

Cyclic Sparse Greedy Deconvolution

Khalid SABRI

STIC laboratory, Faculty of sciences, University Choua b Doukkali, El Jadida, Morocco
E-mail: sabri.k@ucd.ac.ma

Abstract— The 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. 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 increases the performances and reduces significantly the computation cost. Finally, we use simulations to investigate the behavior in deconvolution framework of the algorithms MP, OMP and their respective extensions to cyclic sparsity context, Cyclo-MP and Cyclo-OMP.

Index Terms— Cyclosparsity, Sparsity, Cyclostationary, Deconvolution, Greedy

I. INTRODUCTION

In this study we focus on Cyclostationary (CS) signals made of periodic random impulses. 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, improving 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 [1], a bayesian deconvolution algorithm based on Markov chain Monte Carlo is presented. Cyclic statistics are often used for deconvolution, in [2], [3] the deconvolution is based on cyclic cepstrum, whereas, in [4], [5] the deconvolution is based on cyclic correlation. Actually, signal deconvolution is known to be an ill-posed problem as the IR acts as a low-pass filter and the convolved signal is noisy. Fortunately, regularization techniques allow to retrieve satisfactory solutions accounting for *a priori* information on the original object. Analysing periodic random impulse signals in details uncover another *a priori* information which is sparsity, this is because only few impulses are nonzero. Note that sparsity is interesting as well as cyclostationarity.

Our work consists on extending sparse approximation to CS signals with periodic random impulses, where the aim is to find jointly a sparse approximation of each 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 in the application of cyclosparse approximation to deconvolution.

The paper is organized as follows, section II defines problem statement. In section III, we summarize the statement of sparse and cyclosparse approximation problems. The main contribution of this paper is described in sections IV and V, where we, respectively, introduce the concept of cyclosparsity and address the problem of deconvolution of cyclosparse objects. Also, in section IV, we point-out the link with cyclosparse deconvolution. Such a cyclosparse model is taken into account in the sparse deconvolution by greedy algorithms in section V, which fortunately reduces significantly the computation cost. In this paper, we focus on greedy algorithms and we propose to test some of these algorithms, namely Matching Pursuit (MP) [6] and Orthogonal Matching Pursuit (OMP) [7] on the same statistical basis, i.e. with the same stopping rule deduced from statistical properties of the noise. These algorithms are summarized in the cyclosparse case as well. Also it seems to be necessary to obtain satisfactory deconvolution results, as shown in the simulation results section VI.

II. PROBLEM STATEMENT

Consider the situation where a known system $h(t)$ is excited by a CS signal $x(t)$ consisting of periodic random impulses. By periodic, we mean that the signal can be partitioned 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 i of the i th 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 $d = 5$. Note that the impulse with τ_3 as delay factor is nonzero for the first and second cycles ($x_{1,3}, x_{2,3} \neq 0$) whereas for the third cycle it is of null amplitude $x_{3,3} = 0$.

The output of the system described above can be written as,

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

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 sub-index i stands for the period index, so

$x_{i,k}$ represents the impulse with τ_k as delay factor in the i th-period. $n(t)$ represents the random noise of the system.

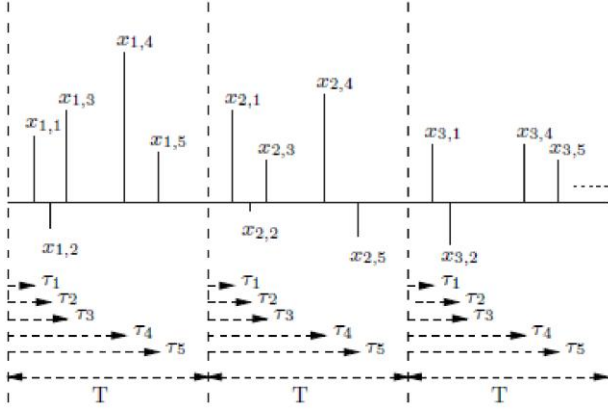


Fig. 1. Example of cyclo sparse signal.

III. FROM SPARSE TO CYCLOSPARSE APPROXIMATION

A. Sparse approximation

The problem of sparse signal approximation consists in approximating a signal as a linear combination of a limited number of elementary signals (atoms) chosen in a redundant collection (dictionary). It can be written as: Find sparse \mathbf{x} such that $\Phi\mathbf{x} \approx \mathbf{y}$, where \mathbf{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. Such compromise results mathematically to minimize the criterion:

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

Where $\|\mathbf{x}\|_0 = \text{Card}\{k, x_k \neq 0\}$, is the number of nonzero components of \mathbf{x} . 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.

Minimizing such a criterion is a combinatory optimization problem which is NP hard. Two approaches are generally used to avoid exploring every combination: Greedy algorithms and convex relaxation algorithms.

B. Cyclo sparse approximation

The extension of sparse approximation for CS and sparse signals will be studied in this article. A cyclo sparse solution is given by minimizing the following criterion:

$$\mathcal{J}(\mathbf{x}) = \sum_{i=1}^T \|\mathbf{y} - \Phi_{i+mT}\mathbf{x}\|_2^2 + \beta K \|\mathbf{x}\|_{\infty,0} \quad (3)$$

with $m = 0, \dots, K-1$.

K and T denote respectively the number of cycles/periods and the cyclic period.

The inner l_∞ norm encourages diversity along the cycles of \mathbf{x} so the $l_{\infty,0}$ mixed-norm measures the cyclo sparsity along the whole signal as illustrated in Fig. 1. The choice of $l_{\infty,0}$ mixed-norm result will be illustrated in the next section.

IV. CONCEPT OF CYCLOSPARSITY

A. Definition

A process \mathbf{x} is said to be cyclo sparse if

$$\|\mathbf{x}\|_0 \leq K \|\mathbf{x}\|_{\infty,0} \quad (4)$$

With \mathbf{x} can be reshaped as $K \times T$ matrix.

B. Proof

Let us first introduce the following notations:

- $\wedge \mathbf{x}_{i,:} = \mathbf{x}((i-1)T + 1 : iT)$ with $1 \leq i \leq K$ sweeps all elements of the i th period (a total of T elements).
- $\wedge \mathbf{x}_{:,k} = \mathbf{x}(k, k+T, k+2T, \dots, k+(K-1)T)$ with $1 \leq k \leq T$ sweeps all k th elements of each period (a total of K elements).

The l_0 -norm of \mathbf{x} is defined as follows

$$\|\mathbf{x}\|_0 = \sum_{k=1}^{KT} |x_k|^0 \quad (5)$$

As is standard, $|x_k|^0 = 1$ if $x_k \neq 0$ and $|x_k|^0 = 0$ if $x_k = 0$.

Using these notations, the equation 5 can be written as,

$$\begin{aligned} \|\mathbf{x}\|_0 &= \|\mathbf{x}_{1,:}\|_0 + \|\mathbf{x}_{2,:}\|_0 + \dots + \|\mathbf{x}_{K,:}\|_0 \\ &= \sum_{k=1}^T |x_{1,k}|^0 + \sum_{k=1}^T |x_{2,k}|^0 + \dots + \sum_{k=1}^T |x_{K,k}|^0 \\ &= |x_{1,1}|^0 + |x_{1,2}|^0 + \dots + |x_{1,T}|^0 \\ &\quad + |x_{2,1}|^0 + |x_{2,2}|^0 + \dots + |x_{2,T}|^0 \\ &\quad + \vdots + \vdots + \dots + \vdots \\ &\quad + |x_{K,1}|^0 + |x_{K,2}|^0 + \dots + |x_{K,T}|^0 \end{aligned}$$

Also, we can write

$$\begin{aligned} \|\mathbf{x}\|_0 &\leq K \left(|\max_i |x_{i,1}||^0 + |\max_i |x_{i,2}||^0 + \dots + \right. \\ &\quad \left. |\max_i |x_{i,T}||^0 \right) \\ &\leq K \sum_{k=1}^T |\max_i |x_{i,k}||^0 = K \left\| \max_i |\mathbf{x}_{i,:}| \right\|_0 = \|\mathbf{x}\|_{\infty,0} \end{aligned}$$

$\max_i |\mathbf{x}_{i,:}|$ stands for l_∞ -norm of \mathbf{x} along dimension of i for each value of the period indexed by k . The composite norm $\left\| \max_i |\mathbf{x}_{i,:}| \right\|_0$ is the mixing norm $l_{\infty,0}$ of \mathbf{x} which the minimization encourages first diversity along k and then sparsity of the resulting vector.

V. DECONVOLUTION IN CYCLOSPARSITY CONTEXT

A. Problem formulation

Let us first specify the boundary condition accounted for in the convolution operator. We assume that the object is zero-valued outside the signal to reconstruct of size L_x . The convolution of the object leads to a signal of size $L_y = L_x + L_h - 1$. Thus, the dictionary is given by the $(L_x + L_h - 1) \times L_x$ Toeplitz matrix \mathbf{H} where the columns are built from the IR \mathcal{H} .

Therefore, Eq. 1 can be written with matrix notations as,

$$\mathbf{y} = \mathbf{H} \mathbf{D} \mathbf{x} + \mathbf{n} \\ = \tilde{\mathbf{H}} \mathbf{x} + \mathbf{n}$$

$\tilde{\mathbf{H}}^T \tilde{\mathbf{H}} = \mathbf{D}(\mathbf{H}^T \mathbf{H})\mathbf{D}$ with \mathbf{D} is the cyclosparsity operator

$$\mathbf{D}^T = \mathbf{D} = \sum_{m=0}^{K-1} \sum_{n=1}^{\|\mathbf{x}\|_{\infty,0}} e_{d_{n+mT}} e_{d_{n+mT}}^T \\ = \text{diag} \left(\sum_{m=0}^{K-1} \sum_{n=1}^{\|\mathbf{x}\|_{\infty,0}} e_{d_{n+mT}} \right)$$

where $e_{d_{n+mT}}$ are the canonical basis vectors and d_n is the index of the n th impulse with τ_n as factor delay. Thus, $\tilde{\mathbf{H}}$ is a 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, $\tilde{\mathbf{H}}$ points out the cyclosparsity property in the convolution case.

It should be noted that 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.

B. Cyclosparse greedy algorithms

Let the sub-matrix \mathbf{H}_{Λ} built-up from the columns of \mathbf{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, $\mathbf{x} = [x_1, \dots, x_{L_x}]$, $\mathbf{y} = [y_1, \dots, y_{L_y}]$,

$\mathbf{n} = [n_1, \dots, n_{L_y}]$ and $\mathbf{r} = [r_1, \dots, r_{L_y}]$ which denotes the residual. L_x , L_y and L_h stand respectively for the length of \mathbf{x} , \mathbf{y} and \mathcal{H} . Finally, let the vector $m = [0, 1, \dots, (K-1)]$ be the vector of period indices.

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 to decide whether to stop or continue the iteration.

Let $\mathbf{x}^{(k)}$ be the solution of the k th 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 a greedy algorithm is: Initialize $k = 0$, $\Lambda^{(k)} = \emptyset$ and $\mathbf{r}^{(k)} = \mathbf{y}$.

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 $\mathbf{x}^{(k)}$, with non-zero elements at indices $\Lambda^{(k)} = \Lambda^{(k-1)} \cup \{i^{(k)}\}$, and the corresponding residual $\mathbf{r}^{(k)}$.

The various algorithms differ on the selection or the updating steps.

Cyclo-Matching Pursuit (Cyclo-MP) is the extension of the MP [6]. The additional atoms (at most K 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: $\Lambda^{(k)} = \Lambda^{(k-1)} \cup \{i_{mT}^{(k)}\}$,
 $i^{(k)} = \arg \max_i \sum_{m=0}^{K-1} |\mathbf{h}_{i+mT}^T \mathbf{r}^{(k-1)}|$ (6)

- ⤴ Update:

solution:

$$\mathbf{x}_{i_{mT}^{(k)}}^{(k)} = \mathbf{x}_{i_{mT}^{(k)}}^{(k-1)} + \left(\mathbf{h}_{i_{mT}^{(k)}}^T \mathbf{h}_{i_{mT}^{(k)}} \right)^{-1} \mathbf{h}_{i_{mT}^{(k)}}^T \mathbf{r}^{(k-1)} \quad (7)$$

residual:

$$\mathbf{r}^{(k)} = \mathbf{r}^{(k-1)} - \mathbf{h}_{i_{mT}^{(k)}} \left(\mathbf{h}_{i_{mT}^{(k)}}^T \mathbf{h}_{i_{mT}^{(k)}} \right)^{-1} \mathbf{h}_{i_{mT}^{(k)}}^T \mathbf{r}^{(k-1)} \quad (8)$$

- ⤴ Stopping criterion

Cyclo-Orthogonal Matching Pursuit (Cyclo-OMP) is the extension of the OMP [7]. 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 (6)

- ⤴ Update:

solution:

$$\mathbf{x}_{\Lambda^{(k)}}^{(k)} = \left(\mathbf{H}_{\Lambda^{(k)}}^T \mathbf{H}_{\Lambda^{(k)}} \right)^{-1} \mathbf{H}_{\Lambda^{(k)}}^T \mathbf{y} \quad (9)$$

residual:

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

- ⤴ Stopping criterion

C. 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 among to decide whether the residual can be distinguished from noise. For Gaussian, centred, independent and identically distributed (*i.i.d.*) noise, of known variance σ^2 , the norm $\|\mathbf{n}\|_2^2/\sigma^2 = \sum_{k=1}^N n_k^2/\sigma^2$ 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 $Pr(\|\mathbf{r}^{(k)}\| \leq \epsilon) = \eta$ for a given probability, e.g. $\eta = 95\%$.

D. 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 $m = 0$.

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 .

Following [7], [8] one can take advantage of the matrix inversion lemma to compute iteratively

$\left(\mathbf{H}_{\Lambda^{(k-1)} \cup \{i_m^{(k)}\}}^T \mathbf{H}_{\Lambda^{(k-1)} \cup \{i_m^{(k)}\}} \right)^{-1}$ at a low cost, from the knowledge of $\left(\mathbf{H}_{\Lambda^{(k-1)}}^T \mathbf{H}_{\Lambda^{(k-1)}} \right)^{-1}$.

VI. SIMULATION

A. Description

We consider simulation example with the following parameters. A CS 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: $H(z) = \frac{1-0,6z^{-1}}{1-0,9z^{-1}+0,6z^{-2}}$. Then *i.i.d.* gaussian noise ($L_y = 270$ as $L_h = 15$) is added to the convolved signal as illustrated by 1, such that the SNR is 14dB.

For the first evaluation we consider the time representations of the reconstructed signals given by each algorithm versus the the true signal. Fig. 2 reports the true signal (blue line) described by (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 figures. Regarding each algorithm and its corresponding extension, we note from Fig. 2 that deconvolution across cyclosparsity hypothesis allows to detect/restore impulses even drowned in noise, reduce false/missing detections and estimate well the amplitude of impulses. Comparing all algorithms, we deduce that the Cyclo-OMP provided the best estimation.

B. Mean Squared Error and Histogram

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 simulation is made with the same parameter as the first example except the SNR which 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 and then *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.

The MSE of the estimated signal $\hat{\mathbf{x}}$ with respect to \mathbf{x} is defined as, $MSE(\hat{\mathbf{x}}) = E[(\hat{\mathbf{x}} - \mathbf{x})^2]$. These MSE will be averaged over the number of MC runs.

Fig. 3 shows the variation of each output's average MSE with the SNR. We note from the trend of the graph that the MSE decreases with increasing SNR. Furthermore, as can be seen from this figure, 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, MP; Cyclo-MP; OMP; Cyclo-OMP. However for SNR less than 6dB, the MSE can be sorted in descending order as, (MP and OMP); Cyclo-MP; Cyclo-OMP.

Fig. 4 shows the variation of each output's average histogram obtained by varying SNR from 1dB to 30dB. As the histogram is been almost periodic, we constrained the plot to the first period in order to avoid overloading figures.

We note from Fig. 4 that false/missing detections increase with decreasing SNR. Consequently, the histogram confirms the MSE behavior of the algorithms and leads to the same sorting of the algorithms.

VII. CONCLUSION

Deconvolution based on cyclosparsity allows to detect and restore impulses even drowned in noise provided that these impulses being significant for the other cycles of the signal. The Cyclic algorithms perform more better than their corresponding classical ones even for low SNR. 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. So when nonzeros

elements are so close and strongly correlated, false detections occur often because the orthogonal projections are made over only the K selected atoms unlike OMP and Cyclo-OMP where the orthogonal projections are made over the whole selected atoms.

REFERENCES

- [1] C. Andrieu, P. Duvaut, and A. Doucet, "Bayesian deconvolution of cyclostationary processes based on point processes," in EUSIPCO, Trieste, Italy, 10-13 Sep 1996.
- [2] 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.
- [3] D. Hatzinakos, "IEEE trans. on sp," Nonminimum phase channel deconvolution using the complex cepstrum of the cyclic autocorrelation, vol. 42, no. 11, pp. 3026–3042, Nov. 1994.
- [4] L. Cerrato and B. Eisenstein, "IEEE trans. on acoustics, speech, and sp," Deconvolution of cyclostationary signals, vol. 25, no. 6, pp. 466–476, Dec. 1977.
- [5] 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.
- [6] S. Mallat and Z. Zhang, "Matching pursuits with time frequency dictionaries," IEEE Trans. on SP., vol. 41, pp. 3397–3415, 1993.
- [7] Y.C. Pati, Ramin Rezaifar, and P.S. Krishnaprasad, "Orthogonal matching pursuit : Recursive function approximation with applications to wavelet decomposition," in the 27 th Annual Asilomar, Pacific Grove, CA, USA, 1993.
- [8] Soussen C., J. Idier, D. Brie, and J. Duan, "From bernoulligaussian deconvolution to sparse signal restoration," IEEE Trans. on SP, vol. 59, no. 10, pp. 4572–4584, oct 2011.

Khalid SABRI was born in Rabat, Morocco on October 28th 1979. He received the B. S. degree in informatic, electronic, electrotechnic and automatic in 2002 and the M.S. degree in informatic and telecommunications from Université Mohammed V-Agdal, Rabat, Morocco in 2004, and the Ph.D. degree, in Signal Processing, in September 2008, jointly from Université Jean Monnet, Saint-Etienne, France and Université Mohammed V-Agdal, Rabat, Morocco. From September 2008 to December 2008, he had a post-doc position at Laboratoire d'analyse des signaux et des processus industriels, Université Jean Monnet, Saint-Etienne, France. From January 2009 to December 2010, he had a post-doc position at Institut de Recherche en Astrophysique et Planéologie, Observatoire Midi-Pyrénées, Toulouse, France. Since 2011, he has been an associate professor at Université Chouaib Doukkali, El Jadida, Morocco. His research interests are in the areas of blind source separation, Sparse deconvolution, Cyclostationarity.

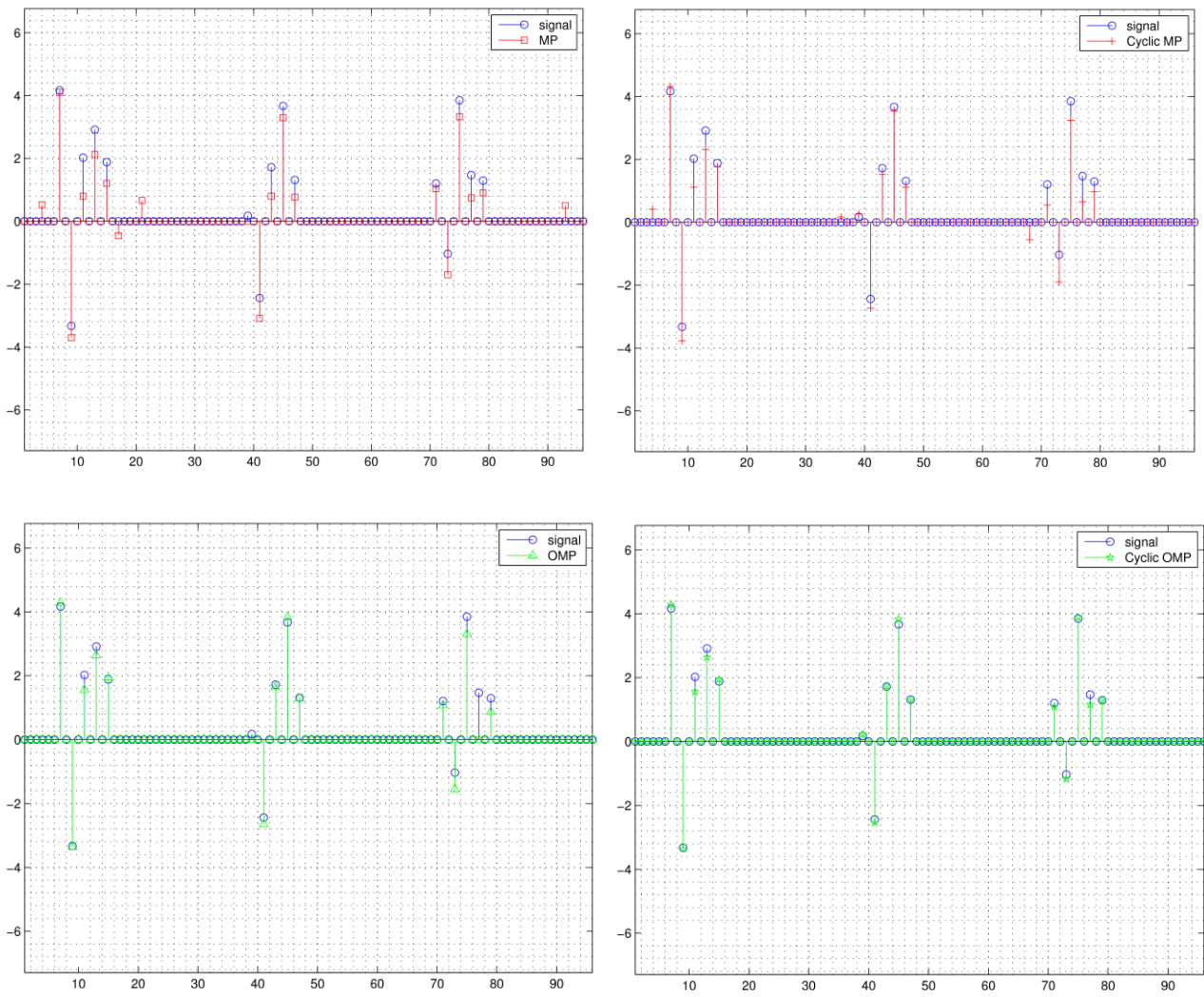


Fig. 2. The reconstructed signal versus the original one (the SNR is 14dB).

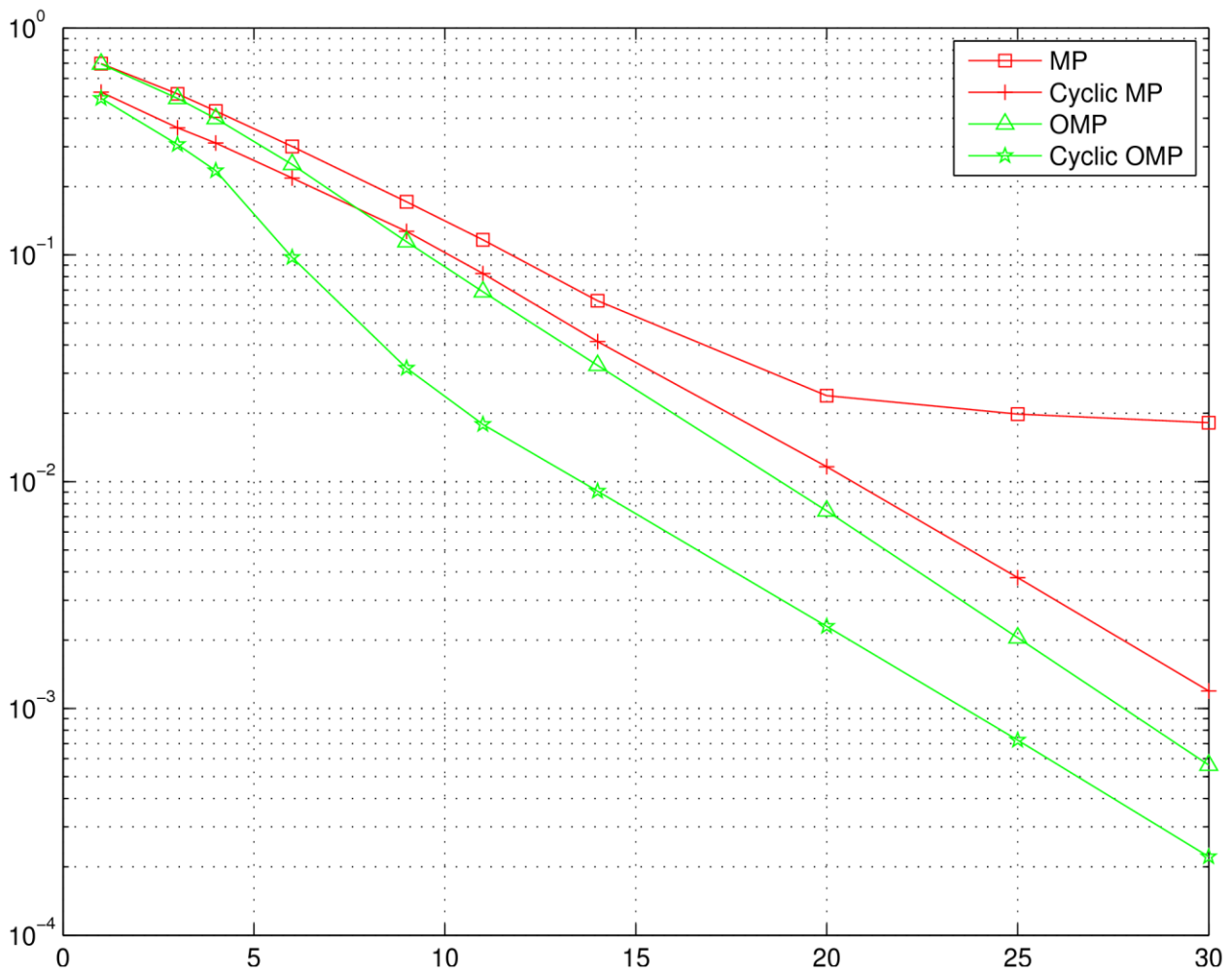


Fig. 3. The effect of varying SNR over MC runs on the MSE.

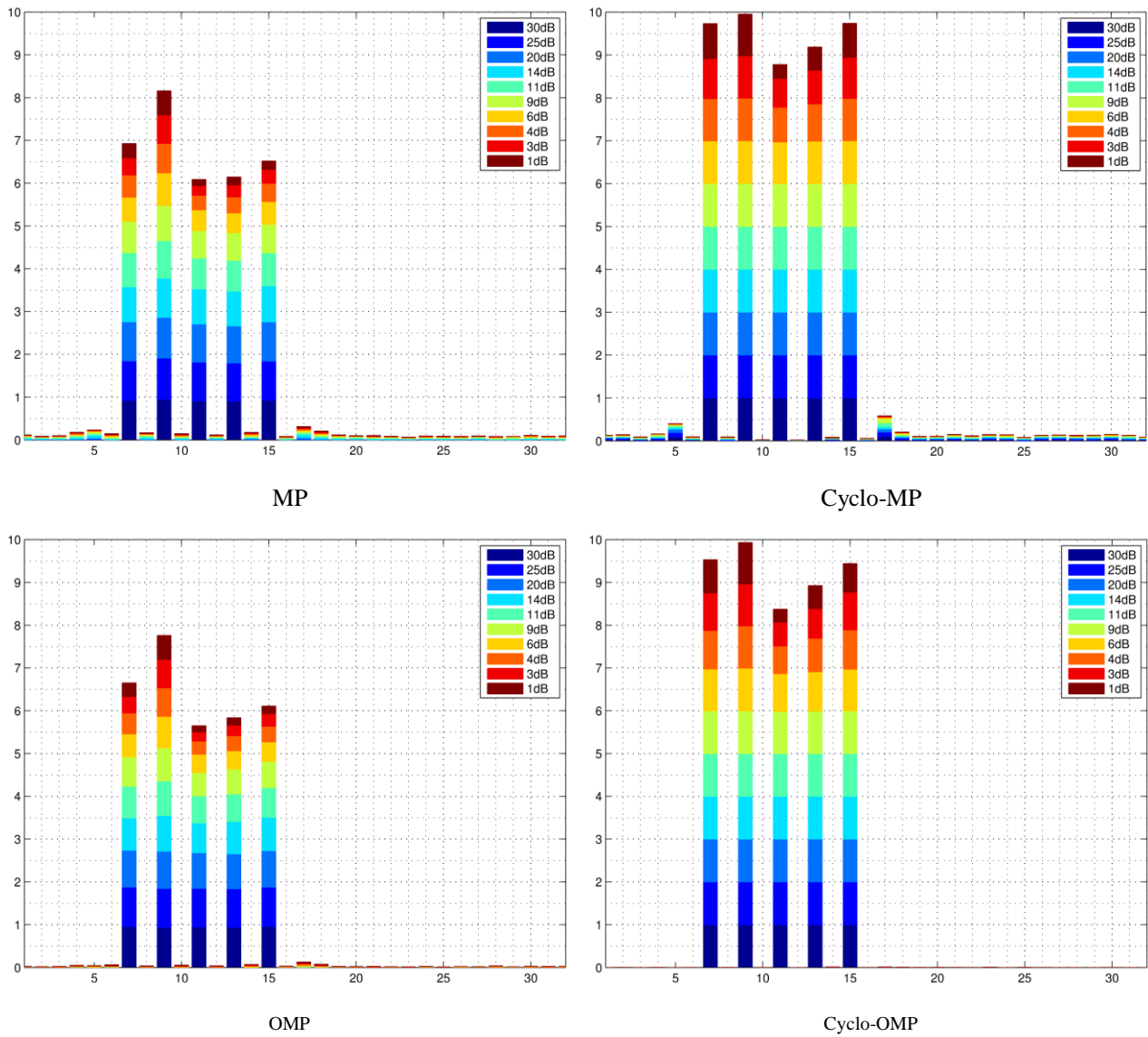


Fig. 4. The effect of varying SNR over MC runs on the histogram.