

Séparation de sources non-négatives par l'approche du maximum de vraisemblance

Saïd MOUSSAOUI¹, David BRIE¹, Cédric CARTERET²

¹Centre de Recherche en Automatique de Nancy (CRAN), UMR 7039 CNRS-UHP-INPL
Université Henri Poincaré, Nancy 1, B.P. 239, Faculté des Sciences et Techniques, 54506 Vandœuvre-lès-Nancy, France

²Laboratoire de Chimie-Physique et Microbiologie pour l'Environnement (LCPME), UMR 7564 UHP-CNRS
Université Henri Poincaré, Nancy 1, 405, rue de Vandoeuvre, 54600, Villers-lès-Nancy, France

{said.moussaoui,david.brie}@cran.uhp-nancy.fr, carteret@lcpme.cnrs-nancy.fr

Résumé – Ce papier traite du problème de la séparation de sources non-négatives en utilisant l'approche du maximum de vraisemblance. Il est montré que cette approche peut être appliquée avec succès si on utilise une fonction non-linéaire dérivée de la modélisation des distributions des signaux sources par des densités à support positif. Dans le cas des signaux de spectroscopie, qui est notre application principale, une bonne loi candidate est la distribution gamma qui permet de prendre en compte conjointement la non-négativité et la parcimonie des signaux sources. Un exemple synthétique est présenté afin d'étudier les performances de l'algorithme résultant.

Abstract – This paper addresses the problem of non-negative source separation using the maximum likelihood approach. It is shown that this approach can be effective by considering that the sources are distributed according to a density having a non-negative support from which an adequate non-linear separating function can be derived. In the particular of spectroscopic data which is our main concern, a good candidate is the Gamma distribution which allows to encode both the non-negativity and the sparsity of the source signals. A synthetic example is used to assess the performances of the resulting algorithm.

1 Introduction

Ce travail est motivé par le besoin d'appliquer les méthodes de séparation de sources au traitement des signaux résultants de l'analyse de substances chimiques multi-composantes par spectroscopie. Les objectifs du traitement sont l'identification de la composition chimique de la substance et l'évaluation de la proportion de chaque composante. D'après la loi de Beer-Lambert, les signaux mesurés sont une combinaison linéaire des spectres des composantes pures inconnues. C'est donc un problème de séparation de sources où les signaux sources correspondent aux spectres pures et les concentrations de chaque espèce sont déduites à partir des coefficients de mélange. La contrainte principale de cette application est que les coefficients de mélange et les signaux sources sont non-négatifs.

Le modèle d'observation suppose que m signaux mesurés, notés $\mathbf{x}_t = [x_1(t), \dots, x_m(t)]^T$, sont un mélange linéaire de p signaux sources non-mesurables $\mathbf{s}_t = [s_1(t), \dots, s_p(t)]^T$

$$\mathbf{x}_t = \mathbf{A} \mathbf{s}_t + \mathbf{e}_t, \quad (1)$$

où \mathbf{A} est la matrice de mélange et de dimension $(m \times p)$. Chaque colonne de cette matrice contient le profil d'évolution de la contribution de chaque source dans les mélanges observés. Le vecteur \mathbf{e}_t représente les erreurs de mesure et les incertitudes du modèle d'observation. Ayant n échantillons, ($t = 1, \dots, n$), et en utilisant une notation matricielle, le modèle de mélange est exprimé par

$$\mathbf{X} = \mathbf{A} \mathbf{S} + \mathbf{E}, \quad (2)$$

où les matrices $\mathbf{X} \in \mathbb{R}^{m \times n}$, $\mathbf{S} \in \mathbb{R}_+^{m \times n}$ et $\mathbf{E} \in \mathbb{R}^{m \times n}$ contiennent, respectivement, les signaux mesurés, les signaux

sources et les séquences de bruits sur chaque ligne. En supposant que le nombre de sources est connu *a priori*, le problème de la séparation de sources non-négatives consiste à estimer les signaux sources et les coefficients de mélanges qui reproduisent aux mieux les données tout en satisfaisant les contraintes de non-négativité.

Afin de résoudre ce problème, plusieurs approches sont possibles. La première est fondée sur une estimation par moindres carrés sous contrainte de non-négativité [1, 2]. Néanmoins, cette approche ne fournit une solution unique que sous certaines conditions restrictives [3, 4]. La sélection d'une solution particulière parmi toutes les solutions admissibles se fait alors par le rajout d'hypothèses supplémentaires. Nous utilisons une approche du maximum de vraisemblance pour tenir compte de l'hypothèse d'indépendance statistique des signaux sources. Sous cette hypothèse, c'est donc un problème d'analyse en composantes indépendantes non-négatives pour lequel une méthode existante permet de faire la séparation dans le cas de sources *well grounded* et orthogonales [5]. En revanche, comme cela est montré dans [6], l'utilisation des méthodes ICA qui imposent l'orthogonalité des sources, tels que JADE [7] ou SOBI [8], peut mener à des estimations négatives lorsque les échantillons disponibles des signaux sources présentent une corrélation spatiale significative. L'approche du maximum de vraisemblance n'impose pas cette contrainte d'orthogonalité mais requiert la spécification d'une fonction non-linéaire afin d'obtenir les sources les plus indépendantes possible et dont la distribution est proche de la densité correspondant à cette fonction non-linéaire. Nous utilisons des distributions gamma qui sont de bons modèles pour tenir compte simultanément de la non-négativité et de la parcimonie des signaux sources [9].

2 Séparation de sources non-négatives par maximum de vraisemblance

2.1 Formulation du problème

Sous l'hypothèse de signaux sources mutuellement statistiquement indépendants et temporellement i.i.d et en tenant compte du modèle de mélange $\mathbf{x}_t = \mathbf{A} \mathbf{s}_t$, dans le cas carré ($m = p$) et non bruité ($\mathbf{e}_t = 0$) et en introduisant la matrice séparante, notée \mathbf{B} , l'estimée de cette matrice par maximum de vraisemblance s'obtient par [10, 11]

$$\hat{\mathbf{B}} = \arg \min_{\mathbf{B}} \left\{ -\log |\det(\mathbf{B})| - \frac{1}{n} \sum_{t=1}^n \sum_{j=1}^p \log p_{s_j}([\mathbf{B} \mathbf{x}_t]_j) \right\}. \quad (3)$$

L'optimisation de ce critère en utilisant la méthode du gradient relatif/naturel [12] conduit à une mise à jour itérative selon

$$\hat{\mathbf{B}}^{(r+1)} = \hat{\mathbf{B}}^{(r)} - \mu^{(r)} \left[\frac{1}{n} \sum_{t=1}^n \psi(\hat{\mathbf{y}}_t^{(r)}) (\hat{\mathbf{y}}_t^{(r)})^T - \mathbf{I} \right] \hat{\mathbf{B}}^{(r)}, \quad (4)$$

où $\hat{\mathbf{y}}_t^{(r)} = \hat{\mathbf{B}}^{(r)} \mathbf{x}_t$ et $\psi^T(\mathbf{z}) = [\psi_1(z_1), \dots, \psi_n(z_n)]$ et les fonctions de séparation $\psi_j(z)$ sont définies par $\psi_j(z) = -\frac{\partial}{\partial z} \log p_{z_j}(z)$. Pour le cas où $m > p$, une réduction de dimension de la matrice des données par transformation aléatoire ($p \times m$) ou par analyse en composantes principales est nécessaire avant l'application de cet algorithme.

Idealement, les fonctions de séparation se déduisent des densités de probabilité des sources recherchées [11, 13], qui dans la plupart des cas ne sont pas connues *a priori*. Il faudra alors estimer ces densités en même temps que la matrice de séparation. Pour cela, la distribution des sources est approchée soit en utilisant les statistiques d'ordre supérieur [10, 14] ou par un modèle paramétrique [11, 15]. Le choix d'un modèle de distribution permet de spécifier des informations préalables sur la structure de la densité de probabilité des sources. Quelques exemples de modèles sont la gaussienne généralisée, la distribution de Student et le mélange de gaussiennes. Par exemple, les deux premiers modèles correspondent à des distributions unimodales symétriques mais ne permettent pas de coder explicitement la contrainte de non-négativité des signaux sources. Pour le cas de sources asymétriques ou multi-modales l'utilisation d'un modèle de mélange de gaussiennes permet de s'affranchir de cette limitation, mais le cas de sources non-négatives reste toujours problématique.

2.2 Synthèse d'une fonction non-linéaire

Afin de prendre en compte explicitement la contrainte de non-négativité des sources, une modélisation de leurs distributions par des densités à support non-négatif est nécessaire. Dans le cadre de l'analyse des signaux de spectroscopie, un modèle possible est la distribution gamma dont la densité de probabilité s'exprime selon

$$p(s|\alpha, \beta) = \frac{\beta^\alpha}{\Gamma(\alpha)} s^{\alpha-1} e^{-\beta s} \mathbb{I}_{s \geq 0}, \quad (5)$$

où α et β sont les paramètres de forme et d'échelle. Le choix de cette densité est motivé par deux avantages ; celle-ci prend en compte explicitement la non-négativité car $p(s < 0) = 0$, et les deux paramètres α et β de cette densité peuvent être utilisés pour ajuster sa forme asymétrique. La fonction non-linéaire correspondante à la loi gamma est exprimée par

$$\psi(s; \alpha, \beta) = \beta - (\alpha - 1) s^{-1}, \quad s > 0. \quad (6)$$

Cette densité de probabilité est à support non-négatif, donc pour des estimées intermédiaires négatives, durant l'optimisation, il faudra redéfinir la fonction non-linéaire selon

$$\psi(s; \alpha, \beta) = \beta - \frac{(\alpha - 1)}{\max(s, \epsilon)}, \quad \forall s \in \mathbb{R} \quad (7)$$

où ϵ est fixé à une petite valeur (de l'ordre de 10^{-3}). Un exemple de distribution gamma et de fonction non-linéaire qui en résulte sont illustrés par la figure 1.

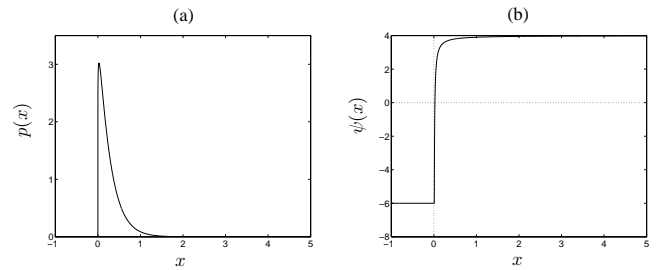


FIG. 1 – (a) Distribution gamma pour $\alpha = 1.1$ et $\beta = 4$ et (b) fonction non-linéaire associée pour $\epsilon = 0.01$.

2.3 Stabilité de l'algorithme

Des conditions suffisantes de stabilité de l'algorithme du gradient relatif/naturel sont exprimées par [16, 17]

$$\mathbb{E}[\psi'(\tilde{s}_j) \tilde{s}_j^2] - \mathbb{E}[\psi(\tilde{s}_j) \tilde{s}_j] > 0 \quad \forall j = 1, \dots, p \quad (8)$$

où $\mathbb{E}[f(\tilde{s}_j)]$ sont les moments non-linéaires des sources centrées

$$\tilde{s}_j = s_j - \mathbb{E}[s_j] = s_j - \alpha_j / \beta_j. \quad (9)$$

Le calcul de ces moments donne

$$\mathbb{E}[\psi'(\tilde{s}_j) \tilde{s}_j^2] = \mathbb{E} \left[\frac{(\alpha_j - 1)}{\tilde{s}_j^2} \cdot \tilde{s}_j^2 \right] = (\alpha_j - 1); \quad (10)$$

et

$$\mathbb{E}[\psi(\tilde{s}_j) \tilde{s}_j] = \mathbb{E}[\beta_j \tilde{s}_j - (\alpha_j - 1)] = (1 - \alpha_j). \quad (11)$$

Par conséquent, une condition suffisante pour la stabilité de l'algorithme est

$$\alpha_j > 1, \quad j = 1, \dots, p. \quad (12)$$

A partir de cette condition de stabilité, il apparaît que seulement des densités gamma ayant un paramètre de forme $\alpha > 1$ peuvent être utilisées, ce qui est un inconvénient de l'algorithme d'optimisation choisi. Nous revenons à ce point dans le paragraphe 3.2.

2.4 Estimation des paramètres des sources

Les paramètres des densités gamma $\theta = \{\alpha_j, \beta_j\}_{j=1}^p$ sont estimés conjointement avec la matrice de séparation par la maximisation de la fonction de vraisemblance conjointe

$$(\hat{B}, \hat{\theta}) = \arg \max_{B, \theta} p(X|B, \theta), \quad (13)$$

La minimisation du critère qui en résulte est réalisée par un algorithme de gradient alterné mettant à jour, à chaque itération, la matrice de séparation en utilisant la dernière valeur du vecteur de paramètres et en utilisant cette nouvelle estimation de la matrice de séparation pour mettre à jour ces paramètres. La mise à jour de la matrice de séparation étant réalisée selon l'équation (4), les paramètres des densités gamma sont mis à jour selon

$$\begin{cases} \alpha_j^{(r+1)} = \alpha_j^{(r)} - \rho_\alpha \nabla_{\alpha_j} \mathcal{J}(\mathbf{s} | \alpha_j^{(r)}, \beta_j^{(r)}), \\ \beta_j^{(r+1)} = \beta_j^{(r)} - \rho_\beta \nabla_{\beta_j} \mathcal{J}(\mathbf{s} | \alpha_j^{(r+1)}, \beta_j^{(r)}), \end{cases} \quad (14)$$

où $(\rho_\alpha, \rho_\beta)$ sont des pas positifs et

$$\begin{cases} \nabla_{\alpha_j} \mathcal{J}(\mathbf{s} | \alpha_j, \beta_j) = -\log \beta_j + \Psi(\alpha_j) - \frac{1}{n} \sum_{t=1}^n \log s_j(t) \\ \nabla_{\beta_j} \mathcal{J}(\mathbf{s} | \alpha_j, \beta_j) = -\frac{\alpha_j}{\beta_j} + \frac{1}{n} \sum_{t=1}^n s_j(t), \end{cases}$$

où $\Psi(\alpha_j) = \frac{d}{d\alpha} \Gamma(\alpha_j)$ est la fonction digamma [18]. En tenant compte des conditions de stabilité (12) et de la définition de la densité gamma, les valeurs de α_j et β_j sont contraintes à $\alpha_j \geq (1 + \epsilon)$ et $\beta_j \geq \epsilon$ où ϵ est fixé à une faible valeur (10^{-3}).

3 Exemple numérique

Les signaux sources et les coefficients de mélange sont simulés de sorte à ce qu'ils soient similaires aux données résultantes de l'analyse de substances chimiques multicomposantes par des techniques de spectroscopie. La figure 2 montre un exemple de $p = 3$ signaux sources normalisés (variance unité) de $n = 1000$ échantillons obtenus à partir d'un modèle simplifié dans lequel chaque signal résulte de la superposition de $K = 15$ motifs lorentziens et gaussiens dont les amplitudes et les paramètres de position et de dispersion sont choisis de façon pseudo-aléatoire. Les profils des coefficients de mélange sont simulés pour $m = 3$ observations.

Afin de mesurer la qualité de la reconstruction des signaux sources, nous utilisons l'erreur quadratique moyenne normalisée, notée \mathcal{E}_{s_j} et définie par

$$\mathcal{E}_{s_j} = \sum_{t=1}^n (s_j(t) - \hat{s}_j(t))^2 / \sum_{t=1}^n (s_j(t))^2, \quad (15)$$

où \hat{s}_j est l'estimée de la j -ème source s_j et les deux signaux sont réduits à une variance unité. Pour comparer les performances globales, nous utilisons l'indice

$$\mathcal{I}_{global} = \frac{1}{2p(p-1)} \sum_{i=1}^p \left\{ \left(\sum_{k=1}^p \frac{|g_{ik}|^2}{\max_{\ell} |g_{i\ell}|^2} - 1 \right) + \left(\sum_{k=1}^p \frac{|g_{ki}|^2}{\max_{\ell} |g_{\ell i}|^2} - 1 \right) \right\}, \quad (16)$$

où g_{ij} est (i, j) -ème élément de la matrice $\mathbf{G} = \hat{B}\mathbf{A}$.

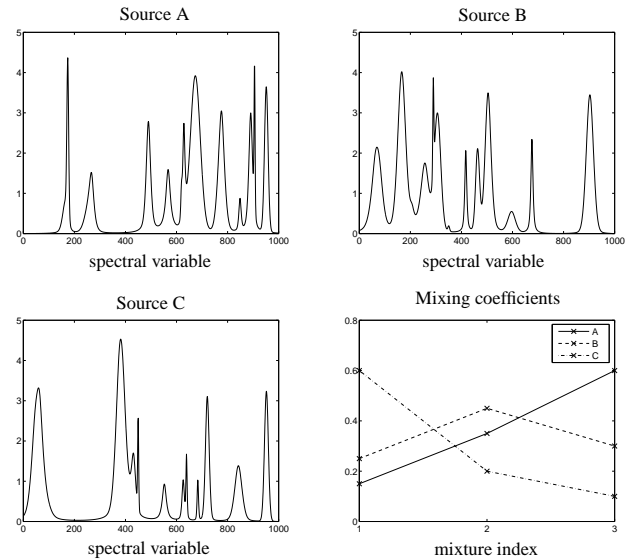


FIG. 2 – Signaux sources et coefficients de mélange simulés.

3.1 Cas sans bruit

La matrice de covariance empirique

$$\hat{\mathbf{R}}_s = \begin{bmatrix} 1.00 & 0.13 & -0.18 \\ 0.13 & 1.00 & -0.22 \\ -0.18 & -0.22 & 1.00 \end{bmatrix}, \quad (17)$$

n'est pas diagonale, ce qui montre que les échantillons disponibles des signaux sources présentent une corrélation spatiale non-négligeable. Ainsi, l'application d'une méthode qui force l'orthogonalité (SOBI ou JADE) va produire une solution incorrecte, de même la méthode NNICA fournit des estimations négatives, même si celle-ci considère la non-négativité des signaux sources. D'autre part, l'application d'une méthode fondée uniquement sur la contrainte de non-négativité (NMF ou ALS) ne fournit pas une solution unique car les conditions nécessaires pour l'unicité de la solution sous la contrainte de non-négativité uniquement ne sont pas satisfaites [4]. Par ailleurs, notons la supériorité de l'approche du maximum de vraisemblance en termes d'indice de séparation global ou de qualité de reconstruction des sources.

TAB. 1 – Comparaison des performances des différentes approches. MLPSS correspond à l'approche proposée et mes indices de performances sont exprimés en décibels (dB).

	JADE	NMF	NNICA	MLPSS
\mathcal{E}_A	-10.58	-16.20	-15.11	-35.88
\mathcal{E}_B	-10.96	-15.88	-14.19	-42.25
\mathcal{E}_C	-17.40	-14.14	-39.41	-40.30
\mathcal{I}_{global}	-12.06	-15.67	-15.90	-38.32

3.2 Cas avec bruit

Afin de discuter les performances de la méthode, un bruit gaussien est ajouté aux observations. Le niveau de bruit est exprimé en termes de rapport signal sur bruit (RSB) défini comme étant le rapport des variances des signaux non-bruités et des séquences de bruit rajoutées. La figure 3 montre l'évolution des performances des méthodes NNICA, MLPSS et BPSS [9] en fonction du niveau de bruit. Pour un RSB très fort, l'utili-

sation de la distribution gamma et l'estimation soit par maximum de vraisemblance ou par approche bayésienne donne de meilleures performances que NNICA. En dépit du fait que BPSS et MLPSS utilisent le même modèle *a priori* sur la distribution des sources, les performances de BPSS sont supérieures, ce qui s'explique par la restriction du paramètre de forme de la densité gamma à une valeur supérieure à 1, afin de garantir la stabilité de l'algorithme du gradient, et de la non-prise en compte de la positivité des coefficients de mélange. En augmentant le niveau de bruit, les performances de MLPSS diminuent puisque cette approche ne prend pas en compte la présence du bruit sur les observations. Afin de palier à ces limitations, une alternative est d'utiliser les algorithmes EM (Expectation-Maximization) afin d'estimer conjointement les coefficients de mélange et la covariance du bruit d'observation et d'envisager des distributions gamma avec des paramètres de forme inférieures à 1.

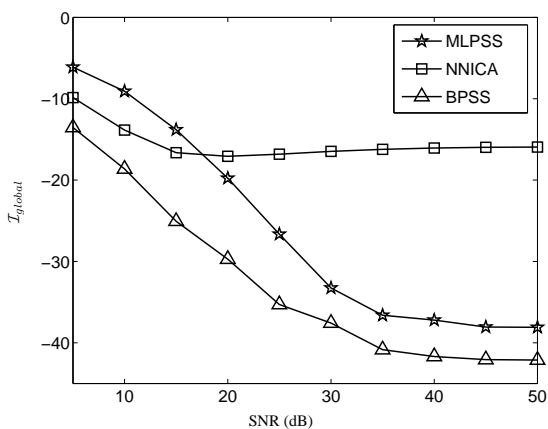


FIG. 3 – Influence du niveau de bruit sur les performances de NNICA, MLPSS et BPSS.

4 Conclusion

L'objectif de ce papier est de montrer que l'approche du maximum de vraisemblance peut être appliquée à la séparation de sources non-négatives à condition de choisir des fonctions non-linéaires adéquates. La non-négativité des sources est prise en compte via la synthèse d'une fonction non-linéaire déduite de la modélisation des sources par des densités à support non-négatif. Dans le cas de données spectrales, la non-négativité et la parcimonie des signaux sources sont incorporées dans la méthode de séparation en utilisant des distributions gamma pour la synthèse de fonctions non-linéaires appropriées. L'analyse des performances dans le cas de données bruitées a montré la nécessité de considérer la présence de bruit dans le modèle d'observation et de prendre en compte conjointement la non-négativité des signaux sources et des coefficients de mélange.

Remerciements

Ce travail est financé par la "Région Lorraine" et le CNRS.

Références

- [1] R. Bro and S. De Jong, "A fast non-negativity constrained least squares algorithm," *Journal of Chemometrics*, vol. 11, pp. 393–401, 1997.
- [2] D. Lee and H. Seung, "Learning the parts of objects by non-negative matrix factorization," *Nature*, vol. 401, pp. 788–791, 1999.
- [3] D. Donoho and V. Stodden, "When does non-negative matrix factorization give a correct decomposition into parts?" in *proceedings of NIPS*, 2003.
- [4] S. Moussaoui, D. Brie, and J. Idier, "Non-negative source separation : Range of admissible solutions and conditions for the uniqueness of the solution," in *proceedings of ICASSP*, 2005, pp. 289–292.
- [5] M. D. Plumbley, "Algorithms for non-negative independent component analysis," *IEEE Transactions on Neural Networks*, vol. 14, no. 3, pp. 534–543, 2003.
- [6] D. Nuzillard and J. Nuzillard, "BSS applied to non-orthogonal signals," in *proceedings of ICA*, 1999.
- [7] J.-F. Cardoso and A. Souloumiac, "Blind beamforming for non Gaussian signals," *IEE Proceedings-F*, vol. 140, no. 6, pp. 362–370, 1993.
- [8] A. Belouchrani and K. Abed-Meraim, "Séparation aveugle au second ordre de sources corrélées," in *actes du Colloque GRETSI sur le Traitement du Signal et des Images (GRETSI'93)*, Juan-les-Pins, France, 1993.
- [9] S. Moussaoui, D. Brie, A. Mohammad-Djafari, and C. Carteret, "Separation of non-negative mixture of non-negative sources using a Bayesian approach and MCMC sampling," à paraître dans *IEEE Transactions on Signal Processing*.
- [10] M. Gaeta and J. Lacoume, "Estimateurs du maximum de vraisemblance étendus à la séparation de sources non-gaussiennes," *Traitement du signal*, vol. 7, no. 5, pp. 419–434, 1990.
- [11] D.-T. Pham, P. Garrat, and C. Jutten, "Separation of a mixture of independent sources through a maximum likelihood approach," in *proceedings of EUSIPCO*, 1992.
- [12] J.-F. Cardoso and B. H. Laheld, "Equivariant adaptive source separation," *IEEE Transactions on Signal Processing*, vol. 44, no. 12, pp. 3017–3030, 1996.
- [13] J.-F. Cardoso, "Infomax and maximum likelihood for source separation," *IEEE Letters on Signal Processing*, vol. 4, pp. 112–114, 1997.
- [14] P. Comon, "Independent component analysis – a new concept?" *Signal Processing*, vol. 36, pp. 287–314, 1994.
- [15] B. Pearlmutter and L. Parra, "A context-sensitive generalization of ICA," in *proceedings of NIPS*, 1996.
- [16] S.-I. Amari, T.-P. Chen, and A. Cichocki, "Stability analysis of adaptive blind source separation," *Neural Networks*, vol. 10, no. 8, pp. 1345–1351, 1997.
- [17] J.-F. Cardoso, "On the stability of source separation algorithms," *Journal of VLSI Signal Processing*, vol. 26, pp. 7–14, 2000.
- [18] M. Abramovitz and I. Stegun, *Handbook of Mathematical Functions*. New York : Dover Publications, 1972.