DICTIONARY LEARNING OF CONVOLVED SIGNALS

Daniele Barchiesi and Mark D. Plumbley

Queen Mary University of London
Centre for Digital Music
School of Electronic Engineering and Computer Science

ABSTRACT

Assuming that a set of source signals is sparsely representable in a given dictionary, we show how their sparse recovery fails whenever we can only measure a convolved observation of them. Starting from this motivation, we develop a block coordinate descent method which aims to learn a convolved dictionary and provide a sparse representation of the observed signals with small residual norm. We compare the proposed approach to the K-SVD dictionary learning algorithm and show through numerical experiment on synthetic signals that, provided some conditions on the problem data, our technique converges in a fixed number of iterations to a sparse representation with smaller residual norm.

Index Terms— Dictionary Learning, Sparse Representation, Convolution, K-SVD.

1. INTRODUCTION

Sparse representation has recently become a very popular research field which is causing a profound impact on the scientific community with applications to sampling, coding, analysis and processing of signals [1].

The general model behind sparse representation consists in decomposing a signal as a linear combination of elementary functions or *atoms* and seeking the representation with the smaller number of active atoms. Generally speaking, when attempting a sparse representation of a variable or group of variables that belong to a common class, one needs to specify a (possibly over-complete) *dictionary*, that is, a collection of atoms in which the observed data are supposed to be sparse. Although a great effort has been spent to design explicit dictionaries that are well suited for images and other types of signals [2], it is often desirable to learn an optimal dictionary from a set of training samples and then utilise it to represent new data of the same class. This problem is known as *dictionary learning* and, over recent years, various techniques have been proposed in order to address it [3]. Some of these methods can be tested and compared using the MATLAB toolbox SMALLBOX[1].

Whenever a physical phenomenon is measured by means of transducers (e.g. recording an audio signal by means of a microphone), the observed variable can be modelled as the convolution of the original signal of interest with the impulse response of the system in which the measurement takes places. Unfortunately, this process causes the sparse representation to be lost, as we will show in Section 2. This fact comes at no surprise if we realise that the signals are no longer sparse in the original dictionary, but in a new dictionary which is a convolved version of the initial one. In Section 3 we propose an algorithm whose aim is to learn a convolved

___
[1]http://small-project.eu/software-data/smallbox/

dictionary that provides a sparse representation with small residual norm and present some numerical tests carried out in order to assess its performance, comparing it to the K-SVD algorithm [4]. Section 4 will conclude the paper with a brief summary of our results and plans for further research.

2. CONVOLUTION AND SPARSE REPRESENTATION

2.1. Model and notation

Suppose that a set of source signals $\{\mathbf{s}_n\}_{n=1}^{N}$ is sparsely representable in a dictionary $\mathbf{\Phi}$, which is a D by K matrix of rank D containing a collection of atoms $\{\boldsymbol{\phi}_k\}_{k=1}^{K}$ in its columns.

$$\mathbf{s}_n = \mathbf{\Phi}\mathbf{x}_n \qquad ||\mathbf{x}_n||_0 \leq S_0 \quad \forall\, n = 1\dots N$$

where $\mathbf{x}_n$ is the vector containing the representation coefficients of the n-th signal and the operator $||\cdot||_0$ is the ℓ_0 pseudo-norm which counts the number of nonzero elements of its argument. This means that each source signal has a sparse representation in the dictionary $\mathbf{\Phi}$ with at most S_0 active atoms.

We do not directly observe the variables $\mathbf{s}_n$, but rather a set of convolved observations

$$y_n[t] = \sum_\tau h[\tau]s_n[t-\tau]$$

that are the output of a single input single output (SISO) convolutive system characterised by the impulse response $\mathbf{h}$. Since convolution is a linear operator, the resulting variables will be no longer sparse in the dictionary $\mathbf{\Phi}$ but in a dictionary $\mathbf{\Psi}$ whose atoms $\boldsymbol{\psi}_k = \mathbf{h} * \boldsymbol{\phi}_k$ are obtained by convolving the original dictionary atoms with the impulse response $\mathbf{h}$.

In many applications the impulse response of the measurement system is unknown; in this case, one could still use the dictionary $\mathbf{\Phi}$ and attempt an approximation of the observed signals. However, as shown in the next section, this would greatly affect the performance of sparse representation.

2.2. Effect of convolution on sparse representation

In order to show the effect of convolution on sparse representation, we tested the performance of both the sparsity-constrained and the error-constrained versions of the orthogonal matching pursuit algorithm [5] which will be labelled OMP-S and OMP-E respectively. The former aims at solving the following optimisation:

$$\underset{\mathbf{X}}{\text{minimise}} \quad ||\mathbf{Y} - \mathbf{\Phi}\mathbf{X}||_{\mathrm{F}}^{2} \qquad (1)$$

$$\text{subject to} \quad ||\mathbf{x}_n||_0 \leq S_0 \quad \forall\, n = 1\dots N$$
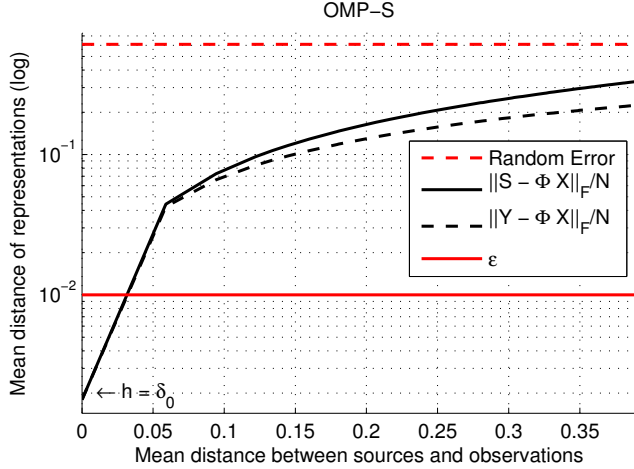
Fig. 1. OMP-S results on convolved signals (averaged values over 100 trials of the experiment). The two black curves represent the mean distance between the reconstructed signals and the sources or the observed variables respectively. For comparison purpose, the red dashed line represents the mean distance between the observed variables and random signals, while the red solid line is the error tolerance defined for OMP-E.

where $||\cdot||_{\mathrm{F}}$ is the Frobenious norm, while the latter attempts to solve the problem

$$\operatorname*{minimise}_{\mathbf{X}} \quad ||\mathbf{X}||_0 \tag{2}$$
$$\text{subject to} \quad ||\mathbf{y}_n - \boldsymbol{\Phi}\mathbf{x}_n||_2^2 \le \epsilon \quad \forall\, n = 1 \dots N$$

The observed signals $\mathbf{y}_n$ were generated according to the model described in Section 2.1. In particular, we firstly defined $N = 500$ source signals of dimension $D = 100$ as sparse linear combinations of the atoms contained in a two times over-complete real Fourier dictionary (one of the standard dictionary ensembles implemented in the SPARSELAB toolbox, available as part of the SMALLBOX). The normalised diversity of the source signals, defined as the ratio between the number of nonzero coefficients of the representations and the number of atoms in the dictionary, was set to $||\mathbf{x}||_0/K = 0.05$. We produced the observations by convolving the sources with a sparse non negative impulse response $\mathbf{h}$ of length $L = 50$. We repeated the experiments varying the number of non-zero elements of $\mathbf{h}$ from 1 to L, thus causing an increasing mean distance between sources and observations measured by $\bar{d}(\mathbf{S}|\mathbf{Y}) = ||\mathbf{S} - \mathbf{Y}||_{\mathrm{F}}/N$. Here the columns of $\mathbf{S}$ and $\mathbf{Y}$ contain the sources and convolved observations respectively.

Figures 1 and 2 depict the results of the experiment for the two versions of OMP. Let us first analyse the results for OMP-S. We ran the algorithm using the dictionary $\boldsymbol{\Phi}$ on the convolved variables $\mathbf{Y}$ and setting the number of active atoms to $S_0 = 15$.

As can be seen in the left side of the plot, when the impulse response is simply $\mathbf{h} = \boldsymbol{\delta}_0$ the source and observed variables are the same and OMP-S is able to represent them with negligible error. However, as the convolution causes the observed $\mathbf{Y}$ to differ from $\mathbf{S}$, the error in the representation quickly increases, to the point where OMP-S becomes almost comparable with a random representation (i.e. a gaussian random signal of unit variance generated for comparison purposes). The behaviour of OMP-E is similar: fixing a tolerance $\epsilon = 10^{-2}$, the algorithm is able to represent the observed signals
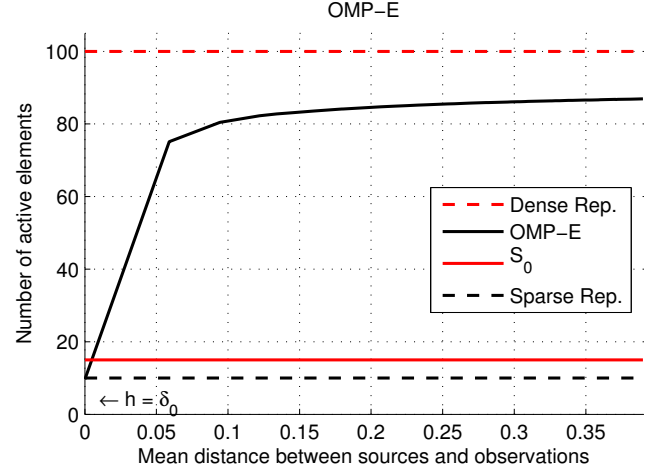
Fig. 2. OMP-E results on convolved signals (averaged values over 100 trials of the experiment). The black and red dashed lines represent the true number of active atoms used to generate the test signals and the number of active atoms of a completely dense representation respectively. S_0 is the constraint parameter used for OMP-S

using the right number of active elements in the trivial case $\mathbf{h} = \boldsymbol{\delta}_0$. However, as soon as the mean distance $\bar{d}(\mathbf{S}|\mathbf{Y})$ increases, the number of active elements needed rapidly rises over 80% of a completely dense representation.

3. DICTIONARY LEARNING OF CONVOLVED SIGNALS

The simple experiment presented in the previous section shows that a sparse representation algorithm fails whenever applied to a convolved observation of sparse signals. Starting from this motivation, we present a dictionary learning algorithm that takes into account the particular structure of the convolved dictionary $\boldsymbol{\Psi}$ and aims to learn the impulse response of the SISO system that produced it. The general optimisation strategy, which will be referred as *block coordinate descent*, is similar to the one employed in the method of optimal directions presented in [6].

3.1. Block Coordinate Descent Optimisation

In order to learn a convolved dictionary for the sparse representation of the observed variables, we seek the solution of the following optimisation problem:

$$\operatorname*{minimise}_{\mathbf{h},\mathbf{X}} \quad ||\mathbf{Y} - \mathbf{H}\boldsymbol{\Phi}\mathbf{X}||_{\mathrm{F}}^2 \tag{3}$$
$$\text{subject to} \quad ||\mathbf{x}_n||_0 \le S_0 \quad \forall\, n = 1 \dots N$$

Here the matrix $\mathbf{H}$ is the Toeplitz convolutive matrix associated with the impulse response $\mathbf{h}$.

Unfortunately, this problem is underdetermined and non-convex; therefore, we choose to adopt an iterative block coordinate descent (BCD) strategy, minimising the cost function alternatively for the variables $\mathbf{X}$ and $\mathbf{h}$. This leads to the following steps:

- **X-Step**: given a fixed impulse response, the above minimisation is a standard sparse representation problem which can be tackled with a variety of methods. In the present formulation, we consider a constraint on the ℓ_0 pseudo-norm of the
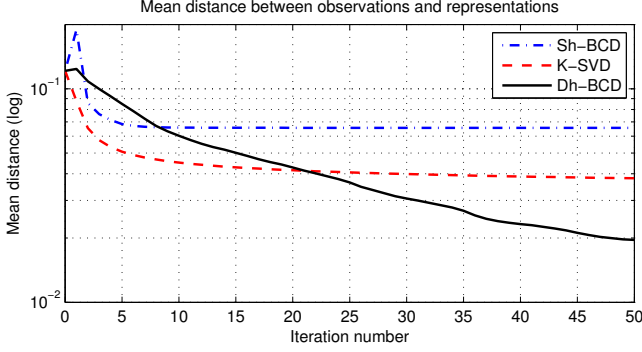
Fig. 3. BCD and K-SVD residuals as a function of the iteration number (average over 100 trials of the experiment)

representation coefficients and employ a suboptimal greedy algorithm such as OMP-S.

- **h**-Step: given the current estimate of the representation coefficients, the minimisation (3) with respect to **h** can be turned into an overdetermined quadratic program which can be solved with any standard convex optimisation method.

At each iteration, the estimated impulse response is re-scaled in order to preserve the Frobenious norm of the original dictionary $\mathbf{\Phi}$ and avoid a scaling ambiguity inherent in the optimisation. For a detailed derivation of the algorithm which includes learning the impulse response in the Fourier domain where the convolution can be interpreted as a simple multiplication, see [7].

The **h**-Step can be modified by constraining the impulse response to be sparse and non-negative. This is realistic for the representation of audio signals in that, according to the image method [8], the early reflections coming from the surfaces of a small size room can be well described by such vectors. Employing or not these constraints during the **h**-Step leads to two different versions of the BCD algorithm that will be labelled *sparse **h*** (S**h**-BCD) and *dense **h*** (D**h**-BCD) and that will be compared in the next section.

3.2. BCD optimisation for dictionary learning

In order to evaluate the proposed dictionary learning algorithms, we generated a set of test signals using the parameters described in Section 2.2 and convolved them with a sparse, non-negative impulse response of normalised diversity $||\mathbf{h}||_0/L = 0.05$. We then initialised a dictionary by convolving $\mathbf{\Phi}$ with a random impulse response and ran S**h**-BCD, D**h**-BCD and the K-SVD algorithm for 50 iterations, calculating the mean distance between observations and representations defined as:

$$\bar{d}(\mathbf{h}, \mathbf{X}|\mathbf{Y}) = \frac{||\mathbf{Y} - \mathbf{H}\mathbf{\Phi}\mathbf{X}||_{\mathrm{F}}}{N}$$

Figure 3 depicts the results averaged over 100 trials of the experiment. As can be seen, D**h**-BCD is the only method which does not seem to get trapped in a local minima of the objective function, improving its value during all the iterations. On the other hand, the convergence of K-SVD and S**h**-BCD is significantly slower, with the latter method performing worse.

Figure 4 offers a more precise comparison between K-SVD and D**h**-BCD by showing the boxplot of the mean distance $\bar{d}(\mathbf{h}, \mathbf{X}|\mathbf{Y})$ as a function of the iteration number. We can note that K-SVD is definitely more robust to outliers. However, the proposed block coordinate descent with dense estimation of the impulse response consis-
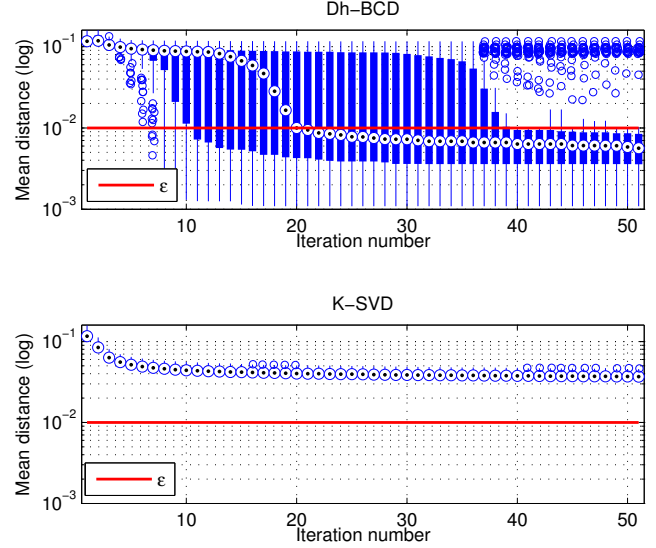

Fig. 4. Boxplot comparison of D**h**-BCD and K-SVD over 100 trials of the experiment. For each iteration, the central mark is the median, the edges of the boxes are the 25th and 75th percentiles, the whiskers extend to the most extreme data points not considered outliers, and outliers are plotted individually.

tently achieves a better result, reaching the error tolerance defined for the experiment described in Section 2.2. These surprising results suggest that constraining the solution to belong to the feasible set from where the test data were generated is not a good strategy, while performing an unconstrained optimisation of the impulse response allows for the necessary flexibility required to minimise the non-convex cost function whenever the initialisation is far from the global minimum. Moreover, the fact that K-SVD is outperformed by D**h**-BCD indicates that taking into account the particular structure of the dictionary and reducing therefore the number of free parameters of the optimisation from the whole set of atoms to the impulse response coefficients leads to significative improvements.

Since the sparsity of the problem data always plays a crucial role in every sparse representation framework, we tested the algorithms varying the normaised diversity of source signals and impulse response between 1% and 25% of the respective dimensions, again comparing the results with the K-SVD algorithm. Figure 5 shows the contour plot of the residual error achieved at the end of the optimisation by the various methods, along with a comparison plot which shows the best performing technique in each point of the sources/impulse response normalised diversity plan.

As we might expect, the two variants of the proposed block coordinate descent method perform well when the source signals and the impulse response are sparse, exhibiting a slightly stronger dependence on the sources normalised diversity. The results for the K-SVD algorithm, on the other hand, seem to depend strongly on the normalised diversity of the impulse response, presenting also a slight drop in correspondence with a source normalised diversity of 0.05. Overall, the comparison plot reveals that, as long as the sources normalised diversity is below 10% of the signals dimension D and the impulse response is sufficiently sparse, then K-SVD is outperformed by D**h**-BCD. This condition is not unrealistic and corresponds to the common assumption $S_0 \ll D$ made throughout most of the literature on sparse representation.
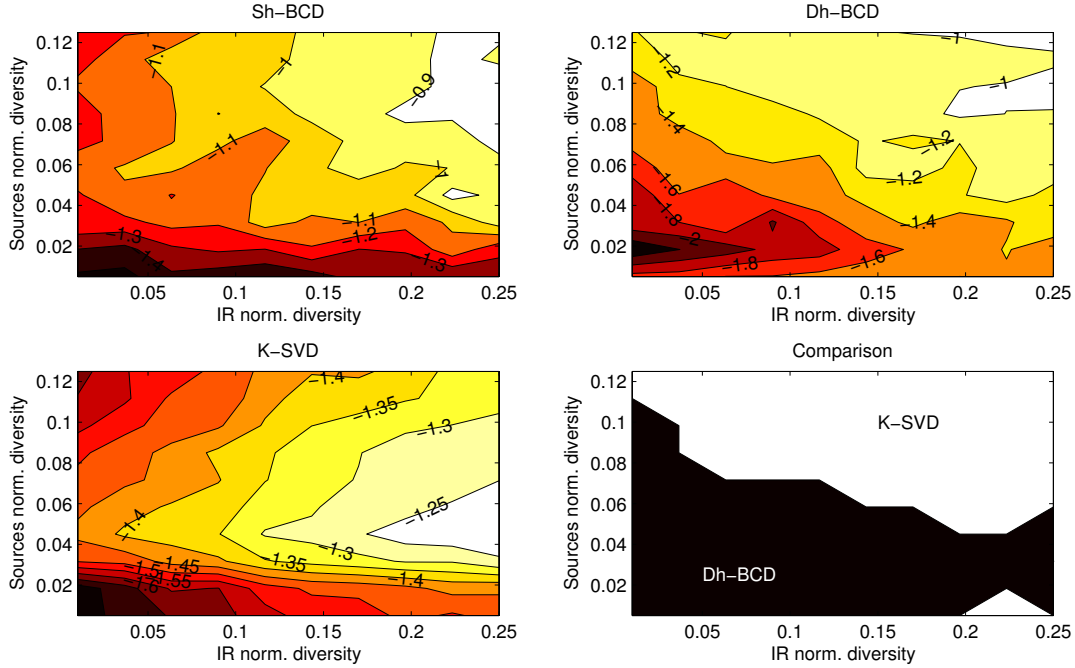
Fig. 5. BCD and K-SVD results for various normalised diversities of source signals and impulse response (average over 20 trials). The values appearing along the contour plots represent the mean distance $\bar{d}(\mathbf{h}, \mathbf{X}|\mathbf{Y})$ achieved after 50 iterations by the various algorithms on a logarithmic scale. Since the dictionary is two times over-complete, a completely dense representation corresponds to a source normalised diversity $||\mathbf{x}||_0/K = .5$

4. CONCLUSIONS AND FURTHER RESEARCH

We presented a dictionary learning method suitable for the sparse representation of convolved signals using a block coordinate descent strategy. We compared it with the K-SVD algorithm and assessed through numerical experiments that, providing some conditions on the sparsity of the problem data, the proposed method leads to improvements over a general purpose dictionary learning algorithm.

The promising results presented in this paper suggest an interesting direction for further investigation. If the algorithm is able to provide a sparse representation with small or negligible residual, this might occur because it identifies the true impulse response that generated the observations, linking therefore two apparently distant problems such as dictionary learning and blind deconvolution. Although preliminary results suggest that this might be the case for short, sparse impulse responses, additional tests with real world data and a deeper understanding of the convergence properties of the optimisation are required and will be object of further research.

5. ACKNOWLEDGMENTS

6. REFERENCES

[1] Richard G. Baraniuk, Emmanuel Candès, Michael Elad, and Yi Ma, "Applications of sparse representation and compressive sensing," *Proceedings of the IEEE*, vol. 98, no. 6, pp. 906–909, Jun. 2010.

[2] Stéphane Mallat, *A Wavelet Tour of Signal Processing, Second Edition (Wavelet Analysis & Its Applications)*, Academic Press, London, UK, 2 edition, Sep. 1999.

[3] Ron Rubinstein, Alfred Bruckstein, and Michael Elad, "Dictionaries for sparse representation modeling," *Proceedings of the IEEE*, vol. 98, no. 6, pp. 1045–1057, Jun. 2010.

[4] Michal Aharon, Michael Elad, and Alfred Bruckstein, "K-SVD: An algorithm for designing overcomplete dictionaries for sparse representation," *IEEE Trans. on Signal Processing*, vol. 54, no. 11, pp. 4311–4322, Nov. 2006.

[5] Y.C. Pati, R. Rezaiifar, and P.S. Krishnaprasad, "Orthogonal matching pursuit: Recursive function approximation with applications to wavelet decomposition," in *Proceedings of the 27th Asilomar Confrence on Signals, Systems and Computers*, Nov. 1993, vol. 1, pp. 40–44.

[6] Kjersti Engan, Sven Ole Aase, and John Håkon Husøy, "Method of optimal directions for frame design," in *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 1999, vol. 5, pp. 2443–2446.

[7] Daniele Barchiesi and Mark D. Plumbley, "Dictionary Learning of Convolved Signals," Tech. Report, available at http://www.eecs.qmul.ac.uk/~danieleb/, Oct. 2010.

[8] Jont B. Allen and David A. Berkley, "Image method for efficiently simulating small-room acoustics," *Journal of the Acoustical Society of America*, vol. 65, no. 4, pp. 943–950, Apr. 1979.