. Instead the sparse sources' statistical properties are used in the filter estimation [17]. Essentially, this approach for blind estimation uses a matrix-based projection pursuit, where windows of the time series are treated as vectors in independent component analysis (ICA) [18]. In this way, the filters can be blindly estimated without estimating the sources or using a reconstruction cost function. Other researchers [19, 20] have demonstrated the ability of FastICA [21] to efficiently estimate the filters. FastICA is particularly suited for this since its objective implicitly estimates the sources through a non-linear function of the projected signals.

We perform experiments on synthetic data to compare shift-invariant sparse coding and ICA across a range of source rates. We consider the limiting case of a single source, which corresponds to blind channel estimation [17]. The single-source case is an important subproblem in some alternating estimation approaches wherein the estimation of a single source is performed assuming the contributions from all other sources have been removed.

The setting for these experiments is a single realization of rel-

This work was supported by DARPA Contract N66001-10-C-2008.

atively long (compared to the system's impulse response) and the sources are active throughout. This is different from 'shift-invariant' dictionary learning where multiple short segments are available and only relatively short shifts are allowed [22].

2. MODELING SYSTEMS EXCITED BY SPARSE SIGNALS

Let $x(t)$ be a signal is created by a multiple-input single-output (MISO) linear system with sparse inputs observed in the presence of noise:

$$x(t) = e(t) + \hat{x}(t) = e(t) + \sum_{p=1}^{P} y_p(t) \tag{1}$$

$$y_p(t) = \int_{-\infty}^{\infty} s_p(t-u)a_p(u)du, \quad p = 1, \ldots, P \tag{2}$$

$$s_p(t) = \sum_i \alpha_{p,i}\delta(t - \tau_{p,i}), \quad p = 1, \ldots, P. \tag{3}$$

Each component, $y_p(t), \quad p = 1, \ldots, P$, is formed by convolving a weighted train of delta functions $s_p(t)$ with a finite impulse response filter $a_p(t)$.

The atomic representation of the noise-free signal, $\hat{x}(t)$, consists of a set of source indices, amplitudes, and timings $\mathcal{S} = \{(p_i, \alpha_i, \tau_i)\}_i$. Using this set, and the model signal can rewritten as:

$$\hat{x}(t) = \sum_i \int_{-\infty}^{\infty} \alpha_i \delta(t - \tau_i - u)a_{p_i}(u)du. \tag{4}$$

3. MATCHING PURSUIT WITH K-SVD

Assuming that number of filters P and source excitations L are known, the blind system identification problem can be posed as a least-squares optimization over $\mathcal{A} = \{a_p(t)\}_{p=1}^P$ and $\mathcal{S} = \{(p_i, \alpha_i, \tau_i)\}_{i=1}^L$

$$\min_{\mathcal{A},\mathcal{S}} J(\mathcal{A}, \mathcal{S}) = \left\| x(t) - \sum_{i=1}^{L} \alpha_i \int_{-\infty}^{\infty} \delta(t-\tau_i)a_{p_i}(u)du \right\|_2^2. \tag{5}$$

Jointly solving for both $\mathcal{A}$ and $\mathcal{S}$ is difficult because of the dependence between them.

Assuming $\mathcal{A}$ is fixed, a greedy solution to $\min_{\mathcal{S}} J(\mathcal{A}, \mathcal{S})$ can be found by using matching pursuit approach for time series. At each iteration the solution is chosen as if only one source is active at only one point in time, and selects a single atom, consisting of the timing, amplitude, and waveform, that explains the most energy remaining in the residual of the signal. This criterion is equivalent to finding the filter with the highest normalized cross-correlation. At the end of each iteration, the residual signal is updated by removing the single-atom reconstruction. This updated residual is used as the input to the next iteration. If after every step the amplitudes are re-estimated this is known as orthogonal matching pursuit.

Given the atomic decomposition $\mathcal{S} = \{(p_i, \alpha_i, \tau_i)\}_{i=1}^L$, either the sources (deconvolution) or the individual components (demixing) can be computed easily via Equation (3) or Equation (2), respectively. How to identify $\mathcal{A}$ is the main question. Mailhé et al. [15] proposed an extension of the dictionary learning algorithm K-SVD [13] to the case of finding $\mathcal{A}$. This consists of an alternating optimization between (orthogonal) matching pursuit and K-SVD. For conciseness, we refer to it as MP-SVD (OMP-SVD).

Let $\mathbf{x}$ denote the discrete time version of $x(t)$. Following the notation in [15], let T_τ denote the linear operator such that $T_\tau \mathbf{v}$ aligns the M-length filter $\mathbf{v}$ within a N length signal. T_τ is a $N \times M$ matrix with the $M \times M$ identity matrix as a submatrix starting at row τ. The transpose of this operator is denoted T_τ^* and is defined such that $T_\tau^* \mathbf{x}$ extracts the M-length window from $\mathbf{x}$ starting at time τ. Using the alignment operator T_τ the cost function can rewritten in terms of vectors as

$$\min_{\{\mathbf{v}_p\}_{p=1}^P, \{(p_i,\alpha_i,\tau_i)\}_{i=1}^L} \left\| \mathbf{x} - \sum_{i=1}^{L} \alpha_i T_{\tau_i} \mathbf{v}_{p_i} \right\|_2^2. \tag{6}$$

To update the filters, we first assume we have an estimate of the components using the current filters. Let $\mathbf{x}^{(p)}$ denote the signal consisting only of the pth component and any error

$$\mathbf{x}^{(p)} = \mathbf{e} + \mathbf{y}_p = \mathbf{x} - \sum_{q \in \{1,\ldots,P\} \setminus p} \mathbf{y}_q \tag{7}$$

where $\mathbf{y}_p = \sum_{j \in \mathcal{I}_p} \alpha_j T_{\tau_j} \mathbf{v}_p$ and $\mathcal{I}_p = \{i : p_i = p\}$. This single-filter signal is used to update each filter, but because of the sparsity in time, only windows of the signal corresponding to source timings are used. These windows are collected into a matrix. The filter is updated to the primary singular vector of this matrix, which minimizes the least-squares error. As the timings may change with the new filter, they are re-estimated and the is alternating optimization problem continues.

To ameliorate the computational complexity of running matching pursuit multiple times, we use a non-overlapping approximation of time-series matching pursuit. As an approximation, we perform only a single cross-correlation for each filter, and extract timings for each source excitation that do not overlap with themselves (different filters are allowed to overlap). Given the timings and filter indices, the least-squares solution for the amplitudes is solved. Additionally, this non-overlapping approximation can be run multiple times in sequence, using the remaining residual from the previous run. As an alternative we compare with the block-based approximation of matching pursuit used in MoTif [23].

4. USING ICA FOR SPARSE MODELING

Using ICA to solve the blind deconvolution problem is motivated by the following reasoning: first, linear filtering induces temporal dependence in the resulting signal; consequently, a deconvolution filter that minimizes this dependence should produce the original signal—up to some indeterminacies. This is especially the case when the source has independent, identically distributed entries. Its well known that as long as the source is not Gaussian, then higher-order statistics can be used to estimate the dependence. Sparse sources are by definition not Gaussian.

The deconvolution filter is a demixing vector applied to a delay line of the signal. If it exists, the optimal deconvolution filter is the filter inverse or a lagged version of it. Unlike the standard case for ICA, when there is single vector for each source, in the time-series case there are many solutions to the single-channel deconvolution problem corresponding to different shifts of the demixing vector. Each shift corresponds to a minimum of the dependence measure [17] and a solution to the deconvolution problem.

4.1. FastICA

In the FastICA algorithm [21], dependence is evaluated as an approximation of negenetropy [24]. Negentropy gauges the non-

Gaussianity of a random variable in terms of its higher-order statistics. Since a Gaussian random variable only has second-order statistics these differences can be quantified by finding the expectation of a non-quadratic function such that the moment expansion will contain higher-order terms. The difference between these higher-order and those of a Gaussian are used to assess the non-Gaussianity.

Approximately, the negentropy of the random variable u is proportional to $E\{G(u)\} - E\{G(\nu)\}$ where $G(\cdot)$ is the contrast function and ν is a Gaussian random variable with the same mean and covariance as u. For random vectors, maximizing the sum of the negentropies, under the constraint of decorrelation, minimizes the mutual information between the elements. This is basic principle of contrast-function-based ICA approaches.

The optimization problem for single-unit FastICA is

$$\underset{\|\mathbf{w}^{\mathrm{T}}\Psi^{-1}\|_2=1}{\arg\max} \left[E\left\{G(\mathbf{w}^{\mathrm{T}}\mathbf{x})\right\} - E\left\{G(\nu)\right\}\right]^2, \tag{8}$$

where $\mathbf{w}$ is the demixing vector, Ψ is the whitening matrix, and $G(\cdot)$ is a suitably chosen contrast function [21]. For sparse sources, the contrast function is typically a symmetric sub-quadratic function such as $G(u) = \log\cosh(u)$, which has the $\tanh(u)$ as its first derivative.

While the vector $\mathbf{w}$ corresponds to a row of the demixing matrix, it also corresponds to the inverse of the filter such that $\mathbf{w}^{\mathrm{T}}\mathbf{x} = \hat{s}$ is a linear estimate of the source signal. The corresponding column of the mixing matrix is denoted $\mathbf{v}$ and is related by $\mathbf{v} = E\{\mathbf{x}\mathbf{x}^{\mathrm{T}}\mathbf{w}\} = E\{\mathbf{x}\hat{s}\}$. Clearly, $\mathbf{v}$ is nothing more than the weighted average of windows of the signal where the weighting assigned to each vector corresponds to the estimated source value.

An approximate solution to the single-unit ICA problem (8) yields an iterative update for $\mathbf{w}$ [21]:

$$\mathbf{w} \leftarrow \Psi^{\mathrm{T}}\Psi E\left\{\mathbf{x}g\left(\mathbf{w}^{\mathrm{T}}\mathbf{x}\right)\right\} - E\left\{g'\left(\mathbf{w}^{\mathrm{T}}\mathbf{x}\right)\right\}\mathbf{w}, \tag{9}$$

$g(u) = \frac{\partial}{\partial u}G(u)$, and $g'(u) = \frac{\partial}{\partial u}g(u)$. For $G(u) = \log\cosh(u)$, $g(u) = \tanh(u)$ is a soft activation function, and $g'(u) = 1 - \tanh^2(u)$ is a symmetric function that peaks at 0. In terms of the mixing vector this update is

$$\mathbf{v} \leftarrow E\left\{\mathbf{x}g\left(\hat{s}\right) - \mathbf{v}g'(\hat{s})\right\}. \tag{10}$$

This matches the interpretation of $\mathbf{v}$ as a weighted average, but gives an interpretation on its update. For a given realization, a large source magnitude $|\hat{s}| > 1$ implies $g(\hat{s}) \approx \mathrm{sign}(\hat{s})$ and $g(\hat{s}) \approx 0$, this yields $\mathbf{v} \approx \mathbf{x}$. When the source coefficient is smaller such that $g'(\hat{s})$ dominates $g(\hat{s})$, the update moves the vector away from its current direction. Thus, single-unit FastICA uses a weighted average to estimate the filter, wherein the source estimates correspond to source that is maximally non-Gaussian.

For the single-source case it would seem unnecessary to use more than on unit. In practice, using a multiple-unit approach with more units than sources performs better. In the time-series setting, the true mixing matrix is overcomplete, and the optimal vector for the single-unit case (8) will not necessarily correspond to a column of the matrix obtained for the multi-unit case.

4.2. Multiple source case

In the multiple source case, an ICA must resolve multiple demixing vectors simultaneously. The projections that yield sparse values correspond to demixing vectors for each source at each permissible lag. Since each source may have arbitrary lag, there is a even greater number of solutions that maximize the measure of independence. This can lead to the redundant estimation of the same filter at different lags. The constraint on correlation that avoids degenerate solutions in instantaneous ICA is inadequate for the convolutive case as different lags of the same source may be uncorrelated. For instance, if the length of the demixing vectors is longer than the filter extent, the 'independence' between filters can be found by simply lagging the filters such that they do not overlap. Sets of demixing vectors often include short filters at multiple shifts, or noisy spurious filters. Consequently, after running a multi-unit ICA, there is a need to select a subset of the unique and 'useful' filters.

4.3. Post-hoc filter subset selection

Assuming the number of estimated filters is higher than the desired (in either the single or multiple source case). The goal is to find the optimal subset that minimizes the reconstruction cost. This problem has a combinatorial number of solutions, and we use a greedy approach to find which set of filters best explains the signal.

The filters need to be correlated with the data, but not redundant copies of each other. To do this, an approximation of the signal is made with each individual filter using time-series matching pursuit. Then these components are treated as the basis vectors, and orthogonal matching pursuit (OMP) [25] is used to select a subset of these components to approximate the signal. In this manner, two filters that are simply shifted versions of each other will not both be selected, since their components will be approximately the same. Furthermore, in this manner it is straightforward to use a model selection criterion or cross-validation to select the total number of filters.

5. SYNTHETIC EXPERIMENTS

Experiments are conducted to compare two basic methodologies for system identification: shift-invariant decomposition with mean-square error cost, and single-channel ICA. The performance of algorithms are gauged on both single-source filter estimation and multiple source filter estimation. In both cases the true underlying filters are chosen as a set of Daubechies 4 (db4) wavelet packets. This ensures the synthetic filters cover a range of frequency space with varying temporal characteristics.

For each run, an output signal with 10,000 samples was created feeding sparse source signals through the MISO system. The wavelet packets were windowed to be 180 time points long. The source signals are independent, marked point processes with a homogeneous Poisson point process for the timing and a bimodal excitation amplitude distribution. The excitation amplitude distribution is a mixture of a Gaussian distributions with mean and variance of $(1, \frac{1}{9})$ and $(-1, \frac{1}{9})$ and equiprobable Bernoulli mixing. In the experiments, the rate of the Poisson process controls the sparsity of the sources. After scaling the amplitude of the filters to have a norm of 180/2, zero-mean, unit-variance white noise was added. This yielded signals with reasonable signal-to-noise ratios.

The algorithms are run with the correct number of filters, but they are not given the sparsity. Matching pursuit with either the block-based or non-overlapping approximation assume a fixed cardinality of the sources, based on how many blocks or non-overlapping filters fit the length of the signal. For the single-channel ICA, FastICA with either single-unit or 40-unit symmetric estimation was used. The $\tanh(\cdot)$ activation function is used for both instances of FastICA. In practice, most of the filters in the 40-unit estimation are meaningless so the OMP-based filter selection is used to select the predefined number of filters. The choice of 40 was arbitrary, but it is around one quarter of a full mixing matrix.

5.1. Single Source, Blind System Identification

For a single source, the source excitation rates between 0.1% and 50% were tested. For instance, at the rate of 0.55% each filter is on average excited once every 180 samples, the same as its length. Thus, this rate is a change point for the performance of MP-SVD with the non-overlapping and block-based approximations.

For each run, the filter estimation performance is quantified as the maximum correlation coefficient, across alignments, between the estimated filter and the true filter. This quantity is averaged across all 32 filters and the results are collected across 8 Monte Carlo generations of source and noise activity. The average computation time for the estimation is also recorded. These results are shown in Fig. 1.

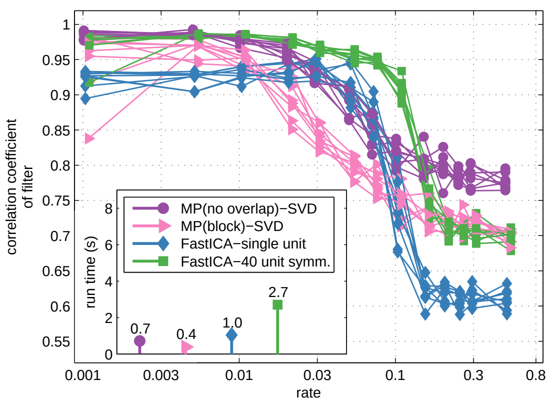

Fig. 1. Single-source blind waveform estimation performance. The average correlation coefficient across the 32 waveforms for each method is recorded across 8 Monte Carlo runs. (Inset) The average run time to estimate one waveform for each method.

At high and low excitation rates, MP-SVD achieves the best performance. However, in the range of rates between 3% and 10% multi-unit ICA outperforms the MP-SVD approach. Across sparsities, the 40-unit symmetric ICA estimation followed by the filter selection outperforms the single-unit ICA estimation.

5.2. Multiple Source, Blind System Identification

A subset of 11 filters was chosen for the multiple source case. The individual source excitation rate were varied between 0.025% and 10%. This corresponds to an overall rate of excitation between 0.275% and 110%.

For each run, the filter estimation performance is quantified using two indices: the first is the average of the correlation coefficient of each estimated filter to its best-matched true filter, and the second is the percentage of the true filters estimated with a correlation coefficient greater than 0.9. The first measure penalizes spurious filters that are not matched to any source. The second measure does not penalize spurious filters.

The results for MP-SVD and the 40-unit symmetric ICA with and without filter subset selection are shown in Fig. 2. The block-based approximation was tested also, but in the multiple source case the block-based approach failed.

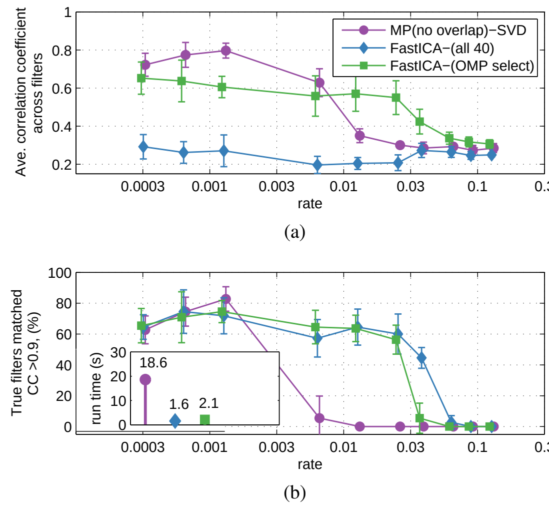

Fig. 2. Multiple waveform, single-channel blind estimation. The average and standard-deviation of the performance measures across 10 Monte Carlo runs are shown as error-bars. (a) Average correlation coefficient of estimated filters to their best-matched true filter. (b) Percentage of the true filters matched (correlation coefficient >0.9).

The MP-SVD approaches achieves the highest performance at select rates, with an average correlation above 0.7 for rates below 1%. The matching performance is more sensitive to the rate, with a peak above 80%, but dropping to zero above 1%. ICA consistently matches 60% of the filters at rates below 3%. The average best-matched correlation coefficient is lower than MP-SVD at low rates, but is consistent across a range of sparsity levels. The ICA-based approach appears to have the best tradeoff in terms of speed and consistent accuracy, as the runtime of MP-SVD is 9× higher.

5.3. Discussion

These results on synthetic data confirm that both single-channel ICA and shift-invariant sparse coding are able to blindly estimate the underlying filters of a sparsely-excited multiple-input single output system. For cases of sparse sources, MP-SVD is the most accurate for blind system identification for both SISO and MISO systems. However, as the source cardinality of non-overlapping MP-SVD was fixed, for higher rate sources its performance suffers. ICA's avoidance of explicit modeling allows it to perform better at higher source rates.

On real signals, the true number of filters and their sparsity is unknown. A model selection criterion can be used to determine the optimal number of filters. For single-channel ICA this can be done during the post-hoc filter selection, but the sparse coding approach this would require the training of multiple models of varying complexity. Single-channel ICA also avoids any pre-specification of the sparsity of the sources. In the sparse coding case, this could also be accomplished using model selection or a generative model able to perform inference of the joint rate-amplitude distribution of each source. These points highlight single-channel ICA with post-hoc filter selection as a viable approach to blindly estimate MISO systems with sparse sources.

6. REFERENCES

[1] J. L. Roux, A. D. Cheveigné, and L. C. Parra, "Adaptive template matching with shift-invariant semi-NMF," in *Advances in Neural Information Processing Systems*, D. Koller, D. Schuurmans, Y. Bengio, and L. Bottou, Eds., number 21, pp. 921–928. 2009.

[2] J. W. Pillow, J. Shlens, E. Chichilnisky, and E. P. Simoncelli, "A model-based spike sorting algorithm for removing correlation artifacts in multi-neuron recordings," *PLoS ONE*, vol. 8, no. 5, pp. e62123, 2013.

[3] C. Ekanadham, D. Tranchina, and E. P. Simoncelli, "A unified framework and method for automatic neural spike identification," *Journal of Neuroscience Methods*, vol. 222, pp. 47–55, 2014.

[4] B. Mailhé, R. Gribonval, F. Bimbot, M. Lemay, P. Vandergheynst, and J.-M. Vesin, "Dictionary learning for the sparse modelling of atrial fibrillation in ECG signals," in *Acoustics, Speech and Signal Processing (ICASSP), IEEE International Conference on*, 2009, pp. 465–468.

[5] A. Papoulis, *Probability & statistics*, vol. 2, Prentice-Hall Englewood Cliffs, 1990.

[6] S. S. Chen, D. L. Donoho, and M. A. Saunders, "Atomic decomposition by basis pursuit," *SIAM Journal on Scientific Computing*, vol. 20, no. 1, pp. 33–61, 1998.

[7] B. A. Olshausen and D. J. Field, "Sparse coding with an overcomplete basis set: A strategy employed by V1?," *Vision Research*, vol. 37, no. 23, pp. 3311–3325, 1997.

[8] M. Lewicki, "Efficient coding of natural sounds," *Nature Neuroscience*, vol. 5, no. 4, pp. 356–363, 2002.

[9] D. C. Balcan and M. S. Lewicki, "Point coding: Sparse image representation with adaptive shiftable-kernel dictionaries," in *SPARS'09-Signal Processing with Adaptive Sparse Structured Representations*, 2009.

[10] C. Ekanadham, D. Tranchina, and E. Simoncelli, "Recovery of sparse translation-invariant signals with continuous basis pursuit," *IEEE Transactions on Signal Processing*, vol. 59, no. 10, pp. 4735–4744, 2011.

[11] T. Blumensath and M. Davies, "Sparse and shift-invariant representations of music," *IEEE Transactions on Audio, Speech and Language Processing*, vol. 14, no. 1, pp. 50–57, 2006.

[12] E. Smith and M. Lewicki, "Efficient auditory coding," *Nature*, vol. 439, no. 7079, pp. 978–982, 2006.

[13] M. Aharon, M. Elad, and A. Bruckstein, "K-SVD: An algorithm for designing overcomplete dictionaries for sparse representation," *IEEE Transactions on Signal Processing*, vol. 54, no. 11, pp. 4311–4322, 2006.

[14] M. Aharon, *Overcomplete dictionaries for sparse representation of signals*, Ph.D. thesis, Technion-Israel Institute of Technology, Faculty of Computer Science, 2006.

[15] B. Mailhé, S. Lesage, R. Gribonval, F. Bimbot, and P. Vandergheynst, "Shift-invariant dictionary learning for sparse representations: Extending K-SVD," in *16th Euro. Signal Process. Conf.*, 2008.

[16] J. Thiagarajan, K. Ramamurthy, and A. Spanias, "Shift-invariant sparse representation of images using learned dictionaries," in *Machine Learning for Signal Processing (MLSP), IEEE Workshop on*, Oct 2008, pp. 145–150.

[17] O. Shalvi and E. Weinstein, "New criteria for blind deconvolution of nonminimum phase systems (channels)," *IEEE Transactions on Information Theory*, vol. 36, no. 2, pp. 312–321, 1990.

[18] A. Bell and T. Sejnowski, "Learning the higher-order structure of a natural sound," *Network: Computation in Neural Systems*, vol. 7, no. 2, pp. 261–266, 1996.

[19] M. Davies and C. James, "Source separation using single channel ICA," *Signal Processing*, vol. 87, no. 8, pp. 1819–1832, 2007.

[20] F. Lucena, A. Barros, J. Príncipe, and N. Ohnishi, "Statistical coding and decoding of heartbeat intervals," *PLoS ONE*, vol. 6, no. 6, pp. e20227, 06 2011.

[21] A. Hyvarinen, "Fast and robust fixed-point algorithms for independent component analysis," *IEEE Transactions on Neural Networks*, vol. 10, no. 3, pp. 626–634, 1999.

[22] C. Rusu, B. Dumitrescu, and S. Tsaftaris, "Explicit shift-invariant dictionary learning," *IEEE Signal Processing Letters*, vol. 21, no. 1, pp. 6–9, Jan 2014.

[23] P. Jost, P. Vandergheynst, S. Lesage, and R. Gribonval, "MoTIF: An efficient algorithm for learning translation invariant dictionaries," in *Acoustics, Speech and Signal Processing (ICASSP), IEEE International Conference on*, May 2006, vol. 5, pp. 857–860.

[24] M. Novey and T. Adali, "Complex ICA by negentropy maximization," *IEEE Transactions on Neural Networks*, vol. 19, no. 4, pp. 596–609, 2008.

[25] Y. Pati, R. Rezaiifar, and P. Krishnaprasad, "Orthogonal matching pursuit: Recursive function approximation with applications to wavelet decomposition," in *Signals, Systems and Computers, 1993. 1993 Conference Record of The Twenty-Seventh Asilomar Conference on*, 1993, pp. 40–44.