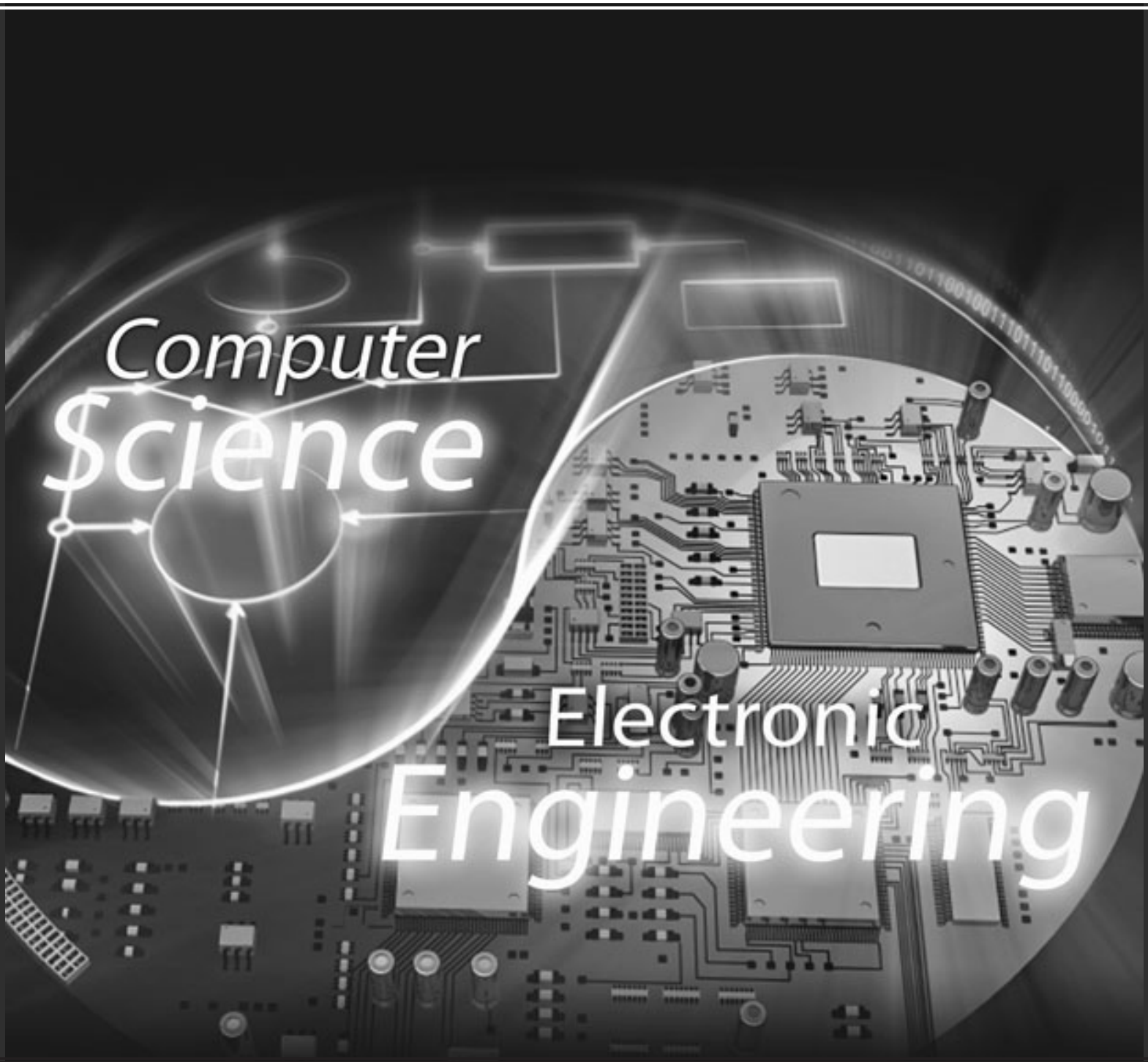
ISSN 2043-0167

Dictionary Learning of Convolved Signals

Daniele Barchiesi and Mark D. Plumbley

EECSRR-10-04 November 2010

Queen Mary
University of London
School of Electronic Engineering
and Computer Science



Dictionary Learning of Convolved Signals

Technical Report

Daniele Barchiesi Mark D. Plumbley

November 23, 2010

Abstract

Assuming that a set of source signals is sparsely representable in a given dictionary, we show how the sparse recovery of such variables 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 to a better representation. The analysis of our results is then extended to the identification of the impulse response that generated the problem data, showing that learning a structured dictionary can be useful for the blind deconvolution of the observed variables.

Keywords: Dictionary Learning, Sparse Representation, Convolution, K-SVD, Blind Deconvolution, Dictionary Identification.

Contents

1 Introduction and Overview

Sparse representations is nowadays a very popular research field which is causing a profound impact on the scientific community and which is likely to revolutionise the way we will acquire, analyse, store and transmit signals in the future. The interested reader can find excellent introductions to the field in the review papers [7, 4, 13] and references therein.

Generally speaking, when seeking a sparse representation of a signal or group of signals that belong to a common class, one needs to specify a (possibly over-complete) dictionary in which the observed variables 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 [10][14, Sec. II-III], it is often desirable to learn an optimal dictionary from a set of training samples and then utilise the learned atoms to represent new data of the same type. This problem is known as *dictionary learning* and, over recent years, various techniques have been proposed in order to address it [14, Sec. IV]. Some of these methods can be tested and compared using the MATLAB toolbox SMALLBOX[1].

One of the advantages of using sparse representations is that they provide an automatic tool for denoising signals, as is shown in the classic paper by Chen et al. [5, Sec. 5]. This happens because a noisy signal, usually modelled as a random i.i.d. gaussian variable, is weakly correlated with the dictionary atoms and, therefore, will not be represented along with the signal of interest.

Unfortunately, whenever a physical phenomenon is observed by means of transducers (e.g. recording an audio signal by means of a microphone), there is another important factor to take into account. The recorded variable can be modelled as the convolution of the original signal of interest (*source signal*) with the impulse response of the system in which the measurement takes places. This process greatly affects the performance of sparse representation algorithms, as is shown in Section 2.

Starting from this motivation, the present work deals with learning a dictionary which can be used to represent convolved observations, given the assumption that the underlying source signals belong to a class for which there exist an explicit or learned optimal dictionary. As is shown in Section 3, a general purpose learning algorithm such as K-SVD [1] can be outperformed by designing an optimisation strategy that takes into account the particular structure of the learned dictionary, that is, a base dictionary convolved with an impulse

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

response. The encouraging results presented in this section lead to a more ambitious goal: if the algorithm is able to provide a sparse representation of the observed data with a small or negligible residual, does it occur because it identified the correct impulse response that generated the convolved variables? Can we therefore link two apparently distant problems such as dictionary learning and blind deconvolution? This topic and its link with the *dictionary identification* task is discussed in Section 4, along with experimental results which show that this is indeed possible under certain conditions on the sparsity of source signals and impulse response. Finally, in Section 5 we will draw some conclusion and present the challenges involved in the further research on these topics. A brief mention of reproducible research (Section 6) will conclude the report.

2 Effect of Convolution on Sparse Representation

2.1 Model and Notation

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

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

Here $\mathbf{x}_n \in \mathbb{R}^K$ is the vector containing the representation coefficients of the n-th signal and the operator $||\cdot||$ is the ℓ_0 pseudo-norm which counts the number of nonzero elements of its argument. The above formula means that each source signal has a sparse representation in the dictionary $\mathbf{\Phi}$ with at most S_0 active elements.
We do not directly observe the variables $\mathbf{s}_n$ but rather a set of convolved observations

$$y_n[i] = \sum_{l=0}^{L-1} h[l] s_n[i-l]$$

that are the output of a single input single output (SISO) convolutive system characterised by the impulse response $\mathbf{h} \in \mathbb{R}^L$.

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 are the convolution of the original atoms $\boldsymbol{\phi}_k$ with the impulse response of the system. This can be written in matrix notation

considering the Toeplitz convolutive matrix $\mathbf{H}$ associated with the impulse response, that is, the matrix whose columns contain shifted versions of $\mathbf{h}$.

$$\mathbf{\Psi} = \mathbf{H}\mathbf{\Phi}.$$

Stacking the convolved variables, the source signals and the representation coefficients in the columns of the matrices $\mathbf{Y}$, $\mathbf{S}$ and $\mathbf{X}$ respectively, we can express the observed data as:

$$\mathbf{Y} = \mathbf{H}\mathbf{S} = \mathbf{H}\mathbf{\Phi}\mathbf{X}.$$

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

2.2 Numerical Evaluation

In order to show the effect of convolution on sparse representations, we tested the performance of both the sparsity constrained and the error constrained versions of the orthogonal matching pursuit (OMP) algorithm [12]. The former aims at solving the following optimisation:

$$\begin{aligned} \underset{\mathbf{X}}{\text{minimise}} \quad & ||\mathbf{Y} - \mathbf{\Phi}\mathbf{X}||_{\mathrm{F}}^2 \\ \text{subject to} \quad & ||\mathbf{x}_n||_0 \leq S_0 \quad \forall\, n = 1 \dots N \end{aligned} \tag{OMP-S}$$

while the latter attempts to solve the problem

$$\begin{aligned} \underset{\mathbf{X}}{\text{minimise}} \quad & ||\mathbf{X}||_0 \\ \text{subject to} \quad & ||\mathbf{y}_n - \mathbf{\Phi}\mathbf{x}_n||_2^2 \leq \epsilon \quad \forall\, n = 1 \dots N. \end{aligned} \tag{OMP-E}$$

The matrix of observed signals $\mathbf{Y}$ was generated according to the model described in Section 2.1 and the parameters used for the simulation are summarised in Table 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 (RST)dictionary, which is one

Variable	Value	Description
Ensemble	RSE	Real Fourier dictionary ensemble
N	500	Number of observed signals
K	200	Number of atoms of the dictionary
$\|\|\mathbf{x}\|\|_0/K$	.05	Mean normalised diversity of source signals
S_0	15	Sparsity constraint for OMP-S
ϵ	10^{-2}	Error constraint for OMP-E

Table 1: *Parameters used to generate the source and convolved signals*

of the standard matrix ensembles implemented in the SparseLab toolbox [2], 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 h of length L = 50. We repeated the experiments varying the number of non-zero elements of h from 1 to L, thus causing an increasing mean distance between sources and observations measured by:

$$\bar{d}(\mathbf{S}|\mathbf{Y}) = \frac{||\mathbf{S} - \mathbf{Y}||_{\mathrm{F}}}{N}$$

Here the columns of $\mathbf{S}$ and $\mathbf{Y}$ contain the sources and convolved observations respectively. The sizes of the variables, which may seem somewhat arbitrary, is within the range of values commonly used for a frame-by-frame processing of audio signals.

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 $\mathbf{\Phi}$ on the convolved variables $\mathbf{Y}$ setting the number of active atoms to $S_0 = 15$. Since under-estimating the signals sparsity improves the performance of OMP in the presence of modelling errors, we allowed for 50% more active atoms than the real sparsity of the sources would.

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.

The behaviour of OMP-E is similar: fixing a tolerance $\epsilon = 10^{-2}$, the algorithm is able

[2] http://sparselab.stanford.edu

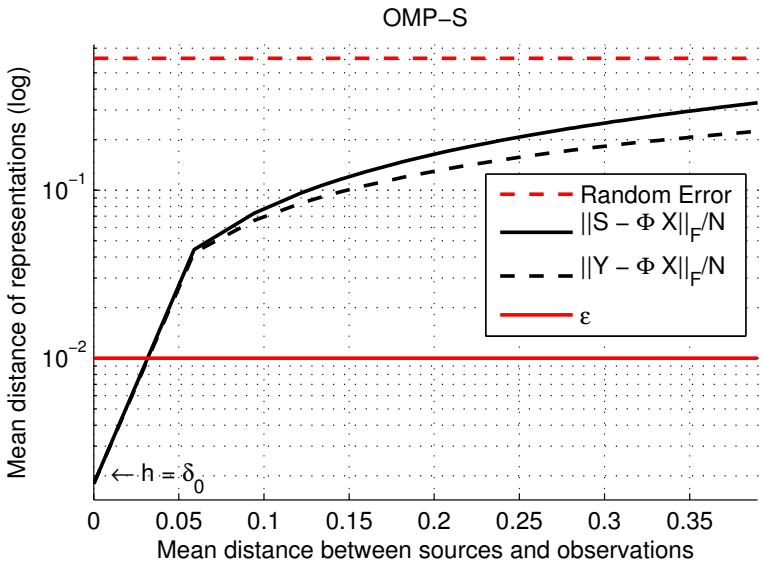


Figure 1: OMP-S results on convolved signals (averaged values over 100 trials of the experiment). The two black curves represent the mean distance between the reconstructed signals and the sources or the observed variables respectively. For comparison purpose, the red dashed line represents the mean distance between the observed variables and random signals, while the red solid line is the error tolerance defined for OMP-E.

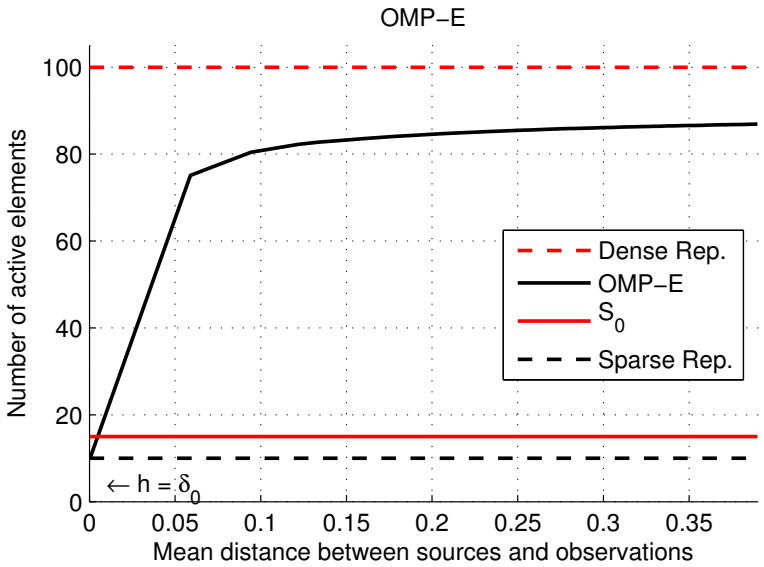


Figure 2: OMP-E results on convolved signals (averaged values over 100 trials of the experiment). The black and red dashed lines represent the true number of active atoms used to generate the test signals and the number of active atoms of a completely dense representation respectively. S_0 is the constraint parameter used for OMP-S

to represent the observed signals 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 described in Section 2.2 demonstrate that a sparse representation algorithm fails whenever applied to a convolved observation of sparse signals. This fact comes as no surprise if we consider the model presented in section 2.1: the observed variables are no longer sparse in the original dictionary $\boldsymbol{\Phi}$, but in the convolved dictionary $\boldsymbol{\Psi} = \mathbf{H}\boldsymbol{\Phi}$. Since the impulse response of the system is unknown, we can develop a dictionary learning algorithm in order to tackle the representation problem. The general strategy employed in the optimisation is analogous to the one adopted for the *method of optimal directions* (MOD) presented in [6].

3.1 Dictionary Learning in the Fourier Domain

Let us first define a cost function $\mathcal{J}(\mathbf{h}, \mathbf{X})$ that represents the total error of the representation:

$$\mathcal{J}(\mathbf{h}, \mathbf{X}) = \frac{1}{2}||\mathbf{Y} - \mathbf{H}\boldsymbol{\Phi}\mathbf{X}||_{\mathrm{F}}^2 \tag{1}$$

We can express (1) in the frequency domain by multiplying both the observations matrix $\mathbf{Y}$ and the estimated signals $\mathbf{H}\boldsymbol{\Phi}\mathbf{X}$ by the DFT matrix $\mathbf{F}^{\mathcal{H}}$. Here the superscript $(\cdot)^{\mathcal{H}}$ indicates the conjugate transpose or hermitian of the matrix $\mathbf{F}$ whose columns contain the Fourier basis.

Being an orthonotmal transform, the Fourier operator does not modify the magnitude of the residual expressed in equation (1) and leads to an equivalent cost function in the Fourier domain[3]:

$$\hat{\mathcal{J}}(\hat{\mathbf{h}}, \mathbf{X}) = \frac{1}{2}||\mathbf{F}^{\mathcal{H}}\mathbf{Y} - \mathbf{F}^{\mathcal{H}}\mathbf{H}\boldsymbol{\Phi}\mathbf{A}||_{\mathrm{F}}^2$$

where the hat indicates that the function $\mathcal{J}$ is evaluated in the frequency domain and will be also employed from now on to identify variables in the Fourier domain, such as $\hat{\mathbf{Y}} \triangleq \mathbf{F}^{\mathcal{H}}\mathbf{Y}$.

Since convolution in the time domain corresponds to an element-wise complex multipli-

[3]here we used the normalised version of the DFT, such that $\mathbf{F}^{\mathcal{H}}\mathbf{F} = \mathbf{I}$.

cation in the frequency domain, we can express the above cost function as

$$\hat{\mathcal{J}}(\hat{\mathbf{h}}, \mathbf{X}) = \frac{1}{2}||\hat{\mathbf{Y}} - \mathcal{D}(\hat{\mathbf{h}})\hat{\mathbf{\Phi}}\mathbf{X}||_{\mathrm{F}}^2 \tag{2}$$

where the operator $\mathcal{D}(\cdot)$ returns a diagonal matrix whose entries are the elements of its argument.

Since the equivalence of equations (1) and (2) only holds for circular convolution, the dictionary $\mathbf{\Phi}$ must be appropriately zero-padded before being transformed in the Fourier domain.

3.1.1 Zero Padding

The linear convolution occurring in the SISO system can be expressed as the circular convolution of the variables $\tilde{\mathbf{S}} = \tilde{\mathbf{\Phi}}\mathbf{A}$ and $\tilde{\mathbf{h}}$, where:

$$\tilde{\mathbf{\Phi}} \triangleq \begin{bmatrix} \mathbf{\Phi} \\ \mathbf{0} \end{bmatrix} \qquad \tilde{\mathbf{h}} \triangleq \begin{bmatrix} \mathbf{h} \\ \mathbf{0} \end{bmatrix}$$

The length of the resulting signals is equal to $D+L-1$, with D being the length of the source signals and L the length of the impulse response. From now on, we will omit for clarity of notation the symbols $\tilde{(\cdot)}$ and assume the variables to be zero padded as a pre-processing stage whenever the estimation occurs in the Fourier domain.

Since we assume the impulse response to be real, its estimation in the frequency domain can be performed only on the first $J \triangleq \lceil (D+L)/2 \rceil$ coefficients, setting the remainders as complex conjugate. The optimisation problem in the frequency domain defined by the cost function (2) becomes:

$$\underset{\hat{\mathbf{h}}^{1:J}, \mathbf{X}}{\text{minimise}} \quad \frac{1}{2}||\hat{\mathbf{Y}}^{1:J} - \mathcal{D}(\hat{\mathbf{h}}^{1:J})\hat{\mathbf{\Phi}}^{1:J}\mathbf{X}||_{\mathrm{F}}^2$$

where the superscripts indicate that only the first J rows of the various variables are taken into account. Once again, from now on we will omit this notation, implicitly assuming that the estimation only considers the first J Fourier coefficients when performed in the frequency domain.

3.2 Block Coordinate Descent Optimisation

The joined minimisation of the cost function (2) over the variables $\hat{\mathbf{h}}$ and $\mathbf{X}$ is an underdetermined problem in that the number of variables is $NK+J$ and the number of observations is JN, with $K \geq D > J$ if we consider an (over)complete dictionary. Moreover, it is easy to show that the problem is not convex if . Therefore, we chose to employ a block coordinate descent strategy (BCD) where the two variables are updated one at a time and the optimisation of one is based on the previous value of the other. This leads to two distinct problems which, as described in the next sections, can be solved reliably using a variety of methods.

3.2.1 Source Signals Optimisation

Given a fixed impulse response $\mathbf{h}^{(0)}$, the cost function (1) (or, its equivalent in the frequency domain (2)) represents a classic sparse representation problem where we seek the matrix $\mathbf{X}$ which minimises the residual norm of the representation over the dictionary $\mathbf{\Psi} = \mathbf{H}^{(0)}\mathbf{\Phi}$, given a constraint on the sparsity of the solution vectors.

$$\begin{aligned} \underset{\mathbf{X}}{\text{minimise}} \quad & ||\mathbf{Y} - \mathbf{\Psi}\mathbf{X}||_{\text{F}}^2 \\ \text{subject to} \quad & ||\mathbf{x}_n||_0 \leq S_0 \quad \forall\, n = 1 \dots N \end{aligned} \tag{OMP-S}$$

This can be tackled using a sub-optimal greedy algorithm such as OMP-S. Alternatively, the sparsity assumption can be relaxed to an ℓ_1 constraint $||\mathbf{x}_n||_1 \leq S_1$, which leads to a convex problem that can be solved with various methods, such as basis pursuit [5] or homotopy [11].

3.2.2 Impulse Response Optimisation

The second step of the block coordinate descent is the optimisation of the impulse response given the current representation matrix $\mathbf{X}$. In this section, we will show how this problem can be turned into a quadratic program.

The cost function defined in equation (2) can be written as:

$$\hat{\mathcal{J}}(\hat{\mathbf{h}}) = \frac{1}{2} \operatorname{Tr}\left[(\hat{\mathbf{Y}} - \mathcal{D}\left(\hat{\mathbf{h}}\right)\hat{\mathbf{\Phi}}\mathbf{X})^{\mathcal{H}}(\hat{\mathbf{Y}} - \mathcal{D}\left(\hat{\mathbf{h}}\right)\hat{\mathbf{\Phi}}\mathbf{X})\right]$$

Omitting the terms that do not depend on $\hat{\mathbf{h}}$, the resulting optimisation problem becomes:

$$\underset{\hat{\mathbf{h}}}{\text{minimise}} \quad \frac{1}{2}\Big[\operatorname{Tr}(\hat{\mathbf{S}}^{\mathcal{H}}\,\mathcal{D}\,(|\hat{\mathbf{h}}|^2)\hat{\mathbf{S}}) - \operatorname{Tr}(\hat{\mathbf{Y}}^{\mathcal{H}}\,\mathcal{D}\,(\hat{\mathbf{h}})\hat{\mathbf{S}}) - \operatorname{Tr}(\hat{\mathbf{S}}^{\mathcal{H}}\,\mathcal{D}\,(\hat{\mathbf{h}}^{*})\hat{\mathbf{Y}})\Big]$$

where the operator $(\cdot)^*$ indicates complex conjugate and the matrix $\hat{\mathbf{S}}$ contains the Fourier transform of the estimated source signals $\mathbf{S} = \mathbf{\Phi X}$.

We can simplify this expression considering a property of diagonal matrices. Let $\mathbf{D}$ be a diagonal matrix, then we have that

$$\operatorname{Tr}(\mathbf{A}^{\mathcal{H}}\mathbf{D}\mathbf{B}) = \mathrm{d}\,(\mathbf{D})^{\mathcal{T}}\,\mathrm{d}\,(\mathbf{B}\mathbf{A}^{\mathcal{H}})$$

where $\mathbf{A}$, $\mathbf{B}$ are arbitrary matrices and the operator $\mathrm{d}\,(\cdot)$ returns a vector whose elements are the diagonal entries of its argument.
Therefore, the unconstrained minimisation of the frequency response $\hat{\mathbf{h}}$ becomes:

$$\underset{\hat{\mathbf{h}}}{\text{minimise}} \quad \frac{1}{2}\Big[\hat{\mathbf{h}}^{\mathcal{H}}\,\mathcal{D}\,(\hat{\mathbf{h}})\,\mathrm{d}\,(\hat{\mathbf{S}}\hat{\mathbf{S}}^{\mathcal{H}}) - \hat{\mathbf{h}}^{\mathcal{T}}\,\mathrm{d}\,(\hat{\mathbf{S}}\hat{\mathbf{Y}}^{\mathcal{H}}) - \hat{\mathbf{h}}^{\mathcal{H}}\,\mathrm{d}\,(\hat{\mathbf{Y}}\hat{\mathbf{S}}^{\mathcal{H}})\Big]$$

Let's introduce for clarity of notation the vectors

$$\boldsymbol{\alpha} \triangleq \mathrm{d}\,(\hat{\mathbf{S}}\hat{\mathbf{S}}^{\mathcal{H}})$$
$$\boldsymbol{\beta} \triangleq \mathrm{d}\,(\hat{\mathbf{Y}}\hat{\mathbf{S}}^{\mathcal{H}})$$

Applying some linear algebra, we can write the optimisation problem as:

$$\underset{\hat{\mathbf{h}}}{\text{minimise}} \quad \frac{1}{2}\left[\hat{\mathbf{h}}^{\mathcal{H}}\,\mathcal{D}\,(\boldsymbol{\alpha})\hat{\mathbf{h}} - \hat{\mathbf{h}}^{\mathcal{T}}\boldsymbol{\beta}^{*} - \hat{\mathbf{h}}^{\mathcal{H}}\boldsymbol{\beta}\right]$$

This is a quadratic program on the complex variable $\hat{\mathbf{h}}$ which can be solved by setting the gradient of the objective function to zero, resulting in the solution:

$$\hat{\mathbf{h}}_{\text{opt}} = \mathcal{D}\,(\boldsymbol{\alpha}^{-1})\boldsymbol{\beta}$$

By considering the impulse response optimisation in the frequency domain, we could turn the initial Frobenious norm minimisation (1) into a simple least-squares problem involving the vectors $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$. Therefore, minimising it with respect to the impulse response can be

explicitly written as a least-squares problem in the time domain as:

$$\underset{\mathbf{h}}{\text{minimise}} \quad \frac{1}{2}|| \mathcal{D}\left(\boldsymbol{\alpha}^{1/2}\right)\mathbf{F}^{\mathcal{H}}\mathbf{h} - \mathcal{D}\left(\boldsymbol{\alpha}^{-1/2}\right)\boldsymbol{\beta}||_2^2 \tag{3}$$

3.2.3 Handling Complex Values

Since many quadratic optimisation solvers and sparse representation algorithms implementations do not handle complex values, the optimisation of (3) must be performed considering separately the real and imaginary parts of the variables. Moreover, if we assume that the impulse response has length $L < D$, then we can reduce the dimension of the problem.
Let's define for clarity of notation the variables

$$\begin{aligned} \boldsymbol{\Gamma} &\triangleq \mathcal{D}\left(\boldsymbol{\alpha}^{1/2}\right)\mathbf{F}^{\mathcal{H}} \\ \boldsymbol{\xi} &\triangleq \mathcal{D}\left(\boldsymbol{\alpha}^{-1/2}\right)\boldsymbol{\beta} \end{aligned}$$

The optimisation (3) can be written as:

$$\underset{\mathbf{h}}{\text{minimise}} \quad \frac{1}{2}\left|\left|\begin{bmatrix} \mathcal{R}(\boldsymbol{\Gamma}_{1:L}) \\ \mathcal{I}(\boldsymbol{\Gamma}_{1:L}) \end{bmatrix}\mathbf{h} - \begin{bmatrix} \mathcal{R}(\boldsymbol{\xi}) \\ \mathcal{I}(\boldsymbol{\xi}) \end{bmatrix}\right|\right|_2^2 \tag{4}$$

where $\mathcal{R}(\cdot)$ and $\mathcal{I}(\cdot)$ return the real and imaginary parts of their arguments respectivelty.

3.2.4 Constrained Optimization

At this point, we can introduce some constraints to the impulse response optimization. Firstly, we can assume that the vector $\mathbf{h}$ is non negative, turning the minimisation (4) into a non negative least squares problem, which can be solved via quadratic programming. We can then constraint the ℓ_1 norm of the solution to be smaller than a fixed value Q_1, inducing sparsity on the impulse response coefficients.
This can be done in a very simple way considering that the ℓ_1 norm of a nonnegative vector

is the sum of its entries. The optimisation problem becomes:

$$\begin{aligned} \underset{\mathbf{h}}{\text{minimise}} \quad & \frac{1}{2}\left|\left|\begin{bmatrix} \mathcal{R}(\boldsymbol{\Gamma}_{1:L}) \\ \mathcal{I}(\boldsymbol{\Gamma}_{1:L}) \end{bmatrix}\mathbf{h} - \begin{bmatrix} \mathcal{R}(\boldsymbol{\xi}) \\ \mathcal{I}(\boldsymbol{\xi}) \end{bmatrix}\right|\right|_2^2 \\ \text{subject to} \quad & \mathbf{h} \geq \mathbf{0} \\ & \mathbf{1}^T\mathbf{h} \leq Q_1 \end{aligned} \tag{5}$$

The choice of sparsity and non-negativity constraints on the impulse response is particularly suited for audio signals in that the early reflections coming from the surfaces of the ambient in which the signals are recorded can be modelled as a sparse non-negative impulse response [3]. The labels D**h**-BCD and S**h**-BCD are the acronyms of *sparse* ***h*** *block coordinate descent* and *dense* ***h*** *block coordinate descent* and will be employed from now on to identify the two versions of the algorithm.

3.2.5 Ambiguities

The cost function (1) contains a norm ambiguity because it is possible to multiply the matrices $\mathbf{H}$ and $\mathbf{X}$ by an arbitrary scalar and its inverse resulting in the same function value. For this reason we need to introduce a constraint on the ℓ_2 norm of the impulse response. In particular, we choose not to modify the initial Frobenious norm of the dictionary:

$$||\mathbf{H}\boldsymbol{\Phi}||_{\mathrm{F}}^2 = ||\boldsymbol{\Phi}||_{\mathrm{F}}^2 = K \tag{6}$$

where we assumed a normalised dictionary $\boldsymbol{\Phi}$ and K is the number of atoms.
Therefore, once the impulse response has been optimised solving the unconstrained problem (4) or the constrained problem (5), we redistribute the energy between the matrix $\mathbf{H}$ and the coefficients $\mathbf{X}$ so that the equality (6) is satisfied.

$$\begin{aligned} R = \quad & \frac{\sqrt{K}}{||\mathbf{H}\boldsymbol{\Phi}||_{\mathrm{F}}} \\ \mathbf{h} \leftarrow \quad & R\mathbf{h} \\ \mathbf{X} \leftarrow \quad & \frac{1}{R}\mathbf{X} \end{aligned}$$

Note that the diagonal and permutation ambiguity present in other problems such as independent component analysis [9] does not occur in this case.

3.3 Numerical Evaluation

In this section we will describe several numerical test performed on synthetic data in order to evaluate the proposed block coordinate descent optimisation and to compare it with the K-SVD dictionary learning algorithm [1]. The source signals were generated according to the model described in section 2.1 using the parameters defined in section 2.2 and were convolved with a sparse non-negative impulse response with normalised diversity $||\mathbf{h}||_0/L = 0.05$.

3.3.1 Sparse vs dense impulse response estimation

While for the sources representation step we used the sparsity-constrained version of OMP (with the sparsity parameter S_0 allowing for a 50% tolerance on the number of active elements), for the impulse response estimation step we may or may not introduce additional constraints, which leads to:

1. S**h**-BCD: at each step of the algorithm, the impulse response $\mathbf{h}$ is updated solving the optimisation problem (5) using a nonnegative version of the LASSO algorithm[4], which constrains the solution to be sparse and nonnegative (again, we set the sparsity constraint to allow for a 50% tolerance on the number of active elements).
2. D**h**-BCD: at each step of the algorithm, the impulse response $\mathbf{h}$ is updated by solving the optimisation problem (4). Note that this is an overdetermined problem whose solution can be derived analytically and corresponds to the least-squares solution of the system.

The two methods, along with the K-SVD dictionary learning algorithm[5] were initialised with the dictionary $\mathbf{\Phi}$ and a random initial impulse response $\mathbf{h}^{(0)}$ and ran for 50 iterations.

Figure 3 shows the mean distance $\bar{d}$ of the representations defined as the Frobenious norm of the residual divided by the number of source signals:

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

the figures are shown on a logarithmic scale and averaged over 100 trials of the experiment.

[4] we used the `SolveLasso` function contained in the SPARSELAB toolbox.
[5] we used the `ksvd` function available as part of the SMALLBOX toolbox

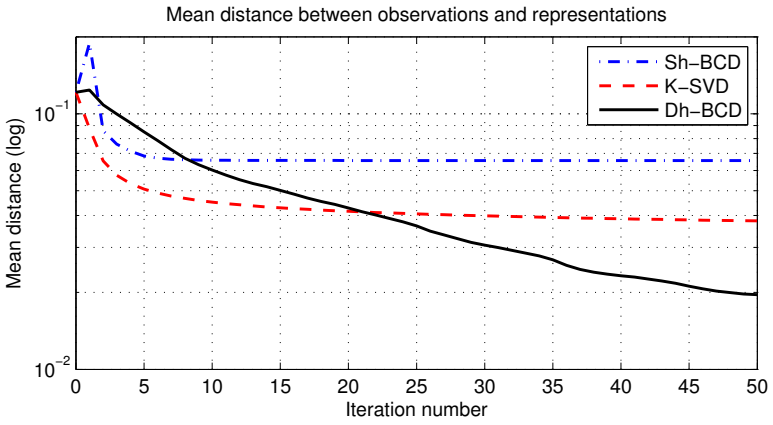


Figure 3: BCD and K-SVD for dictionary learning of convolved signals (average over 100 trials of the experiment)

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 performing worse.

Figure 4 offers a more precise comparison between K-SVD and D**h**-BCD by showing the boxplot of the mean distance d(h, X—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 consistently 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.

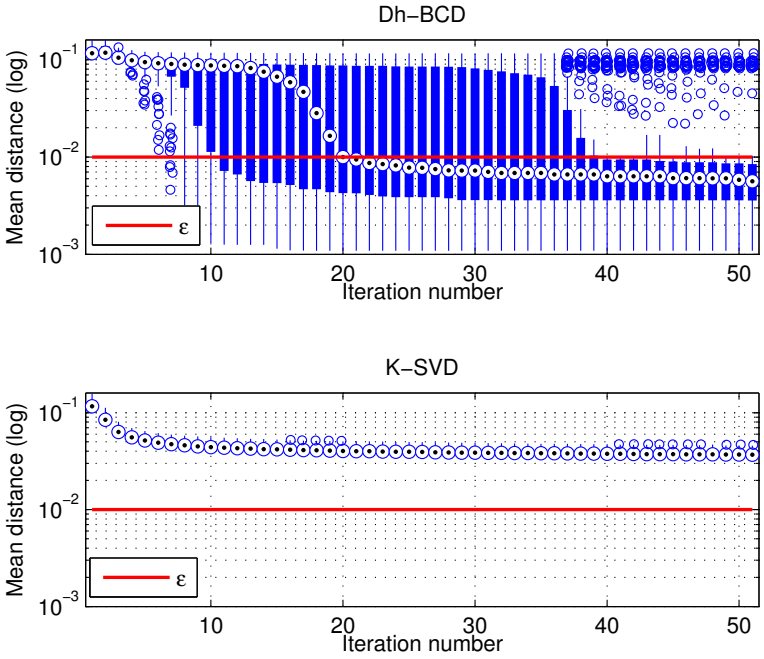


Figure 4: Boxplot comparison of D**h**-BCD and K-SVD over 100 trials of the experiment. For each iteration, the central mark is the median, the edges of the boxes are the 25th and 75th percentiles, the whiskers extend to the most extreme data points not considered outliers, and outliers are plotted individually.

3.3.2 Sparsity Phase-Transition

The block coordinate descent strategy described in section 3.2 proved to be effective in learning a dictionary for a problem with sparse source signals and a sparse non-negative impulse response. In particular, the version employing an unconstrained estimation of the impulse response seems not to get trapped in a local minima of the objective function and is able to provide a representation of the observed signals with small residual error. However, as in every sparse representation problem, the sparsity of the input data plays a crucial role. For this reason, we tested the algorithms varying the normalised 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 show 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.

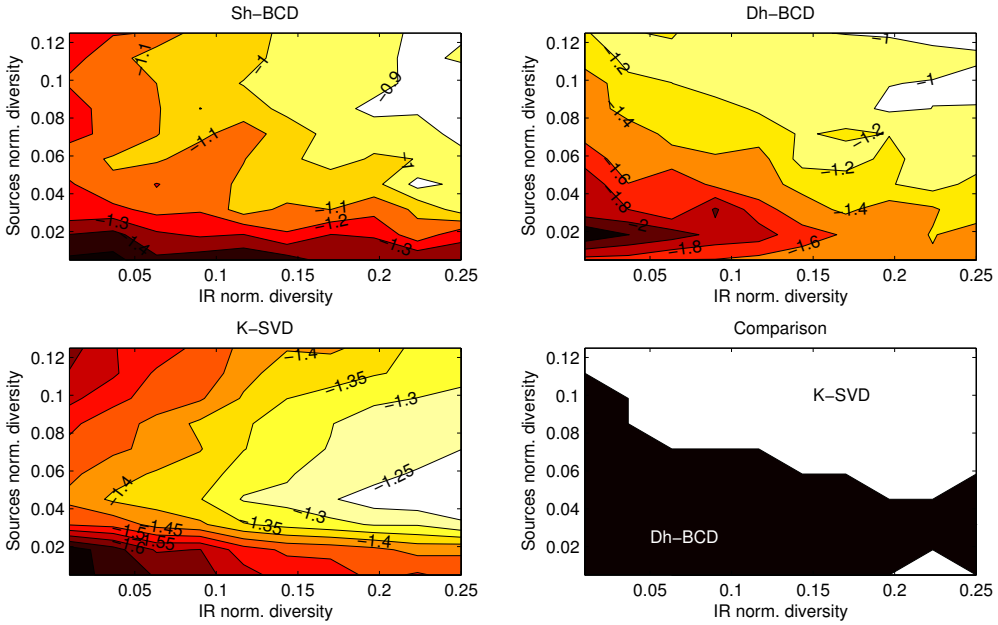


Figure 5: BCD and K-SVD results for various densities of source signals and impulse response (average over 20 trials). The values appearing along the contour plots represent the mean distance $\bar{d}(\mathbf{h}, \mathbf{X}|\mathbf{Y})$ achieved after 50 iterations by the various algorithms on a logarithmic scale. Since the dictionary is two times over-complete, a completely dense representation corresponds to a source normalised diversity $||\mathbf{x}||_0/K = 0.5$

4 Dictionary Identification and Blind Deconvolution

The promising results shown in the previous section suggest that dictionary learning of convolved signals could be used to tackle a more ambitious problem, such as the blind deconvolution of the observed variables.

The main idea is as follows: if dictionary learning is able to provide a representation of the matrix $\mathbf{Y}$ with small or negligible residual norm, then this may occur because it identifies the dictionary $\mathbf{\Psi} = \mathbf{H\Phi}$ that generated the observed data, retrieving therefore the impulse response $\mathbf{h}$ of the system. In this section we will highlight the difference between dictionary learning and dictionary identification and we will analyse the results of the experiments presented in sections 3.3.1 and 3.3.2 on the light of the impulse response identification.

4.1 Dictionary Learning and Dictionary Identification

As for a standard sparse representation problem, the dictionary learning optimisation can be expressed in different ways by performing constrained or unconstrained optimisations and by considering the ℓ_0 pseudo-norm or by relaxing it to the ℓ_1 norm of the sparse representation coefficients. One valid formulation of the problem is the ℓ_0 sparsity constrained version that we have considered throughout this technical report:

$$\begin{aligned} \underset{\mathbf{\Phi},\mathbf{X}}{\text{minimise}} \quad & \frac{1}{2}||\mathbf{Y} - \mathbf{\Phi X}||_{\text{F}}^2 \\ \text{subject to} \quad & ||\mathbf{x}_n||_0 \leq S_0 \quad \forall\, n = 1 \dots N \end{aligned} \tag{7}$$

This somewhat naive formulation is often complemented by a constraint on the norm of the dictionary atoms in order to avoid a scaling ambiguity. There is also an inherent permutation ambiguity due to the fact that, given an arbitrary permutation matrix $\mathbf{P}$, we have:

$$\mathbf{\Phi P P}^{-1}\mathbf{X} = \mathbf{\Phi X}$$

It is important to stress that the problem (7) is highly underdetermined: the number of available data is $\mathcal{O}(ND)$, while the number of unknowns is $\mathcal{O}((D+N)K)$ with $K \geq D$ and non convex[6].

[6] it is still non convex even if we relax the constraints to the ℓ_1 norm of the coefficient vectors.

The goal of dictionary identification is even more ambitious and can be expressed as follows: suppose that the observed data $\mathbf{Y}$ were generated from a dictionary $\mathbf{\Phi}^\star$ and a matrix of coefficients $\mathbf{X}^\star$. Then a dictionary learning algorithm which attempts to solve the optimisation (7) should converge to the matrices $(\mathbf{\Phi}^\star, \mathbf{X}^\star)$, up to any ambiguity inherent to the optimisation. Theoretical warranties about the success of dictionary identification have been studied in [2, 8]. However, research in this area is still in its infancy and additional effort is required in order to understand this topic in depth. In this work, we will not attempt a convergence analysis of the proposed algorithm, but we will rather rely on empirical results. By learning a structured dictionary and, in particular, by optimising the representation coefficients contained in the matrix $\mathbf{X}$ and the impulse response $\mathbf{h}$, we can obtain a more tractable cost function described by equation (1) and avoid the mentioned ambiguities, as described in section 3.2.5. The resulting problem:

$$\begin{aligned} \underset{\mathbf{h},\mathbf{X}}{\text{minimise}} \quad & \frac{1}{2}||\mathbf{Y} - \mathbf{H}\mathbf{\Phi}\mathbf{X}||_{\text{F}}^2 \\ \text{subject to} \quad & ||\mathbf{x}_n||_0 \leq S_0 \quad \forall\, n = 1 \dots N \end{aligned}$$

is still underdetermined and non convex. In the next section we will analyse the results of the numerical tests described in section 3.3 in order to assess whether the proposed block coordinate descent algorithm converges to the original impulse response that generated the test data.

4.2 Numerical Results

The upper plot in Figure 6 shows the log SNR between the estimated impulse response $\mathbf{h}$ resulting from S$\mathbf{h}$-BCD and D$\mathbf{h}$-BCD and the true impulse response $\mathbf{h}^\star$ that was used to generate the test data.

$$\text{SNR}(\mathbf{h}|\mathbf{h}^\star) = 10 \log \frac{||\mathbf{h}^\star||_2}{\min\limits_{\alpha} ||\mathbf{h}^\star - \alpha\mathbf{h}||_2}$$

the minimisation appearing at the denominator is introduced in order to take into account the scaling ambiguity present in the optimisation and results in considering the projection of the true impulse response $\mathbf{h}^\star$ along the direction determined by the estimated impulse response $\mathbf{h}$. The figures are obtained from the numerical test described in section 3.3.1 and are displayed as a boxplot which depicts the distribution of the SNR over 100 trials of the experiment. The constrained version of the BCD algorithm fails to retrieve the correct

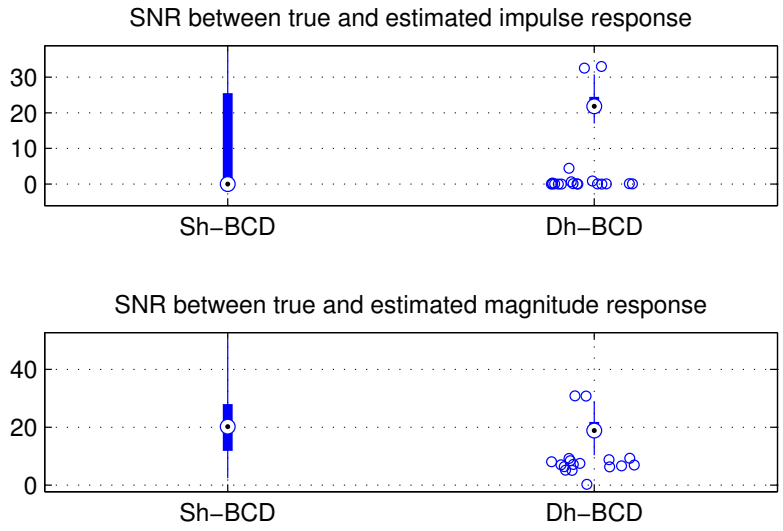


Figure 6: Impulse response and magnitude response identification

impulse response most of the times, as revealed by the median of the SNR. On the other hand, unconstrained BCD consistently achieves a SNR between 20dB and 25dB with only a few outliers. The results are different if we consider the SNR between the magnitudes of the Fourier transforms of the true and estimated impulse responses. In this case we are discarding any error related to the phase of the variables and find that constrained and unconstrained optimisation achieve similar results, with the latter being slightly more consistent over the trials of the experiment. This result suggests that, if we allow for a shifting ambiguity in the identification of the impulse response, both algorithms are successful in retrieving the true response $\mathbf{h}^\star$.

In order to better visualise what discussed so far, figure 7 shows the true and estimated impulse and magnitude responses for one of the trials of the experiment. The shift error can be easily seen in the upper plot, while the two methods lead to undistinguishable results in the lower graph where, if we consider the convolution of audio signals, the well-known comb filtering pattern is correctly identified by both S**h**-BCD and D**h**-BCD.

At this point, we can analyse the results of the sparsity phase-transition described in Section 3.3.2 on the light of the impulse and magnitude response identification. Figure 8 shows a surf plot of the RMS values achieved by the algorithms at the end of the optimisation. The results for the IR identification by the S**h**-BCD do not show any particular trend, and this is due to the fact that this method is basically unreliable. The identification is only

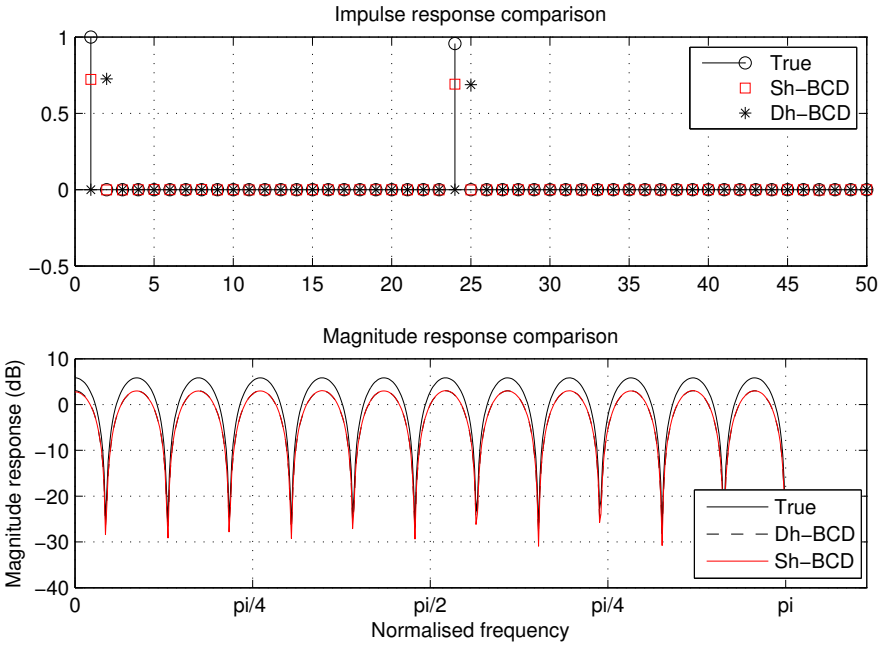


Figure 7: Impulse response and magnitude response comparison

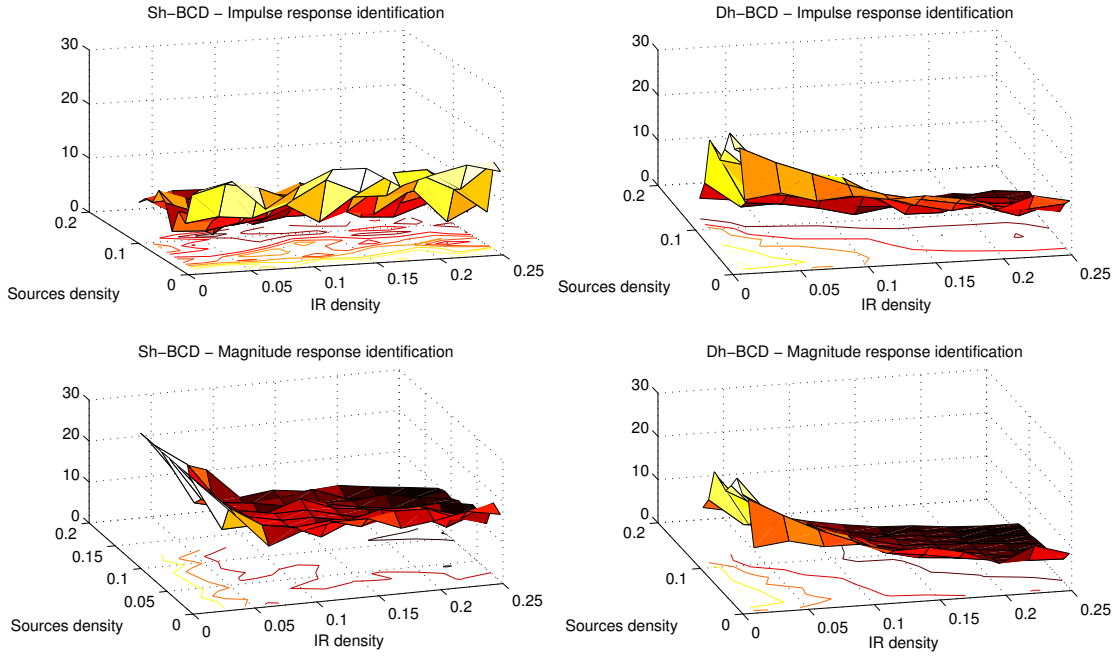


Figure 8: Impulse response and magnitude response identification as a function of source signals and impulse response normalised diversity.

sometimes correct and plotting the mean RMS over 20 trials of the algorithm results in a random pattern. On the other end, the magnitude response identification seems to be more accurate. In particular, we can observe a peak in correspondence with the trivial case $\mathbf{h} = \boldsymbol{\delta}_0$, and a drop in the identification performance as the normalised diversity of sources and impulse response increases.

A similar trend can be observed for both the IR and magnitude identification obtained by D$\mathbf{h}$-BCD. In this case the surface appears to be smoother, with a good SNR for sparse problems and worse results as the normalised diversity of the variables increases. The contour plots depicted at the bottom of the graphs reveal a similar pattern to the ones observed for the dictionary learning objective, which suggests that, whenever the algorithm is able to provide a sparse representation of the observed variables, that indeed occurs because it identified the correct impulse response.

5 Conclusions and Further Research

In this report, we presented a dictionary learning method suitable for the sparse representation of convolved signals using a block coordinate descent strategy. We assessed through numerical experiments that, providing some conditions on the sparsity of the problem data, the proposed method leads to improvements over K-SVD.

We then briefly described the difference between dictionary learning and dictionary identification, suggesting a link between the latter problem and blind deconvolution. By analysing the results of our numerical experiments in the light of this consideration, we assessed an empirical link between the success of our technique for the dictionary learning problem and the identification of the impulse response or the magnitude response of the system that generated the problem data. This promising result needs to be further investigated in the future by performing additional tests with real world data and by gaining a more deep understanding of the convergence properties of our algorithm. Moreover, the main idea that inspired this work can be extended to other cases. For example, the *Double Sparsity* dictionary learning algorithm presented in [15] can be viewed as an instantaneous blind source separation problem. The link between learning structured dictionary and blind system identification will be the long term goal of our research.

6 Reproducible Research

Over recent years the signal processing community has become more and more aware of the importance of making research reproducible [16]. This is a crucial step for increasing the reliability of the results presented in research papers and a great means for disseminate new ideas in a way that allows other researchers to further develop on them.

The MATLAB code used to generate the figures presented in this report can be downloaded from http://www.eecs.qmul.ac.uk/ danieleb/, along with a brief documentation.

References

[1] M. Aharon, M. Elad, and A. Bruckstein. K-SVD: An algorithm for designing overcomplete dictionaries for sparse representation. *IEEE Trans. on Signal Processing*, 54(11):4311–4322, Nov. 2006.

[2] M. Aharon, M. Elad, and A. M. Bruckstein. On the uniqueness of overcomplete dictionaries, and a practical way to retrieve them. *Linear Algebra and its Applications*, 416(1):48–67, Jul. 2006.

[3] J. B. Allen and D. A. Berkley. Image method for efficiently simulating small-room acoustics. *Journal of the Acoustical Society of America*, 65(4):943–950, Apr. 1979.

[4] R. G. Baraniuk, E. Candès, M. Elad, and Y. Ma. Applications of sparse representation and compressive sensing. *Proceedings of the IEEE*, 98(6):906–909, Jun. 2010.

[5] S. S. Chen, D. L. Donoho, and M. A. Sauders. Atomic decomposition by basis pursuit. *SIAM Review*, 43(1):129–159, Mar. 2001.

[6] K. Engan, S. O. Aase, and J. H. Husøy. Method of optimal directions for frame design. In *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, volume 5, pages 2443–2446, 1999.

[7] J. M. Fadili, J.-L. Starck, J. Bobin, and Y. Moudden. Image decomposition and separation using sparse representations: An overview. *Proceedings of the IEEE*, 98(6):983–994, Jun. 2010.

[8] R. Gribonval and K. Schnass. Dictionary identification: Sparse matrix-factorisation via ℓ_1-minimisation. *IEEE Trans. on Information Theory*, 56(7):3523–3539, Jul. 2010.

[9] A. Hyvärinen, J. Karhunen, and E. Oja. *Independent Component Analysis*. John Wiley & Sons, New York, NY, USA, 2001.

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

[11] M. R. Osborne, B. Presnell, and B. A. Turlach. A new approach to variable selection in least squares problems. *IMA Journal of Numerical Analysis*, 20(3):389–404, Jul. 2000.

[12] Y. Pati, R. Rezaiifar, and P. Krishnaprasad. Orthogonal matching pursuit: Recursive function approximation with applications to wavelet decomposition. In *Proceedings of the 27th Asilomar Confrence on Signals, Systems and Computers*, volume 1, pages 40–44, Nov. 1993.

[13] M. D. Plumbley, T. Blumensath, L. Daudet, R. Gribonval, and M. Davies. Sparse representations in audio and music: From coding to source separation. *Proceedings of the IEEE*, 98(6):995–1005, Jun. 2010.

[14] R. Rubinstein, A. Bruckstein, and M. Elad. Dictionaries for sparse representation modeling. *Proceedings of the IEEE*, 98(6):1045–1057, Jun. 2010.

[15] R. Rubinstein, M. Zibulevsky, and M. Elad. Double sparsity: Learning sparse dictionaries for sparse signal approximation. *IEEE Trans. on Signal Processing*, 58(3):1553–1564, Mar. 2010.

[16] P. Vanderwalle, J. Kovačević, and M. Vetterli. Reproducible research in signal processing. *IEEE Signal Processing Magazine*, 26(3):37–47, May 2009.