Cyclosparsity: A New Concept for Sparse Deconvolution

By Khalid Sabri

University of Chouaib Doukkali, Morroco

Abstract- Periodic random impulse signals are appropriate tools for several situations of interest and are a natural way for modeling highly localized events occuring randomly at given times. Nevertheless, the impulses are generally hidden and swallowed up in noise because of unwanted convolution. Thus, the resulting signal is not legible and may lead to erroneaous analysis, and hence, the need of deconvolution to restore the random periodic impulses. The main purpose of this study is to introduce the concept of cyclic sparsity or cyclosparsity in deconvolution framework for signals that are jointly sparse and cyclostationary like periodic random impulses. Indeed, all related works in this area exploit only one property, either sparsity or cyclostationarity and never both properties together. Although, the key feature of the cyclosparsity concept is that it gathers both properties to better characterize this kind of signals. We show that deconvolution based on cyclic sparsity hypothesis increases the performances and reduces significantly the computation cost as well. Finally, we use computer simulations to investigate the behavior in deconvolution framework of the algorithms Matching Pursuit (MP) [13], Orthogonal Matching Pursuit (OMP) [14], Orthogonal Least Square (OLS) [15], Single Best Replacement (SBR), [19, 20, 21] and the proposed extensions to cyclic sparsity context: Cyclo-MP, Cyclo-OMP, Cyclo-OLS and Cyclo-SBR.

Keywords: cyclosparsity, sparsity, cyclostationary, periodic random impulses, deconvolution, greedy.

GJCST-F Classification : G.1.3

CYCLOSPARSITYANEWCONCEPTFORSPARSEDECONVOLUTION

Strictly as per the compliance and regulations of:

Cyclosparsity: A New Concept for Sparse Deconvolution

Khalid Sabri

Abstract- Periodic random impulse signals are appropriate tools for several situations of interest and are a natural way for modeling highly localized events occurring randomly at given times. Nevertheless, the impulses are generally hidden and swallowed up in noise because of unwanted convolution. Thus, the resulting signal is not legible and may lead to erroneous analysis, and hence, the need of deconvolution to restore the random periodic impulses. The main purpose of this study is to introduce the concept of cyclic sparsity or cyclosparsity in deconvolution framework for signals that are jointly sparse and cyclostationary like periodic random impulses. Indeed, all related works in this area exploit only one property, either sparsity or cyclostationarity and never both properties together. Although, the key feature of the cyclosparsity concept is that it gathers both properties to better characterize this kind of signals. We show that deconvolution based on cyclic sparsity hypothesis increases the performances and reduces significantly the computation cost as well. Finally, we use computer simulations to investigate the behavior in deconvolution framework of the algorithms Matching Pursuit (MP) [13], Orthogonal Matching Pursuit (OMP) [14], Orthogonal Least Square (OLS) [15], Single Best Replacement (SBR), [19, 20, 21] and the proposed extensions to cyclic sparsity context: Cyclo-MP, Cyclo-OMP, Cyclo-OLS and Cyclo-SBR.

Keywords: *cyclosparsity, sparsity, cyclostationary, periodic random impulses, deconvolution, greedy.*

I. INTRODUCTION

Cyclostationarity is very useful tool for studying periodically correlated signals by means of cyclic statistics [1].

As cyclostationary signals often encountered in practice, cyclostationary modeling being used in various domains such as mechanics [2], telecommunications [3], biomechanics [5] and allows good performances.

In this study we focus on particular case of cyclostationary signals which are wide sense second order cyclostationary i.e. first order and second order cyclostationary. Also, the signals are assumed to be made of periodic impulses with random amplitudes namely few nonzero impulses per period. Given the Impulse Response (IR), the aim is to retrieve the original object which has been distorted by passage through a known linear and time-invariant system in presence of noise. Indeed, enhancing the resolution of the signal and the Signal to Noise Ratio (SNR) from the knowledge of the IR corresponds to a deconvolution problem.

The deconvolution of cyclostationary signals has been addressed by several authors with different approaches. In [6], a bayesian deconvolution algorithm based on Markov chain Monte Carlo is presented. Cyclic statistics are often used for deconvolution, in [7, 8] the deconvolution is based on cyclic cepstrum, whereas, in [9, 10] the deconvolution is based on cyclic correlation. The main drawback of these methods is their inability to detect and restore impulses drowned in noise.

Actually, signal deconvolution belongs to inverse problems and is particularly well-known to be an ill-posed problem since the IR acts as a low-pass filter and the convolved signal is always noisy. Fortunately, regularization methods lead to acceptable solutions accounting for a priori information on the original object [38].

Analyzing periodic random impulse signals in details uncover another a priori information which is sparsity, this is because only few impulses are nonzero. Thus, data are sparse on the direct domain.

In this study, we will not exploit cyclic statistics but only the periodic character jointly with sparsity of periodic random impulses. This is possible thanks to the new concept of cyclosparsity that gathers both properties to better characterize these signals. Thus, cyclosparse deconvolution can be performed taking benefit of the correlation between the signal at different cycles (or periods).

Lately, in a different framework, sparse approximation of signals has drawn significant interest in many areas. The key idea is that a signal can be very well approximated with only few elementary signals (hereinafter referred to as atoms) taken from a redundant family (often referred to as dictionary), while its projection onto a basis of elementary signals may lead to a larger number of nonzero coefficients. Such a basic idea is the origin of recent theoretical development and many practical applications in denoising, compression, blind source separation and inverse problems [35, 11, 12].

Contrarily to orthogonal transforms, a redundant dictionary leads to non-unique representations of a given signal and several methods and algorithms have been developed to find the sparse approximations, i.e. the approximation with the smaller number of nonzero coefficients. In other words, minimizing the number of nonzero coefficients in a linear combination

Author : STIC Laboratory, Faculty of sciences, University Chouaïb Doukkali, El Jadida, Morocco; (email: sabri.k@ucd.ac.ma)

approximating the data leads to an exhaustive search which is a NP-hard problem. Various methods and algorithms have been proposed to attempt to solve this problem and some sufficient conditions for these algorithms to reach the sparse solution have been established. Algorithms can be roughly classified in two approaches: greedy pursuit algorithms and convex relaxation. Greedy pursuit algorithms iteratively improve the approximation selecting at each iteration an additional elementary signal and many algorithms have been proposed based on such a scheme [13, 14, 15].

The principle of convex relaxation methods is to replace the minimization of the number of elements by the minimization of another functional which can be minimized more easily and still guarantee the solution to have a large number of zero coefficients. A ℓ_1-norm is mainly used to this end [16, 17, 18].

Our work consists on extending sparse approximation to cyclostationary signals with periodic random impulses, where the aim is to find jointly a sparse approximation of each cyclic period (or cycle), accounting for the same elementary signals in each approximation, but shifted with a multiple of the cyclic period and with different coefficients. However, we insist on the application of cyclosparse approximation to deconvolution i.e. the dictionary is given by the Toeplitz matrix formed by the IR.

The paper is organized as follows; section 2 defines the problem statement and motivations of the study. The main contribution of this paper is described in sections 3, 4 and 5. The concept of cyclosparsity is introduced in section 3. In section 4, we summarize the statement of sparse and cyclosparse approximation problems. We also point-out the link with cyclosparse deconvolution and we insist on the differences with joint sparsity. At the end of this section, we came to the conclusion that none of the usual algorithms ensure to reach the actual solution or even to reach the same solution. Thus some numerical experiments have to be performed to compare these algorithms. Such a

cyclosparse model is taken into account in the sparse deconvolution by greedy algorithms in section 5, which fortunately reduces significantly the computation cost. In this paper, we focus on greedy algorithms namely Matching Pursuit (MP) [13], Orthogonal Matching Pursuit (OMP) [14], Orthogonal Least Square (OLS) [15] and Single Best Replacement (SBR), [19, 20, 21] (which, despite a different aim, has a structure very similar to greedy algorithms). And we propose to generalize these algorithms to the cyclosparse context: Cyclo-MP, Cyclo-OMP, Cyclo-OLS and Cyclo-SBR. We propose furthermore to test all the algorithms on the same statistical basis, i.e. with the same stopping rule deduced from statistical properties of the noise. Also it seems to be necessary to obtain satisfactory deconvolution results, as shown in the simulation results section 6.

II. Background

a) Problem Formulation

Consider the situation where a known system $\mathcal{H}(t)$ is excited by a cyclostationary signal $x(t)$ consisting of periodic random impulses. By periodic, we mean that the signal can be divided into portions of length T (which is known as the cyclic period of the signal) with d impulses in each portion. Moreover, the delay factor τ_i of the *ith* impulse x_i is constant for all portions. Note that in general τ_i will be different for different i although in most of the cases they may be integral multiples of a constant τ. An example of $x(t)$ is shown in Fig. 1 with $d = 5$.

Reconsider the system as described above. Since the impulses are periodic, we can consider a period of time T and write that portion of the output as $\sum_{k=1}^{d} x_k \, \mathcal{H}(t - \tau_k) + n(t), \; 0 \le t < T$. This relationship can be generalized to cover the whole signal as,

$$y(t) = \sum_{i=0}^{K-1} \sum_{k=1}^{d} x_{i,k} \, \mathcal{H}(t - (\tau_k + iT)) + n(t), \tag{1}$$
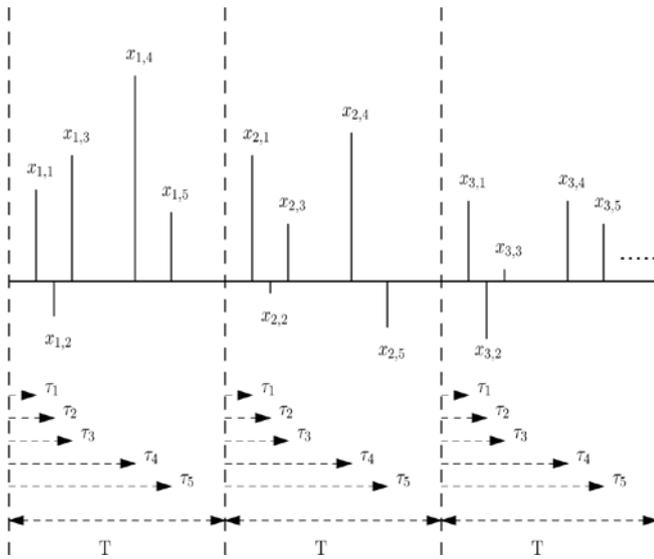
Figure 1: Example of cyclosparse signal $x(t)$

Where d is the number of effective impulses in the period T with $x_{i,k}$ and τ_k being their amplitude and delay factors respectively. K denotes the number of period per signal and the sub-index i stands for the period index, so $x_{i,k}$ represents the impulse with τ_k as delay factor in the *ith*-period. $n(t)$ represents the random noise of the system.

b) Motivation of the study

The choice of the signal modeling 1 is not arbitrary but can be justified in practice. Periodic random impulse processes are suitable tools that allow physicists to model many situations of interest, indeed they are a natural way for modeling highly localized events occurring randomly at given times or points of the state space [7, 8, 10, 6].

Another additional property of cyclostationary signals with periodic random impulses is sparsity, as only few impulses are nonzero i.e. $K \times d$ nonzero impulses. Hence, these signals are sparse on the direct domain. Unfortunately, all related works in this area exploit only one property, either sparsity or cyclostationarity and never both properties jointly. Convinced that combining simultaneously sparsity and cyclostationarity may lead to an enhancement of performances. So we wondered whether it is possible to build up an approach based on this idea. These were our motivations for introducing the new concept of cyclosparsity that gathers both properties to better characterize this kind of signals. The contributions of this article with respect to the short one [4], lie in:

- studying and generalizing the cyclosparsity concept to the algorithms OLS and SBR, which gives rise to Cyclo-OLS and Cyclo-SBR.
- providing insights into how the cyclosparsity works, and reviewing its implications in improving the selection step and reducing the computational cost of the four generalized greedy algorithms.

- jointly comparing the performances of all algorithms by varying the SNR for different number of cycles (K=2, 4, 8 and 16).

III. Concept of Cyclosparsity

The signal object of this study is assumed to be cyclostationary with random impulses i.e. consists of d periodic random impulses with d represents the number of impulses by cycle. On the one hand, only few elements are nonzero by cycle, so the signal is considered to be sparse. Furthermore, the positions of these nonzero elements (impulses) are cyclic/periodic i.e. they keep the same positions whatever the cycle. Hence, the ℓ_0-norms of the signal for each cycle are equal. On the other hand, the amplitudes of these nonzero elements are different and assumed to be random (see Fig. 1). Thus, the way to better characterize cyclostationary random impulses is to combine both properties i.e. cyclostationarity and sparsity. This hypothesis is the key idea of the cyclosparsity concept and is often satisfied in practice.

Before announcing the theorem defining cyclosparse process, we first introduce the following notations:

- $x = [x_1, \ldots, x_{L_x}]^T$ being the column vector of length L_x constructed from the signal $x(t)$.
- $x_i = x(i, i+T, i+2T, \ldots, i+(K-1)T)$ being a column vector with $1 \le i \le T$ sweeps all *ith* elements of each period (a total of K elements).
- $x_{i,:} = x((i-1)T+1, : iT)$ being a column vector with $1 \le i \le K$ sweeps all elements of the *ith* period (a total of T elements).
- $X = [x_1, \ldots, x_T] = [x_{1,:}, \ldots, x_{K,:}]^T$ being a $K \times T$ matrix where the elements are $x_{i,j}$ with *i* and *j* stand respectively for the period/cycle index and the position on each period.

The superscript T stands for the transpose of a vector or a matrix.

Theorem. A periodic random impulse signal $x(t)$ of period *T* is cyclosparse if the cardinal of nonzero elements of the whole signal is equal to the cardinal of nonzero elements of the signal over any period multiplied by the number of cycles *K*. This can be formulated into the following mathematical relationship:

$$\|x\|_0 \le K\|X\|_{\infty,0} \qquad (2)$$

the $\ell_{\infty,0}$-norm is applied to the column vector x after being reshaped as $K \times T$ matrix.

Proof. The ℓ_0-norm of x is defined as follows

$$\|x\|_0 = \sum_{k=1}^{KT} |x_k|^0 \qquad (3)$$

As is standard, $|x_k|^0 = 1$ if $x_k \ne 0$ and $|x_k|^0 = 0$ if $x_k = 0$. Thus, $\|x\|_0 = Card\{k, x_k \ne 0\}$ which indicates

the number of nonzero components of the column vector $\boldsymbol{x}$. Using these notations, Eq. 3 can be written as,

$$\|\boldsymbol{x}\|_0 = \|\boldsymbol{x}_{1,:}\|_0 + \|\boldsymbol{x}_{2,:}\|_0 + \cdots + \|\boldsymbol{x}_{K,:}\|_0$$

$$= \sum_{k=1}^{T} |x_{1,k}|^0 + \sum_{k=1}^{T} |x_{2,k}|^0 + \cdots + \sum_{k=1}^{T} |x_{K,k}|^0$$

Reorganizing the previous sum leads to,

$$\|\boldsymbol{x}\|_0 = |x_{1,1}|^0 + |x_{1,2}|^0 + \cdots + |x_{1,T}|^0$$

$$+ |x_{2,1}|^0 + |x_{2,2}|^0 + \cdots + |x_{2,T}|^0$$

$$+ \quad \vdots \quad + \quad \vdots \quad + \cdots + \quad \vdots$$

$$+ |x_{K,1}|^0 + |x_{K,2}|^0 + \cdots + |x_{K,T}|^0$$

It is easy to see that each vertical sum corresponds to a ℓ_∞-norm along the dimension of its corresponding column multiplied by the scalar K.

$$\|\boldsymbol{x}\|_0 \le K \left(\left| \max_i |\boldsymbol{x}_1| \right|^0 + \left| \max_i |\boldsymbol{x}_2| \right|^0 + \cdots + \left| \max_i |\boldsymbol{x}_T| \right|^0 \right)$$

$$\le K \left\| \max_i |\boldsymbol{x}_i| \right\|_0$$

$$\le \|\boldsymbol{X}\|_{\infty,0}$$

Where $\max_i |\boldsymbol{x}_i|$ stands for ℓ_∞-norm of $\boldsymbol{x}_i$ and the composite norm $\|\max_i |\boldsymbol{x}_i|\|_0$ represents the mixing norm $\ell_{\infty,0}$ of $\boldsymbol{X}$. The minimization of this norm encourages first diversity along i and then sparsity of the resulting vector.

The cyclosparsity in the case of deconvolution involves that two kinds of atoms participate to the linear combination in order to build data $\boldsymbol{y}$. Thus, we distinguish,

- Fundamental atoms, which correspond to nonzero impulses of the first cycle.
- Harmonic atoms, which are deduced from fundamental atoms by shifting with multiple of the cycle i.e. atoms corresponding to nonzero impulses of the remaining cycles.

Before studying in detail the cyclosparse approximation we first recall the principle of sparse approximation.

IV. From Sparse Approximation to Cyclosparse Approximation

a) Sparse approximation

The problem of sparse signal approximation consists in approximating a signal as a linear combination of a restricted number of elementary signals selected in a redundant collection (dictionary). It can be written as:

Find sparse $\boldsymbol{x}$ such that $\boldsymbol{\Phi x} \approx \boldsymbol{y}$,

where $\boldsymbol{y}$ corresponds to measured data and $\boldsymbol{\Phi}$ is a known matrix with atoms $\{\boldsymbol{\phi}_k\}_{k=1\ldots Q}$ as columns. Sparse approximations have to deal with a compromise between a good approximation and the number of involved elementary signals. Mathematically such compromise arises from minimizing the following criterion:

$$\mathcal{J}(\boldsymbol{x}) = \|\boldsymbol{y} - \boldsymbol{\Phi x}\|_2^2 + \beta \|\boldsymbol{x}\|_0 \qquad (4)$$

The parameter β controls the trade-off between the sparsity of the solution and the quality of the approximation. The lower is β, less sparse is the solution and better is the approximation. Hence, β is the key parameter to reach the compromise [38].

Of course, minimizing such a criterion corresponds to a combinatory optimization problem which is widely known to be NP hard. However, two approaches are usually used to avoid sweeping every combination: 1) Greedy algorithms, which iteratively ripen the approximation by successively identifying additional elementary signals that improves the approximation quality [13, 22]; 2) Convex relaxation algorithms i.e. based on the relaxation of the criterion (4), which replace the combinatorial problem with an easier optimization problem often chosen convex [16]. In the latter, the ℓ_0-norm is often relaxed with a ℓ_p-norm, where $\|\boldsymbol{x}\|_p = (\sum_k |x_k|^p)^{\frac{1}{p}}$ (note that $p = \infty$ is a limiting case for which $\|\boldsymbol{x}\|_\infty = \max_k |x_k|$). For $p = 1$ this problem corresponds to the Least Absolute Shrinkage And Selection Operator Regression (LASSO) [18] or Basis Pursuit DeNoising (BPDN) in signal processing [16].

Serious efforts have been made over many years to relate the solution of the original problem and the results of these two approaches and related algorithms. Some theoretical results give sufficient conditions of equivalence, depending on the dictionary and eventually of the solution [23, 22]. One such sufficient condition, not depending on the elementary signals of solution, is related to the coherence parameter μ of a dictionary which corresponds to the maximum absolute inner product between two distinct atoms in the dictionary:

$$\mu \stackrel{\text{def}}{=} \max_{j \ne k} |\langle \boldsymbol{\phi}_j, \boldsymbol{\phi}_k \rangle| = \max_{j \ne k} |(\boldsymbol{\Phi}^{\mathrm{T}} \boldsymbol{\Phi})_{jk}|.$$

In particular, it has been shown [24, 25, 22] that recovery condition for which the solution is unique and can be computed with various algorithms (BP, OMP, . . .) is

$$\|x\|_0 < \frac{1}{2}\left(1 + \frac{1}{\mu}\right) \text{ with } \mu < 1. \quad (5)$$

This is a sufficient condition under which both ℓ_1 criterion and greedy approaches can recover an exactly sparse signal.

b) *Cyclosparse approximation*

The extension of sparse approximation for cyclostationary and sparse signals is the main aim of this study. A cyclosparse solution is given by minimizingthe following criterion:

$$\mathcal{J}(x)\sum_{i=1}^{T}\left\|y - \Phi_{\{i+mT;\; m=0,\dots,K-1\}}x\right\|^2 \beta K\|X\|_{\infty,0}; \quad (6)$$

with K and T denote respectively the number of cycles/periods and the cyclic period. Actually, the inner ℓ_∞-norm encourages diversity along the cycles so the $\ell_{\infty,0}$ mixed-norm measures the cyclosparsity along the whole signal. Both approaches used to avoid exploring every combination of the sparse approximation problem can be extended to cyclosparse context: greedy algorithms and convex relaxation. In this paper, we focus on greedy algorithms.

c) *Difference between joint sparse approximation and cyclosparse deconvolution*

In spite of the fact that the joint sparse approximation [26, 27, 28, 29] and cyclosparse deconvolution seem to have similar formulations, they are completely different, and the most basic differences are listed as follows:

1. In the case of deconvolution, the dictionary is imposed by the IR and cannot be chosen unlike sparse approximation where the dictionary is generally chosen as union of bases or wavelet dictionary. Also, the dictionary size is of the same order as data size, so the dictionary does not correspond to a redundant set of elementary signals as in sparse approximation.

2. The deconvolution aims to retrieve the object x, i.e., to find the coefficients associated to the actual elementary signals forming the data. Whereas, the problem of sparse approximation searches for a good approximation of the object involving few atoms.

3. The joint sparsity is used for multi-dimensional signals or images that share almost the same dictionary. In the context of our study, we deal with mono-dimensional signals which cannot be split to apply joint sparsity.

The first point is a matter of prime importance in terms of computation of the solution. Indeed, as for deconvolution, the dictionary atoms correspond to shifted versions of the IR, hence they are highly correlated. So, the theoretical properties which guarantee the solution of greedy algorithms or convex relaxation to correspond to the sparse approximation are often not satisfied. For example, the coherence is generally very high $(\mu > \frac{1}{2})$ so theorems based on such quantities guarantee to reach the solution only if it is composed of a single elementary signal. This can be demonstrated by considering a cyclosparse object x with $\|x\|_0 = K\|X\|_{\infty,0} = Kd,$ relationship 5 becomes $\mu < \frac{1}{2Kd-1}$. To be in conformity with $\mu > \frac{1}{2}$, Kd must be set to 1, which means that the signal is composed of single period with one impulse i.e. a single elementary signal. More precisely, it can easily be seen in this case that the scalar product $\langle \mathbf{h}_j, \mathbf{h}_k \rangle$ roughly corresponds (up to the boundary conditions taken into account for the convolution) to a value of the auto-correlation of the IR: $\langle \mathbf{h}_j, \mathbf{h}_k \rangle = \langle \mathcal{H} * \delta_j, \mathcal{H} * \delta_k \rangle = \langle \delta_j, \mathcal{H}^- * \mathcal{H} * \delta_k \rangle = \mathcal{H}^- * \mathcal{H}(k-j),$ where $\mathcal{H}^-(j) = \mathcal{H}(-j)$ and δ represents the Kronecker symbol. For example, if the IR has a Gaussian-like shape, the autocorrelation of the IR also has a Gaussian-like shape with a twice as large standard deviation, so a correct sampling of the Gaussian leads to $\mu > \frac{1}{2}$. On the other hand, it is obvious that two spikes are easy to detect if they are sufficiently far away from each other. A sufficient recovery condition has been established on the basis of such distance for sparse deconvolution [30] and also for super-resolution [37], but this condition may not be carried out for real data. Therefore, no theoretical results guarantee that the algorithms converge towards the sparse solution. So the various algorithms have to be tested on realistic data. Moreover, as each algorithm could give different solutions, one should compare the efficiency of each algorithm on a coherent basis, in particular with respect to the tuning parameters.

The second point points-out the issues of sparse deconvolution and sparse approximation. Indeed, the objective of sparse deconvolution is to recover the spike-like objects, i.e. to detect the unresolved objects and to estimate their amplitude, and not only to have a sparse solution which gives a good approximation of the data. So if the algorithms fail to reach the real solution, this leads to false alarms and missing detection of objects. Moreover, when dealing with real data, it is not guaranteed that the sparse solution corresponds to the true objects. To avoid such a problem in this study, we will only compare hereafter the algorithms on simulated data for which the true solution is known.

Finally, the third point establishes that joint sparsity modeling is not suitable for periodic random

impulses. Actually, the major difference between both concepts arises in the fact that joint sparsity assumes that multi-dimensional signals are disjointed on the border. This means that each signal does not contribute on its neighbors because the convolution is independently made between each mono-dimensional signal and the IR. However, for cyclosparsity, the cycles are related and connected because of the convolution with the IR. Hence, the cycles are not independent anymore and cannot be split as for joint sparsity.

Consequently, we came to the conclusion that the joint sparsity cannot be used to characterize and model cyclostationary signals with periodic random impulses. These were the reasons behind developing the concept of cyclosparsity.

In addition, as the algorithms may give different solutions, one should test all the existing algorithms. In this study we focus on the greedy algorithms. The first reason is that greedy algorithms are classically used in many domains for a long time, and even before the first publications on sparse representations. So the use of greedy algorithms is quite natural. Moreover, one important aspect of this study is to be able to compare the algorithms on the same basis in terms of parameters tuning. The noise level of real data can often be modeled or at least estimated from the data so the noise variance can be considered as a known parameter. In this study, we will use the noise variance to build the condition to stop the iterations of the various greedy algorithms on the same statistical basis. Such a condition is not as easy to build on the same basis using convex relaxation.

V. Cyclosparse Deconvolution

a) Structure of the dictionary H

Let us first specify the boundary condition accounted for in the convolution operator. We assume that the convolution (Eq. 1) is computed with the zero-padded edges. Using this option the resulting signal has length $L_y = L_x + L_h - 1$ where L_x and L_h stand respectively for the length of the signal to reconstruct and the Point Spread Function (PSF). Of course, such boundary hypothesis influences size and structure of the dictionary H formed from the IR. In particular, for physical reasons, the IR is normalized such that $\sum_{j=1}^{L_h} \mathcal{H}_j^2 = 1$ so the columns of the matrix which correspond to shifted versions of the IR should have a constant norm $\mathbf{h}_i^T \mathbf{h}_i = \sum_{j=1}^{L_h} \mathcal{H}_j^2$.

Note that H is a sparse matrix of dimension $L_y \times L_x$ with $L_h \times L_x$ nonzero elements (the length of the PSF is generally largely smaller than the length of the signal). But as the L_x atoms of the dictionary correspond to shifted versions of the PSF, matrix H is composed only with the L_h elements of the PSF.

Moreover, as matrix H models a convolution operator, it has a Toeplitz structure (diagonal-constant matrix) and each operation involving H may be computed as a result of a convolution.

b) Problem statement

Before discussing the general case modeling of cyclosparse deconvolution, let us consider an example of cyclosparse object x, with two periodic random impulses of positions $d_1 + mT$ and $d_2 + mT$, in order to make clear the formulation. Thus, the atoms that participate to build-up y are $\mathbf{h}_{d_1+mT}$ and $\mathbf{h}_{d_2+mT}$ with $m = 0, \ldots, K - 1$. In addition, the atoms $\mathbf{h}_{d_1+mT}$ are not correlated (likewise $\mathbf{h}_{d_2+mT}$) (i.e. the scalar product of these vectors for different values of m is null) as the IR length L_h is generally smaller than the cyclic period T. If it is not the case, the IR $\mathcal{H}$ can be truncated such as $L_h < T$.

Consequently, for any cyclosparse object x, the model 1 can be written with matrix notations as follows,

$$y = H\,D\,x + n$$
$$= \widetilde{H}\,x + n$$

$\widetilde{H}^T \widetilde{H} = D(H^T H)D$ with D is the cyclosparsity operator which is a $L_x \times L_x$ diagonal matrix

$D^T = D = \sum_{m=0}^{K-1} \sum_{n=1}^{\|x\|_{\infty,0}} e_{d_n+mT}\, e_{d_n+mT}^T =$ $\mathrm{diag}\left(\sum_{m=0}^{K-1} \sum_{n=1}^{\|x\|_{\infty,0}} e_{d_n+mT} \right)$ where e_{d_n+mT} are the canonical basis vectors and d_n is the index of the *nth* impulse with τ_n as factor delay. Thus, $\widetilde{H}$ is a $L_y \times L_x$ matrix with particular structure i.e. the nonzero columns for each cycle are deduced by shifting the columns of the first cycle with a multiple of the cyclic period T. Hence, $\widetilde{H}$ points out the cyclosparsity property in the convolution case.

c) Cyclosparse greedy algorithms

Let the sub-matrix $\mathbf{H}_\Lambda$ built-up from the columns of H where the indices are in Λ, $\mathbf{h}_i = \mathbf{H}_{\{i\}}$, and $\Lambda^{(k)}$ is the set of the selected indices at iteration k. The vectors are defined as follows, $x = [x_1, \ldots, x_{L_x}]^T$, $y = [y_1, \ldots, y_{L_y}]^T$, $n = [n_1, \ldots, n_{L_y}]^T$ and $r = [r_1, \ldots, r_{L_y}]^T$ which denotes the residual. L_x, L_y and L_h stand respectively for the length of x, y and $\mathcal{H}$. Finally, let the vector $m = [0, 1, \ldots, (K-1)]$ be the vector of period indices.

In this study the interest is focused on the extension of greedy sparse approximation algorithms to cyclosparsity context for deconvolution. The dictionary is given by the Toeplitz matrix H deduced from the IR $\mathcal{H}$. Greedy algorithms are iterative algorithms composed of two major steps at each iteration: 1) the selection of an additional elementary signal in the dictionary; 2) the update of the solution and the corresponding

approximation. A stopping rule helps decide whether to stop or continue the iteration.

Let us consider $x^{(k)}$ be the solution of the *kth* iteration, $x_\Lambda^{(k)}$ being its coefficients at indices Λ and $r^{(k)} = y - H\,x^{(k)}$ the residual corresponding to this solution (approximation error). The typical structure of a greedy algorithm is:

Initialize $k = 0$, $\Lambda^{(k)} = \emptyset$ and $r^{(k)} = y$.

Iterate on $k = k + 1$ until the stopping rule is satisfied:

- Select the index $i^{(k)}$ corresponding to an atom $\mathbf{h}_i$ improving the approximation.
- Update the solution $x^{(k)}$, with nonzero elements at indices $\Lambda^{(k)} = \Lambda^{(k-1)} \cup \{i^{(k)}\}$, and the corresponding residual $r^{(k)}$.

The various algorithms differ on the selection or the updating steps. The most popular greedy algorithm must be the Matching Pursuit [13] and its orthogonal version OMP [14]. Both algorithms and two other algorithms will be extended to cyclosparsity context.

Cyclo-Matching Pursuit (Cyclo-MP)

The Cyclo-MP (or Cyclic-MP) is the extension of the Matching Pursuit (MP) [13]. The K additional atoms jointly maximize the scalar product with the residual. The update corresponds to an orthogonal projection of the residual on the selected atoms, so only the solution at the selected indices is updated. Note that with such a scheme it is possible to select already selected atoms.

In order to avoid overloading equations, the index $i^{(k)} + mT$ will be replaced by $i_{mT}^{(k)}$ in the following.

- Selection of K atoms: $\Lambda^{(k)} = \Lambda^{(k-1)} \cup \{i_{mT}^{(k)}; m = 1, \dots, K-1\}$,

$$i^{(k)} = \underset{i}{\operatorname{argmax}} \sum_{m=0}^{K-1} \left| \mathbf{h}_{i+mT}^{\mathrm{T}}\, r^{(k-1)} \right| \qquad (7)$$

- Update:
solution:

$$x_{i_{mT}^{(k)}}^{(k)} = x_{i_{mT}^{(k)}}^{(k-1)} + \left(\mathbf{H}_{i_{mT}^{(k)}}^{\mathrm{T}} \mathbf{H}_{i_{mT}^{(k)}} \right)^{-1} \mathbf{H}_{i_{mT}^{(k)}}^{\mathrm{T}}\, r^{(k-1)} \qquad (8)$$

residual:

$$r^{(k)} = r^{(k-1)} - \mathbf{H}_{i_{mT}^{(k)}} \left(\mathbf{H}_{i_{mT}^{(k)}}^{\mathrm{T}} \mathbf{H}_{i_{mT}^{(k)}} \right)^{-1} \mathbf{H}_{i_{mT}^{(k)}}^{\mathrm{T}}\, r^{(k-1)} \qquad (9)$$

- Stoping criterion

where the matrix $\mathbf{H}_{i_{mT}^{(k)}} = [\mathbf{h}_i,\ \mathbf{h}_{i+T}, \dots, \mathbf{h}_{i+(K-1)T}]$ gathers the K selected atoms. Appendix A explains why the selection step of the Cyclo-MP (Eq. 7) is more efficient?

Cyclo-Orthogonal Matching Pursuit (Cyclo-OMP)

The Cyclo-OMP (or Cyclic-OMP) is the extension of the Orthogonal Matching Pursuit (OMP) [14]. The Cyclo-OMP differs from the Cyclo-MP on the updating step as an orthogonal projection of the data on the whole selected atoms is performed. This avoids the selection of already selected atoms but increases the computation cost as the amplitudes associated to all the selected atoms are updated.

- Selection: same as for the Cyclo-MP (7)
- Update:
solution:

$$x_{\Lambda^{(k)}}^{(k)} = \left(\mathbf{H}_{\Lambda^{(k)}}^{\mathrm{T}} \mathbf{H}_{\Lambda^{(k)}} \right)^{-1} \mathbf{H}_{\Lambda^{(k)}}^{\mathrm{T}} y \qquad (10)$$

residual:

$$r^{(k)} = y - \mathbf{H}_{\Lambda^{(k)}} x_{\Lambda^{(k)}}^{(k)}$$

- Stopping criterion

where the matrix $\mathbf{H}_{i_{mT}^{(k)}} = [\mathbf{h}_i,\ \mathbf{h}_{i+T}, \dots, \mathbf{h}_{i+(K-1)T}]$ gathers the K selected atoms. Appendix A explains why the selection step of the Cyclo-MP (Eq. 7) is more efficient?

Cyclo-Orthogonal Matching Pursuit (Cyclo-OMP)

The Cyclo-OMP (or Cyclic-OMP) is the extension of the Orthogonal Matching Pursuit (OMP) [14]. The Cyclo-OMP differs from the Cyclo-MP on the updating step as an orthogonal projection of the data on the whole selected atoms is performed. This avoids the selection of already selected atoms but increases the computation cost as the amplitudes associated to all the selected atoms are updated.

- Selection: same as for the Cyclo-MP (7)
- Update:
solution:

$$x_{\Lambda^{(k)}}^{(k)} = \left(\mathbf{H}_{\Lambda^{(k)}}^{\mathrm{T}} \mathbf{H}_{\Lambda^{(k)}} \right)^{-1} \mathbf{H}_{\Lambda^{(k)}}^{\mathrm{T}} y \qquad (10)$$

residual:

$$r^{(k)} = y - \mathbf{H}_{\Lambda^{(k)}} x_{\Lambda^{(k)}}^{(k)}$$

- Stopping criterion

Cyclo-Orthogonal Least Squares (Cyclo-OLS)

The Cyclo-OLS (or Cyclic-OLS) is the extension of the Orthogonal Least Squares (OLS) [15]. The Cyclo-OLS differs from the Cyclo-OMP on the selecting step as the selected atoms minimize the approximation error. The Cyclo-OLS is the more coherent greedy algorithm as both in the selection and the updating steps it aims to minimize the approximation error. However, the computation cost of the Cyclo-OLS highly increases compared to the Cyclo-OMP.

- Selection: $\Lambda^{(k)} = \Lambda^{(k-1)} \cup \{i_{mT}^{(k)}; m = 1, \dots, K-1\}$,

$$i^{(k)} = \underset{i}{\operatorname{argmin}} \| y - \mathbf{H}_\Lambda \left(\mathbf{H}_\Lambda^{\mathrm{T}} \mathbf{H}_\Lambda \right)^{-1} \mathbf{H}_\Lambda^{\mathrm{T}} y \|^2 \qquad (11)$$

with $\Lambda = \Lambda^{(k-1)} \cup \{i_{mT}^{(k)}; m = 1, \dots, K-1\}$,

- Update: same as for the Cyclo-OMP solution (10)
- Stopping criterion

Appendix B shows why the selection step of the Cyclo-OLS (Eq. 11) is more efficient?

Cyclo-Single Best Replacement (Cyclo-SBR):

The Cyclo-SBR (or Cyclic-SBR) is the extension of the Single Best Replacement (SBR) algorithm [19, 20, 21]. The SBR algorithm is not strictly speaking a greedy algorithm. It is an iterative algorithm which aims to minimize criterion (4). The SBR has been inspired by the Single Most Likely Replacement (SMLR) algorithm [31] proposed for Bernoulli-Gaussian deconvolution. Of course, the SBR is not guaranteed to converge towards the global minimum of (4), i.e. to the sparse approximation, but it is an interesting alternative to greedy algorithms as it has a very similar iterative

scheme and has been shown to give better results for deconvolution [19, 20, 21].

For a given parameter ⬚, the selection and update steps of the Cyclo-SBR can be written as:

- Selection:

$$\Lambda^{(k)} = \underbrace{\Lambda^{(k-1)} \cup \{i_{mT}^{(k)}; m = 1, \dots, K-1\}}_{\text{Addition}}$$

or

$$\Lambda^{(k)} = \underbrace{\Lambda^{(k-1)} \setminus \{i_{mT}^{(k)}; m = 1, \dots, K-1\}}_{\text{Removal}}$$

$$i^{(k)} = \arg\min_i \mathcal{J}_\Lambda = \arg\min_i \|y - H_\Lambda (H_\Lambda^T H_\Lambda)^{-1} H_\Lambda\, y\|^2 + \beta \,\#\,\{\Lambda\} \quad (12)$$

$$\text{with} \begin{cases} \Lambda = \Lambda^{(k-1)} \cup \left\{i_{mT}^{(k)}; m = 1, \dots, K-1\right\} \\ \text{or} \\ \Lambda = \Lambda^{(k-1)} \setminus \{i_{mT}^{(k)}; m = 1, \dots, K-1\} \end{cases} \quad (13)$$

- Update: same as for the Cyclo-OMP solution (10)
- Stopping criterion: no replacement is accepted

This means that the selection step compares both selecting K new atoms and deleting K previously selected atoms, the replacement which minimizes criterion (4) being selected. In case of suppression, the set of selected atoms is updated as $\Lambda^{(k)} = \Lambda^{(k-1)} \setminus \left\{i_{mT}^{(k)}; m = 1, \dots, K-1\right\}$. Note that the selection step for adding K new atoms is strictly identical to the Cyclo-OLS one, the updating step being identical to that of the Cyclo-OMP and the Cyclo-OLS. The iterations of the Cyclo-SBR stops when no replacement allows decreasing the criterion. Appendix C illustrates the reason why the selection step of the Cyclo-SBR (Eq. 12) is more efficient?

d) Stopping rule

The only parameter to set for using these greedy algorithms is the stopping rule. In terms of sparse approximation, comparing the norm of the residual to a threshold is a natural stopping rule, as it corresponds to an expected quality of approximation. On the other hand, for deconvolution, the residual for the true object corresponds to the noise. So a statistical test on the residual may be used as stopping rule, which decides whether the residual can be distinguished from noise. For Gaussian, centered, independent and identically distributed (*i.i.d.*) noise, of known variance σ^2, the norm $\|n\|_2^2/\sigma^2 = \sum_{k=1}^N n_k^2/\sigma^2$ follows a Chi-square distribution with N degrees of freedom. So, the Chi-square distribution may be used to determine the value of ϵ for which $Pr(\|r^{(k)}\| \leq \epsilon) = \eta$ for a given probability, e.g. $\eta = 95\%$.

e) Discussion

- Note that for the four algorithms, the selection step is made jointly over the whole set of cycle indices m thanks to cyclosparsity. Also note that the original version of the algorithms can be retrieved taking into account a single period for $m = 0$.

- An objective and meaningful comparison of the greedy algorithms and the Cyclo-SBR requires a common tuning of the parameters. This means that the parameter ⬚ of the Cyclo-SBR has to be set in agreement with the stopping rule of the greedy algorithms. This can be done using a continuation path technique for the SBR [19, 20, 21] named CSBR (Continuation SBR). Roughly, the SBR algorithm is executed for decreasing values of $\beta = \beta_q$ and stopped at the higher value of ⬚ for which the approximation is acceptable, i.e. with the same stopping rule as the greedy algorithms presented § 5.4. Note that the critical values β_q for which the selected atoms may change have been shown to be a by-product of the SBR algorithm [19, 20, 21] and do not require additional computation.

- Another advantage of cyclic greedy algorithms is the significant reduction of the computation cost. Cyclic greedy algorithms select K atoms at time utilizing the residual at iteration k i.e. one scalar product with atoms. Unlike greedy algorithms that select one atom at time utilizing the residual at iteration k, to select an additional atom, the residual at iteration $(k + 1)$ must be calculated and then the scalar product with atoms. In consequence, to select K atoms, K scalar product must be performed. Thus, the computation cost for the

selection step is roughly divided by K. For that matter Appendix E summarizes the computational cost of the studied greedy algorithms.

- Proceeding [32, 14, 19, 20, 21] one can use the matrix inversion lemma to compute iteratively

$$(\mathbf{H}^{\mathrm{T}}_{\Lambda^{(k-1)} \cup \{i_{mT}^{(k)}; m=1,\dots,K-1\}} \mathbf{H}_{\Lambda^{(k-1)} \cup \{i_{mT}^{(k)}; m=1,\dots,K-1\}})^{-1} \text{ at a}$$

low cost, from the knowledge of $(\mathbf{H}^{\mathrm{T}}_{\Lambda^{(k-1)}} \mathbf{H}_{\Lambda^{(k-1)}})^{-1}$ (see Appendix D for more details).

- The cyclic greedy algorithms have been presented with matrix notations, which are very useful to understand the algorithms but may not be used directly for their implementation in the case of large size data. For the deconvolution case, according to the dictionary and matrix structures some efficient implementation has to be accounted for to reduce the computation cost and the memory storage. As matrix $\mathbf{H}$ models a convolution operator, it has a Toeplitz structure and each operation involving $\mathbf{H}$ (Matrices and vectors operation) may be computed as a result of a convolution. An efficient implementation is proposed in [33], based on the convolution operator and not on vector and matrix products as is usually done for sparse approximations. We used therefore the same practical implementation for cyclic greedy algorithms for the deconvolution.

VI. SIMULATION

a) Description

The proposed methods are tested using synthetic signals in order to evaluate their effectiveness. To do so, we consider simulation example with the following parameters. A cyclostationary signal based on periodic random impulses $L_x = 256$ and $T = 32$, so the number of periods is $K = 8$). This input signal consists of $d = 5$ periodic random impulses of the same positions (7, 9, 11, 13 and 15) in each cycle. The signal is then filtered by an ARMA system where the transfer function is given as: $H(z) = \frac{1+b_1 z^{-1}}{1+a_1 z^{-1}+a_2 z^{-2}}$ where $b_1 = -0.6$, $a_1 = -0.9$ and $a_2 = 0.6$; the time representation of the IR ($L_h = 15$) is reported in Fig. 2-b.

Then *i.i.d.* Gaussian noise ($L_y = 270$) is added to the convolved signal, as illustrated by (1), such that the SNR is 14dB. The resulting signal is reported in Fig. 2-a.

For the first evaluation we consider the time representations of the reconstructed signals given by each algorithm versus the true signal. Fig. 3 reports the true signal (blue line) described by the relationship (1) and the estimated signal (colored line) for each method. We note that we constrained the plot to the three first periods in order to avoid overloading Fig. 3. Regarding each algorithm and its corresponding extension, we

note from Fig. 3 that deconvolution across cyclosparsity hypothesis allows to:
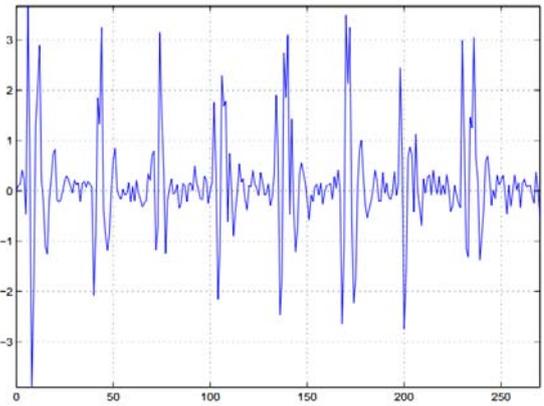
(a)
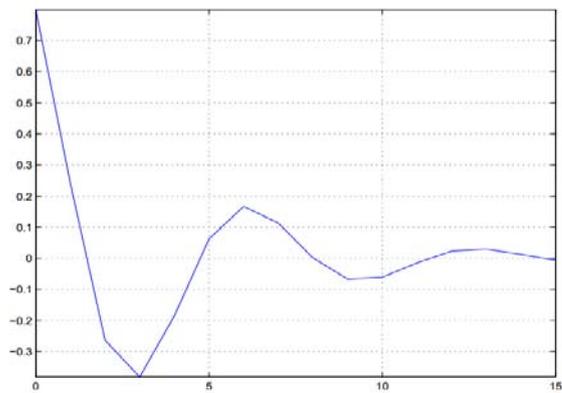
(b)

Figure 2 (a) : The observed signal (SNR= 14dB).
(b): The IR used for all simulations

- Detect and restore impulses even drowned in noise
- Reduce false and missing detections
- Estimate well the amplitude of impulses

Comparing all algorithms, we deduce that the Cyclo-OMP, the Cyclo-OLS and the Cyclo-SBR provided the best estimations. This point will be examined in detail using other evaluations (mean squared error and histogram).

b) Mean Squared Error

We provide here a comparison between the proposed approaches against the original ones. The aim is to show the performance of cyclosparse greedy deconvolution in various *i.i.d.* noisy environments. The simulation is made with the same parameters as the first example except SNR. Actually, the SNR will vary from 1dB to 30dB. And for each value of the SNR, 500 Monte Carlo (MC) runs will be implemented. Thus, for each MC run,

- The periodic random impulses keep the same positions but with random amplitudes

- The input signal is filtered by the IR of Fig. 2-b
- *i.i.d.* Gaussian noise is added to the convolved signal, as illustrated by (1), such that the SNR is set to the desired value

Another measure adopted to evaluate our system involved changing the SNR while every other variable used in MC simulations remained constant. We aim to examine the effects of increasing noise on the performances of these methods. The evaluation quantities for our simulation study, comparing the performances of these methods, were average Mean Squared Error (MSE) and average histogram.

The MSE provides a measure of the quality of the reconstructed signal. The MSE of the estimate $\hat{x}$ with respect to x is defined as, $\mathrm{MSE}(\hat{x}) = \mathrm{E}[(\hat{x} - x)]$. These MSE will be averaged over the number of MC runs. Fig. 4 shows the variation of each output'MSE, for the proposed methods and the original ones as well, with the SNR. We note from the trend of the graph that the MSE decreases with increasing SNR. This is because higher SNR implies lower noise effect on observed data y. This effect allows fewer amplitude estimation errors after detection takes place; hence, good performances of the algorithms. However, we note from Fig. 4 (for $K = 8$) that higher MSE occurs for lower SNR. This is also as a result of higher noise effect for lower SNR. Also, the MSE is nearly the same for all cyclo-algorithms except Cyclo-MP, with highest MSE occurring for lower SNR.
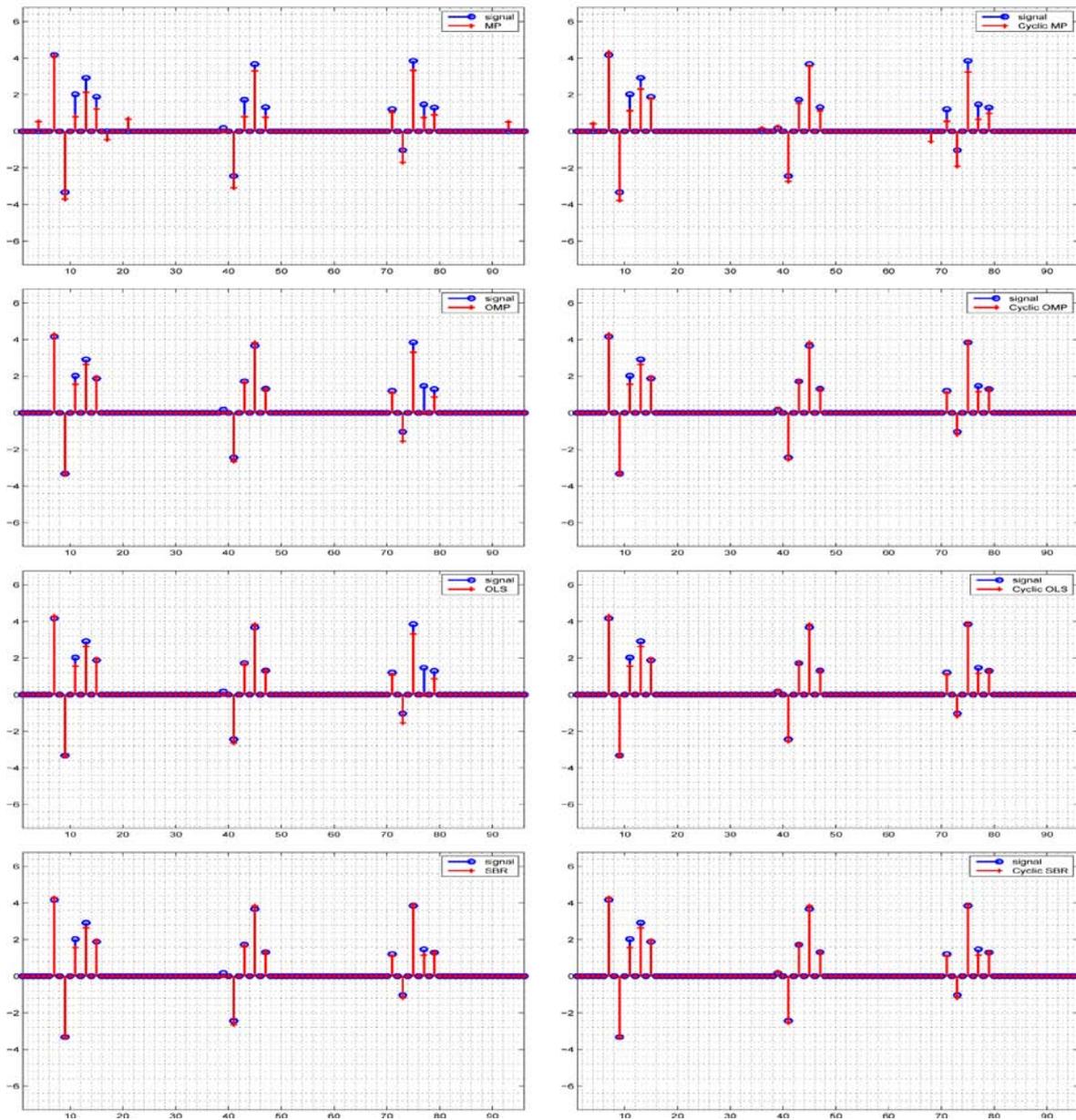
Figure 3 : The reconstructed signal versus the original one (the SNR is set to 14dB)

Furthermore, as can be seen from the same figure (Fig. 4, for $K = 8$), the algorithms' behavior with respect to the MSE against noise can be decomposed into two parts: SNR less or greater than 6dB. For SNR greater than 6dB, the MSE of the algorithms can be sorted in descending order as follow, MP; Cyclo-MP; (OMP, OLS and SBR); (Cyclo-OMP, Cyclo-OLS and Cyclo-SBR). However for SNR less than 6dB, the MSE of the algorithms can be sorted as, (MP, OMP, OLS and SBR); Cyclo-MP; (Cyclo-OMP and Cyclo-OLS); Cyclo-SBR. We conclude therefore, that cyclo-algorithms perform well even for lower SNR

c) Histogram

The histogram shows the distribution of data values. Thus, performing the histogram to the reconstructed signal will show the number of the true

impulses and false/missing detections that happen within the true impulses as well. Therefore, this will help us know how much detection is in error for each algorithm. We made MC simulations in order to help us determine an average histogram over the number of 500 MC runs. Fig. 5 shows the variation of each output' average histogram obtained by varying SNR from 1dB to 30dB for the proposed methods and the original ones as well. As the histogram is almost periodic, we constrained the plot to the first period in order to avoid overloading Figures 5, 6, 7 and 8.

We note from Fig. 5 that false detections increase with decreasing SNR. This is because higher SNR implies lower noise effect on observed data y. The

histogram is nearly the same for all cyclo-algorithms except Cyclo-MP, with highest missing detections occurring even for higher SNR. However, we note from Fig. 5 that higher missing/false detections occur mainly for lower SNR. This is also as a result of higher noise effect for lower SNR.

The histogram indicates good detection, false detection/alarms and missing detection. Based on these three criteria, we can classify the histogram similarly to the MSE. As can be seen from this Fig. 5, the algorithms' behavior with respect to the histogram against noise can be decomposed into two parts: SNR less or greater than 6dB. Consequently, the histogram confirms the MSE behavior of the algorithms and leads to the same sorting of the algorithms.

d) *Influence of the number of cycles*

Another parameter which can influence the performances of the cyclo-algorithms is the number of cycles/periods K. To examine this, we performed three simulations in which K was gradually increased. The simulations are made with the same parameters as the second example except data size. Actually, changing the number of cycles means changing data size as well. So, for K equal to 2, 4, 8 and 16, correspond respectively to data size 64, 128, 256 and 512. Simulation results for each data size set were reported in

data size set were reported in Fig. 4 for the average MSE and Figures 6, 7, 5 and 8 for the average histogram. As expected from theory, increasing the number of cycles leads to good performances with less false/missing detections and errors in the estimation of the impulses amplitudes, for cyclo-algorithms in comparison with their corresponding algorithms. We note also that the Cyclo-OLS and Cyclo-OMP converge gradually (especially for lower SNR) to the Cyclo-SBR as K increases. This is because less error occurred in the selection step for adding new atoms, so no need for the Cyclo-SBR to correct any error on the selection step by removing already added atoms.

VII. Discussion

The objective of the previous simulations is to evaluate the contribution of cyclosparsity for greedy deconvolution. It is apparent therefore, that deconvolution across cyclosparsity hypothesis allows detecting and restoring impulses even drowned in noise provided that these impulses being significant for the other cycles of the signal. Also, increasing the number of cycles i.e. more atoms involved in the average leads to a

(K=2)   (K=4)

(K=8)   (K=16)
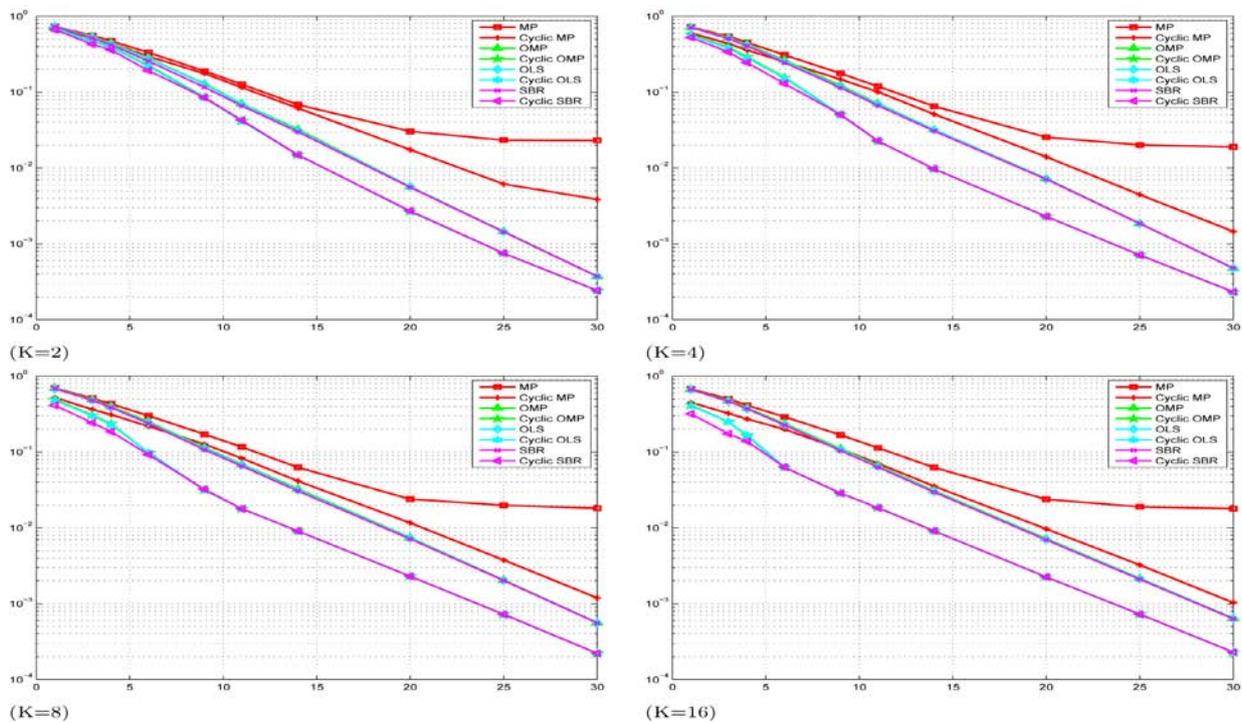
Figure 4 : The effect of varying SNR from 1dB to 30dB over MC runs on the MSE

considerable enhancement of the performances of cyclo-algorithms.

The Cyclic algorithms perform better than their corresponding classical ones even for lower SNR. The Cyclo-OLS and Cyclo-OMP reach the performances of

the Cyclo-SBR, especially for higher SNR, because averaging over m reduces false/missing detections, so the Cyclo-SBR seldom if ever removes already added atoms. The Cyclo-MP has the bad performances. What happened to the behavior of the Cyclo-MP can be explained by the distance between adjacent impulses. When nonzero elements are so close and strongly correlated, false detections occur often because the orthogonal projections are made over only the K selected atoms unlike the other algorithms where the orthogonal projections are made over the whole selected atoms. However, by increasing the parameter K, the Cyclo-MP behavior converges to the behavior of OMP, OLS and SBR for higher SNR, whereas for lower SNR the Cyclo-MP behavior converges to the behavior of Cyclo-OMP and Cyclo-OLS. This means that cyclosparsity allows to the Cyclo-MP to overcome the drawback of the MP.

VIII. Conclusion

- The objective of the paper is the introduction of the concept of cyclosparsity for cyclostationary signals based on periodic random impulses. Then, integrate this concept for greedy sparse algorithms in order to increase the performances of the deconvolution and reduce significantly the computation cost as well.
- The performance of the new algorithms using computer simulated cyclostationary signals was demonstrated. It is apparent therefore that the proposed methods compare favorably with the original ones. Reasons for the improved performance of the proposed methods over the original ones include the following: the cyclosparsity model makes possible the exploitation of the information given by the periodicity which allows less false alarms and missing detections as well.
- The unique additional information required by cyclic greedy algorithms is the cyclic period T. In general, the cyclic period is related to the studied system. For rotating machines, the cyclic period is a multiple of the shaft rotation. Furthermore, the problem of estimating the cyclic period or the cyclic frequency has been addressed in several articles as [34].
- We investigate to apply the proposed algorithms to vibratory signals namely bearing signals for diagnostic. Actually, bearing with inner rice or outer rice default signals are known to be random periodic impulse signals. These signals are convolved by the IR of the mechanical structure of the rotating machine and then noise is added to the convolved signal. Thus, the resulting signal is not legible, and hence, the need of deconvolution to restore the random periodic impulses in order to estimate the degree of the default.

Appendix A. Comparison between the selection step of the Cyclo-MP (Eq. 7) and the one of the MP

The selection step of the Cyclo-MP and Cyclo-OMP is given by,

$$i^{(k)} = \arg\max_i \sum_{m=0}^{K-1} |\mathbf{h}_{i+mT}^{\mathrm{T}} \boldsymbol{r}^{(k-1)}| \qquad (A.1)$$

Let develop the term $\sum_{m=0}^{K-1} |\mathbf{h}_{i+mT}^{T} \boldsymbol{r}^{(k-1)}|$,

$$\sum_{m=0}^{K-1} |\mathbf{h}_{i+mT}^{\mathrm{T}} \boldsymbol{r}^{(k-1)}| = \sum_{m=0}^{K-1} |C_{r^{(k-1)}h_i}(i+mT)| = \sum_{m=0}^{K-1} |\mathcal{H}^- * \boldsymbol{r}^{(k-1)}(i+mT)|$$

$$= |C_{r^{(k-1)}h_i}(i)| + |C_{r^{(k-1)}h_i}(i+T)| + \cdots + |C_{r^{(k-1)}h_i}(i+(K-1)T)|$$

$$= |\mathcal{H}^- * \boldsymbol{r}^{(k-1)}(i)| + |\mathcal{H}^- * \boldsymbol{r}^{(k-1)}(i+T)| + \cdots + |\mathcal{H}^- * \boldsymbol{r}^{(k-1)}(i+(K-1)T)|$$

where $C_{r^{(k-1)}h_i}^T$ represents the correlation between $r^{(k-1)}$ and $\mathbf{H}$ and $\mathcal{H}^-(j) = \mathcal{H}(-j)$.

The selection step of the MP and OMP is exclusively based on the term $|C_{r^{(k-1)}h_i}^{\mathrm{T}}(i)|$ (equivalent to $|\mathcal{H}^- * \boldsymbol{r}^{(k-1)}(i)|$). So the selected atom is the one who maximizes it. Whereas, the Cyclo-MP and Cyclo-OMP are based on the correlation at i for the multiple of the cyclic period T, i.e.

$$|C_{r^{(k-1)}h_i}(i)| + |C_{r^{(k-1)}h_i}(i+T)| + \cdots + |C_{r^{(k-1)}h_i}(i+(K-1)T)|$$

equivalent to

$$|\mathcal{H}^- * \boldsymbol{r}^{(k-1)}(i)| + |\mathcal{H}^- * \boldsymbol{r}^{(k-1)}(i+T)| + \cdots + |\mathcal{H}^- * \boldsymbol{r}^{(k-1)}(i+(K-1)T)|$$

for the step selection. Thus, the K selected atoms should maximize jointly the sum.

Obviously, this has the advantage to avoid penalizing atoms associated to impulses with small amplitude thanks to the sum over periods that allows a joint selection of K atoms at once. And hence, the need to involve more atoms in the sum. So, when K increases the sum covers more atoms, therefore the chance to have atoms that bear on the sum increase significantly. Then less errors occur (especially when impulses are close) in the selection step.

Appendix B. Comparison between the selection step of the Cyclo-OLS (Eq. 11) and the one of the OLS

The selection step of the Cyclo-OLS is given by,

$$i^{(k)} = \arg\min_i \|\boldsymbol{y} - \mathbf{H}_{\Lambda^{(k)}}\big(\mathbf{H}_{\Lambda^{(k)}}^{\mathrm{T}}\mathbf{H}_{\Lambda^{(k)}}\big)^{-1}\mathbf{H}_{\Lambda^{(k)}}^{\mathrm{T}}\boldsymbol{y}\|^2 \quad (B.1)$$

where $\mathbf{H}_{\Lambda^{(1)}} = [\mathbf{h}_i, \mathbf{h}_{i+T}, \dots, \mathbf{h}_{i+(K-1)T}]$.

For this case, $\mathbf{H}_{\Lambda^{(1)}}^{\mathrm{T}}\mathbf{H}_{\Lambda^{(1)}} = \mathbf{I}_K$ with $\mathbf{I}_K$ stands for the $K{\times}K$ identity matrix. This is because, the IR is normalized such that $\sum_{j=1}^{L_h}\mathcal{H}_j^2 = 1$ so the columns of the matrix which correspond to shifted versions of the IR should have a constant norm $\mathbf{h}_{i+mT}^{\mathrm{T}}\mathbf{h}_{i+mT} = \sum_{j=1}^{L_h}\mathcal{H}_j^2$. Also, for a given i the atoms $\mathbf{h}_{i+mT}$ (for all m) are not correlated (this is because the scalar product of these vectors for different values of m is null) as the IR length

Proof

L_h is assumed to be smaller than the cyclic period T. Thus, Eq. B.2 becomes,

$$i^{(1)} = \arg\min_i \|\boldsymbol{y} - \mathbf{H}_{\Lambda^{(1)}}\mathbf{H}_{\Lambda^{(1)}}^{\mathrm{T}}\boldsymbol{y}\|^2$$

Link between the selections criteria of the Cyclo-MP/Cyclo-OMP and Cyclo-OLS

$$i^{(1)} = \arg\max_i \sum_{m=0}^{(K-1)} |\mathbf{h}_{i+mT}^{\mathrm{T}}\boldsymbol{r}^{(0)}|$$

$$= \arg\max_i \|\mathbf{H}_{\Lambda^{(1)}}^{\mathrm{T}}\mathbf{r}^{(0)}\|^2$$

Which is equivalent to

$$i^{(1)} = \arg\min_i \min_\alpha \|\mathbf{r}^{(0)} - \mathbf{H}_{\Lambda^{(1)}}\alpha\|^2$$

$$\min_\alpha \|\boldsymbol{r}^{(0)} - \mathbf{H}_{\Lambda^{(1)}}\alpha\|^2 = \min_\alpha \boldsymbol{r}^{(0)\mathrm{T}}\boldsymbol{r}^{(0)} - 2\boldsymbol{r}^{(0)\mathrm{T}}\mathbf{H}_{\Lambda^{(1)}}\alpha + \alpha^{\mathrm{T}}\mathbf{H}_{\Lambda^{(1)}}^{\mathrm{T}}\mathbf{H}_{\Lambda^{(1)}}\alpha$$

$$= \min_\alpha - 2\,\boldsymbol{r}^{(0)\mathrm{T}}\mathbf{H}_{\Lambda^{(1)}}\alpha + \alpha^{\mathrm{T}}\alpha \quad (\text{as } \mathbf{H}_{\Lambda^{(1)}}^{\mathrm{T}}\mathbf{H}_{\Lambda^{(1)}} = \mathbf{I}_K)$$

The solution is $\alpha = \mathbf{H}_{\Lambda^{(1)}}^{\mathrm{T}}\boldsymbol{r}^{(0)}$. Therefore,

$$\min_i \|\boldsymbol{r}^{(0)} - \mathbf{H}_{\Lambda^{(1)}}\mathbf{H}_{\Lambda^{(1)}}^{\mathrm{T}}\boldsymbol{r}^{(0)}\|^2 = \min_i \boldsymbol{r}^{(0)\mathrm{T}}\boldsymbol{r}^{(0)} - 2\boldsymbol{r}^{(0)\mathrm{T}}\mathbf{H}_{\Lambda^{(1)}}\mathbf{H}_{\Lambda^{(1)}}^{T}\boldsymbol{r}^{(0)} + \boldsymbol{r}^{(0)\mathrm{T}}\mathbf{H}_{\Lambda^{(1)}}\mathbf{H}_{\Lambda^{(1)}}^{\mathrm{T}}\boldsymbol{r}^{(0)} = \min_i -\boldsymbol{r}^{(0)\mathrm{T}}\mathbf{H}_{\Lambda^{(1)}}\mathbf{H}_{\Lambda^{(1)}}^{\mathrm{T}}\boldsymbol{r}^{(0)}$$

$$= \max_i \|\mathbf{H}_{\Lambda^{(1)}}^{\mathrm{T}}\boldsymbol{r}^{(0)}\|^2$$

As for the first iteration $\boldsymbol{y} = \boldsymbol{r}^{(0)}$ the selection step of the Cyclo-OLS for the first K atoms is identical to the Cyclo-MP/Cyclo-OMP and is based on the sum of the correlation between $\boldsymbol{y}$ and $\mathcal{H}$ at $i + mT$.

For the other iterations, the selection step of the Cyclo-OLS is given by the minimization of $\|\boldsymbol{y} - \mathbf{H}_{\Lambda^{(k)}}\widehat{\boldsymbol{x}}^{(k)}\|^2$ with $\widehat{\boldsymbol{x}}^{(k)} = (\mathbf{H}_{\Lambda^{(k)}}^{\mathrm{T}}\mathbf{H}_{\Lambda^{(k)}})^{-1}\mathbf{H}_{\Lambda^{(k)}}^{\mathrm{T}}\boldsymbol{y}$ being the orthogonal projection of $\boldsymbol{y}$ on the atoms of index in $\Lambda^{(k)} = \Lambda^{(k-1)} \cup \{i_{mT}^{(k)}; m = 1, \dots, K-1\}$, and $\mathbf{H}_{\Lambda^{(k)}}\widehat{\boldsymbol{x}}^{(k)}$ represents the contribution of the estimated impulses (till iteration k) $\widehat{\boldsymbol{x}}^{(k)}$ on $\boldsymbol{y}$.

Hence, the K selected atoms $\{i_{mT}^{(k)}\}$ should minimize jointly the MSE between $\boldsymbol{y}$ and $\mathbf{H}_{\Lambda^{(k)}}\widehat{\boldsymbol{x}}^{(k)}$. Since the minimization is made simultaneously, the impulses of small amplitudes are not penalized if the remaining atoms (for other value of m) bear on the minimization of the MSE.

Consequently, the selection step of the Cyclo-OLS is more efficient than the one of the OLS which minimizes independently the MSE for each atom.

Appendix C. Comparison between the selection step of the Cyclo-SBR (Eq. 12) and the one of the SBR

As the Cyclo-SBR has a similar selection criterion (Eq. 12) as the Cyclo-OLS except the second term, $\beta \,\#\, \{\Lambda^{(k)}\}$, which does not really depend on the selected atoms as it indicates how many atoms are added. Note that for $\beta = 0$ the Cyclo-SBR is reduced to the Cyclo-OLS. Also, for the first iteration ($k{=}1$), there is no added atoms as $\Lambda^{(0)} = \emptyset$ and hence, removal test is inconceivable. Only addition test is possible and is

identical to the Cyclo-OLS with an additional value $\beta \,\#\, \{\Lambda^{(1)}\} = \beta K$. If we leave out the second term of the criterion (Eq. 12), we conclude that Cyclo-SBR behaves in the same way as the Cyclo-OLS for the selection step (mainly for adding K new atoms). Unfortunately, this is not the case for the SBR as it behaves as the OLS (for addition test). In consequence, the selection step of the Cyclo-SBR (Eq. 12) is more efficient than the one of the SBR.

Appendix D. Use of the Matrix inversion lemma

In addition to the matrix-vector products, the Cyclo-OMP, Cyclo-OLS and Cyclo-SBR algorithms require the inversion of the matrix $\mathbf{F}_\Lambda = (\mathbf{H}_\Lambda^{\mathrm{T}}\mathbf{H}_\Lambda)^{-1}$, (for the sake of simplicity the superscript (k) is omitted hereafter) with a growing set of indices [], in particular in the updating step (10) of the Cyclo-OMP and even in the selecting steps (11) and (12) of the Cyclo-OLS and Cyclo-SBR respectively. Following [32, 14, 19, 20, 21] one can take advantage of the matrix inversion lemma to compute iteratively $\mathbf{F}_{\Lambda\cup\{i,\dots,i+(K-1)T\}}$ at a low computation cost, from the knowledge of $\mathbf{F}_\Lambda = (\mathbf{H}_\Lambda^{\mathrm{T}}\mathbf{H}_\Lambda)^{-1}$. Indeed, using a block matrices notation, it can be shown that as $\mathbf{H}_{\widehat{\Lambda}\cup\{i+mT\}} = [\mathbf{H}_{\widehat{\Lambda}}|\mathbf{h}_{(i+mT)}]$ and $\widehat{\Lambda} = \Lambda$, $\mathbf{F}_{\widehat{\Lambda}\cup\{i+mT\}} = \left[\frac{\mathbf{F}_{11}|\mathbf{f}_{12}}{\mathbf{f}_{12}^{\mathrm{T}}|\mathbf{f}_{22}}\right]$:

$$\text{with}\begin{cases} \mathbf{f}_{22} = (\mathbf{h}_{i+mT}^{\mathrm{T}}\mathbf{h}_{i+mT} - \mathbf{h}_{i+mT}^{\mathrm{T}}\mathbf{H}_{\widehat{\Lambda}}\mathbf{F}_{\widehat{\Lambda}}\mathbf{H}_{\widehat{\Lambda}}^{\mathrm{T}}\mathbf{h}_{i+mT})^{-1}, \\[4pt] \mathbf{f}_{12} = -\mathbf{f}_{22}\mathbf{F}_{\widehat{\Lambda}}\mathbf{H}_{\widehat{\Lambda}}^{\mathrm{T}}\mathbf{h}_{i+mT}. \qquad\qquad (D.1) \\[4pt] \mathbf{F}_{11} = \mathbf{F}_{\widehat{\Lambda}} + \mathbf{f}_{22}\mathbf{F}_{\widehat{\Lambda}}\mathbf{H}_{\widehat{\Lambda}}^{\mathrm{T}}\mathbf{h}_{i+mT}\mathbf{h}_{i+mT}^{\mathrm{T}}\mathbf{H}_{\widehat{\Lambda}}\mathbf{F}_{\widehat{\Lambda}}, \end{cases}$$

Such notation can also be used to compute $\mathbf{F}_{\widehat{\Lambda}}$ from $\mathbf{F}_{\widehat{\Lambda}\cup\{i+mT\}}$, which is required in the selection step (12) of the Cyclo-SBR, as

$$\mathbf{F}_{\widehat{\Lambda}} = \mathbf{F_{11}} - f_{22}^{-1}\mathbf{f}_{12}\,\mathbf{f}_{12}^{\mathrm{T}} \qquad (D.2)$$

It should be noted that the relations D.1 and D.2 will be repeated iteratively for each value of $m = [0, 1, \dots, (K-1]$ with of course an updated set $\widehat{\Lambda} = \widehat{\Lambda} \cup \{i + mT\}$ when addition or $\widehat{\Lambda} = \widehat{\Lambda}\backslash\{i + mT\}$ when removal.

Using these relations, the computation of the matrix $\mathbf{F}_{\widehat{\Lambda}\cup\{i+mT\}}$, with an increasing (or eventually

$$\mathcal{J}_{\widehat{\Lambda}\cup\{i,\dots,i+(K-1)T\}} - \mathcal{J}_{\Lambda} = \sum_{m=0}^{K-1}\{-f_{22}(\boldsymbol{y}^{\mathrm{T}}\mathbf{H}_{\widehat{\Lambda}}\mathbf{F}_{\widehat{\Lambda}}\mathbf{H}_{\widehat{\Lambda}}^{\mathrm{T}}\mathbf{h}_{i+mT} - \boldsymbol{y}^{\mathrm{T}}\mathbf{h}_{i+mT})^2 + \beta\} \quad \text{with } \widehat{\Lambda} = \widehat{\Lambda}\cup\{i+mT\} \quad (D.3)$$

$$\mathcal{J}_{\widehat{\Lambda}\backslash\{i,\dots,i+(K-1)T\}} - \mathcal{J}_{\Lambda} = \sum_{m=0}^{K-1}\{-f_{22}^{-1}([\mathbf{f}_{12}^{\mathrm{T}}|f_{22}]\mathbf{H}_{\widehat{\Lambda}}^{\mathrm{T}}\boldsymbol{y})^2 - \beta\} \quad \text{with } \widehat{\Lambda} = \widehat{\Lambda}\backslash\{i+mT\} \qquad (D.4)$$

where $f_{22}, \mathbf{f}_{12}, \mathbf{F}_{\widehat{\Lambda}}$, and $\mathbf{H}_{\widehat{\Lambda}}$ are updated with the relations D.1 and D.2 for each value of m.

Appendix E. Computational cost

Since the computational cost of the studied cyclo-algorithms is roughly the one of their corresponding greedy algorithm divided by K, therefore, the computational cost is given for MP, OMP, OLS and SBR, which can be respectively retrieved from Cyclo-MP, Cyclo-OMP, Cyclo-OLS and Cyclo-SBR taking into account a single period for $m=0$. As the cost of an addition operation is generally negligible compared to a multiplication operation, only multiplication operation is considered in the computation cost. The multiplications

decreasing for the Cyclo-SBR) set of indices $\widehat{\Lambda}$ can be performed at a relative low cost.

Furthermore, the selection steps of the Cyclo-OLS and Cyclo-SBR do not require the computation of the solution but only the computation of the criterion $\mathcal{J}_{\Lambda}$ of (12) (note that eq. (11) is identical to eq. (12) for parameter $\beta = 0$) which can be updated using the previous block matrix notation, in case of addition or removal of the K atoms:

required for each algorithm at a given iteration k are summarized in table E.1. For the SBR, we suppose only addition of atoms (atom removal does not happen), this corresponds to the worst case. It should be noted that this computational cost is founded on the efficient implementation (proposed in [33]) which is based on the convolution operator and not on vector and matrix products as is usually done for sparse approximations. In other respects, many applications in signal and image processing where the computations are expensive from the execution time and from memory storage point of view use parallel approach as [36].

Table E.1 : multiplication required for each algorithm at a given iteration k

		Algorithms			
	Multi.	MP	OMP	OLS	SBR
Full convolution	$L_x L_h$	1 (7)	1 (7)	0	0
Sparse convolution	kL_h	1 (9)	k (11)	0	0
Vector-Matrix product	k^2	0	1 (10)	0	0
Update $\mathbf{F}$ (addition)	$3k^2 + k$	0	1 (10)	1 (11)	1/0 (12)
Addition test	$k^2 + k + L_h$	0	0	$L_x - k + 1$ (D.3)	$L_x - k + 1$ (D.3)
Removal test	k	0	0	0	$k - 1$ (D.4)
Update $\mathbf{F}$ (removal)	$k^2 + k$	0	0	0	0/1 (D.2)

A rough estimate of the maximum number of multiplications of the algorithms for a number M of iterations is given in table E.2

Table E.2 : Maximum number of multiplications of the various algorithms for M of iterations

MP	$L_x L_h M + \frac{1}{2}(M+1)L_h M$
OMP	$L_x L_h M + (\frac{1}{3}M^2 + \frac{1}{2}M + \frac{1}{6})L_h M + (\frac{4}{3}M^2 + \frac{5}{2}M + \frac{7}{6})M$
OLS	$L_x L_h M + (\frac{1}{2}M^2 + M + \frac{1}{2})L_x M + \frac{1}{2}(-M+1)L_h M + \frac{1}{2}(-\frac{1}{3}M^3 + M^2 + 4M + 3)M$
SBR	$L_x L_h M + (\frac{1}{2}M^2 + M + \frac{2}{3})L_x M + \frac{1}{2}(-M+1)L_h M + \frac{1}{2}(-\frac{2}{3}M^3 + \frac{5}{2}M^2 + 4M + \frac{7}{3})M$

The MP has a low computation cost, but may select several times the same atom as the amplitudes are not computed from a joint orthogonal projection. Compared to the MP, the OMP just adds, for each iteration, an orthogonal projection step to compute the amplitude of the selected atoms and *k*-sparse convolution for the residual update, so the additional computation cost is relatively low. The selection step of the OLS is based on the orthogonal projection used in the update step of the OMP, but the computation cost is dramatically reduced thanks to the use of the block matrix inversion and eq. (D.3). Finally, the SBR has a computation cost very similar to the OLS as the removing tests can be computed at a low cost using eq. (D.4).

REFERENCES RÉFÉRENCES REFERENCIAS

1. W. A. Gardner, *Introduction to random processes: with application to signals and systems,* 1986.
2. J. Antoni and R. B. Randall, "The spectral kurtosis: application to the vibratory surveillance and diagnostics of rotating machines," *Mechanical Systems and Signal Processing,* vol. 20, no. 2, pp. 308–331, Feb. 2006.
3. W. A. Gardner, *Cyclostationarity in communications and signal processing,* IEEE PRESS, 1993.
4. Khalid Sabri, "Cyclic Sparse Greedy Deconvolution," IJIGSP, vol. 4, no. 12, pp. 1–8, 2012.
5. K. Sabri, M. El Badaoui, F. Guillet, A. Belli, G. Millet and J.-B. Morin, "Cyclostationary modeling of ground reaction force signals," Signal Processing, vol. 90, no. 4, pp. 1146–1152, April 2010.
6. C. Andrieu, P. Duvaut, and A. Doucet, "Bayesian deconvolution of cyclostationary processes based on point processes,"in EUSIPCO, Trieste, Italy, 10-13 Sep 1996.
7. D. Hatzinakos J. Ilow, "Recursive least squares algorithm for blind deconvolution of channels with cyclostationary inputs,"in IEEE Military Communications Conference, Boston, Massachusetts, 1993, pp. 123–126.
8. D. Hatzinakos, "Nonminimum phase channel deconvolution using the complex cepstrum of the cyclic autocorrelation,"IEEE Trans. on SP, vol. 42, no. 11, pp. 3026–3042, Nov. 1994.
9. L. Cerrato and B. Eisenstein, "Deconvolution of cyclostationary signals," IEEE Trans. on Acoustics, Speech, and SP, vol. 25, no. 6, pp. 466–476, Dec. 1977.
10. K.M. Cheung and S.F. Yau, "Blind deconvolution of system with unknown response excited by cyclostationary impulses,"in IEEE, ICASSP, Detroit, MI,, 9-12 May 1995, pp. 1984–1987.
11. R. Gribonval and M. Zibulevsky, Sparse Component Analysis, chapter Handbook of Blind Source Separation: Independent Component Analysis and Applications, pp. 367–420, Pierre Comon and Christian Jutten (Editors), ELSEVIER, 2010.
12. J. Bobin, J-L. Starck, Y. Moudden, and M.J. Fadili, Advances in Imaging and Electron Physics, chapter Blind Source Separation: The Sparsity Revolution, pp. 221–298, Peter Hawkes Ed., Academic Press, ELSEVIER, 2008.
13. S. Mallat and Z. Zhang, "Matching pursuits with time frequency dictionaries," IEEE Trans. on SP., vol. 41, pp. 3397–3415, 1993.
14. Y.C. Pati, Ramin Rezaiifar, and P.S. Krishnaprasad, "Orthogonal matching pursuit: Recursive function approximation with applications to wavelet decomposition," in the 27th Annual Asilomar, Pacific Grove, CA, USA, 1993.
15. S. Chen, S. A. Billings, and W. Luo, "Orthogonal least squares methods and their application to non-linear system identification," International Journal of Control, vol. 50, no. 5, pp. 1873–1896, 1989.
16. S. Shaobing Chen, David L. Donoho, and Michael A. Saunders, "Atomic decomposition by basis pursuit," SIAM Journal on Scientific Computing, vol. 20, no. 1, pp. 33–61, 1998.
17. Bashkar Rao and Kenneth Kreutz-Delgado, "An affine scaling methodology for best basis selection," IEEE Transaction On Signal Processing, vol. 47, no. 1, pp. 187200, Jan. 1999.
18. Robert Tibshirani, "Regression shrinkage and selection via the lasso," Journal of the Royal Statistical Society Serie B, vol. 58, no. 1, pp. 267–288, 1996.
19. J. Duan, C. Soussen, D. Brie, and J. Idier, "On the properties of the solution path of the constrained and penalized l2-l0 problems," Tech. Rep., Centre de Recherche en Automatique de Nancy, Feb. 2009.
20. Soussen C., J. Idier, D. Brie, and J. Duan, "From bernoulli-gaussian deconvolution to sparse signal restoration," IEEE Trans. on SP, vol. 59, no. 10, pp. 4572–4584, oct 2011.
21. Duan J., Soussen C., Brie D., and Idier J., "Détection conjointe de discontinuités d'ordres différents dans un signal par minimisation de crit`ere L2-L0," in Actes 22ecoll. GRETSI, Dijon, Sep. 2009.
22. Joel A. Tropp, "Greed is good: Algorithmic results for sparse approximation," IEEE Trans. on Information Theory, vol. 50, no. 10, pp. 2231–2242, Oct. 2004.
23. D.L. Donoho, M. Elad, and V.N. Temlyakov, "Stable recovery of sparse overcomplete representations in the presence of noise," vol. 52, no. 1, pp. 6–186, Jan. 2006.
24. J.J. Fuchs, "On sparse representations in arbitrary redundant basis," vol. 50, pp. 1341–1344, 2004.

25. D. L. Donoho and X. Huo, "Uncertainty principles and ideal atomic decomposition," vol. 47, pp. 28452862, 2001.
26. Massimo Fornasier and Holger Rauhut, "Recovery algorithm for vector-valued data with joint sparsity constraints," SIAM Journal on Numerical Analysis, vol. 46, no. 2, pp. 577-613, 2008.
27. J. A. Tropp, A. C. Gilbert, and M. J. Strauss, "Algorithms for simultaneous sparse approximation part I: Greedy pursuit,"*Signal Processing, special issue* "Sparse approximations in signal and image processing,", vol. 86, pp. 572–588, Apr. 2006.
28. A. J. Tropp, "Algorithms for simultaneous sparse approximation. part II: Convex relaxation," vol. 86, pp. 589–602, Apr. 2006.
29. M. Kowalski and B. Torrésani, "Sparsity and persistence: mixed norms provide simple signal models with dependent coefficients," Signal Image and Video Processing, vol. 3, no. 3, pp. 251–264, Sep. 2009.
30. C. Dossal and S. Mallat, "Sparse spike deconvolution with minimum scale," in SPARS05, Rennes, France, Nov. 2005.
31. J. J. Kormylo and J. M. Mendel, "Maximum-likelihood detection and estimation of bernoulli-gaussian processes," vol. 28, no. 3, pp. 482–488, May 1982.
32. Y. Goussard, G. Demoment, and G. Idier, "A new algorithm for iterative deconvolution of sparse spike trains," in IEEE ICASSP, Albuquerque, NM, April 1990, pp. 1547–1550.
33. Khalid Sabri and Hervé Carfantan, "High resolution sparse spike hyperspectral images deconvolution," Submitted to IEEE Journal of Selected Topics in Signal Processing, 2014.
34. A.V. Dandawate and G.B. Giannakis, "Statistical tests for presence of cyclostationarity," IEEE Trans. on Signal Processing, vol. 42, no. 9, pp. 2355 – 2369, Sep. 1994.
35. M. Ehler, "Shrinkage rules for variational minimization problems and applications to analytical ultracentrifugation," Journal of Inverse and Ill-Posed Problems, vol.19, no. 4-5, pp. 593–614, 2011.
36. D. M. Onchis, "Signal Reconstruction in Multi-Windows Spline-Spaces Using the Dual System," IEEE Signal Processing Letters, Vol. 19, no. 11, 729–732, Nov. 2012.
37. E. J. Candes and C. Fernandez-Granda, "Towards a Mathematical Theory of Super-Resolution," journal CoRR, vol.abs/1203.5871, 2012.
38. Bayesian Approach to Inverse Problems, jISTE Ltd and John Wiley & Sons Inc, April 2008.
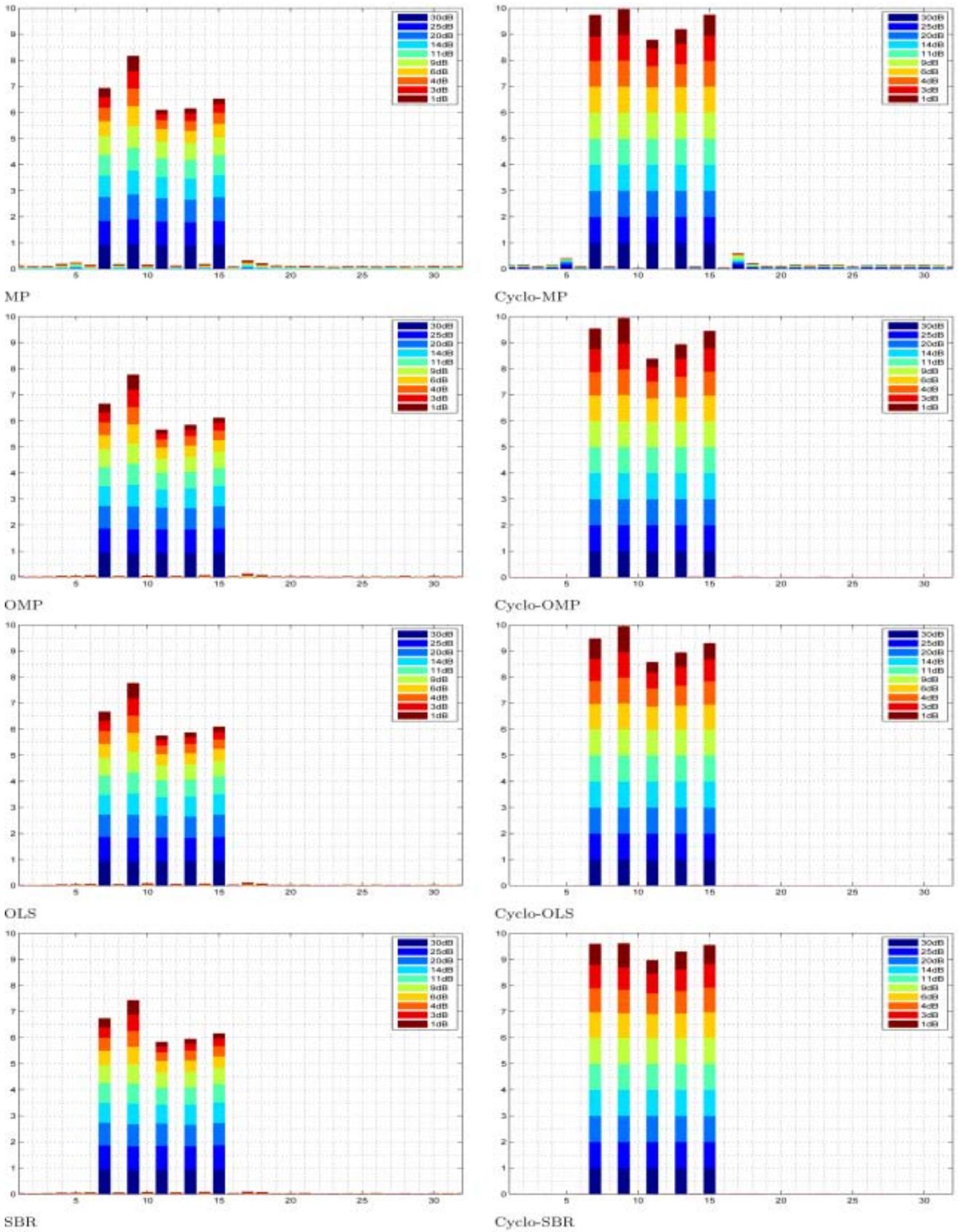
Figure 5 : The effect of varying SNR from 1dB to 30dB over MC runs on the histogram (K=8)
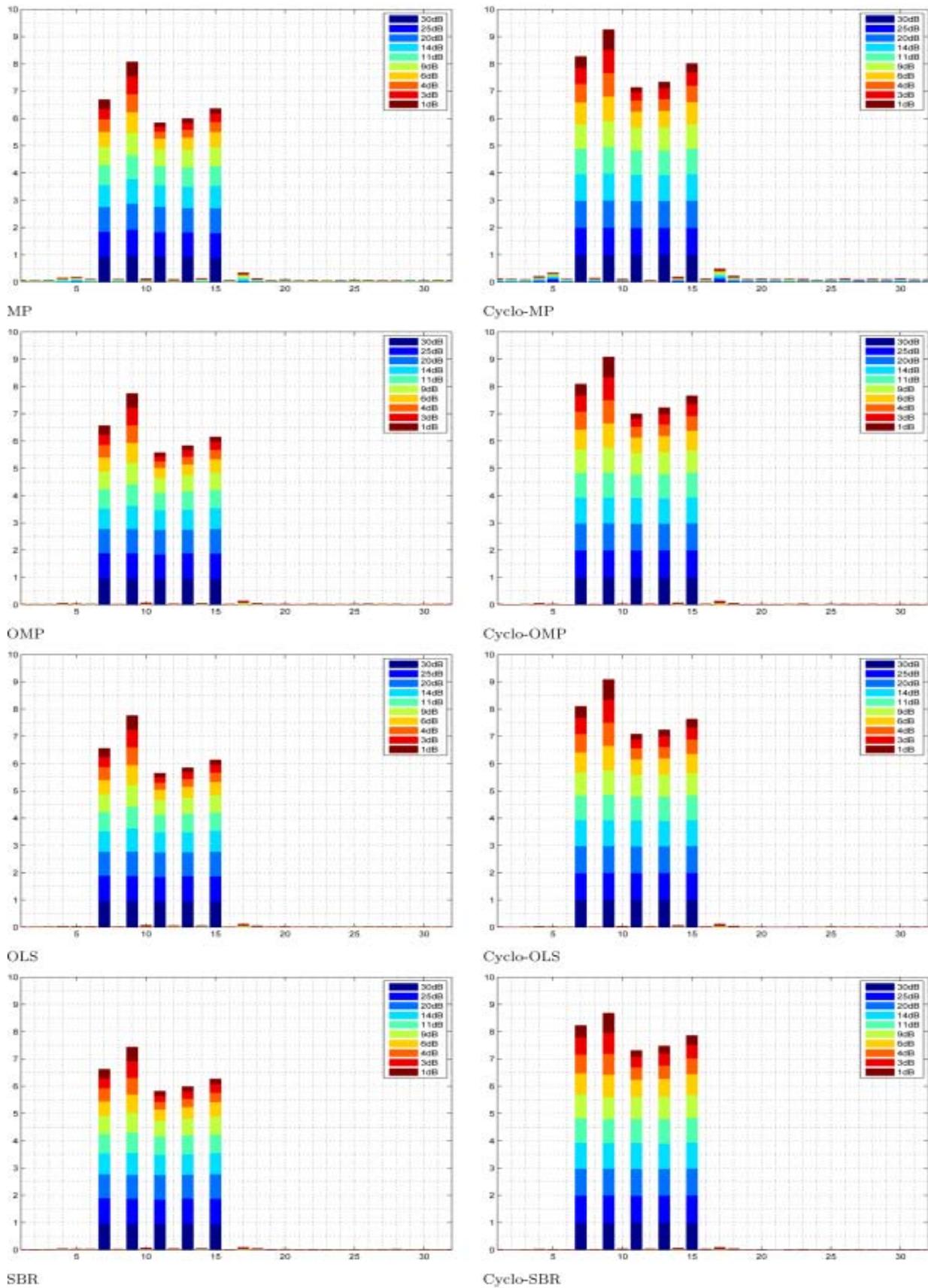
Figure 6 : The effect of varying SNR from 1dB to 30dB over MC runs on the histogram (K=2)
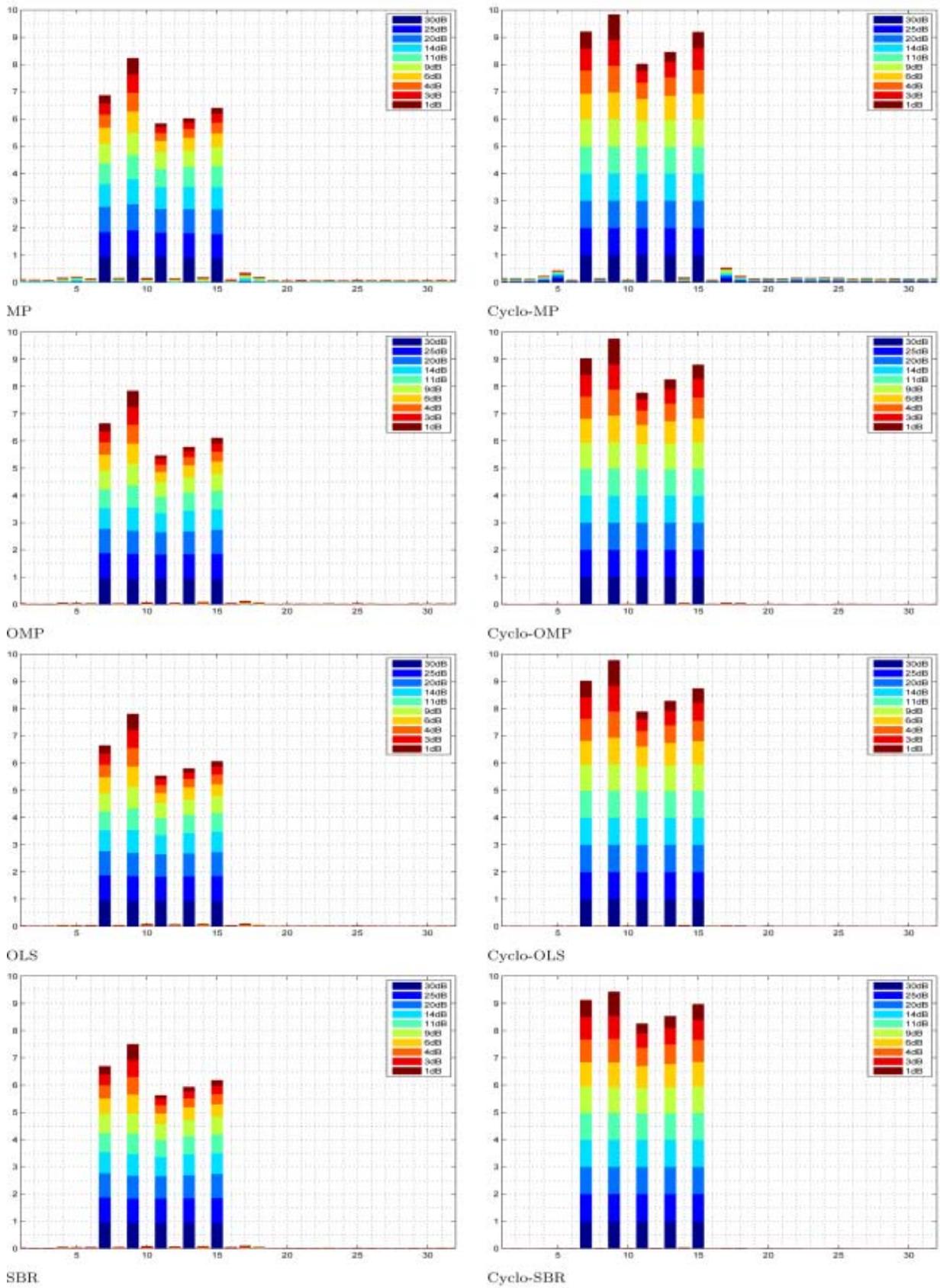
Figure 7 : The effect of varying SNR from 1dB to 30dB over MC runs on the histogram (K=4)
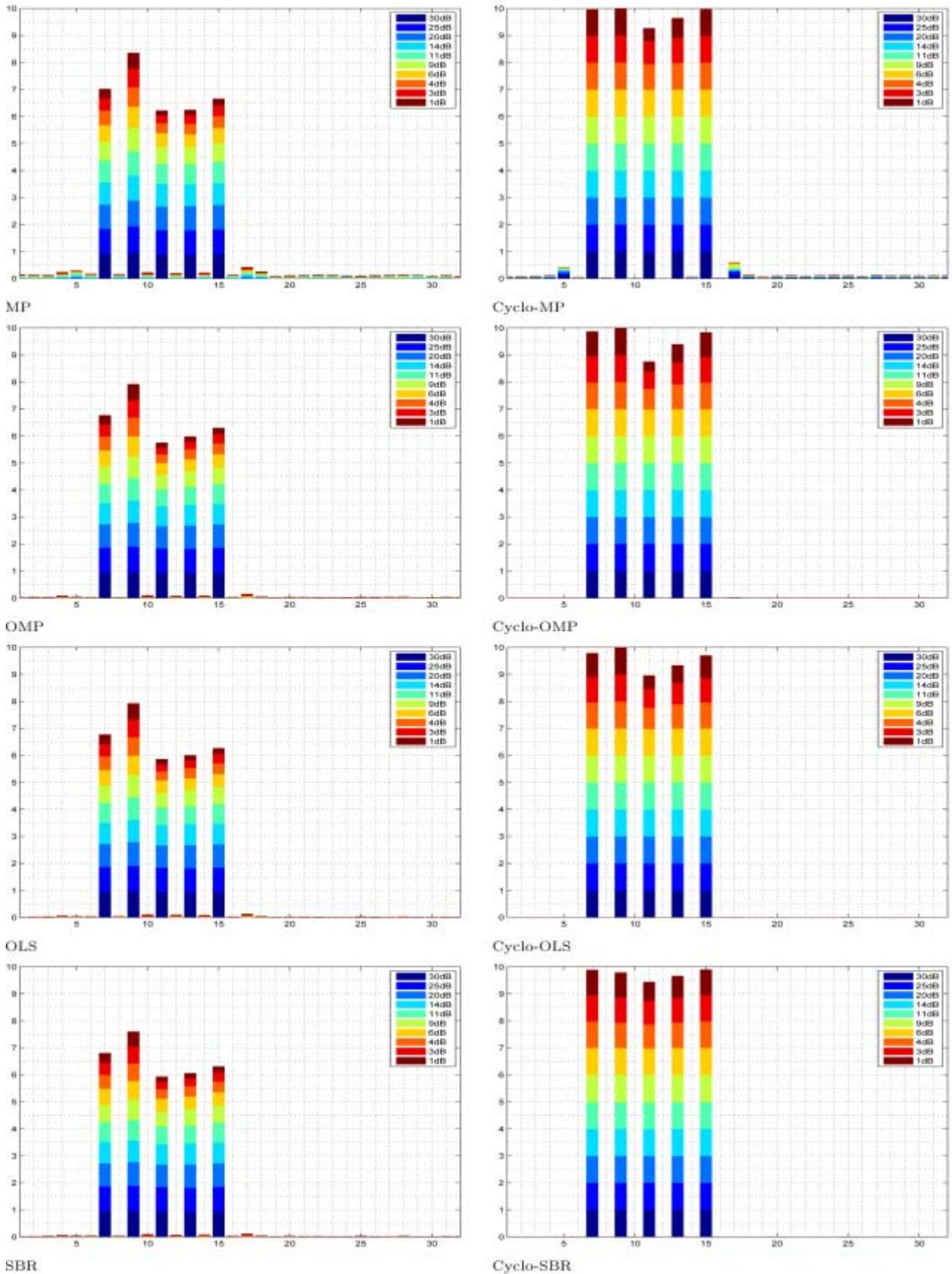
Figure 8 : The effect of varying SNR from 1dB to 30dB over MC runs on the histogram (K=16)

This page is intentionally left blank