

# Morphological diversity and sparsity in blind source separation

J. Bobin<sup>1</sup>, Y. Moudden<sup>1</sup>, J. Fadili<sup>2</sup> and J.-L. Starck<sup>1</sup>

<sup>1</sup> [jerome.bobin@cea.fr](mailto:jerome.bobin@cea.fr), [ymoudden@cea.fr](mailto:ymoudden@cea.fr), [jstarck@cea.fr](mailto:jstarck@cea.fr) -  
CEA-DAPNIA/SEDI, Service d'Astrophysique,  
CEA/Saclay, 91191 Gif sur Yvette, France

<sup>2</sup> [jalal.fadili@greyc.ensicaen.fr](mailto:jalal.fadili@greyc.ensicaen.fr) - GREYC CNRS UMR 6072, Image Processing  
Group, ENSICAEN 14050, Caen Cedex, France

**Abstract.** This paper describes a new blind source separation method for instantaneous linear mixtures. This new method coined GMCA (Generalized Morphological Component Analysis) relies on morphological diversity. It provides new insights on the use of sparsity for blind source separation in a noisy environment. GMCA takes advantage of the sparse representation of structured data in large overcomplete signal dictionaries to separate sources based on their morphology. In this paper, we define morphological diversity and focus on its ability to be a helpful source of diversity between the signals we wish to separate. We introduce the blind GMCA algorithm and we show that it leads to good results in the overdetermined blind source separation problem from noisy mixtures. Both theoretical and algorithmic comparisons between morphological diversity and independence-based separation techniques are given. The effectiveness of the proposed scheme is confirmed in several numerical experiments.

## Introduction

Hereafter, we address the classical blind source separation problem. The  $m \times t$  data matrix  $\mathbf{X}$  is the concatenation of  $m$  mixtures  $\{x_i\}_{i=1, \dots, m}$  each of which being the instantaneous linear combination of  $n$  sources  $\{s_i\}_{i=1, \dots, n}$  stored in the  $n \times t$  matrix  $\mathbf{S}$ :

$$\mathbf{X} = \mathbf{AS} + \mathbf{N} \tag{1}$$

where  $\mathbf{A}$  is the mixing matrix and  $\mathbf{N}$  models noise or model imperfections. In this setting, the aim of blind source separation (BSS) techniques is to estimate both the sources  $\mathbf{S}$  and the mixing matrix  $\mathbf{A}$ . BSS is clearly an ill-posed inverse problem which requires additional prior information in order to be solved. Previous work addressing BSS issues clearly emphasized on the need for diversity between the sources to be separated. From a statistical point of view, ICA-like source separation methods use statistical independence (more precisely mutual information) as a kind of “diversity measure” to distinguish between the sources. In [1], the authors proved that maximizing any measure of independence is equivalent to minimizing mutual information. ICA algorithms are then devised

according to particular approximations of mutual information.

Recently, sparsity has raised interest in a wide range of applications. Briefly, a signal is said to be sparse in representation  $\Phi$  if most of the entries of  $\alpha$  such that  $x = \alpha\Phi$  are almost zero and only a few have significant amplitudes. Sparsity-based BSS methods have recently been devised. In [2], a BSS algorithm is described in which it is taken advantage of sparsity to enhance the diversity between independent sources. Several studies (see [3] and references therein) have explored the extreme sparse case as they considered sources with strictly disjoint (and thus orthogonal) supports. In Section 1, we define a particular sparsity-based diversity measure coined morphological diversity. We propose a new effective BSS algorithm coined GMCA which separates the mixed sources based on their morphological diversity. In Section 2, numerical experiments are given showing how GMCA performs well to separate sources from noisy mixtures.

## 1 The GMCA framework

### Notations and definitions

Let  $x$  be a  $1 \times t$  signal and  $\Phi$  a signal dictionary. For the sake of simplicity, we will first assume that  $\Phi$  is orthonormal. In this case,  $x$  has a unique representation  $\alpha$  in  $\Phi$  such that  $x = \alpha\Phi$  readily obtained as  $\alpha = x\Phi^T$ . The support  $\mathcal{S}_{0,\Phi}(x)$  of  $x$  in  $\Phi$  is defined as  $\mathcal{S}_{0,\Phi}(x) = \{t \mid |\alpha[t]| > 0\}$  where  $\alpha[t]$  is the  $t$ -th entry of  $\alpha$ . Let us also define the  $\delta$ -support of  $x$  in  $\Phi$  as :  $\mathcal{S}_{\delta,\Phi}(x) = \{t \mid |\alpha[t]| > \delta\|x\|_\infty\}$ . We then say that two sources  $s_1$  and  $s_2$  are  $\delta$ -disjoint in  $\Phi$  if  $\mathcal{S}_{\delta,\Phi}(s_1) \cap \mathcal{S}_{\delta,\Phi}(s_2) = \emptyset$ . Sources with strictly disjoint supports in  $\Phi$  are obviously  $\delta$ -disjoint with  $\delta = 0$ .

### 1.1 Generalized Morphological Component Analysis

*Sparse coding* : Let us first assume that the mixing matrix  $\mathbf{A}$  is known. In the GMCA framework, the data are modelled as a linear combination of several sources as in Equation 1. Furthermore, the sources  $\{s_i\}_{i=1,\dots,n}$  are assumed to be the linear combination of so-called morphological components (see [4]) :  $s_i = \sum_{k=1}^D \varphi_{ik}$ . By definition, those morphological components are assumed to be sparse in different orthonormal bases  $\{\Phi_k\}_{k=1,\dots,D}$ . Based on these assumptions, the GMCA algorithm endeavors to estimate the sources *via* the estimation of those morphological components :

$$\{\varphi_{ik}\} = \text{Arg min}_{\{\varphi_{ik}\}} \|\mathbf{X} - \mathbf{A}\mathbf{S}\|_2^2 + 2\lambda \sum_{i=1}^n \sum_{k=1}^D \|\varphi_{ik}\Phi_k^T\|_{\ell_1} \quad (2)$$

In [5], we proposed solving this optimization problem by estimating iteratively and alternately each multichannel morphological component  $\{\varphi_{ik}\}$  *via* a ‘‘block-coordinate’’-like algorithm (see [6]). The product  $\mathbf{A}\mathbf{S}$  is then split into  $n \times D$

terms. Introducing the data residual  $\mathbf{X}_{ik} = \mathbf{X} - \sum_{\{j,l\} \neq \{i,k\}} a^i \varphi_{jl}$ , where  $a^i$  is the  $i$ -th column of  $\mathbf{A}$ , the morphological components are estimated one at a time according to :  $\varphi_{ik} = \text{Arg min}_{\varphi_{ik}} \|\mathbf{X}_{ik} - a^i \varphi_{ik}\|_2^2 + 2\lambda \|\varphi_{ik} \Phi_k^T\|_{\ell_1}$ . This equation has an exact solution known as soft-thresholding (see [7]). This sparse decomposition is closely linked to a *sparse coding* stage as already exposed in [8].

*Dictionary learning* : In the previous paragraph, we assumed that the mixing matrix was known and we showed that estimating the morphological components (and thus the sources) boils down to a *sparse coding* step. We consider now that the morphological components are fixed and we want to learn the mixing matrix  $\mathbf{A}$ . This *dictionary* learning issue has already been addressed by extensive work for a wide range of applications. Refer to [8] and references therein for more on that question. Following the same estimation scheme we introduced previously, we propose to estimate each column of  $\mathbf{A}$  assuming the morphological components are fixed as follows:  $a^i = \text{Arg min}_{a^i} \|\mathbf{X} - \sum_{j \neq i} a^j s_j - a^i s_i\|_2^2$ . This update clearly leads to a least-squares estimate of the columns of  $\mathbf{A}$  :  $a^i = \left( \mathbf{X} - \sum_{j \neq i} a^j s_j \right) s_i^T / \|s_i\|_2^2$ .

## 1.2 The GMCA algorithm for blind source separation :

Owing to the “block-coordinate”-like structure of our optimization scheme, for a fixed threshold  $\lambda$ , the *blind* GMCA algorithm estimates alternately the different parameters in the model *i.e.* the columns of  $\mathbf{A}$  and the morphological components. The *blind* GMCA algorithm is as follows.

1. Set the number of iterations  $I_{\max}$  and threshold  $\lambda^{(0)}$
2. While  $\lambda^{(h)}$  is higher than a given lower bound  $\lambda_{\min}$  (e.g. can depend on the noise variance),
  - For  $i = 1, \dots, n$ 
    - For  $k = 1, \dots, D$ 
      - Compute the residual term  $r_{ik}^{(h)}$  assuming the current estimates of  $\varphi_{\{pq\} \neq \{ik\}}$ ,  $\tilde{\varphi}_{\{pq\} \neq \{ik\}}^{(h-1)}$  are fixed:
 
$$r_{ik}^{(h)} = \tilde{a}^{i(h-1)T} \left( \mathbf{X} - \sum_{\{p,q\} \neq \{i,k\}} \tilde{a}^{p(h-1)} \tilde{\varphi}_{\{pq\}}^{(h-1)} \right)$$
      - Estimate the current coefficients of  $\tilde{\varphi}_{ik}^{(h)}$  by Thresholding with threshold  $\lambda^{(h)}$ :
 
$$\tilde{\alpha}_{ik}^{(h)} = \Delta_{\lambda^{(h)}} \left( r_{ik}^{(h)} \Phi_k^T \right)$$
      - Get the new estimate of  $\varphi_{ik}$  by reconstructing from the selected coefficients  $\tilde{\alpha}_{ik}^{(h)}$ 

$$\tilde{\varphi}_{ik}^{(h)} = \tilde{\alpha}_{ik}^{(h)} \Phi_k$$
  - Update  $a^i$  assuming  $a^{p \neq k^{(h)}}$  and the morphological components  $\tilde{\varphi}_{pq}^{(h)}$  are fixed :
 
$$\tilde{a}^{i(h)} = \frac{1}{\|\tilde{s}_i^{(h)}\|_2^2} \left( \mathbf{X} - \sum_{p \neq i} \tilde{a}^{p(h-1)} \tilde{s}_p^{(h)} \right) \tilde{s}_i^{(h)T} \text{ where } \tilde{s}_i^{(h)} = \sum_{k=1}^D \tilde{\varphi}_{ik}^{(h)}$$
  - Decrease the thresholds  $\lambda^{(h)}$  following a given strategy

Note that the value of  $\lambda$  fixes a certain *sparsity level* in the *sparse coding* stage. When  $\lambda$  is “high”, the *sparse coding* step will select the most “significant” features in the data which are very likely to belong to the true morphological components. As already introduced in [5] and [7], the threshold  $\lambda$  decreases towards  $\lambda_{(min)}$  progressively incorporating new features. The purpose of such a thresholding scheme is twofold : i) it provides numerical stability to the algorithm, ii) it gives robustness to noise as the morphological components  $\{\varphi_{ik}\}$  are first estimated from their most significant coefficients in  $\{\Phi_k\}$ . The *sparse coding* step is quite similar to a thresholding-based “denoising”. Handling noisy mixtures then boils down to fixing the final threshold  $\lambda_{min}$ . Typically, in the white Gaussian noise case,  $\lambda_{min} = 3\sigma$  where  $\sigma$  is the noise standard deviation.

### 1.3 A fast GMCA algorithm :

*When  $\Phi$  is orthogonal :* Note that the above GMCA algorithm is a multichannel extension of MCA (Morphological Component Analysis - see [9, 4]) which has been devised in the single channel case. In [4], we showed that MCA is likely to solve the  $\ell_0$  decomposition (and thus the  $\ell_1$  decomposition when the two problems are equivalent - see [10] and references therein) of sparse signals in the overcomplete dictionary  $\Phi = [\Phi_1, \dots, \Phi_D]$ . Nevertheless, in the multichannel case, this *sparse coding* step requires the use of  $D$  transforms for each of the  $n$  sources leading to a prohibitive computational cost. Interestingly, if we restrict ourselves to the case where  $\Phi$  is orthonormal, the problem in Equation 2 becomes simpler:

$$\Theta_S = \text{Arg min}_{\Theta_S} \|\Theta_X - \mathbf{A}\Theta_S\|_2^2 + 2\lambda \sum_{i=1}^n \|\Theta_S\|_{\ell_1} \quad (3)$$

where  $\Theta_S = \mathbf{S}\Phi^T$ . The sources and the mixing matrix can be estimated in the sparse domain  $\Phi$  which drastically reduces the computational burden. Assuming that  $\mathbf{A}$  is nearly orthonormal<sup>3</sup> leads to a very simple *sparse coding* step:

$$\Theta_S = \Delta_\lambda (\mathbf{A}^\dagger \Theta_X) \quad (4)$$

where  $\Delta_\lambda (\cdot)$  is the soft-thresholding operator with threshold  $\lambda$  and  $\mathbf{A}^\dagger$  is the pseudo-inverse of  $\mathbf{A}$ . In that setting, estimating  $\mathbf{A}$  leads straightforwardly to a simple least-squares estimate :

$$\mathbf{A} = \Theta_X \Theta_S^T (\Theta_S \Theta_S^T)^{-1} \quad (5)$$

Interestingly, we show in [5] that alternating the updates in Equation 4 and Equation 5 provides a fixed-point algorithm the convergence condition of which is the following:  $\Theta_S \Delta_\lambda (\Theta_S)^T = (\Delta_\lambda (\Theta_S) \Delta_\lambda (\Theta_S)^T)$

In [5], we give heuristics supporting the good convergence of our algorithm. In the same paper, we also show that the same *fast blind* GMCA algorithm can be used with non-orthonormal  $\Phi$ .

<sup>3</sup> In practice, even if this assumption is very stringent and seldom true, the algorithm performs well.

## 1.4 Morphological diversity

At the beginning of this section we introduced a particular sparsity-based diversity measure to distinguish the sources. A classical sparsity-based diversity measure (see [3]) leads to separate sources with strictly disjoint supports in a sparsifying representation  $\Phi$ . Nevertheless, most *natural* signals seldom have strictly disjoint supports in most practical dictionaries  $\Phi$  (*e.g.* discrete cosine, wavelets, bandlets ... etc.). In [7], we slightly relaxed this assumption by considering sources with disjoint supports in an overcomplete signal dictionary made of a union of orthonormal bases  $\Phi = [\Phi_1^T, \Phi_2^T]^T$ . In this paper, we introduce a new sparsity-based diversity measure which relaxes the strict disjoint support assumption.

*The genesis - a deterministic diversity measure :* In the next section, we switch from the deterministic point of view we adopted in the above, and examine the concept of morphological diversity from the statistical side. In the former viewpoint, separable sources are such that there exists a sparse representation  $\Phi$  in which these signals have  $\delta$ -disjoint supports. Heuristically, given sources *e.g.* images which are “visually” and in that sense morphologically different, there exists a dictionary  $\Phi$  and a value of  $\delta$  for which these sources are sparse and have  $\delta$ -disjoint supports. In that setting, the way  $\Phi$  is chosen specifies which signals are distinguishable. In practice, in image processing, taking  $\Phi$  to be the union of the curvelet frame [11] and the local discrete cosine representation leads to good separation results for a wide range of images.

*From a probabilistic viewpoint :* The minimization problem in Equation 2 can also be interpreted as Maximum A Posteriori estimator of the morphological components assuming : 1) the coefficients of the morphological components are generated independently from the same Laplacian law with zero mean and precision  $\lambda$ , 2) the entries of the mixing matrix are uniformly distributed, 3) the additive noise follows a Gaussian distribution with zero mean and identity covariance matrix. Then what is the meaning of morphological diversity in a statistical framework? The point is that sources generated independently from the same *iid* sparse stochastic process are very likely to have  $\delta$ -disjoint supports for some value of  $\delta$ . For instance, let us assume that the sources  $s_1$  and  $s_2$  are independently generated from the same Laplacian probability density in the sparse  $\Phi$ -domain. Indeed, each entry of the coefficient vector  $\alpha_{i=1,2}$  is drawn according to :  $P_\alpha(\alpha_i[k]) = \frac{\mu}{2} \exp(-\mu|\alpha_i[k]|)$ . We would like to assess the probability for such sources to have  $\delta$ -disjoint supports. Define the proposition  $H_\tau = “|\alpha_1[t^*]| = \|\alpha_1\|_\infty > \tau, |\alpha_2[t^*]| = \|\alpha_2\|_\infty > \tau$  and  $\forall t \neq t^*, |\alpha_{i=1,2}[t]| \leq \tau”$ .  $H_\tau$  states that  $s_1$  and  $s_2$  have strictly  $\tau$ -joint supports; otherwise, if  $H_\tau$  is false then  $s_1$  and  $s_2$  have at least  $\tau$ -disjoint supports. We define  $P_{|\alpha|>\tau} = P_\alpha(|\alpha| > \tau)$  and  $P_{|\alpha|\leq\tau} = P_\alpha(|\alpha| \leq \tau)$ . As the entries of each vector  $\alpha_{i=1,2}$  are independently generated from the same probability density function  $P_\alpha$ , then  $P(H_\tau)$  is such that :  $P(H_\tau) = T \exp(-2\mu\tau) (1 - \exp(-\mu\tau))^{2(T-1)}$ . As, in practice,  $T = dN \gg 1$ , sources generated independently from the same sparse probability density function are  $\tau$ -disjoint. In other words, from a statistical viewpoint,

sparse independent sources are morphologically diverse with very high probability. In that sense a separation technique based on morphological diversity is closely related to ICA in a statistical framework.

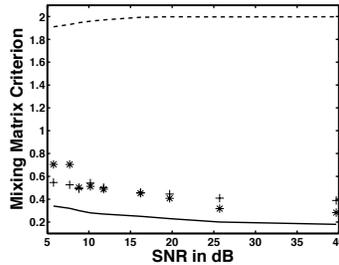
*ICA and GMCA from an algorithmic viewpoint* : The *fast blind* GMCA method introduced in section 1.3 can be expressed as a fixed-point algorithm the convergence condition of which asks that the matrix  $\Theta_{\mathbf{S}}\Delta_{\lambda}(\Theta_{\mathbf{S}})^T$  be symmetric, for all values of  $\lambda$ . Interestingly, as summarized in [12], this condition can be related to the convergence condition of some ICA algorithms which require the symmetry of matrix  $\mathbb{E}\{f(\mathbf{B}\mathbf{X})\mathbf{B}\mathbf{X}\}$  where  $\mathbf{B}$  is a demixing matrix and  $f(\cdot)$  is the so-called *score* function. The thresholding operator  $\Delta_{\lambda}(\cdot)$  in *blind* GMCA is similar in its role to the *score* function of ICA algorithms. A specific and important feature of the thresholding  $\Delta_{\lambda}(\cdot)$  is that it evolves as the threshold  $\lambda$  decreases from one iteration to the next of the *blind* GMCA algorithm. In [5], we give heuristics showing that this “evolving” *score* function is likely to avoid local *false* minima of the objective thus providing some numerical stability to the algorithm. A clear difference lies in the estimation of  $\mathbf{A}$  instead of a demixing matrix  $\mathbf{B}$ ; this distinction is important as GMCA is also designed to handle data in a noisy environment.

## 2 Results

The last paragraph emphasized on sparsity as the key for very efficient source separation methods. In this section, we will compare several BSS techniques with GMCA in an image separation context. We choose 3 different reference BSS methods: i) JADE : the well-known ICA (Independent Component Analysis) based on fourth-order statistics (see [13]), ii) Relative Newton Algorithm : the separation technique we already mentioned. This seminal work (see [14]) paved the way for sparsity in Blind Source Separation. In the next experiments, we used the Relative Newton Algorithm (RNA) on the data transformed by a basic orthogonal bidimensional wavelet transform (2D-DWT), iii) EFICA : this separation method improves the FastICA algorithm for sources following generalized Gaussian distributions (which can be well-suited for some sparse signals). EFICA was also applied after a 2D-DWT of the data where the assumptions on the source distributions are appropriate. Figure 1 shows the original sources (top pictures) and the 2 mixtures (bottom pictures). The original sources  $s_1$  and  $s_2$  have unit variance. The matrix  $\mathbf{A}$  that mixes the sources is such that  $x_1 = 0.25s_1 + 0.5s_2 + n_1$  and  $x_2 = -0.75s_1 + 0.5s_2 + n_2$  where  $n_1$  and  $n_2$  are Gaussian noise vectors (with decorrelated samples) such that the SNR equals 10dB. The noise covariance matrix  $\Gamma_{\mathbf{N}}$  is diagonal. Figure 2 depicts the behavior of the mixing matrix criterion  $\Omega_{\mathbf{A}} = \|\mathbf{I}_n - \mathbf{P}\tilde{\mathbf{A}}^{\dagger}\mathbf{A}\|_1$  ( $\tilde{\mathbf{A}}$  is the estimate of  $\mathbf{A}$ ) as the signal-to-noise ratio (SNR in dB) increases. When the mixing matrix is perfectly estimated,  $\Omega_{\mathbf{A}} = 0$ , otherwise  $\Omega_{\mathbf{A}} > 0$ . First, JADE does not perform well; it points out the importance of choosing an appropriate diversity measure to separate the sources. Thus, fourth-order statistics are not well suited



**Fig. 1.** **Left pictures** : the  $256 \times 256$  source images. **Right pictures** : two different mixtures. Gaussian noise is added such that the SNR is equal to 10dB.



**Fig. 2.** Evolution of the mixing matrix criterion  $\Delta_A$  as the noise variance varies: *solid* line : GMCA, *dashed* line : JADE, (\*) : EFICA, (+) : RNA. **Abscissa** : SNR in dB. **Ordinate** : mixing matrix criterion value.

to the source images in these experiments. Secondly, RNA and EFICA behave rather similarly. Finally, GMCA provides the best results giving mixing matrix criterion values that are approximately 2 times lower than RNA and EFICA. These results clearly show that i) sparsity enhances the distinguishability of the sources, ii) morphological diversity is a well-performing diversity measure and GMCA is well suited to account for that measure.

### 3 Conclusion

In this paper we introduced a new *diversity measure* to distinguish between sources : the morphological diversity. It states that morphologically diverse signals should be separated in so-called sparse representations. The recent advances in harmonic analysis and overcomplete representation theory make morphological diversity a practical way to disentangle source processes from mixtures. We proposed a new algorithm coined *blind* GMCA (Generalized Morphological Component Analysis) to address the blind source separation problem based on morphological diversity. Numerical experiments show that GMCA performs notably well. Furthermore, GMCA is an effective algorithm designed to handle noisy mixtures. Further work will focus on extending the algorithm to the underdetermined blind source separation issue.

## References

1. Lee, T.W., Girolami, M., Bell, A.J., Sejnowski, T.J.: A unifying information-theoretic framework for independent component analysis (1998)
2. Zibulevsky, M., B.B.Pearlmutter: Blind source separation by sparse decomposition. *Neural Computations* **13/4** (2001)
3. Li, Y., S.Amari, A.Cichocki, D.W.C.Ho, Xie, S.: Underdetermined blind source separation based on sparse representation. *IEEE Transactions on signal processing* **54** (2006) 423–437
4. J.Bobin, J-L.Starck, J.Fadili, Y.Moudden, D.L.Donoho: Morphological component analysis: new results. *IEEE Transactions on Image Processing* - revised - available at <http://perso.orange.fr/jbobin/pubs2.html> (2006)
5. J.Bobin, J-L.Starck, J.Fadili, Y.Moudden: Sparsity and morphological diversity in blind source separation. *IEEE Transactions on Image Processing* - submitted - available at <http://perso.orange.fr/jbobin/pubs2.html> (2007)
6. Tseng, P.: Convergence of a block coordinate descent method for nondifferentiable minimizations. *J. of Optim. Theory and Appl.* **109(3)** (2001) 457–494
7. Bobin, J., Moudden, Y., Starck, J.L., Elad, M.: Morphological diversity and source separation. *IEEE Signal Processing Letters* **13(7)** (2006) 409–412
8. M.Aharon, M.Elad, A.Bruckstein: k-svd: An algorithm for designing overcomplete dictionaries for sparse representation. *IEEE Transactions on Signal Processing* **54(11)** (2006) 4311–4322
9. Elad, M., Starck, J.L., Donoho, D., Querre, P.: Simultaneous cartoon and texture image inpainting using morphological component analysis (MCA). *ACHA* **19(3)** (2005) 340–358
10. Tropp, T.: Greed is good : algorithmic results for sparse approximation. *IEEE Transactions on Information Theory* **50(10)** (2004) 2231–2242
11. E.Candes, L.Demanet, D.Donoho, L.Ying: Fast discrete curvelet transforms. *SIAM Multiscale Model. Simul.* **5/3** (2006) 861–899
12. Cichocki, A.: New tools for extraction of source signals and denoising. In: *Proc. SPIE 2005, Bellingham*. Volume 5818. (2005) 11–24
13. Cardoso, J.F.: Blind signal separation: statistical principles. *Proceedings of the IEEE. Special issue on blind identification and estimation* **9(10)** (October 1998) 2009–2025
14. Zibulevski, M.: Blind source separation with relative newton method. *Proceedings ICA2003* (2003) 897–902