# VARIATIONS AROUND GRADIENT LIKE ALGORITHMS FOR JOINT DIAGONALIZATION OF HERMITIAN MATRICES

*Tual TRAININI and Eric MOREAU*

Université du Sud Toulon Var
LSIS, UMR CNRS 7296
Av. G. Pompidou, BP 56,
83162 La Valette du Var, Cedex, France
Email: {trainini,moreau}@univ-tln.fr

**ABSTRACT**

In this paper, we address the problem of joint diagonalization of hermitian complex matrix sets, which arises in many signal processing problems (telecommunications, radioastronomy, biology). We present different gradient based algorithms using an optimal step size multiplicative update. Computer simulations are provided to illustrate the comparative behavior of those algorithms together with an application to source separation.

***Index Terms***— Joint Diagonalization, Independant Component Analysis, Source Separation

## 1. INTRODUCTION

Approximate joint diagonalization of matrix sets is an important tool in Blind Source Separation (BSS) or Independant Component Analysis (ICA). It has been used, to name a few, in Joint Approximate Diagonalization of Eigenmatrices (JADE) algorithm for fourth-order cumulants [1], generalized to any order of cumulants in [2] and in Second-Order Blind Identification (SOBI) [3]. In the above approaches, only the case of decompositions under the constraint of a unitary joint diagonalizer, was considered. Other methods were designed in the non-unitary case, as e.g. Alternating Columns-Diagonal Centers (ACDC) in [4], Diagonalization Of Matrices Using a Natural Gradient (DOMUNG) in [5], or others as in [6, 7, 8, 9, 10]. A comparative study of some of these algorithms has been performed in [11].

For the approximate joint diagonalization problem of hermitian complex matrices, we consider gradient based algorithms for their simplicity and ability for real time implementation. Our main goal consists in the derivation of algorithms based on different approximations of the considered criterion before the gradient derivation. All in one, although existing algorithms are encountered, new ones are enlightened. Besides, all algorithms are based on the derivation of an optimal step size, see e.g. [12, 5, 13, 14] which also depends on the chosen approximate criterion.

We are particularly interested in gradient algorithms because of their simplicity and their capability to be easily modified for tracking.

The paper is organized as follows. First, the joint diagonalization problem associated to the classical cost function is reminded. Then the overall iterative procedure is detailed and the derivations leading to different levels of approximate criteria are explained. Finally we derive the associated optimal step size and computer simulations are provided to illustrate the respective performances of the different proposed algorithms.

## 2. JOINT MATRIX DIAGONALIZATION

### 2.1. Problem Statement

The problem is stated as follows. A set of $N$ complex square matrices $\mathbf{M}_i$ is considered. These matrices all admit the following decomposition

$$\mathbf{M}_i = \mathbf{A}\mathbf{D}_i\mathbf{A}^H\,, \qquad i = 1, \ldots, N \tag{1}$$

where $\mathbf{A} \in \mathbb{C}^{N_o \times N_s}$ is the mixing matrix, $\mathbf{D}_i \in \mathbb{R}^{N_s \times N_s}$ are diagonal matrices for all $i$ and $(\cdot)^H$ stands for the transpose conjugate operator. The joint diagonalization problem mainly consists in estimating the mixing matrix $\mathbf{A}$, which is assumed to be full-column rank, or its pseudo inverse denoted by $\mathbf{B} = \mathbf{A}^\dagger$ where $(\cdot)^\dagger$ is the pseudo-inverse operator in the Moore-Penrose sense.

### 2.2. Cost Function

Two classical cost functions can be used. The first one corresponds to the so-called subspace fitting criterion which allows to directly estimate matrix $\mathbf{A}$ in (1) and the second one is a quadratic measure of the diagonality of a matrix defined as

$$\mathcal{J}\left(\mathbf{B}\right) = \sum_{i=1}^{N} \| \mathsf{ZDiag}\left\{\mathbf{B}\mathbf{M}_i\mathbf{B}^H\right\} \|^2 \tag{2}$$

where $\|\cdot\|$ is the Frobenius norm and $\mathsf{ZDiag}\{\cdot\}$ sets to zero the diagonal terms of the matrix argument. We consider the criterion in (2) for the direct estimation of matrix $\mathbf{B}$ and we now consider iterative gradient algorithms for the estimation.

### 2.3. Iterative scheme

The considered iterative scheme was already used in [5]. It is based on two updating stages. The first stage is concerned by the searched matrix and is drawn according to the following multiplicative update

$$\mathbf{B}^{(n+1)} = \left(\mathbf{I} + \mathbf{Z}^{(n)}\right)\mathbf{B}^{(n)} \tag{3}$$

where $\mathbf{I}$ is the identity matrix and $\mathbf{Z}^{(n)} \in \mathbb{C}^{N\times N}$ will be constrained to be a zero diagonal matrix.

The second stage is concerned by the matrices diagonalization. At each iteration, the initial matrices set is updated according to

$$\mathbf{M}_i^{(n+1)} = \mathbf{B}^{(n+1)}\mathbf{M}_i\mathbf{B}^{(n+1)^H} \tag{4}$$

By using (3) in (4), one directly has

$$\begin{aligned}\mathbf{M}_i^{(n+1)} &= \left(\mathbf{I} + \mathbf{Z}^{(n)}\right)\mathbf{B}^{(n)}\mathbf{M}_i\mathbf{B}^{(n)^H}\left(\mathbf{I} + \mathbf{Z}^{(n)}\right)^H \\ &= \left(\mathbf{I} + \mathbf{Z}^{(n)}\right)\mathbf{M}_i^{(n)}\left(\mathbf{I} + \mathbf{Z}^{(n)}\right)^H \end{aligned} \tag{5}$$

Hence the criterion in (2) based on (4) can be seen as only depending on the zero diagonal matrix $\mathbf{Z}^{(n)}$ when (5) is directly considered. This is written as

$$\mathcal{I}(\mathbf{Z}) = \sum_{i=1}^{N} \|\mathsf{ZDiag}\left\{(\mathbf{I}+\mathbf{Z})\,\mathbf{M}_i\,(\mathbf{I}+\mathbf{Z})^H\right\}\|^2 \tag{6}$$

where we drop the iterative index $n$ for simplicity. We will do this in the following each time it appears as not strictly necessary.

In a classical gradient based approach, the matrix $\mathbf{Z}$ can be derived according to $\mathbf{Z} = -\mu\nabla\mathcal{I}(\mathbf{Z})$ where $\mu$ is a small positive step size and the complex gradient is given by $\nabla\mathcal{I}(\mathbf{Z}) = \partial\mathcal{I}(\mathbf{Z})/\partial\mathbf{Z}^*$ where $(\cdot)^*$ is the complex conjugate operator.

However, it was shown in [5] that the matrix $\mathbf{Z}$ can be constrained to be zero diagonal. Here we propose to consider the same constraint written as

$$\mathbf{Z} = -\mu\mathsf{ZDiag}\{\nabla\mathcal{I}(\mathbf{Z})\} = \mu\mathbf{F} \tag{7}$$

thus leading to an approximate gradient algorithm. The explicit derivation of the gradient $\nabla\mathcal{I}(\mathbf{Z})$ can be easily deduced from results established in [5].

But now in order to reduce the computational cost, we are going to consider different approximations of the criterion. The validation of these approximations will be verified thanks to computer simulations in section 3.

### 2.4. Approximate gradient derivations

Using the above gradient approximation, the norm of the matrix $\mathbf{Z}$ clearly decreases with the iterations. Hence we suppose that $\|\mathbf{Z}\| \ll 1$, near a stationary point of the algorithm, in such a way that the matrix product in (6) can be approximated at first order by

$$(\mathbf{I}+\mathbf{Z})\,\mathbf{M}_i\,(\mathbf{I}+\mathbf{Z})^H \approx \mathbf{M}_i + \mathbf{Z}\mathbf{M}_i + \mathbf{M}_i\mathbf{Z}^H \tag{8}$$

leading to the first approximate criterion $\mathcal{I}_1(\mathbf{Z}) \approx \mathcal{I}(\mathbf{Z})$ given by

$$\mathcal{I}_1(\mathbf{Z}) = \sum_{i=1}^{N} \|\mathsf{ZDiag}\left\{\mathbf{M}_i + \mathbf{Z}\mathbf{M}_i + \mathbf{M}_i\mathbf{Z}^H\right\}\|^2 \tag{9}$$

Now let us consider that matrix $\mathbf{M}_i = \mathbf{P}_i + \mathbf{Q}_i$ where

$$\mathbf{P}_i = \mathsf{Diag}\{\mathbf{M}_i\} \quad \text{and} \quad \mathbf{Q}_i = \mathsf{ZDiag}\{\mathbf{M}_i\} \tag{10}$$

where $\mathsf{Diag}\{\cdot\}$ stands for the diagonal matrix built from the diagonal of the matrix argument. Since the matrices $\mathbf{M}_i$ are updated in such a way that they tend to be diagonal, then we also suppose that $\|\mathbf{Q}_i\| \ll 1$, near a separating point, $\forall i$. Thus, using this decomposition into the right term of (8), leads to

$$\mathbf{M}_i + \mathbf{Z}\mathbf{M}_i + \mathbf{M}_i\mathbf{Z}^H \approx \mathbf{P}_i + \mathbf{Q}_i + \mathbf{Z}\mathbf{P}_i + \mathbf{P}_i\mathbf{Z}^H \tag{11}$$

This leads to another approximate criterion $\mathcal{I}_2(\mathbf{Z}) \approx \mathcal{I}(\mathbf{Z})$ given by

$$\mathcal{I}_2(\mathbf{Z}) = \sum_{i=1}^{N} \|\mathsf{ZDiag}\left\{\mathbf{Q}_i + \mathbf{Z}\mathbf{P}_i + \mathbf{P}_i\mathbf{Z}^H\right\}\|^2 \tag{12}$$

Notice that since only the non diagonal terms are considered into the criterion, the matrix $\mathbf{P}_i$ in (11) is useless.

At this stage, the two above approximations $\mathcal{I}_1(\mathbf{Z})$ and $\mathcal{I}_2(\mathbf{Z})$ can be written thanks to a unified formulation as

$$\mathcal{J}(\mathbf{Z}) = \sum_{i=1}^{N} \|\mathsf{ZDiag}\left\{\mathbf{R}_i + \mathbf{Z}\mathbf{S}_i + \mathbf{S}_i\mathbf{Z}^H\right\}\|^2 \tag{13}$$

where $\mathbf{R}_i = \mathbf{S}_i = \mathbf{M}_i$ when (9) is considered and $\mathbf{R}_i = \mathbf{Q}_i$ and $\mathbf{S}_i = \mathbf{P}_i$ when (12) is considered.

Now by using $\|\mathsf{ZDiag}\{\cdot\}\|^2 = \mathsf{tr}\left\{(\cdot)^H\mathsf{ZDiag}\{\cdot\}\right\}$, where $\mathsf{tr}\{\cdot\}$ is the trace of the matrix in argument, we have

$$\begin{aligned}\mathcal{J}(\mathbf{Z}) = \sum_{i=1}^{N} \mathsf{tr}\Big\{&\left(\mathbf{R}_i + \mathbf{Z}\mathbf{S}_i + \mathbf{S}_i\mathbf{Z}^H\right)^H \\ &\mathsf{ZDiag}\left\{\mathbf{R}_i + \mathbf{Z}\mathbf{S}_i + \mathbf{S}_i\mathbf{Z}^H\right\}\Big\}\end{aligned} \tag{14}$$

Considering once again that $\|\mathbf{Z}\| \ll 1$, we have the following approximation $\mathcal{J}_a(\mathbf{Z})$ of $\mathcal{J}(\mathbf{Z})$

$$\begin{aligned}\mathcal{J}_a(\mathbf{Z}) = \sum_{i=1}^{N} \mathsf{tr}\Big\{&\left(\mathbf{R}_i + \mathbf{Z}\mathbf{S}_i + \mathbf{S}_i\mathbf{Z}^H\right)^H \mathsf{ZDiag}\{\mathbf{R}_i\} \\ &+ \mathbf{R}_i^H\mathsf{ZDiag}\left\{\mathbf{R}_i + \mathbf{Z}\mathbf{S}_i + \mathbf{S}_i\mathbf{Z}^H\right\}\Big\}\end{aligned} \tag{15}$$

Finally, the derivation of the gradient using this last approximation is very easy, and is given by

$$\nabla \mathcal{J}_a(\mathbf{Z}) = \sum_{i=1}^{N} \left\{ \mathbf{Q}_i \mathbf{S}_i^H + \mathbf{Q}_i^H \mathbf{S}_i \right\} \tag{16}$$

When considering $\mathbf{S}_i = \mathbf{M}_i$ in the above expression, the overall algorithm corresponds to the DOMUNG algorithm proposed in [5] in the real case.

### 2.5. Optimal step size

In practice, the algorithm performances depend on the choice of the parameter $\mu$ in (7). However this parameter can be derived in an optimal way by searching, at each iteration, its value minimizing the considered criterion.

Hence using (7) in the non approximate criterion (6) leads to the following fourth order polynomial in $\mu$

$$\mathcal{I}(\mathbf{F}) = \sum_{l=0}^{4} \mu^l \sum_{i=1}^{N} a_{i,3}^l(\mathbf{F})$$

where the coefficients $a_i^l(\mathbf{F})$ are given by

$$\begin{aligned}
a_{i,3}^0(\mathbf{F}) &= \mathsf{tr}\left\{\mathbf{M}_i^H \mathbf{Q}_i\right\} \\
a_{i,3}^1(\mathbf{F}) &= \mathsf{tr}\left\{\mathbf{O}_{1,i}^H \mathbf{Q}_i + \mathbf{M}_i^H \mathsf{ZDiag}\left\{\mathbf{O}_{1,i}\right\}\right\} \\
a_{i,3}^2(\mathbf{F}) &= \mathsf{tr}\left\{\mathbf{M}_i^H \mathsf{ZDiag}\left\{\mathbf{O}_{2,i}\right\} + \mathbf{O}_{2,i}^H \mathbf{Q}_i \right. \\
&\qquad \left. + \mathbf{O}_{1,i}^H \mathsf{ZDiag}\left\{\mathbf{O}_{1,i}\right\}\right\} \\
a_{i,3}^3(\mathbf{F}) &= \mathsf{tr}\left\{\mathbf{O}_{1,i}^H \mathsf{ZDiag}\left\{\mathbf{O}_{2,i}\right\} + \mathbf{O}_{2,i}^H \mathsf{ZDiag}\left\{\mathbf{O}_{1,i}\right\}\right\} \\
a_{i,3}^4(\mathbf{F}) &= \mathsf{tr}\left\{\mathbf{O}_{2,i}^H \mathsf{ZDiag}\left\{\mathbf{O}_{2,i}\right\}\right\}
\end{aligned} \tag{18}$$

with

$$\mathbf{O}_{1,i} = \mathbf{F}\mathbf{M}_i + \mathbf{M}_i \mathbf{F}^H \qquad \mathbf{O}_{2,i} = \mathbf{F}\mathbf{M}_i \mathbf{F}^H \tag{19}$$

We can also propose the use of an approximate criterion. Thus, using (7) in the approximate criterion (13) leads to the following second order polynomial in $\mu$

$$\mathcal{J}(\mathbf{F}) = \sum_{l=0}^{2} \mu^l \sum_{i=1}^{N} a_{i,1}^l(\mathbf{F}) \tag{20}$$

where the coefficients $b_i^l(\mathbf{F})$ are given by

$$\begin{aligned}
a_{i,1}^0(\mathbf{F}) &= \mathsf{tr}\left\{\mathbf{R}_i^H \mathsf{ZDiag}\left\{\mathbf{R}_i\right\}\right\} \\
a_{i,1}^1(\mathbf{F}) &= \mathsf{tr}\left\{\mathbf{O}_{3,i}^H \mathsf{ZDiag}\left\{\mathbf{R}_i\right\} + \mathbf{R}_i^H \mathsf{ZDiag}\left\{\mathbf{O}_{3,i}\right\}\right\} \\
a_{i,1}^2(\mathbf{F}) &= \mathsf{tr}\left\{\mathbf{O}_{3,i}^H \mathsf{ZDiag}\left\{\mathbf{O}_{3,i}\right\}\right\}
\end{aligned} \tag{21}$$

with

$$\mathbf{O}_{3,i} = \mathbf{F}\mathbf{S}_i + \mathbf{S}_i \mathbf{F}^H \tag{22}$$

Derivating (17) (respectively (20)) w.r.t. $\mu$ leads to a third (respectively first) degree polynomial. Then, the root of this polynomial minimizing (17) or (20) yields the used optimal step size $\mu$. Notice that in the case of a polynomial of degree one, the root is clearly directly calculated.

### 2.6. Proposed algorithms

All the proposed algorithms are based on the approximate gradient given in (16). They depend on whether matrices $\mathbf{S}_i = \mathbf{M}_i$ or $\mathbf{S}_i = \mathbf{P}_i$, for all $i$, and whether the step size is derived from the original criterion in (6) or from the approximate criterion in (13). When $\mathbf{S}_i = \mathbf{M}_i$ and the step size is derived from (6), the algorithm is called CDOMUNG since it directly corresponds to the DOMUNG algorithm proposed in [5], but in the complex case. When $\mathbf{S}_i = \mathbf{P}_i$ and the step size is derived from (6), the algorithm is called AJDuAG for Approximate Joint Diagonalization using Approximate Gradient. In the two above cases for $\mathbf{S}_i$, when the step size is derived from (13), the two algorithms are called CDOMUNG-OS and AJDuAG-OS. The last three algorithms can be considered as novel.

## 3. SIMULATIONS

In this section, we compare the relative performances of the four above algorithms thanks to computer simulations. Only the square case ($N_o = N_s$) is considered and the used Performance Index [15, 16] is defined according to

$$\begin{aligned}
I(\mathbf{G}) = &\frac{1}{N_s(N_s-1)} \sum_{i=1}^{N_s} \left( \sum_{j=1}^{N_s} \frac{|G_{i,j}|^2}{\max\limits_l |G_{i,l}|^2} - 1 \right) \\
&+ \frac{1}{N_s(N_s-1)} \sum_{j=1}^{N_s} \left( \sum_{i=1}^{N_s} \frac{|G_{i,j}|^2}{\max\limits_l |G_{l,j}|^2} - 1 \right)
\end{aligned} \tag{23}$$

where $\mathbf{G} = \mathbf{B}\mathbf{A}$ is the so-called global matrix. In all the charts, a mean index over 100 Monte Carlo trials is presented.

### 3.1. Joint Diagonalization

We first consider the direct joint diagonalization problem where matrices $\mathbf{M}_i$ are directly built from the model given in (1) where all complex components (the real and the imaginary parts independently) of matrix $\mathbf{A}$ are derived from a uniform distribution onto $[-2.5, 2.5]$ and where all components of all real diagonal matrices $\mathbf{D}_i$ are derived from a uniform distribution onto $[-1, 1]$. The number of matrices $\mathbf{M}_i$ is fixed to $N = 10$ and $N_o = 3$.

We also consider the case of additive $\mathbf{N}_i$ noise matrices onto matrices $\mathbf{M}_i$. The complex components (the real and the imaginary parts independently) of the noise matrices $\mathbf{N}_i$ are derived from a zero mean normal distribution. In this case, we define a Signal over Noise Ratio (SNR) as

$$\mathrm{SNR}_{dB} = 10 \log_{10} \left( \frac{\sigma_s^2}{\sigma_b^2} \right) \tag{24}$$

where $\sigma_s^2$ is the power of each component of the matrices $\mathbf{M}_i$ and $\sigma_b^2$ is the power of each component of the matrices $\mathbf{N}_i$.

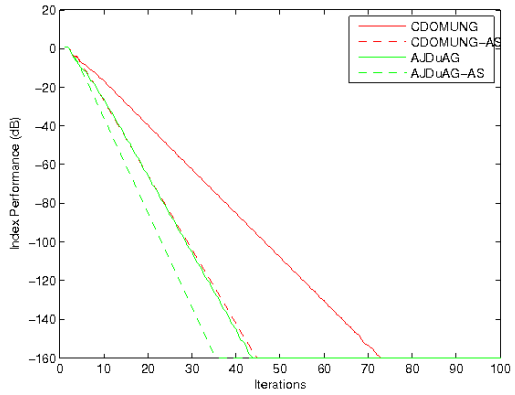


**Fig. 1**. Convergence speed in the noiseless case

The initial guess is the same for all the algorithms: $\mathbf{Z}^{(0)} = 0$ and $\mathbf{B}^{(0)} = \mathbf{I}$.

Fig.1 illustrates the mean convergence speed of the performance index in the noiseless case. One can remark that algorithm AJDuAG − OS has the best convergence speed.

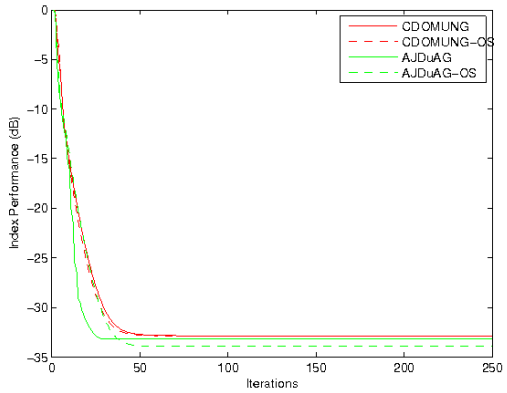


**Fig. 2**. Convergence speed in the noisy case, SNR = 20dB

Fig.2 illustrates the mean convergence speed of the performance index in the noisy case where the SNR is equal to 20dB. In this case, one can remark that the AJDuAG algorithm has the best convergence speed although not associated with the best performances after convergence. After convergence, the AJDuAG − OS algorithm has the best performances.

Fig.3 illustrates the performances (after convergence) of each algorithm with respect to the SNR. One can see that algorithm AJDuAG − OS has globally the best performances.

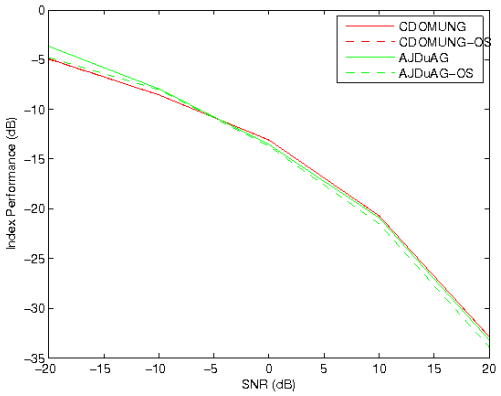


**Fig. 3**. Performance Index vs. SNR

### 3.2. Source Separation

We now consider an application to the source separation problem. We suppose that an observation $\mathbf{x}$ vector is built as

$$\mathbf{x} = \mathbf{A}\mathbf{s} + \mathbf{n} = \mathbf{y} + \mathbf{n} \tag{25}$$

where $\mathbf{s}$ is the so-called source vector, $\mathbf{y}$ stands for the noiseless observations and $\mathbf{n}$ is a gaussian noise vector. We consider fourth order cumulant statistics of the observations defined by $C_x^{a,b,c,d} = \text{Cum}\{x_a, x_b^*, x_c, x_d^*\}$ where Cum$\{\cdot\}$ denotes the cumulant. We now simply build a set of matrices as $\left(\mathbf{M}(c,d)\right)_{a,b} = C_x^{a,b,c,d}$ and it is well known that the matrices $\mathbf{A}^{-1}\mathbf{M}(c,d)\mathbf{A}^{-H}$ are diagonal for all $c$ and $d$. For simulations, we consider the square case where we have 3 sources and 3 observations, hence 9 matrices for the set to be joint diagonalized. The matrix $\mathbf{A}$ is fixed as in the above section. The source signals are complex and their real and imaginary parts follow a uniform distribution onto $[-1, 1]$. The number of data is fixed to $2^{16}$. The SNR is defined in directly using the ratio of the power of the source signals and the power of the noises.

Fig.4 illustrates the mean convergence speed of the performance index in the noisy case when the SNR is equal to 20dB. In this particular application, one can also remark that the AJDuAG − OS algorithm has the best convergence speed and that all algorithms nearly have the same performances after convergence.

Finally, Fig.5 illustrates the performances (after convergence) of each algorithm with respect to the SNR. Nearly all algorithms have the same behavior.

## 4. CONCLUSIONS

In this paper, we present new variations of gradient like algorithms for the joint diagonalization problem of a hermitian matrix set. These variations are based on different levels of approximations that can be fixed for the considered criterion.

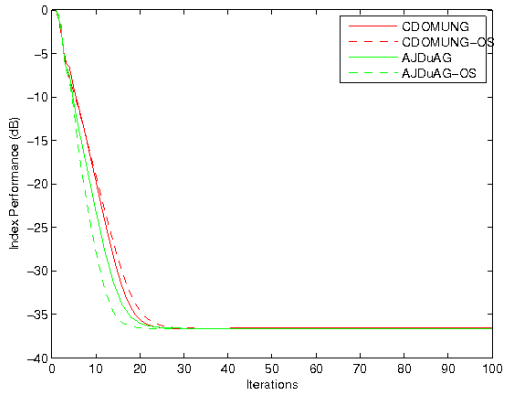


**Fig. 4**. Convergence speed in the noisy case, SNR = 20dB

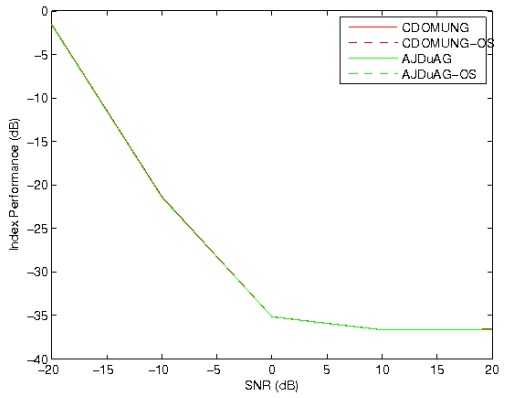


**Fig. 5**. Performance Index vs. SNR

Computer simulations illustrate that a gain of convergence speed is clearly obtained when the highest level of approximation is used. The application to the source separation problem illustrates the same fact.

## 5. REFERENCES

[1] J.-F. Cardoso and A. Souloumiac, "Blind beamforming for non gaussian signal," *IEE Proceedings-F*, vol. 140, pp. 362 – 370, 1993.

[2] E. Moreau, "A generalization of joint-diagonalization criteria for source separation," *IEEE Trans. on Signal Processing*, vol. 49, no. 3, pp. 530 – 541, March 2001.

[3] A. Belouchrani, K. Abed-Meraïm, J.-F. Cardoso, and E. Moulines, "A blind source separation technique using second-order statistics," *IEEE Trans. on Signal Processing*, vol. 45, no. 2, pp. 434 – 444, 1997.

[4] A. Yeredor, "Non-orthogonal joint diagonalization in the least-squares sense with application in blind source separation," *IEEE Trans. on Signal Processing*, vol. 50, no. 7, pp. 1545 – 1553, July 2002.

[5] A. Yeredor, A. Ziehe, and K.-R. Müller, "Approximate joint diagonalization using a natural gradient approach," in *Proc. ICA*, September 2004, vol. 3195, pp. 89 – 96.

[6] P. Comon, "Independent component analysis, a new concept?," *Signal Processing*, vol. 36, pp. 287 – 314, 1994.

[7] D. T. Pham, "Joint approximate diagonalization of positive definite hermitian matrices," *SIAM J. Matrix Anal. Appl.*, vol. 22, no. 4, pp. 1136 – 1152, 2001.

[8] M. Joho and H. Mathis, "Joint diagonalization of correlation matrices by using gradient methods with application to blind signal separation," in *Proc. SAM*, August 2002, pp. 273 – 277.

[9] M. Joho and K. Rahbar, "Joint diagonalization of correlation matrices by using newton methods with application to blind signal separation," in *Proc. SAM*, August 2002, pp. 403 – 407.

[10] A. Ziehe, P. Laskov, G. Nolte, and K.-R. Müller, "A fast algorithm for joint diagonalization with non-orthogonal transformations and its application to blind source separation," *Journal of Machine Learning Research*, vol. 5, pp. 777 – 800, 2004.

[11] S. Degérine and E. Kane, "A comparative study of approximate joint diagonalization algorithms for blind source separation in presence of additive noise," *IEEE Trans. on Signal Processing*, vol. 55, no. 6, pp. 3022 – 3031, June 2007.

[12] R. Bro, *Multi-way Analysis in the Food Industry : Models, Algorithms, and Applications*, Ph.D. thesis, University of Amsterdam, 1998.

[13] Y.H. Dai, "A new gradient method with an optimal stepsize property," *Computational Optimization and Applications*, vol. 33, pp. 73 – 88, 2006.

[14] T. Trainini, X.-L. Li, E. Moreau, and T. Adali, "A relative gradient algorithm for joint decompositions of complex matrices," in *Proc. EUSIPCO*, August 2010, pp. 1073–1076.

[15] E. Moreau and O. Macchi, "A one stage self-adaptive algorithm for source separation," in *Proc. ICASSP*, April 1994, vol. 3, pp. 49–52.

[16] E. Moreau and O. Macchi, "High order contrasts for self-adaptive source separation," *International Journal of Adaptive Control and Signal Processing*, vol. 10, no. 1, pp. 19–46, January 1996.