CrossMark

New Cyclic Sparsity Measures for Deconvolution Based on Convex Relaxation

Khalid Sabri

Abstract Periodic random impulse signals are powerful tools for various situations of interest and are a natural way for modeling highly localized events occurring randomly at given times. However, the measured impulses are often masked because of unwanted convolution and eventually drowned in noise. Thus, the resulting signal is not legible and may lead to poor or even erroneous analysis and, hence, the need of deconvolution to recover the random periodic impulses. As a matter of fact, periodic random impulse signals are sparse and the sparse coefficients are periodically correlated. It has thus emerged as to how to combine the data structure and the sparsity jointly for a best description. The originality of this study lies in the design of new measures of cyclic sparsity property for the deconvolution of signals that are simultaneously sparse and cyclostationary. To our knowledge, all related works in this field exploit only one property, either sparsity or cyclostationarity and never both properties together. The key feature of cyclic sparsity deconvolution is that it combines the cyclic structure and the sparsity together which implies a significantly enhanced performance. Finally, we include examples of computer simulations to illustrate the behavior in deconvolution context of the proposed algorithms against an ℓ_1 sparse deconvolution through convex optimization. We show that deconvolution based on cyclic sparsity hypothesis increases the performance and reduces significantly the computation cost as well.

Keywords Deconvolution · Periodic random impulses · Cyclic sparsity · Cyclostationary · Convex relaxation

K. Sabri (✉)
STIC Laboratory, Department of Physics, Faculty of Sciences, University Chouaïb Doukkali, El Jadida, Morocco
e-mail: sabri.k@ucd.ac.ma

Birkhäuser

1 Introduction

The past 50 years have seen increasing theoretical and practical advances in cyclo-stationary analysis. Thus, cyclostationarity has become an intense area of research in signal processing and statistics. It has since been shown that cyclostationary frame-work is appropriate for any physical phenomenon that gives rise to data with periodic statistical characteristics [15] such as mechanics [2], telecommunications [16], biome-chanics [24]. In this article, we focus on wide-sense second-order cyclostationary, i.e., first-order and second-order cyclostationary. The signals of interest are assumed to be made of periodic impulses with random amplitudes, namely a 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 sys-tem in the presence of noise. Indeed, enhancing the resolution of the signal and the signal-to-noise ratio (SNR) from the knowledge of the IR matches to a deconvolution problem.

The deconvolution of cyclostationary signals has been addressed by several authors with different approaches. In [1], a Bayesian deconvolution algorithm based on Markov chain Monte Carlo is presented. Cyclic statistics are often used for deconvolution, in [18,19] the deconvolution is based on cyclic cepstrum, whereas in [6,8] the deconvo-lution is based on cyclic correlation. Despite this encouraging perspective, actual con-tributions of cyclostationarity to deconvolution have remained very limited. The main drawback of these methods is their inability to detect and restore impulses drowned in noise.

As a matter of fact, signal deconvolution belongs to inverse problems and is par-ticularly well known to be an ill-posed problem since the IR acts generally as a low-pass filter and the convolved signal is always noisy. Fortunately, regularization meth-ods lead to acceptable solutions accounting for a priori information on the original object.

Analyzing periodic random impulse signals in detail uncovers another a priori infor-mation which is sparsity, since only few impulses are nonzero. Thus, data are sparse on the direct domain. The biggest issue that we are dealing now is how to exploit the periodic character jointly with sparsity of periodic random impulses. This is possi-ble thanks to the proposed criterions that gather both properties to better characterize these signals. Thus, cyclic sparse deconvolution can be performed taking benefit of the correlation between the impulses at neighborhood cycles.

Recently, in a different framework, sparse approximation has become a subject of active research. 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 [4,12,17].

In contrast 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 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 classes: convex relaxation and greedy pursuit algorithms. 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 [22,29,30]. 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 [7,20,21].

Our work consists of extending sparse approximation to cyclostationary signals with periodic random impulses, where the main aim is to design appropriate cyclic sparsity measures. Analyses and algorithms for the traditional ℓ_1 norm can readily be extended to these new norms, making them more efficient and flexible, whereas we insist on the application of cyclic sparse approximation to deconvolution, i.e., the dictionary is obviously given by the Toeplitz matrix formed by the IR.

This correspondence is organized as follows, Sect. 2 defines the problem statement and motivations of the study. In Sect. 3, we summarize the statement of sparse approximation problem. Such a sparse model is taken into account in sparse deconvolution by convex relaxation algorithms in Sect. 4. The main contributions of this paper are described in Sects. 5, 6 and 7. New cyclic sparsity measures are taken into account in sparse deconvolution by convex relaxation; this leads to the Cyclic LASSO, Referenced Cyclic LASSO and Joint Cyclic LASSO problems which are actually extensions of the LASSO. Finally, we propose to test the solutions of the LASSO and the three proposed algorithms on the same statistical basis, i.e., with the same stopping rule deduced from statistical properties of the noise. Thus, some numerical experiments have been performed in order to compare these algorithms as shown in the simulation results in Sect. 8.

2 Problem Formulation

2.1 Problem Statement

Consider the situation where a known system $\mathscr{H}(t)$ 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 $\sum_{k=1}^{d} x_k \mathscr{H}(t -$
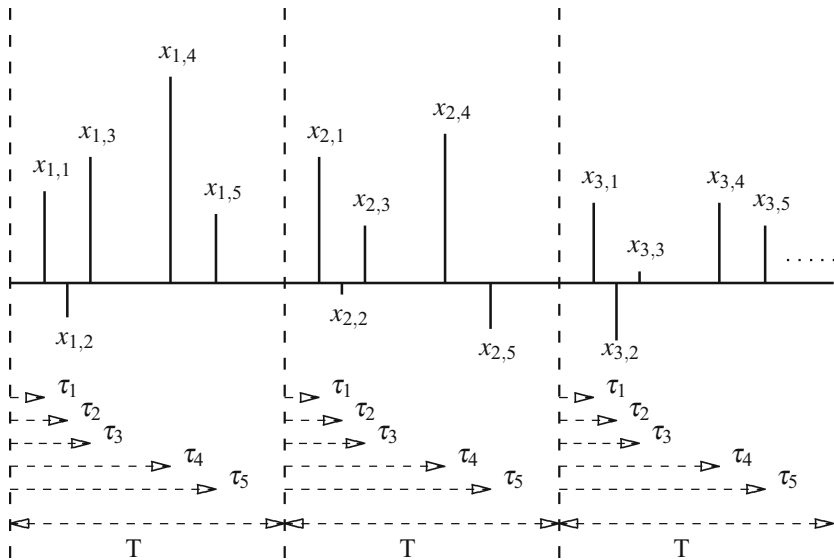
Fig. 1 Example of cyclic sparse signal $x(t)$

$\tau_k) + n(t), \ 0 \leq t < T$. This relationship can be generalized to cover the whole signal as

$$y(t) = \sum_{i=0}^{K-1} \sum_{k=1}^{d} x_{i,k} \, \mathscr{H}(t - (\tau_k + iT)) + n(t), \tag{1}$$

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 subindex 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.

2.2 Motivations 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 [1,8,18,19].

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, as far as we are aware, 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 performance and reduction in the computation cost as well. 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 cyclic sparsity that gathers both properties to better characterize this kind of signals.

3 Reminder of 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 x such that $\Phi x \approx y$, where y corresponds to measured data and Φ is a known matrix with atoms $\{\phi_k\}_{k=1\dots Q}$ 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 with respect to x:

$$||y - \Phi x||_2^2 + \beta_1 ||x||_0, \tag{2}$$

where

- $||x||_0 = \mathrm{Card}\{k, x_k \neq 0\}$ is the number of nonzero components of the column vector x,
- $x = [x_1, \dots, x_{L_x}]^{\mathrm{T}}$ being the column vector of length L_x constructed from the signal $x(t)$,
- the superscript T stands for the transpose of a vector or a matrix.

The parameter β_1 controls the trade-off between the sparsity of the solution and the quality of the approximation. The lower is β_1, less sparse is the solution and better is the approximation. Hence, β_1 is the key parameter to reach the compromise.

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 improve the approximation quality [20,31] and (2) convex relaxation algorithms, i.e., based on the relaxation of the criterion (2), which replace the combinatorial problem with an easier optimization problem often chosen convex [29]. In the latter, the ℓ_0 norm is often relaxed with a ℓ_p norm, where $||x||_p = (\sum_k |x_k|^p)^{\frac{1}{p}}$ (note that $p = \infty$ is a limiting case for which $||x||_\infty = \max_k |x_k|$). For $p = 1$, this problem corresponds to the least absolute shrinkage and selection operator (LASSO) regression [30] or basis pursuit denoising (BPDN) in signal processing [29].

In this work, the interest is focused in the use of sparse approximation techniques in deconvolution framework for signals that are jointly sparse and cyclostationary. Both approaches used to avoid exploring every combination of the sparse approximation problem can be extended to cyclosparse context: greedy algorithms [25,26] and convex relaxation. In this paper, we focus on convex relaxation algorithms. Before announcing and describing the proposed criterions, let us first recall the LASSO problem in deconvolution context.

4 Sparse Approximation for Deconvolution

4.1 Problem Formulation

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 IR. Of course, such boundary hypothesis influences size and structure of the dictionary $\mathbf{H}$ formed from the IR. In particular, the columns of the matrix which correspond to shifted versions of the IR should have a constant ℓ_2 norm $\mathbf{h}_i^{\mathrm{T}} \mathbf{h}_i = \sum_{j=1}^{L_h} \mathscr{H}_j^2 = \|\mathscr{H}\|_2^2$.

Note that $\mathbf{H}$ is a sparse matrix of dimension $L_y \times L_x$ with $L_h \times L_x$ nonzero elements (the length of the IR 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 IR, matrix $\mathbf{H}$ is composed only with the L_h elements of the IR. Moreover, as matrix $\mathbf{H}$ models a convolution operator, it has a Toeplitz structure (diagonal-constant matrix) and each operation involving $\mathbf{H}$ may be computed as a result of a convolution. Consequently, the model (1) can be written with matrix notations as:

$$\mathbf{y} = \mathbf{H}\mathbf{x} + \mathbf{n},$$

where the vectors are defined as follows: $\mathbf{x} = [x_1, \ldots, x_{L_x}]^{\mathrm{T}}$, $\mathbf{y} = [y_1, \ldots, y_{L_y}]^{\mathrm{T}}$ and $\mathbf{n} = [n_1, \ldots, n_{L_y}]^{\mathrm{T}}$.

4.2 LASSO Problem

Estimating least squares (LS) parameters subject to an ℓ_1 penalty was presented and popularized independently under the names LASSO [30] and BPDN [29]. The LASSO has become the dominant expression describing this problem, and for the remainder of the paper, we will use the term LASSO to denote the LS problem with ℓ_1 regularization.

$$
\begin{aligned}
\mathscr{J}_l(\mathbf{x}) &= \frac{1}{2} \|\mathbf{y} - \mathbf{H}\mathbf{x}\|_2^2 + \beta_1 \|\mathbf{x}\|_1 \\
&= \frac{1}{2} \sum_{i=1}^{L_y} \left(y_i - \sum_{j=1}^{L_x} H_{i,j} x_j \right)^2 + \beta_1 \sum_{j=1}^{L_x} |x_j|
\end{aligned}
\tag{3}
$$

with $H_{i,j} = (\mathbf{H})_{i,j}$. $\hat{\mathbf{x}}(\beta_1) = \arg\min_{\mathbf{x}} \mathscr{J}_l(\mathbf{x})$ corresponds to the LASSO problem in the equivalent Lagrangian form. Computing the optimal LASSO parameters is a convex optimization problem, and thus, any local minimum found is guaranteed to be a global minimum.

Tibshirani [30] presented several different methods for optimizing the LASSO, each of which differed significantly from the method used in [29]. Since these original works, there have been a wide variety of approaches proposed for the LASSO minimization problem. Many have simply been minor variations on methods already

proposed, but many offer unique ways of viewing the problem. In addition, as the algorithms may give different solutions, one should test all the existing algorithms. Schmidt et al. [27,28] review and compare state-of-the-art optimization techniques for solving the problem of minimizing a twice-differentiable loss function subject to ℓ_1 regularization. The methods are roughly divided into their approach for handling the non-differentiability of the objective function:

- sub-gradient methods: methods that directly optimize (3) and use sub-gradients at points of non-differentiability.
- unconstrained approximations: methods that circumvent the lack of smoothness by optimizing a smooth surrogate objective function.
- constrained optimization methods: the unconstrained non-differentiable problem is transformed into an equivalent differentiable problem with constraints.

A powerful and famous approach for computing the LASSO solution is the simple coordinate descent. This idea was proposed by Fu [14] and Daubechies et al. [10], and later studied and generalized by Friedman et al. [13], Wu and Lange [33] and others. The idea is to fix the penalty parameter β_1 in the Lagrangian form (3) and optimize successively over each parameter, holding the other parameters fixed at their current values. Even though the second term is not differentiable, we can easily compute a simple closed-form solution to this problem using sub-differential calculus with respect to x_j; We consider selecting (coordinate-wise) the following sub-gradient $\nabla_j^s \mathscr{J}_l(\boldsymbol{x})$ (the superscript s denotes sub-gradient and the index j stands for the one of x_j) for each x_j,

$$
\nabla_j^s \mathscr{J}_l(\boldsymbol{x}) = \begin{cases} \|\mathscr{H}\|_2^2 x_j + \mathbf{h}_j^{\mathrm{T}} \mathbf{H}_{\neq j} \boldsymbol{x}_{\neq j} - \mathbf{h}_j^{\mathrm{T}} \boldsymbol{y} + \beta_1 \operatorname{sign}(x_j), & |x_j| > 0 \\ |\|\mathscr{H}\|_2^2 x_j + \mathbf{h}_j^{\mathrm{T}} \mathbf{H}_{\neq j} \boldsymbol{x}_{\neq j} - \mathbf{h}_j^{\mathrm{T}} \boldsymbol{y}| \leq \beta_1, & x_j = 0 \end{cases}
$$

where $\mathbf{H}_{\neq j}$ and $\boldsymbol{x}_{\neq j}$ denote, respectively, $\mathbf{H}$ and $\boldsymbol{x}$ without the jth component and $\mathbf{h}_j = \mathbf{H}_{\{j\}}$; hence, their sizes are $L_y \times L_x - 1$ for $\mathbf{H}_{\neq j}$ and $L_x - 1$ for $\boldsymbol{x}_{\neq j}$. The signum function takes on the sign of x_j if x_j is nonzero, and if x_j is zero, then the signum function can take any value in the range $[-1; 1]$.

More precisely,

$$
\nabla_j^s \mathscr{J}_l(\boldsymbol{x}) = \begin{cases} \|\mathscr{H}\|_2^2 x_j + \mathbf{h}_j^{\mathrm{T}} \mathbf{H}_{\neq j} \boldsymbol{x}_{\neq j} - \mathbf{h}_j^{\mathrm{T}} \boldsymbol{y} + \beta_1, & x_j > 0 \\ \|\mathscr{H}\|_2^2 x_j + \mathbf{h}_j^{\mathrm{T}} \mathbf{H}_{\neq j} \boldsymbol{x}_{\neq j} - \mathbf{h}_j^{\mathrm{T}} \boldsymbol{y} - \beta_1, & x_j < 0 \\ |\|\mathscr{H}\|_2^2 x_j + \mathbf{h}_j^{\mathrm{T}} \mathbf{H}_{\neq j} \boldsymbol{x}_{\neq j} - \mathbf{h}_j^{\mathrm{T}} \boldsymbol{y}| \leq \beta_1, & x_j = 0 \end{cases}
$$

Explicitly, the solution is

$$
x_j = \begin{cases} \frac{1}{\|\mathscr{H}\|_2^2} (\mathbf{h}_j^{\mathrm{T}} \boldsymbol{y} - \mathbf{h}_j^{\mathrm{T}} \mathbf{H}_{\neq j} \boldsymbol{x}_{\neq j} - \beta_1), & \mathbf{h}_j^{\mathrm{T}} \boldsymbol{y} - \mathbf{h}_j^{\mathrm{T}} \mathbf{H}_{\neq j} \boldsymbol{x}_{\neq j} > \beta_1 \\ \frac{1}{\|\mathscr{H}\|_2^2} (\mathbf{h}_j^{\mathrm{T}} \boldsymbol{y} - \mathbf{h}_j^{\mathrm{T}} \mathbf{H}_{\neq j} \boldsymbol{x}_{\neq j} + \beta_1), & \mathbf{h}_j^{\mathrm{T}} \boldsymbol{y} - \mathbf{h}_j^{\mathrm{T}} \mathbf{H}_{\neq j} \boldsymbol{x}_{\neq j} < -\beta_1 \\ 0, & |\mathbf{h}_j^{\mathrm{T}} \boldsymbol{y} - \mathbf{h}_j^{\mathrm{T}} \mathbf{H}_{\neq j} \boldsymbol{x}_{\neq j}| \leq \beta_1 \end{cases}
$$

Or more elegantly,

$$x_j = \frac{1}{\|\mathscr{H}\|_2^2} \mathscr{S}_{\beta_1} \left(\mathbf{h}_j^{\mathrm{T}} (\mathbf{y} - \mathbf{H}_{\neq j} \mathbf{x}_{\neq j}) \right) \tag{4}$$

The first argument of the soft thresholding operator $\mathscr{S}(.)$ (see Appendix 1) is the simple LS coefficient of the partial residual on the atom $\mathbf{h}_j^{\mathrm{T}}$. Repeated iteration of (4), cycling through each variable in turn until convergence, yields the LASSO estimate $\hat{\mathbf{x}}(\beta_1)$.

5 Deconvolution in Cyclic Sparsity Context Through Convex Relaxation

5.1 Background

We now focus on the extension of sparse convex relaxation approximation to cyclic sparsity context for deconvolution. The dictionary is naturally given by the Toeplitz matrix $\mathbf{H}$ deduced from the IR $\mathscr{H}$. The principle of the proposed cyclic sparse convex relaxation algorithms is to relax the ℓ_0 norm penalty function by convex penalty functions that encourage periodic sparsity of coefficients. To do so, we propose to combine two penalties with LS; this leads to three cost functions which the optimization encourage cyclostationarity and sparsity and are given under the following terms: Cyclic LASSO problem, Referenced Cyclic LASSO problem and Joint Cyclic LASSO problem.

5.2 Cyclic LASSO Problem

The function to optimize is given by,

$$\mathscr{J}_{cl}(\mathbf{x}) = \frac{1}{2} \|\mathbf{y} - \mathbf{H}\mathbf{x}\|_2^2 + \beta_1 \|\mathbf{x}\|_1 + \beta_2 \|\mathbf{A}_{L_x,T} |\mathbf{x}| \|_1, \tag{5}$$

where $\mathbf{A}_{L_x,T}$ is an $L_x \times L_x$ circulant square and sparse matrix with $A_{i,i} = -1$ and $A_{i,\mathrm{modulo}(i+T,Lx)} = 1$ (see Appendix 2 for more details). Of course, the cyclic sparsity measure is given by the term $\|\mathbf{A}_{L_x,T} |\mathbf{x}| \|_1$ and corresponds to,

$$\|\mathbf{A}_{L_x,T} |\mathbf{x}| \|_1 = (\mathbf{A}_{L_x,T} \mathbf{S} \mathbf{x})^{\mathrm{T}} \mathrm{sign}(\mathbf{A}_{L_x,T} |\mathbf{x}|)$$
$$= \sum_{j=1}^{L_x} (|x_{j+T}| - |x_j|) * \mathrm{sign}(|x_{j+T}| - |x_j|)$$
$$= \cdots + (|x_{j+T}| - |x_j|) * \mathrm{sign}(|x_{j+T}| - |x_j|) + \cdots$$
$$+ (|x_j| - |x_{j-T}|) \mathrm{sign}(|x_j| - |x_{j-T}|) + \cdots \tag{6}$$

where $\mathbf{S} = \mathrm{diag}(\mathrm{sign}(\mathbf{x}))$ is a diagonal matrix.

Criterion (5) can actually be more conveniently rephrased as:

$$\mathscr{J}_{cl}(\boldsymbol{x}) = \frac{1}{2}\sum_{i=1}^{L_y}\left(y_i - \sum_{j=1}^{L_x}H_{i,j}x_j\right)^2 + \beta_1\sum_{j=1}^{L_x}|x_j|$$

$$+ \beta_2\sum_{j=1}^{L_x}(|x_{j+T}| - |x_j|) * \mathrm{sign}(|x_{j+T}| - |x_j|) \tag{7}$$

It should be noted that indices $j \pm T$ are computed circularly, i.e., modulo$(j \pm T, Lx)$. The first penalty function encourages sparsity in the coefficients, whereas the second penalty encourages sparsity in their cyclic differences at neighborhood cycles.

5.3 Referenced Cyclic LASSO Problem

We propose the following cost function:

$$\mathscr{J}_{rcl}(\boldsymbol{x}) = \frac{1}{2}||\boldsymbol{y} - \mathbf{H}\boldsymbol{x}||_2^2 + \beta_1||\boldsymbol{x}||_1 + \beta_2||\mathbf{Q}_{L_x,T}|\boldsymbol{x}|\,||_1, \tag{8}$$

where $\mathbf{Q}_{L_x,T}$ (with $L_x = KT$) is an $(L_x - T) \times L_x$ circulant sparse matrix. The T first elements of the main diagonal are nonzero and are equal to -1; also, the $-T$th and Tth off-diagonal elements are, respectively, set to -1 and 1 (see Appendix 3 for more details). In this case, the cyclic sparsity measure is given by the term $||\mathbf{Q}_{L_x,T}|\boldsymbol{x}|\,||_1$ and corresponds to,

$$||\mathbf{Q}_{L_x,T}|\boldsymbol{x}|\,||_1 = (\mathbf{Q}_{L_x,T}\mathbf{S}\boldsymbol{x})^{\mathrm{T}}\mathrm{sign}(\mathbf{Q}_{L_x,T}|\boldsymbol{x}|)$$

$$= \frac{1}{K-1}\sum_{j=1}^{T}\sum_{k=1}^{K-1}|\,|x_{j+kT}| - |x_j|\,|$$

$$= \frac{1}{K-1}\sum_{j=1}^{T}\sum_{k=1}^{K-1}(|x_{j+kT}| - |x_j|)\,\mathrm{sign}(|x_{j+kT}| - |x_j|) \tag{9}$$

Moreover, relationship (8) can be more conveniently expressed as:

$$\mathscr{J}_{rcl}(\boldsymbol{x}) = \frac{1}{2}\sum_{i=1}^{L_y}\left(y_i - \sum_{j=1}^{L_x}H_{i,j}x_j\right)^2 + \beta_1\sum_{j=1}^{L_x}|x_j|$$

$$+ \beta_2\frac{1}{K-1}\sum_{j=1}^{T}\sum_{k=1}^{K-1}(|x_{j+kT}| - |x_j|)\,\mathrm{sign}(|x_{j+kT}| - |x_j|) \tag{10}$$

The second penalty encourages sparsity of coefficients in their cyclic differences simultaneously according to a reference cycle, the first cycle for example, for the whole remaining cycles. The reference cycle can be any cycle of the signal and not necessarily the first one.

5.4 Joint Cyclic LASSO Problem

We consider the criterion given below:

$$\mathscr{J}_{jcl}(\boldsymbol{x}) = \frac{1}{2}||\boldsymbol{y} - \mathbf{H}\boldsymbol{x}||_2^2 + \beta_1||\boldsymbol{x}||_1 + \beta_2||\mathbf{N}_{L_x,T}|\boldsymbol{x}|\,||_1, \tag{11}$$

where $\mathbf{N}_{L_x,T}$ is an $L_x \times L_x$ circulant square and sparse matrix. The main diagonal elements are set to $\frac{K-1}{K}$, and the $\pm kT$th off-diagonal elements ($k = 1, \ldots, K - 1$) are set to $-\frac{1}{K}$ (see Appendix 4 for more details). The third and last cyclic sparsity measure $||\mathbf{N}_{L_x,T}|\boldsymbol{x}|\,||_1$ corresponds to,

$$||\mathbf{N}_{L_x,T}|\boldsymbol{x}|\,||_1 = (\mathbf{N}_{L_x,T}\mathbf{S}\boldsymbol{x})^{\mathrm{T}}\mathrm{sign}(\mathbf{N}_{L_x,T}|\boldsymbol{x}|)$$

$$= \frac{1}{K-1}\sum_{j=1}^{T}\sum_{k=0}^{K-1}|\,|x_{j+kT}| - m_j|$$

$$= \frac{1}{K-1}\sum_{j=1}^{T}\sum_{k=0}^{K-1}(|x_{j+kT}| - |x_j|)\,\mathrm{sign}(|x_{j+kT}| - m_j), \tag{12}$$

where $\boldsymbol{m} = [m_1, \ldots, m_j, \ldots, m_T] = \frac{1}{K}\left[\sum_{l=0}^{K-1}|x_{1+lT}|, \ldots, \sum_{l=0}^{K-1}|x_{j+lT}|, \ldots, \sum_{l=0}^{K-1}|x_{T+lT}|\right]$ represents the synchronous mean of the absolute value of $\boldsymbol{x}$.

In addition, criterion (11) can be more conveniently rewritten as:

$$\mathscr{J}_{jcl}(\boldsymbol{x}) = \frac{1}{2}\sum_{i=1}^{L_y}\left(y_i - \sum_{j=1}^{L_x}H_{i,j}x_j\right)^2 + \beta_1\sum_{j=1}^{L_x}|x_j|$$

$$+ \beta_2\frac{1}{K-1}\sum_{j=1}^{T}\sum_{k=0}^{K-1}(|x_{j+kT}| - |x_j|)\,\mathrm{sign}(|x_{j+kT}| - m_j) \tag{13}$$

The second penalty encourages sparsity of coefficients in their cyclic differences simultaneously according to the absolute synchronous mean. The absolute synchronous mean gives somewhat a common and joint reference to the whole cycles better than choosing a reference cycle.

Note that all of the proposed cost functions are convex and, hence, have a unique minimum. The hyperparameters β_1 and β_2 control the trade-off between the cyclic sparsity of the solution and the quality of the approximation.

6 Proposed Solutions

Let $\mathscr{J}(\boldsymbol{x})$ denotes criterions (5, 8 or 11) and consider the sub-gradient $\nabla_j^s\mathscr{J}(x_j)$ with respect to x_j. This leads to,

$$\nabla_j^s \mathscr{J}(\boldsymbol{x}) = \begin{cases} \|\mathscr{H}\|_2^2 x_j + \mathbf{h}_j^{\mathrm{T}} \mathbf{H}_{\neq j} \boldsymbol{x}_{\neq j} - \mathbf{h}_j^{\mathrm{T}} \boldsymbol{y} + \mathrm{sign}(x_j)\,[\beta_1 + \beta_2 \mu], & |x_j| > 0 \\ |\|\mathscr{H}\|_2^2 x_j + \mathbf{h}_j^{\mathrm{T}} \mathbf{H}_{\neq j} \boldsymbol{x}_{\neq j} - \mathbf{h}_j^{\mathrm{T}} \boldsymbol{y}| \leq \beta_1 + \beta_2 \mu, & x_j = 0 \end{cases}$$

where μ depends on the chosen cost function,

- Cyclic LASSO
    $\mu_{cl} = \big(\mathrm{sign}(|x_j| - |x_{j-T}|) - \mathrm{sign}(|x_{j+T}| - |x_j|)\big),$
    μ_{cl} takes values in the range $[-2, -1, 0, 1, 2]$.
- Referenced cyclic LASSO
    $\mu_{rcl} = -\frac{1}{K-1} \sum_{k=1}^{K-1} \mathrm{sign}(|x_{j+kT}| - |x_j|),$
    μ_{rcl} takes values in the discrete set $\frac{1}{K-1}[-(K-1), \ldots, -1, 0, 1, \ldots, (K-1)]$.
- Joint Cyclic LASSO
    $\mu_{jcl} = \frac{1}{K} \Big((K-1)\mathrm{sign}(|x_j| - m_j) - \sum_{k=1}^{K-1} \mathrm{sign}(|x_{j+kT}| - m_j)\Big),$
    μ_{jcl} takes values in the discrete set $\frac{1}{K}[-2(K-1), \ldots, -1, 0, 1, \ldots, 2(K-1)]$.

More specifically,

$$\nabla_j^s \mathscr{J}(\boldsymbol{x}) = \begin{cases} \|\mathscr{H}\|_2^2 x_j + \mathbf{h}_j^{\mathrm{T}} \mathbf{H}_{\neq j} \boldsymbol{x}_{\neq j} - \mathbf{h}_j^{\mathrm{T}} \boldsymbol{y} + \beta_1 + \beta_2 \mu, & x_j > 0 \\ \|\mathscr{H}\|_2^2 x_j + \mathbf{h}_j^{\mathrm{T}} \mathbf{H}_{\neq j} \boldsymbol{x}_{\neq j} - \mathbf{h}_j^{\mathrm{T}} \boldsymbol{y} - \beta_1 - \beta_2 \mu, & x_j < 0 \\ |\|\mathscr{H}\|_2^2 x_j + \mathbf{h}_j^{\mathrm{T}} \mathbf{H}_{\neq j} \boldsymbol{x}_{\neq j} - \mathbf{h}_j^{\mathrm{T}} \boldsymbol{y}| \leq \beta_1 + \beta_2 \mu, & x_j = 0 \end{cases}$$

Explicitly, the solution of the (Referenced/Joint) Cyclic LASSO problem is

$$x_j = \begin{cases} \frac{1}{\|\mathscr{H}\|_2^2}(\mathbf{h}_j^{\mathrm{T}} \boldsymbol{y} - \mathbf{h}_j^{\mathrm{T}} \mathbf{H}_{\neq j} \boldsymbol{x}_{\neq j} - \beta_1 - \beta_2 \mu), & \mathbf{h}_j^{\mathrm{T}} \boldsymbol{y} - \mathbf{h}_j^{\mathrm{T}} \mathbf{H}_{\neq j} \boldsymbol{x}_{\neq j} > \beta_1 + \beta_2 \mu \\ \frac{1}{\|\mathscr{H}\|_2^2}(\mathbf{h}_j^{\mathrm{T}} \boldsymbol{y} - \mathbf{h}_j^{\mathrm{T}} \mathbf{H}_{\neq j} \boldsymbol{x}_{\neq j} + \beta_1 + \beta_2 \mu), & \mathbf{h}_j^{\mathrm{T}} \boldsymbol{y} - \mathbf{h}_j^{\mathrm{T}} \mathbf{H}_{\neq j} \boldsymbol{x}_{\neq j} < -\beta_1 - \beta_2 \mu \\ 0, & |\mathbf{h}_j^{\mathrm{T}} \boldsymbol{y} - \mathbf{h}_j^{\mathrm{T}} \mathbf{H}_{\neq j} \boldsymbol{x}_{\neq j}| \leq \beta_1 + \beta_2 \mu \end{cases}$$

Or more elegantly,

$$x_j = \frac{1}{\|\mathscr{H}\|_2^2} \mathscr{S}_{\beta_1 + \beta_2 \mu}(\mathbf{h}_j^{\mathrm{T}} \boldsymbol{y} - \mathbf{h}_j^{\mathrm{T}} \mathbf{H}_{\neq j} \boldsymbol{x}_{\neq j}). \tag{14}$$

The major advantage of the Referenced/Joint Cyclic LASSO over the LASSO and the Cyclic LASSO is due to the fact that Eq. (14) will be evaluated for j varying from 1 to T instead of L_x. This is mainly because for a given cycle, x_j gathers all the information given by its multiple, so there is no need to evaluate Eq. (14) for the multiples of j. Contrarily, the estimation x_j of the LASSO explicitly contains no information about the other impulses and hence the need of evaluating Eq. (4) for whole indices. Furthermore, the estimation x_j of the Cyclic LASSO contains only information about the neighborhood (up to $\pm T$) impulses, which is not enough to ignore Eq. (14) for the multiples of j. Consequently, for the Referenced/Joint Cyclic LASSO, indices of impulses for the remaining cycles will be deduced as follows: $\Lambda^{(k)} = \big\{\{j_1^{(k)}, \ldots, j_{N_k}^{(k)}\} + mT\big\}$, where $m = 0, \ldots, K-1$ and $j_1^{(k)}, \ldots, j_{N_k}^{(k)}$ are the N_k selected indices at iteration k.

7 Implementation

7.1 Non-separability of the Penalty Functions

Unfortunately, coordinate-wise descent does not work for (5, 8 or 11). Bertsekas [3] shows that every limit point of successive coordinate-wise minimization of a continuously differentiable function is a stationary point for the overall minimization, provided that the minimum is uniquely obtained along each coordinate. However, the cost functions of (5, 8 or 11) are not continuously differentiable, which means that coordinate-wise descent can get stuck. Moreover, Tseng [32] has established that coordinate descent works in problems like the following. He considers minimizing functions of the form $f(z_1, \ldots, z_p) = g(z_1, \ldots, z_p) + \sum_{j=1}^{p} g_j(z_j)$, where $g(0)$ is differentiable and convex and the $g_j(.)$ are convex. Here, each z_j can be a vector, but the different vectors cannot have any overlapping members. He shows that coordinate descent converges to the minimizer of f. The key to this result is the separability of the penalty function $\sum_{j=1}^{p} g_j(z_j)$, a sum of functions of each individual parameter. This result implies that the coordinate-wise algorithms could not converge to the optimal solution as the penalty functions of (5, 8 or 11) are not separable, and hence, Tseng's theorem cannot be applied in that case.

Despite this, it turns out that the coordinate-wise descent procedure can be modified to work for (5, 8 or 11), yielding an algorithm that is powerful.

The key idea is to make an orthogonal projection of the measurements $\boldsymbol{y}$ over the atoms $\mathbf{H}_{\Lambda^{(k)}}$ corresponding to the set of nonzero elements $\Lambda^{(k)}$ at the iteration k given by (14). Obviously, this avoids the algorithm to get stuck, but increases the computation cost as the orthogonal projection is performed for each iteration.

7.2 Regularization Parameters β_1 and β_2

Running the algorithm involves choosing the hyperparameter pair (β_1 and β_2). Because of the trade-off between the sparsity and cyclostationarity of the solution and the quality of the approximation. This is not a trivial problem and its analytical resolution is still an open problem, which would occupy both theoreticians and practitioners in the coming years.

For the LASSO, we can use (4) to efficiently compute solutions at a grid of values of β_1. We start with the smallest value $\beta_{1,\max}$ for which $\hat{\boldsymbol{x}}(\beta_{1,\max}) = 0$, decrease it a little and cycle through the variables until convergence. Then, β_1 is decreased again and the process is repeated, using the previous solution as a warm start for the new value of β_1. Furthermore, we find that we can compute the entire regularization path for the LASSO in not much more time than it takes to solve a single problem instance [11,23]. This is because we can determine the critical values of β_1 for which the solution changes. It is shown [5] that $\hat{\boldsymbol{x}}(\beta_1)$ is identically zero if $\beta_1 > \beta_{1,\max} = \max_j |\mathbf{h}_j^{\mathrm{T}} \boldsymbol{y}|$. Once the algorithm have converged, starting it with this converged value, it is not possible to move towards zero. To make it possible, we need to decrease β_1 by a small value $\Delta \beta_1$, till the new value for which move toward zero is possible.

Another problem to fix for the (Referenced/Joint) Cyclic LASSO is the estimation of β_2 as the solution depends on it. We limit our interest to a one-dimensional line in (β_1, β_2) space, i.e., $\beta_1 = a\beta_2$, $a \in \mathbb{N}^*$; we get a piecewise linear solution path. For our simulations, we took $a \geq 4$.

7.3 Modified Coordinate-Wise Algorithm

Let the sub-matrix $\mathbf{H}_\Lambda$ built up from the columns of $\mathbf{H}$ where the indices are in Λ, $\Lambda^{(k)}$ is the set of the selected indices at iteration k and the vector $\mathbf{r} = [r_1, \ldots, r_{L_y}]^{\mathrm{T}}$ denotes the residual. Recall that all methods are biased toward smaller nonzero entries, due to the ℓ_1 penalty. Thus, adopting only the support of the solution and solving for the nonzero values by LS, we may obtain improved results, and hence, all methods gain from this modification.

The modified coordinate-wise algorithm for solving (5, 8 or 11) is iterative and composed of three major steps at each iteration: (1) the selection of the indices $\Lambda^{(k)}$ corresponding to the set of nonzero elements using (14); however, for the Referenced Cyclic LASSO and the Joint Cyclic LASSO, the selection step will be restricted to indices j, from 1 to T instead of L_x and the indices of impulses for the rest of cycles will be deduced as follows: $\Lambda^{(k)} = \left\{ \{j_1^{(k)}, \ldots, j_{N_k}^{(k)}\} + mT \right\}$, where $m = 0, \ldots, K-1$ and $j_1^{(k)}, \ldots, j_{N_k}^{(k)}$ are the N_k selected indices at iteration k; (2) update the solution $\mathbf{x}^{(k)}$, with nonzero elements at indices $\Lambda^{(k)}$ and the corresponding residual $\mathbf{r}^{(k)}$; (3) update the parameters (β_1, β_2); a stopping rule helps decide whether to stop or continue the iteration; let us consider $\mathbf{x}^{(k)}$ the solution of the kth iteration, $\mathbf{x}_\Lambda^{(k)}$ being its coefficients at indices Λ and $\mathbf{r}^{(k)} = \mathbf{y} - \mathbf{H}\mathbf{x}^{(k)}$ the residual corresponding to this solution (approximation error). The typical structure of the proposed algorithm is: Initialize $k = 0$, $\Lambda^{(k)} = \emptyset$, $\mathbf{x}^{(k)} = 0$, $\beta_1 = \beta_{1,\max}$, $\beta_2 = \beta_1/a$ and $\mathbf{r}^{(k)} = \mathbf{y}$. Iterate on $k = k + 1$ until the stopping rule is satisfied:

- Selection: $\Lambda^{(k)}$ is the set of nonzero elements given by (14)
- Update:

$$\text{solution: } \mathbf{x}_{\Lambda^{(k)}}^{(k)} = (\mathbf{H}_{\Lambda^{(k)}}^{\mathrm{T}} \mathbf{H}_{\Lambda^{(k)}})^{-1} \mathbf{H}_{\Lambda^{(k)}}^{\mathrm{T}} \mathbf{y}$$

$$\text{residual: } \mathbf{r}^{(k)} = \mathbf{y} - \mathbf{H}_{\Lambda^{(k)}} \mathbf{x}_{\Lambda^{(k)}}^{(k)} \tag{15}$$

- Update β_1: $\beta_1 = \beta_1 - \Delta\beta_1$
- Update β_2: $\beta_2 = \beta_1/a$
- Stopping criterion

Note that the above algorithm can be used to solve the LASSO problem by setting β_2 to zero.

7.4 Stopping Rule

As the algorithms LASSO and (Joint/Referenced) Cyclic LASSO may give different solutions, one should compare them using the same 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. 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. Thus, 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 $||\boldsymbol{n}||_2^2/\sigma^2 = \sum_{k=1}^{N} \frac{n_k^2}{\sigma^2}$ 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 $Pr(||\boldsymbol{r}^{(k)}|| \leq \epsilon) = \eta$ for a given probability, e.g., $\eta = 95\%$.

7.5 Discussion

- The parameter $\beta_2 \mu_{cl} = \beta_2 \left(\text{sign}(|x_j| - |x_{j-T}|) - \text{sign}(|x_{j+T}| - |x_j|) \right)$ can take any value in the range $\{-2\beta_2, -\beta_2, 0, \beta_2, 2\beta_2\}$. This makes the threshold for the Cyclic LASSO fluctuating up to $\beta_1 \pm 2\beta_2$ (which is not the case for the LASSO as $\beta_2 = 0$). Consequently, this allows the Cyclic LASSO to move to zero components, where the LASSO would not do, even if the condition is not satisfied (for the LASSO $|\mathbf{h}_j^\mathrm{T} \boldsymbol{y} - \mathbf{h}_j^\mathrm{T} \mathbf{H}_{\neq j} \boldsymbol{x}_{\neq j}| > \beta_1$) as long as it is for $|\mathbf{h}_j^\mathrm{T} \boldsymbol{y} - \mathbf{h}_j^\mathrm{T} \mathbf{H}_{\neq j} \boldsymbol{x}_{\neq j}| \leq \beta_1 + \beta_2 \mu_{cl}$ whenever $\mu_{cl} > 0$, and fish out components, where the LASSO would move to zero, even if the condition is satisfied (for the LASSO $|\mathbf{h}_j^\mathrm{T} \boldsymbol{y} - \mathbf{h}_j^\mathrm{T} \mathbf{H}_{\neq j} \boldsymbol{x}_{\neq j}| \leq \beta_1$) as long as $|\mathbf{h}_j^\mathrm{T} \boldsymbol{y} - \mathbf{h}_j^\mathrm{T} \mathbf{H}_{\neq j} \boldsymbol{x}_{\neq j}| > \beta_1 + \beta_2 \mu_{cl}$ whenever $\mu_{cl} < 0$. This summarizes the key feature of the Cyclic LASSO against the LASSO thanks to the combination of cyclostationarity and sparsity.
- In addition to the advantage of the Cyclic LASSO over the LASSO, the Referenced/Joint Cyclic LASSO presents another important advantage. Actually, the Referenced/Joint Cyclic LASSO makes a kind of *sampling* of the threshold β_2 to several levels, as $\beta_2 \mu_{rcl}$ and $\beta_2 \mu_{jcl}$, respectively, take values in the discrete sets $\frac{\beta_2}{K-1}[-(K-1), \ldots, -1, 0, 1, \ldots, (K-1)]$ and $\frac{\beta_2}{K}((K-1)\text{sign}(|x_j| - m_j) - \sum_{k=1}^{K-1} \text{sign}(|x_{j+kT}| - m_j))$. As a consequence, this makes the threshold fluctuating up to $\beta_1 \pm \beta_2$ for the Referenced Cyclic LASSO and approximately up to $\beta_1 \pm 2\beta_2$ for the Joint Cyclic LASSO. Although the Joint Cyclic LASSO roughly doubles the number of possible levels to $(4K - 3)$ as compared to that of the Referenced Cyclic LASSO, which does no exceed $(2K - 1)$ levels. The larger the K, more are the levels possible and better is the approximation. Consequently, less false alarms and missing detections are made even for impulses drowned in noise. This summarizes the key feature of the Referenced Cyclic LASSO and the Joint Cyclic LASSO against the Cyclic LASSO and the LASSO, thanks to the combination of cyclostationarity and sparsity in addition to the integration in one cycle of the information given by the remaining cycles.
- The differences between the three proposed methods lie exclusively in the second penalty of criteria. Actually, the second penalty encourages sparsity of coefficients in their cyclic differences,

(a)

(b)
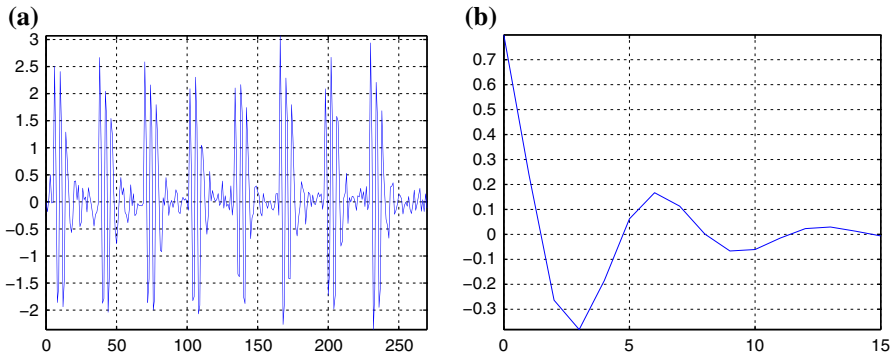
**Fig. 2** **a** Observed signal (SNR = 14 dB). **b** IR used for all simulations

- at neighborhood cycles, i.e., left $((j-1)$th cycle) and right $((j+1)$th cycle) ones, for Cyclic LASSO. Thus, for the jth cycle, there is no connection with the remaining cycles.
- simultaneously according to a reference cycle for Referenced Cyclic LASSO. Hence, the choice of the reference cycle is important as all detections depend on it.
- simultaneously according to the absolute synchronous mean for Joint Cyclic LASSO. The absolute synchronous mean represents a common and joint reference to the whole cycles better than choosing a reference cycle as for Referenced Cyclic LASSO.

8 Simulation

8.1 Description

The objective of these simulations is to evaluate the contribution of cyclic sparsity for convex relaxation deconvolution. The proposed methods and the modified LASSO 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. 2b. Then, i.i.d. Gaussian noise ($L_y = 270$) is added to the convolved signal, as illustrated by (1), such that the SNR is 14 dB. The resulting signal is reported in Fig. 2a.

For the first evaluation, we consider the time representations of the reconstructed signals given by each algorithm versus the true signal. Figure 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
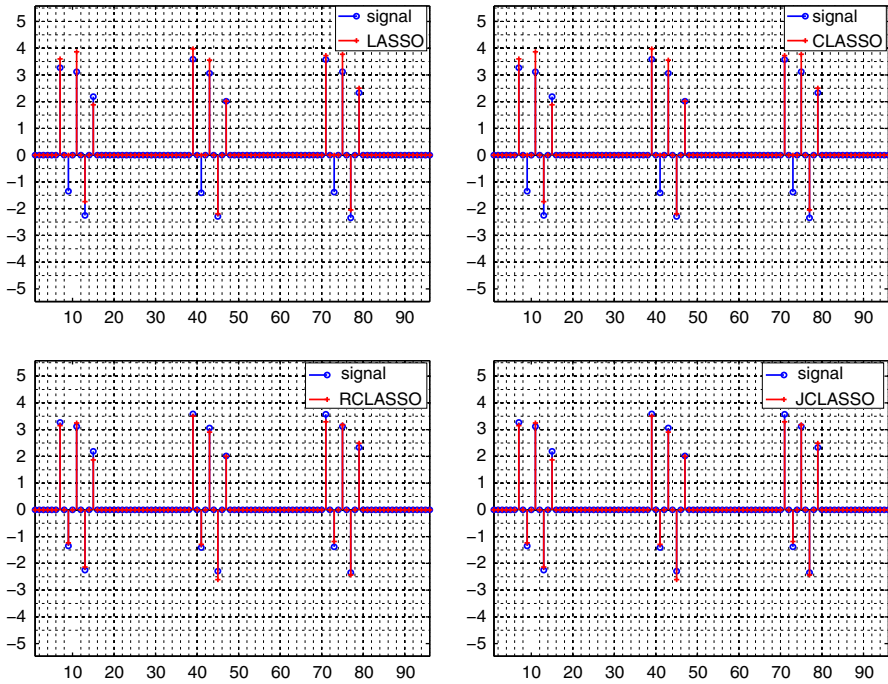
Fig. 3 Time representations of the reconstructed signal (*colored line*) for each method versus the original one (*blue line*) described by Eq. (1) (the SNR is set to 14 dB)

avoid overloading Fig. 3. Comparing all algorithms based on several simulation tests as the one of Fig. 3, we note that the Referenced Cyclic LASSO and the Joint Cyclic LASSO are able to detect and restore almost all impulses. This is not always the case for the Cyclic LASSO and the LASSO algorithms, although the Cyclic LASSO can sometimes do better than the LASSO algorithm. Therefore, we conclude that deconvolution across cyclic sparsity hypothesis allows detecting and restoring impulses even drowned in noise. This point will be examined in detail using other evaluations (mean-squared error and histogram). The reason why the Referenced Cyclic LASSO and the Joint Cyclic LASSO have the same amplitudes for the restored signals (Fig. 3) can be clearly explained by the fact that the estimation step for both algorithms is based on LS over the selected indices. Thus, when both algorithms select the same indices (which can happen for all algorithms), the estimated objects will have exactly the same amplitudes.

8.2 Mean-Squared Error

We provide here a comparison between the proposed approaches against the original one. The aim is to show the performance of cyclosparse convex relaxation deconvolutions in various i.i.d. noisy environment.

The simulation is made with the same parameter as the first example except SNR. Actually, the SNR will vary from 1 to 30 dB. And for each value of the SNR, 500 Monte Carlo (MC) runs will be implemented. Thus, for each MC run,
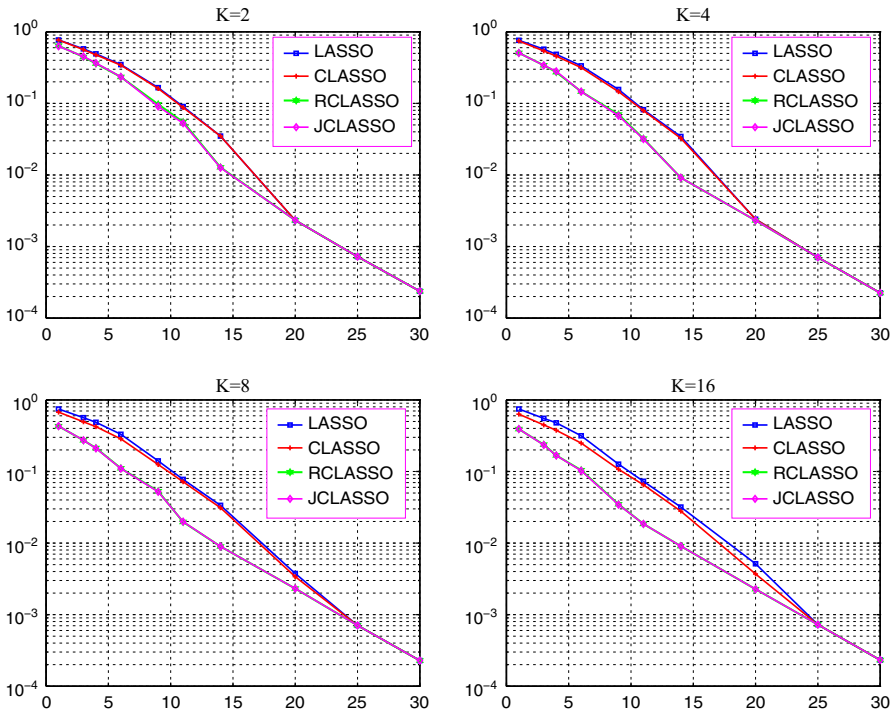
Fig. 4 Effect of varying SNR from 1 to 30 dB over MC runs on the MSE for each method with different number of cycles K

- the periodic random impulses keep the same positions but with random amplitudes
- the input signal is filtered by the IR of Fig. 2b
- 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 performance 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: $\text{MSE}(\hat{x}) = \mathbb{E}\big[(\hat{x} - x)^2\big]$. These MSE will be averaged over the number of MC runs.

Figure 4 shows the variation of each output's MSE, for the proposed methods and the original one 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 the Referenced Cyclic LASSO and the Joint Cyclic LASSO, with highest MSE occurring for lower SNR. 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 25 dB. For SNR greater than 25 dB, the MSE of the algorithms is almost the same. However, for SNR less than 25 dB, the MSE of the algorithms can be sorted in descending order as, LASSO, Cyclic LASSO and Referenced/Joint Cyclic LASSO. Therefore, we conclude that cyclic algorithms perform well even for lower SNR.

8.3 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 many detections are 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. Figure 5 (for $K = 8$) shows the variation of each output's average histogram obtained by varying SNR from 1 to 30 dB for the proposed methods and the original one as well. As the histogram is almost periodic, we constrained the plot to one period in order to avoid overloading figures. To do so, we considered the mean over K periods of the computed average histograms as shown in Fig. 5.

We note from Fig. 5 (for $K = 8$) that false detections increase with decreasing SNR. This is because higher SNR implies lower noise effect on observed data $\mathbf{y}$. The histogram is nearly the same for the Referenced Cyclic LASSO and the Joint Cyclic LASSO. However, we note 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 Fig. 5, the algorithms' behavior with respect to the histogram against noise can be decomposed into two parts: SNR less or greater than 25 dB. Consequently, the histogram confirms the MSE behavior of the algorithms and leads to the same sorting of the algorithms.

8.4 Influence of the Number of Cycles

Another parameter which can influence the performances of cyclic algorithms is the number of cycles/periods K. To examine this, we performed three additional simulations in which K was gradually increased ($K = 2$, 4 and 16). The simulations are made with the same parameters as the second example except for the data size. Actually, changing the number of cycles means changing the data size as well. So, K equal to 2, 4, 8 and 16, corresponds, respectively, to data size 64, 128, 256 and 512. Simulation results for each data size set were reported in Fig. 4 for the average MSE and Fig. 5 for the average histogram.

We note that for the LASSO algorithm, the histograms are almost identical whatever data size with a small increase in false/missing detections. However, as expected
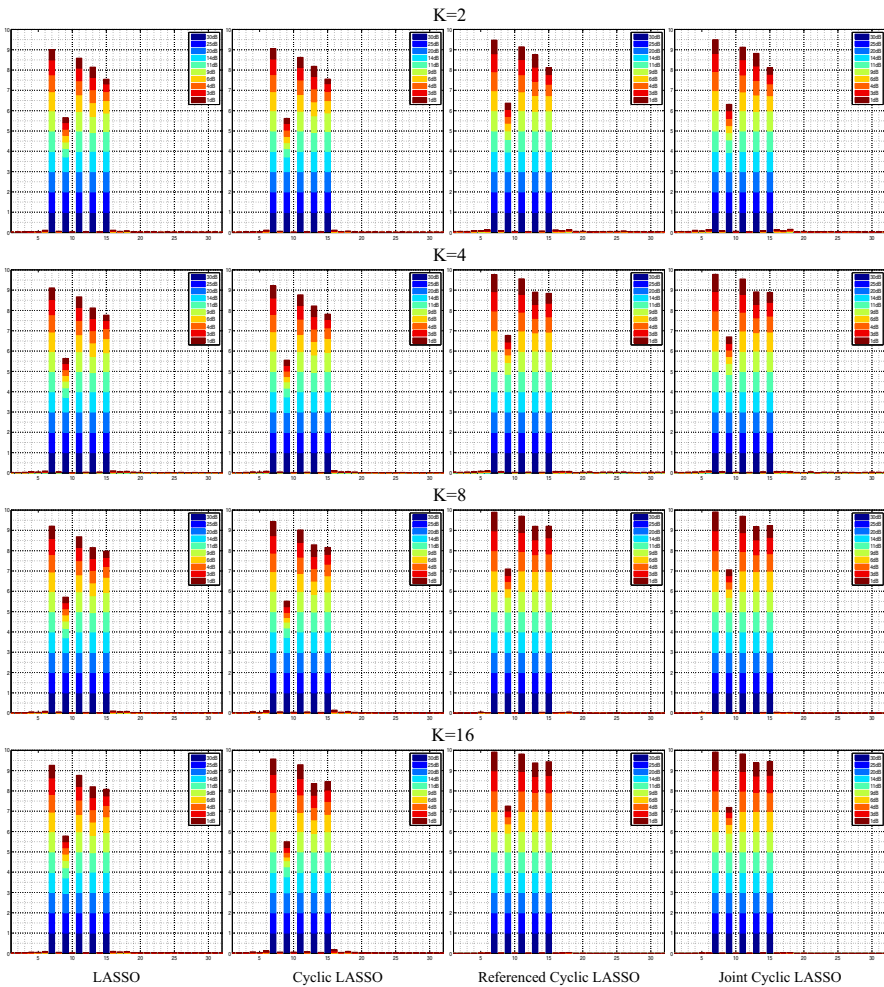
Fig. 5 Effect of varying SNR from 1 to 30 dB over MC runs on the histogram for different number of cycles K. *Each column* corresponds to one algorithm' results. *Each color* denotes an SNR level. The x axis represents the first period and, hence, the impulse positions. For a given impulse and a given color (SNR level), the maximum detection is reached when the bar height is 1 (500 successful detections); however, the minimum detection is reached when the bar height is 0 (500 missing detections). For a zero element and a given color, the maximum false detection is reached when the bar height is 1 (500 false detections); however, the minimum false detection is reached when the bar height is 0 (no false detections)

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 cyclic algorithms in comparison with the LASSO algorithm. The algorithms' performances can be summarized as follows: For $K = 2, 4$, cyclic algorithms become more powerful than the LASSO for SNR less than 20 dB; for $K = 8, 16$, cyclic algorithms become more powerfull than the LASSO for SNR less than 25 dB. This enhancement is more significant specifically for the Referenced Cyclic LASSO and the Joint Cyclic

LASSO even for lower SNR. Finally, we come to the conclusion that deconvolution across cyclic sparsity hypothesis allows to detect and restore impulses even drowned in noise provided that these impulses being significant for the other cycles of the signal. Also, increasing the number of cycles leads to a considerable enhancement of the performances of cyclic algorithms.

9 Conclusion

This study aims to introduce the concept of cyclic sparsity for signals based on periodic random impulses. The findings obtained from the integration of this concept for convex relaxation sparse algorithms will help 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 follows that the proposed methods compare favorably with the original one. Reasons for the improved performance of the proposed methods over the original one include the following: The cyclic sparsity model makes possible the exploitation of the information given by the periodicity which allows less false alarms and missing detections as well. Moreover, for the Referenced/Joint Cyclic LASSO, the purpose is to jointly find 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. The unique additional information required by cyclic convex relaxation algorithms is the cyclic period T. In general, the cyclic period is related to the studied system and the problem of its estimation has been addressed in several articles as [9].

Furthermore, we intend to apply the proposed algorithms to mechanical signals for bearing fault detection. 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 [2]. 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. In addition, the fluctuation of the cyclic period, i.e., the positions of random impulses may lightly fluctuate (oscillate around average positions), will dramatically reduce the performances of the proposed methods as the K selected atoms may not be exactly periodic and, hence, the need to develop additional criteria to fix this problem.

Appendix 1: The Soft Thresholding Operator

The soft thresholding operator $\mathscr{S}(.)$ is defined as:

$$\mathscr{S}_\beta(a) = \begin{cases} a - \beta, & a > \beta \\ a + \beta, & a < -\beta \\ 0, & |a| \leq \beta \end{cases}$$

or equivalently,

$$\mathscr{S}_\beta(a) = (a - \beta)_+ - (-a - \beta)_+$$

where $(.)_+$ denotes the positive part operator (defined as $(a)_+ = a$ if $a \geq 0$, and $(a)_+ = 0$ if $a < 0$). Yet another formula, which shows that the soft thresholding operator is a shrinkage operator (i.e., moves a point toward zero), is $\mathscr{S}_\beta(a) = (1 - \beta/|a|)_+ a$.

Appendix 2: Structure of the Matrix $\mathbf{A}_{L_x,T}$

The square matrix $\mathbf{A}_{L_x,T}$ is sparse with only $2L_x$ nonzero elements. The main diagonal elements are set to -1, whereas the $-(L_x - T)$th and Tth off-diagonal elements are set to 1. Also $\mathbf{A}_{L_x,T}$ has zero row and column sums. An example of the matrix $\mathbf{A}_{L_x,T}$ for $L_x = 8$ and $T = 3$ is given below:

$$\mathbf{A}_{8,3} = \begin{bmatrix} -1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & -1 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & -1 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & -1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 & 0 & -1 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & -1 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & -1 \end{bmatrix}$$

Appendix 3: Structure of the Matrix $\mathbf{Q}_{L_x,T}$

The matrix $\mathbf{Q}_{L_x,T}$ (with $L_x = KT$) is an $(L_x - T) \times L_x$ sparse with only $2(L_x - T)$ nonzero elements. The T first elements of the main diagonal are nonzero and are equal to -1; also, the $-T$th and Tth off-diagonal elements are, respectively, set to -1 and 1. An example of the matrix $\mathbf{Q}_{L_x,T}$ for $L_x = 9$ and $T = 3$ is given below:

$$\mathbf{Q}_{9,3} = \frac{1}{2} \begin{bmatrix} -1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & -1 & 0 & 0 & 1 & 0 & 0 & 0 \\ -1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & -1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & -1 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

Appendix 4: Structure of the Matrix $\mathbf{N}_{L_x,T}$

The square matrix $\mathbf{N}_{L_x,T}$ (with $L_x = KT$) is an $L_x \times L_x$ circulant sparse with only $L_x K$ nonzero elements. It is easy to show that $\mathbf{N}_{L_x,T}$ can be written as:

$$\mathbf{N}_{L_x,T} = \mathbf{I}_{L_x} - \frac{1}{K}\mathbf{B}_{L_x,T}$$

where $\mathbf{I}_{L_x}$ is the identity matrix.

Let us first consider the matrix form of the synchronous mean of the absolute value of x,

$$\boldsymbol{m} = \mathbf{M}_{L_x,T}|\boldsymbol{x}|$$

with $\mathbf{M}_{L_x,T}$ is an $T \times L_x$ circulant sparse matrix of L_x nonzero elements. More precisely, the main diagonal and the kTth off-diagonal elements $(k = 1, \ldots, K - 1)$ are nonzero and are equal to 1. An example of the matrix $\mathbf{M}_{L_x,T}$ for $L_x = 9$ and $T = 3$ is given below:

$$\mathbf{M}_{9,3} = \begin{bmatrix} 1 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 & 1 & 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 1 \end{bmatrix}$$

The circulant square matrix $\mathbf{B}_{L_x,T}$ can be deduced from $\mathbf{M}_{L_x,T}$ as below:

$$\mathbf{B}_{L_x,T} = \begin{bmatrix} \mathbf{M}_{L_x,T} \\ \mathbf{M}_{L_x,T} \\ \vdots \\ \mathbf{M}_{L_x,T} \end{bmatrix}$$

It is clear that the main diagonal and the $\pm kT$th off-diagonal, elements $(k = 1, \ldots, K - 1)$ are nonzero and are equal to 1. An example of the circulant matrix $\mathbf{N}_{L_x,T}$ for $L_x = 9$, $K = 3$ and $T = 3$ is given below:

$$\mathbf{N}_{9,3} = \frac{1}{3}\begin{bmatrix} 2 & 0 & 0 & -1 & 0 & 0 & -1 & 0 & 0 \\ 0 & 2 & 0 & 0 & -1 & 0 & 0 & -1 & 0 \\ 0 & 0 & 2 & 0 & 0 & -1 & 0 & 0 & -1 \\ -1 & 0 & 0 & 2 & 0 & 0 & -1 & 0 & 0 \\ 0 & -1 & 0 & 0 & 2 & 0 & 0 & -1 & 0 \\ 0 & 0 & -1 & 0 & 0 & 2 & 0 & 0 & -1 \\ -1 & 0 & 0 & -1 & 0 & 0 & 2 & 0 & 0 \\ 0 & -1 & 0 & 0 & -1 & 0 & 0 & 2 & 0 \\ 0 & 0 & -1 & 0 & 0 & -1 & 0 & 0 & 2 \end{bmatrix}$$

Of course, $\mathbf{N}_{L_x,T}$ has the same structure than $\mathbf{Q}_{L_x,T}$ except the values of nonzero elements. The main diagonal elements are set to $\frac{K-1}{K}$, and the $\pm kT$th off-diagonal elements $(k = 1, \ldots, K - 1)$ are set to $-\frac{1}{K}$.

References

1. C. Andrieu, P. Duvaut, A. Doucet, Bayesian Deconvolution of Cyclostationary Processes Based on Point Processes, in *EUSIPCO*, Trieste, Italy, 10–13 Sep 1996
2. J. Antoni, R.B. Randall, The spectral kurtosis: application to the vibratory surveillance and diagnostics of rotating machines. Mech. Syst. Signal Process. **20**(2), 308–331 (2006)
3. D. Bertsekas, *Nonlinear Programming* (Athena Scientific, Belmont, 1999)
4. J. Bobin, J.-L. Starck, Y. Moudden, M.J. Fadili, Blind source separation: the sparsity revolution, in *Advances in Imaging and Electron Physics*, ed. by P. Hawkes (Academic Press, Elsevier, Amsterdam, 2008), pp. 221–298
5. S. Bourguignon, H. Carfantan, T. Böhm, SparSpec: a new method for fitting multiple sinusoids with irregularly sampled data. Astronom. Astrophys. **462**(1), 379–387 (2007)
6. L. Cerrato, B. Eisenstein, Deconvolution of cyclostationary signals. IEEE Trans. Acoust. Speech SP **25**(6), 466–476 (1977)
7. S. Chen, S.A. Billings, W. Luo, Orthogonal least squares methods and their application to non-linear system identification. Int. J. Control **50**(5), 1873–1896 (1989)
8. K.M. Cheung, 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
9. A.V. Dandawate, G.B. Giannakis, Statistical tests for presence of cyclostationarity. IEEE Trans. Signal Process. **42**(9), 2355–2369 (1994)
10. I. Daubechies, M. Defrise, C. De Mol, An iterative thresholding algorithm for linear inverse problems with a sparsity constraint. Commun. Pure Appl. Math. **57**, 1413–1457 (2004)
11. B. Efron, T. Hastie, I. Johnstone, R. Tibshirani, Least angle regression. Ann. Stat. **32**(2), 407–499 (2004)
12. M. Ehler, Shrinkage rules for variational minimization problems and applications to analytical ultracentrifugation. J. Inverse Ill-Posed Probl. **19**(4–5), 593–614 (2011)
13. J. Friedman, T. Hastie, H. Hoefling, R. Tibshirani, Pathwise coordinate optimization. Ann. Appl. Stat. **2**(1), 302–332 (2007)
14. W. Fu, Penalized regressions: the bridge vs. the LASSO. J. Comput. Graph. Stat. **7**(3), 397–416 (1998)
15. W.A. Gardner, *Introduction to Random Processes: With Application to Signals and Systems* (Macmillan, New York, 1986)
16. W.A. Gardner, *Cyclostationarity in Communications and Signal Processing* (IEEE Press, New York, 1993)
17. R. Gribonval, M. Zibulevsky, Chapter Handbook of blind source separation: independent component analysis and applications, in *Sparse Component Analysis*, ed. by P. Comon, C. Jutten (Elsevier, Amsterdam, 2010), pp. 367–420
18. D. Hatzinakos, J. Ilow, Recursive least squares algorithm for blind deconvolution of channels with cyclostationary inputs, in *IEEE Military Communications Conference*, Boston, MA, 1993, pp. 123–126
19. D. Hatzinakos, Nonminimum phase channel deconvolution using the complex cepstrum of the cyclic autocorrelation. IEEE Trans. SP **42**(11), 3026–3042 (1994)
20. S. Mallat, Z. Zhang, Matching pursuits with time frequency dictionaries. IEEE Trans. SP **41**, 3397–3415 (1993)
21. Y.C. Pati, R. Rezaiifar, P.S. Krishnaprasad, Orthogonal matching pursuit: recursive function approximation with applications to wavelet decomposition, in *The 27th Annual Asilomar*, Pacific Grove, CA, USA, 1993
22. B. Rao, K. Kreutz-Delgado, An affine scaling methodology for best basis selection. IEEE Trans. Signal Process. **47**(1), 187–200 (1999)
23. S. Rosset, *Topics in Regularization and Boosting*. PhD Thesis, Aug 2003
24. K. Sabri, M. El Badaoui, F. Guillet, A. Belli, G. Millet, J.-B. Morin, Cyclostationary modeling of ground reaction force signals. Signal Process. **90**(4), 1146–1152 (2010)
25. K. Sabri, Cyclic sparse greedy deconvolution. IJIGSP **4**(12), 1–8 (2012)
26. K. Sabri, Cyclosparsity: a new concept for sparse deconvolution. Global J. Comput. Sci. Technol. **4**(4), 19–40 (2014)
27. M. Schmidt, *Least Squares Optimization with L1-Norm Regularization*. cole Normale Suprieure-CS542B Project Report, Dec 2005

28. M. Schmidt, G. Fung, R. Rosaless, *Optimization Methods for L1-Regularization*. UBC Technical Report TR-2009-19, 2009
29. S. Shaobing Chen, David L. Donoho, Michael A. Saunders, Atomic decomposition by basis pursuit. SIAM J. Sci. Comput. **20**(1), 33–61 (1998)
30. R. Tibshirani, Regression shrinkage and selection via the LASSO. J. R. Stat. Soc. Ser. B **58**(1), 267–288 (1996)
31. J.A. Tropp, Greed is good: algorithmic results for sparse approximation. IEEE Trans. Inf. Theory **50**(10), 2231–2242 (2004)
32. P. Tseng, Convergence of block coordinate descent method for nondifferentiable maximization. J. Optim. Theory Appl. **109**(3), 474–494 (2001)
33. T. Wu, K. Lange, Coordinate descent procedures for LASSO penalized regression. Ann. Appl. Stat. **2**(1), 224–244 (2008)