



## Low cost and efficient kurtosis-based deflationary ICA method: application to MRS sources separation problem.

Majd Saleh, Ahmad Karfoul, Amar Kachenoura, Lotfi Senhadji, Laurent Albera

### ► To cite this version:

Majd Saleh, Ahmad Karfoul, Amar Kachenoura, Lotfi Senhadji, Laurent Albera. Low cost and efficient kurtosis-based deflationary ICA method: application to MRS sources separation problem.. Conference proceedings: . Annual International Conference of the IEEE Engineering in Medicine and Biology Society. IEEE Engineering in Medicine and Biology Society. Annual Conference, Institute of Electrical and Electronics Engineers (IEEE), 2016, Feb 2016, Orlando, United States. pp.3191-3194, 10.1109/EMBC.2016.7591407 . hal-01476136

**HAL Id: hal-01476136**

**<https://hal.science/hal-01476136>**

Submitted on 24 Feb 2017

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Low Cost and Efficient Kurtosis-Based Deflationary ICA Method: Application to MRS Sources Separation Problem

M. Saleh\*, A. Karfoul<sup>†‡</sup>, A. Kachenoura<sup>†‡</sup>, L. Senhadji<sup>†‡</sup> and L. Albera<sup>†‡</sup>  
<sup>\*</sup>Faculty of Mechanical and Electrical Engineering, AL-Baath University, Homs, Syria  
<sup>†</sup>INSERM, UMR 1099, Rennes, France  
<sup>‡</sup>LTSI, Université de Rennes 1, F-35000, Rennes, France

**Abstract**—Improving the execution time and the numerical complexity of the well-known kurtosis-based maximization method, the RobustICA, is investigated in this paper. A Newton-based scheme is proposed and compared to the conventional RobustICA method. A new implementation using the nonlinear Conjugate Gradient one is investigated also. Regarding the Newton approach, an exact computation of the Hessian of the considered cost function is provided. The proposed approaches and the considered implementations inherit the global plane search of the initial RobustICA method for which a better convergence speed for a given direction is still guaranteed. Numerical results on Magnetic Resonance Spectroscopy (MRS) source separation show the efficiency of the proposed approaches notably the quasi-Newton one using the BFGS method.

## I. INTRODUCTION

Independent Component Analysis (ICA) consists of estimating statistically independent source signals from a mixture of them. This mixture can be linearly modeled as follows:

$$\mathbf{x}[m] = \mathbf{A}\mathbf{s}[m] + \mathbf{v}[m] \quad (1)$$

where  $m$  is a natural number and  $\mathbf{x}[m]$  is a realization of an  $N$ -dimensional random vector process  $\{\mathbf{x}[m]\}$  of the observations,  $\mathbf{A}$  ( $N \times P$ ) denotes a tall mixing matrix,  $\mathbf{s}[m]$  stands for a realization of a  $P$ -dimensional random process of sources  $\{\mathbf{s}[m]\}$  with components assumed to be statistically independent and  $\mathbf{v}[m]$  denotes a realization of an  $N$ -dimensional process of noise assumed to be independent from the source. Note that all processes considered in this paper are real-valued. ICA algorithms estimate sources by maximizing a measure of nongaussianity such as the kurtosis (i.e. Fourth order cumulants) of the data [3], [4], [6], [7], [13], [16] to cite a few. Some ICA approaches [3], [13] recover independent sources in only one single-step, these are the so-called *symmetric* approaches. Others, referred to as *deflationary* approaches extract one source at a time [4], [6], [7], [16]. The key advantage of the deflationary scheme is that contrasts for single-source extraction can easily be proven to be free from spurious local solutions, so that all their local optima are associated with valid solutions [5]–[7], [16]. Besides, deflationary scheme can be stopped early once the sources of interest has already been extracted. Hence, a considerable reduction in processing time is guaranteed [15].

RobustICA algorithm is a common deflationary kurtosis-based ICA algorithm. It performs kurtosis maximization

in an efficient way using a gradient-ascent algorithm with global plane search scheme [16]. However both the execution time and the numerical complexity of the RobustICA could be further improved while maintaining a good estimation quality especially for situations where both the previous issues are crucial. For example, in the case of high dimensional problems often encountered in practice such as in High Resolution-ElectroEncephaloGraphic (HR-EEG) data [14], in the context of Magnetic Resonance Spectroscopy (MRS) [9], etc.

The purpose of this paper is to investigate the possible ways of improving both the numerical complexity and the execution time of the RobustICA approach by i) comparing the relative speed of different optimization schemes used in the RobustICA method and ii) investigating the strategies to accelerate the optimization process (by assessing the validity of these strategies). With these aims in mind, a natural extension of the original RobustICA is first given by resorting to the first order optimization scheme, namely the nonlinear Conjugate Gradient (CG) method. Next, beyond this first order implementation, two second order-based approaches are proposed. The Newton method wherein the Hessian is exactly computed and the quasi-Newton using the BFGS (Named after its inventors, Broyden, Fletcher, Goldfarb, and Shanno) scheme where an approximation of the Hessian is considered instead. The efficiency of the proposed approaches is evaluated in the context of the MRS sources separation and compared to both the conventional RobustICA and the deflation-based FastICA [7] methods.

## II. ALGORITHMS

Both the first order and the second order-based RobustICA approaches are considered in this section. Regarding the Newton-based variant, an exact computation of the Hessian is also provided.

*A. First order-based optimization scheme (beyond the classical RobustICA implementation)*

Basically, the RobustICA algorithm solves the real-valued ICA problem (1) by maximizing the following contrast function [16]:

$$F(\mathbf{w}^{(p)}) = \frac{|C_4(y_p)|}{E[y_p^2]^2} = \left| \frac{E[y_p^4]}{E[y_p^2]^2} - 3 \right| \quad (2)$$

where  $C_4(y_p)$  denotes the fourth-order cumulant (kurtosis) of the  $p$ -th estimated source,  $y_p$ , such that  $y_p = \mathbf{w}^{(p)\top} \mathbf{x}$  where  $\mathbf{w}^{(p)}$  is the associated  $p$ -th separation vector. Besides,  $\mathbb{E}[\cdot]$  stands for the mathematical expectation and  $|\cdot|$  denotes the absolute value of its argument.

Although data pre-whitening is not mandatory for RobustICA, employing it would improve the convergence especially with high-dimensional data [8]. Therefore, the observations  $\mathbf{x}$  are considered to be pre-whitened in the sequel. The considered optimization problem is then defined as:

$$\begin{aligned} \max_{\mathbf{w}^{(p)}} F(\mathbf{w}^{(p)}) &= |\mathbb{E}[y_p^4] - 3| \\ \text{s. t. } \|\mathbf{w}^{(p)}\| &= 1 \end{aligned} \quad (3)$$

Originally, a first-order optimization scheme is adopted to maximize  $F$  (2) with the following update rule [16]:

$$\mathbf{w}_{k+1}^{(p)} \leftarrow \mathbf{w}_k^{(p)} + \mu_{opt}^{(k)} \mathbf{g}_k^{(p)} \quad (4)$$

where  $\mathbf{g}_k^{(p)}$  is typically the gradient direction at the current iteration  $k$  associated to the  $p$ -th source and  $\mu_{opt}^{(k)}$  is the optimal step size associated to the considered direction  $\mathbf{g}_k^{(p)}$ . Note that from now on and for the sake of clarity, the subscript  $p$  will be omitted in the rest of this paper. It is noteworthy that an optimal computation of  $\mu$  showed i) a very low computational complexity and ii) a robustness to the presence of spurious local extrema and saddle points in the contrast  $F$  (2) [5], [16]. However, the gradient-ascent algorithm often finds itself taking steps in the same direction as the earlier steps. Hence more iterations are needed for convergence and consequently a considerable numerical complexity. A natural way to improve the convergence speed of the gradient-ascent based RobustICA is to use the nonlinear Conjugate Gradient (CG) instead. Indeed, computing nonlinear CG directions has the same complexity as the gradient descent/ascent directions, but in general, they are more effective [12]. The CG-based RobustICA algorithm, denoted by RobustICA<sub>CG</sub>, based on the Polak and Ribière scheme, is summarized as follows:

- 1) Initialization: random initialization of  $\mathbf{w}_0$ ,  $k \leftarrow 0$  and  $\mathbf{g}_0 \leftarrow \nabla \mathbf{w}_0 F$ , where  $\nabla \mathbf{w}_0 F$  is the gradient of  $F$  with respect to  $\mathbf{w}$  evaluated at  $\mathbf{w} = \mathbf{w}_0$ .
- 2) Repeat until convergence or a maximum number of iterations is reached
  - a) Calculate the optimal step size  $\mu_{opt}$  in the search direction  $\mathbf{g}_k$  as in [16].
  - b) Update  $\mathbf{w}_{k+1} \leftarrow \mathbf{w}_k + \mu_{opt} \mathbf{g}_k$ .
  - c) Normalize  $\mathbf{w}_{k+1} \leftarrow \mathbf{w}_{k+1} / \|\mathbf{w}_{k+1}\|$ .
  - d) Calculate  $\nabla \mathbf{w}_{k+1} F$  as given by [16].
  - e) Calculate  $\beta_{k+1}^{PR} = \frac{\nabla \mathbf{w}_{k+1} F^\top (\nabla \mathbf{w}_{k+1} F - \nabla \mathbf{w}_k F)}{\|\nabla \mathbf{w}_k F\|^2}$ .
  - f) Update the search direction:  $\mathbf{g}_{k+1} \leftarrow \nabla \mathbf{w}_{k+1} F + \beta_{k+1}^{PR} \mathbf{g}_k$ .
  - g)  $k \leftarrow k + 1$ .
- 3) Deflate the observations using the Gram Schmidt approach [16] and go back to step (2).

Note that deflation step can be also performed using a linear regression strategy [16]. For the sake of clarity, only the Gram Schmidt approach is considered in this paper. CG method typically converges in few iterations thanks to the conjugate direction strategy and shows very low numerical complexity and hence small memory allocation compared to its gradient ascent counterpart. However, the convergence speed of the RobustICA is expected to be further improved when second order information about the objective function is incorporated. In this respect, a second-order optimization scheme such as the Newton [2] could be adopted giving rise to the Newton-based RobustICA.

### B. Second order optimization scheme

Newton method exploits further information about the surface of the objective function such as the second order derivative (i.e. the Hessian) which is a good way to accelerate the local convergence [10]. Thanks to the global plane search combined with the Hessian computation, the proposed second order implementation of the RobustICA using the Newton method would provide a faster convergence.

Using the Lagrangian of  $F$  (3), denoted by  $L(\mathbf{w})$  we obtain:

$$L(\mathbf{w}) = |\mathbb{E}[y^4] - 3| + \frac{\beta}{2}(1 - \|\mathbf{w}\|^2) \quad (5)$$

First, the gradient of (5) is obtained as follows:

$$\nabla \mathbf{w} L(\mathbf{w}) = 4\epsilon \mathbb{E}[y^3 \mathbf{x}] - \beta \mathbf{w} \quad (6)$$

where  $\epsilon = \text{sign}(\mathbb{E}[y^4] - 3)$ . The Hessian is then obtained as follows:

$$H(\mathbf{w}) = 12\epsilon \mathbb{E}[y^2 \mathbf{x} \mathbf{x}^\top] - \beta \mathbf{I} \quad (7)$$

Subject to  $\|\mathbf{w}\| = 1$ , the constant Lagrange multiplier  $\beta$  is calculated at the extrema where  $\nabla \mathbf{w} L(\mathbf{w}) = 0$  as follows:

$$\begin{aligned} 4\epsilon \mathbb{E}[y^3 \mathbf{x}] - \beta \mathbf{w} &= 0 \Rightarrow \mathbf{w}^\top \mathbf{w} \beta = 4\epsilon \mathbf{w}^\top \mathbb{E}[y^3 \mathbf{x}] \\ \Rightarrow \|\mathbf{w}\|^2 \beta &= 4\epsilon \mathbf{w}^\top \mathbb{E}[y^3 \mathbf{x}] \Rightarrow \beta = 4\epsilon \mathbb{E}[\mathbf{w}^\top y^3 \mathbf{x}] \\ &\Rightarrow \beta = 4\epsilon \mathbb{E}[y^4] \end{aligned} \quad (8)$$

By substituting (8) in (6) and (7), the gradient and the exact Hessian expressions of  $L$  (5) are respectively obtained as follows:

$$\nabla \mathbf{w} L(\mathbf{w}) = 4\epsilon (\mathbb{E}[y^3 \mathbf{x}] - \mathbb{E}[y^4] \mathbf{w}) \quad (9)$$

$$H(\mathbf{w}) = 4\epsilon (3\mathbb{E}[y^2 \mathbf{x} \mathbf{x}^\top] - \mathbb{E}[y^4] \mathbf{I}) \quad (10)$$

The following pseudo-code summarizes the Newton-based RobustICA approach, denoted by RobustICA<sub>N</sub>:

- 1) Initialization: random initialization of  $\mathbf{w}_0$ ,  $k \leftarrow 0$ , Hessian initialization  $\mathbf{H}_0 = \mathbf{I}$  (where  $\mathbf{I}$  is the identity matrix) and calculate  $\nabla \mathbf{w}_0 L$ .
- 2) Repeat until convergence or a maximum number of iterations is reached
  - a) Compute the search direction  $\mathbf{g}_k = -\mathbf{H}_k^{-1} \nabla \mathbf{w}_k L$  where  $\mathbf{H}_k$  is the Hessian of  $L$  (5).
  - b) Compute the optimal step size  $\mu_{opt}$  in the search direction  $\mathbf{g}_k$  as in [16]

- c) Update  $\mathbf{w}_{k+1} \leftarrow \mathbf{w}_k + \mu_{opt} \mathbf{g}_k$ .
- d) Normalize  $\mathbf{w}_{k+1} \leftarrow \mathbf{w}_{k+1} / \|\mathbf{w}_{k+1}\|$ .
- e) Compute  $\nabla \mathbf{w}_{k+1} L$  using (9).
- f) Compute  $\mathbf{H}_{k+1}$  using (10).
- g)  $k \leftarrow k + 1$ .

3) Deflate the observations using the Gram Schmidt approach [16] and go back to step (2).

Despite its high convergence speed (few iterations), the Newton direction scheme requires an exact computation of the Hessian which is time consuming and computationally demanding per iteration. Therefore, quasi-Newton directions have been proposed wherein the Hessian is rather well approximated using the gradient of the considered objective function. The BGFS approach is one of the most popular quasi-Newton methods. It enjoys the Hessian's auto-correction property. That is to say, the approximated Hessian tends to correct itself in few iterations once an adequate line search defined by Wolfe conditions [12] is adopted. It is worth noted that thanks to the global plane search used in the RobustICA, those conditions are intrinsically verified when implementing the BFGS-based RobustICA approach, denoted by RobustICA<sub>BFGS</sub>. The following pseudo-code summarizes the proposed RobustICA<sub>BFGS</sub> approach:

- 1) Initialization: random initialization of  $\mathbf{w}_0$ . Set  $k \leftarrow 0$ ,  $\mathbf{H}_0^{-1} = \mathbf{I}$  and calculate  $\nabla_{\omega_0} F$ .
- 2) Repeat until convergence or a maximum number of iterations is reached
  - a) Compute the search direction  $\mathbf{g}_k = -\mathbf{H}_k^{-1} \nabla \mathbf{w}_k F$ .
  - b) Compute the optimal step size  $\mu_{opt}$  in the search direction  $\mathbf{g}_k$  as in [16].
  - c) Update  $\mathbf{w}_{k+1} \leftarrow \mathbf{w}_k + \mu_{opt} \mathbf{g}_k$ .
  - d) Normalize  $\mathbf{w}_{k+1} \leftarrow \mathbf{w}_{k+1} / \|\mathbf{w}_{k+1}\|$ .
  - e) Calculate  $\nabla \mathbf{w}_{k+1} F$  as given by [16].
  - f) Define  $\mathbf{s}_k = \mathbf{w}_{k+1} - \mathbf{w}_k$  and  $\mathbf{y}_k = \nabla \mathbf{w}_{k+1} F - \nabla \mathbf{w}_k F$ .
  - g) calculate  $\rho_k = 1 / (\mathbf{y}_k^\top \mathbf{s}_k)$ .  
if  $k = 0$ ,  $\mathbf{H}_k^{-1} = \frac{\mathbf{y}_k \mathbf{s}_k^\top}{\mathbf{y}_k^\top \mathbf{y}_k} \mathbf{I}$ ; end if.  
 $\mathbf{H}_{k+1}^{-1} \leftarrow (\mathbf{I} - \rho_k \mathbf{s}_k \mathbf{y}_k^\top) \mathbf{H}_k^{-1} (\mathbf{I} - \rho_k \mathbf{y}_k \mathbf{s}_k^\top) + \rho_k \mathbf{s}_k \mathbf{s}_k^\top$
  - h)  $k \leftarrow k + 1$ .
- 3) Deflate the observations using the Gram Schmidt approach [16] and go back to step (2).

### III. EXPERIMENTAL ANALYSIS

This section is devoted to compare the performance of the proposed implementations of the RobustICA, namely, RobustICA<sub>CG</sub> and the other proposed Newton-based schemes, the RobustICA<sub>N</sub> and RobustICA<sub>BFGS</sub> to the RobustICA method in its typical gradient-based implementation denoted by RobustICA<sub>G</sub> [16] and the FastICA algorithm in its deflationary version [7]. The FastICA algorithm indeed stands for the most popular method for ICA, due to its simplicity, convergence speed, and satisfactory results in numerous applications. Experiments are carried out on synthetic in vivo MRS data acquired using an array of 32

TABLE I  
THE NUMERICAL COMPLEXITY PER ITERATION IN FLOPS FOR THE FIVE CONSIDERED KURTOSIS-BASED METHODS.  $N$  IS THE NUMBER OF OBSERVATIONS,  $T$  IS THE SAMPLE SIZE AND  $K_p$  IS THE TOTAL NUMBER OF ITERATIONS NEEDED TO EXTRACT THE  $P$ -TH SOURCE.

	Numerical complexity per iteration [flops]
RobustICA <sub>G</sub>	$(5N + 12)T$
RobustICA <sub>CG</sub>	$(5N + 12)T$
RobustICA <sub>BFGS</sub>	$(5N + 12)T$
RobustICA <sub>N</sub>	$(\frac{N^3}{2K_p} + \frac{N^2}{2} + 5N + 12)T$
FastICA	$(2N + 2)T$

sensors. Particularly, a realistic MRS of two metabolites (Creatine and Myo-inositol) is generated using Lorentzian and Gaussian functions [11]. The performance of the considered implementations is evaluated in terms of the Normalized Mean Square Error (NMSE) [1], the execution time, the global numerical complexity in floating point operations (flops) and the total number of iterations required to extract the  $P$  sources. Table I shows the numerical complexity per iteration of the considered methods. The aforementioned performance criteria are evaluated for different SNR values, i.e.  $\text{SNR} \in \{0, 5, 10, 15, 20, 25, 30, 35, 40\}$  dB. Results are averaged over 40 Monte Carlo realizations wherein the noise and the mixing matrix were generated randomly.

Figure 1 shows a mixture of the two above metabolites for SNR equal to 5 dB (Figure 1 (Top)) and the estimated sources using the considered approaches compared to the original sources (red line) (Figure 1 (bottom)). For the sake of clarity i) signals in this figure are slightly vertically shifted with respect to each other and ii) only the extracted signals using the RobustICA<sub>G</sub>, the RobustICA<sub>CG</sub>, the RobustICA<sub>BFGS</sub> are shown where similar behavior for the rest of the considered methods was observed. As shown in figure 1 (bottom), all the proposed RobustICA's variants show quasi similar behavior with comparable estimation accuracy. This claim is justified by the NMSE values shown in figure 2 for all SNR ones. Furthermore, the latter figure shows the benefit behind extending the gradient ascent scheme to the conjugate gradient one in terms of numerical complexity and execution time. Figure 2 shows also the efficiency of the proposed RobustICA<sub>BFGS</sub>. Indeed, it provides a competitive estimation quality, a lower execution time and numerical complexity (reduction around 50%) compared to the original RobustICA<sub>G</sub>. Generally the new RobustICA<sub>CG</sub> implementation and the proposed RobustICA<sub>BFGS</sub> provide a reduction of both the execution time and the numerical complexity over the conventional RobustICA<sub>G</sub> and the conventional FastICA algorithms. Note that, although the FastICA has the lowest numerical complexity per iteration, the entier numerical complexity of both the RobustICA<sub>CG</sub> and the RobustICA<sub>BFGS</sub> stands for the lowest as they converge in much less iterations. Regarding the RobustICA<sub>N</sub>, it shows the smallest number of iterations for convergence but the highest numerical complexity among the all considered methods. This high numerical complexity is due to the necessary exact

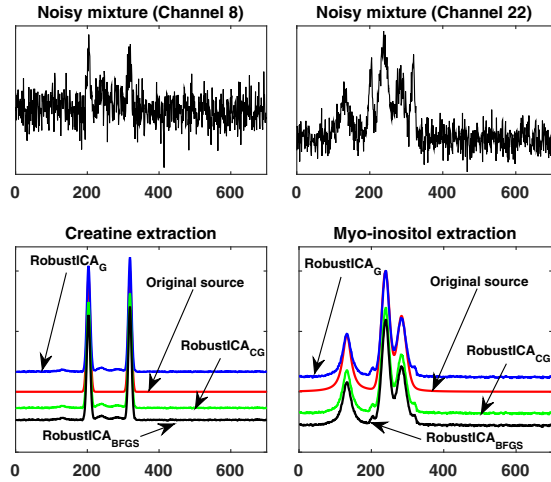


Fig. 1. Top: 2 observations of the Mixture of two metabolites (Creatine and Myo-inositol) for SNR = 5 dB. Bottom: Original sources (Red) vs. estimated sources using RobustICA<sub>G</sub> (Blue), RobustICA<sub>CG</sub> (Green) and RobustICA<sub>BFGS</sub> (Black).

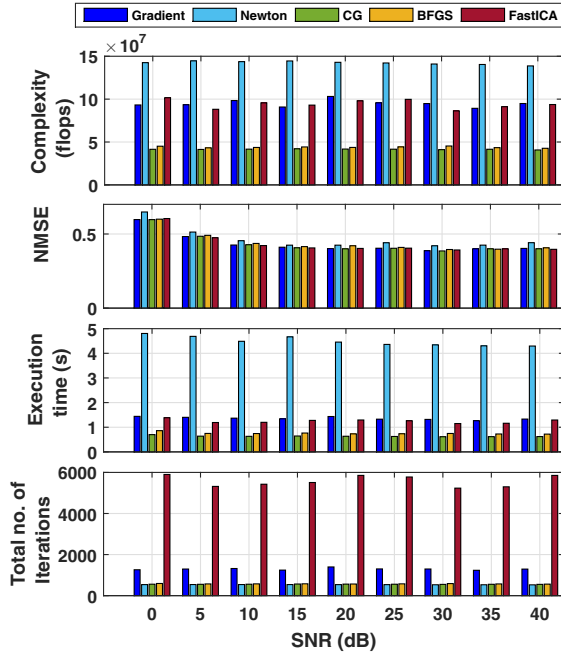


Fig. 2. Global numerical complexity, NMSE, execution time and total number of iterations vs. SNR for RobustICA<sub>G</sub>, RobustICA<sub>N</sub>, RobustICA<sub>CG</sub>, RobustICA<sub>BFGS</sub> and FastICA.

Hessian computation step.

#### IV. CONCLUSION

In this paper, two Newton-based deflationary ICA algorithms have been investigated as a variant of the exact line search kurtosis-based algorithm, namely RobustICA. They are either based on an exact Hessian computation, RobustICA<sub>N</sub> or on an approximation of this latter using the BFGS scheme giving rise to the RobustICA<sub>BFGS</sub> approach. Both the second order optimization-based approaches were

also compared to the deflation-based FastICA and the original RobustICA methods. Besides, the nonlinear CG-based implementation, the RobustICA<sub>CG</sub>, was considered in this study as a natural extension of the gradient ascent-based implementation initially adopted in the RobustICA method. All of the considered variants inherit the exact line search property from the original RobustICA algorithm. Experimental results on MRS data showed around 50% reduction in both of execution time and numerical complexity using the proposed quasi-Newton (i.e. the BGFS method) and the CG-based schemes compared to the original gradient ascent procedure for the RobustICA method. This is while maintaining a good estimation accuracy. Regarding the Newton-based implementations, it showed, despite its fast convergence (i.e. very small number of iterations), a higher execution time and numerical complexity compared to the conventional RobustICA method and the other proposed implementations notably its quasi-Newton counterpart, the RobustICA<sub>BFGS</sub>. This behavior was due to the heavy numerical operations required to compute the exact Hessian.

#### REFERENCES

- [1] L. Albera, A. Kachenoura, P. Comon, A. Karfoul, F. Wendling, L. Senhadji, and I. Merlet. ICA-based EEG demonising: a comparative analysis of fifteen methods. *Special Issue of the Bulletin of the Polish Academy of Sciences - Technical sciences*, 60(3):407–418, 2012.
- [2] S. Boyd and L. Vandenberghe. *Convex optimization*. Cambridge University Press, 2004.
- [3] J.-F. Cardoso and A. Souloumiac. Jacobi angles for simultaneous diagonalization. *SIAM Journal Matrix Analysis and Applications*, 17(1):161–164, 1996.
- [4] P. Comon. Independent component analysis - a new concept? *Signal Processing*, 36:287–314, 1994.
- [5] P. Comon and C. Jutten, editors. *Handbook of Blind Source Separation*. Academic Press, New York, 2010.
- [6] N. Delfosse and P. Loubaton. Adaptive blind separation of independent sources: a deflation approach. *Signal Processing*, 45:59–83, 1995.
- [7] A. Hyvärinen. Fast and robust fixed-point algorithms for independent component analysis. *IEEE Transactions on Neural Networks*, 10(3):626–634, May 1999.
- [8] A. Hyvärinen, J. Karhunen, and E. Oja. *Independent Component Analysis*. John Wiley & Sons, New York, 2001.
- [9] J.F. Jansen, W.H. Backes, K. Nicolay, and M.E. Kooi. 1h mr spectroscopy of the brain: absolute quantification of metabolites. *Radiology*, 240(2):318–332, 2006.
- [10] K. Madsen, H. B. Nielsen, and O. Tingleff. *Methods for non-linear least squares problems. Informatics and Mathematical Modeling*, Technical University of Denmark, 2004.
- [11] S. Moussaoui, D. Brie, and C. Carteret. Non-negative source separation using the maximum likelihood approach. In *IEEE Workshop on Statistical Signal Processing (SSP05)*, pages 1114 – 1119, Novosibirsk, 2005.
- [12] J. Nocedal and S. Wright. *Numerical Optimization*. Springer, New York, 2 edition, 2006.
- [13] E. Oja and Z. Yuan. The fastica algorithm revisited: convergence analysis. *IEEE Transactions on Neural Networks*, 17(6):1370–1381, November 2006.
- [14] R. Oostenveld and P. Praamstra. The five percent electrode system for high-resolution EEG and ERP measurements. *Clinical Neurophysiology*, 112:713–719, 2001.
- [15] S. H. Sardouie, L. Albera, M. B. Shamsollahi, and I. Merlet. An efficient jacob-like deflationary ICA algorithm: Application to EEG denoising. *IEEE Signal Processing Letters*, 22(8):1198–1202, August 2015.
- [16] V. Zarzoso and P. Comon. Robust independent component analysis by iterative maximization of the kurtosis contrast with algebraic optimal step size. *IEEE Transactions on Neural Networks*, 21(2):248–261, 2010.