

Séparation aveugle de sources par minimisation des alpha-divergences

Amor KEZIOU¹, Hassan FENNIRI², M. S. OULD MOHAMED² et Georges DELAUNAY²

¹Laboratoire de Mathématique FRE 3111 CNRS,
Université de Reims et LSTA-Université Paris 6, France.

²Laboratoire CReSTIC-DéCOM, Université de Reims, France
amor.keziou@univ-reims.fr, hassan.fenniri@univ-reims.fr

Résumé – Nous proposons un nouvel algorithme de séparation aveugle de signaux exploitant la minimisation des α -divergences. La méthode préconisée généralise l’approche de l’Information Mutuelle (IM). Nous proposons une classe de critères contenant l’IM. Nous montrons par simulations que l’algorithme utilisant la divergence de Hellinger a de bonnes propriétés de robustesse en présence du bruit en comparaison avec l’IM. Les méthodes proposées s’appliquent dans le cas des mélanges instantanés et convolutifs. Nous illustrons la performance de l’algorithme par simulations.

Abstract – In this paper, we propose a new algorithm for blind source separation based on the minimization of α -divergences. The method generalizes the Mutual Information (MI) approach. The proposed method applies in the case of instantaneous or convolutive mixtures. We show by simulation that the algorithm using Hellinger’s divergence has better properties in terms of efficiency-robustness, for noisy data.

1 Introduction

La méthode de séparation aveugle de sources que nous proposons dans ce papier est basée sur la minimisation des α -divergences entre lois de probabilités. Nous allons considérer le cas des mélanges convolutifs. La méthode proposée généralise celle basée sur le critère de l’IM (c.f. e.g. Pham (2002)). Nous montrons que la méthode utilisant la divergence de Hellinger possède de bonnes propriétés en terme d’efficacité-robustesse. D’autres classes de divergences ont été utilisées dans la littérature. Dans Erdogmus *et al.* (2001), les auteurs présentent des algorithmes de séparation de sources basé sur la minimisation des divergences de Renyi; voir aussi Vrins *et al.* (2007). L’utilisation des α -divergences que nous proposons est justifiée par les raisons suivantes. Elle généralise l’approche de l’IM. Elle s’applique dans les cas de mélange instantané ou convolutif. La classe des α -divergences contient la divergence de Hellinger qui a des bonnes propriétés d’efficacité-robustesse en estimation statistique, comme il a été démontré dans Beran (1977), Lindsay (1994) et Jiménez et Shao (2001). Par ailleurs, l’estimation de ces α -divergences est faite par injection des estimateurs à noyau des densités dans leurs définitions. Le reste de ce papier est organisé comme suit. Dans la section 2, nous rappelons les divergences entre lois de probabilités et leurs propriétés. La section 3 présente le problème de la séparation de sources pour des mélanges convolutifs. Dans la section 4, nous introduisons de nouveaux critères de séparations utilisant les divergences. Dans la section 5, nous comparons les différentes méthodes par simulations en terme d’efficacité et robustesse.

2 Les divergences entre lois de probabilités

Soit φ une fonction convexe quelconque définie sur $[0, +\infty]$ à valeurs dans $[0, +\infty]$ telle que $\varphi(1) = 0$. Pour toutes lois de probabilité Q et P telles que Q est absolument continu par rapport à P , la I_φ -divergence entre Q et P est définie par

$$I_\varphi(Q, P) := \int \varphi\left(\frac{dQ}{dP}\right) dP. \quad (2.1)$$

Si Q n’est pas absolument continue par rapport à P , on pose $I_\varphi(Q, P) = +\infty$. Pour toute loi de probabilité P , l’application $Q \mapsto I_\varphi(Q, P)$ est convexe et est positive. Si $Q = P$, alors $I_\varphi(Q, P) = 0$. De plus, si la fonction $x \mapsto \varphi(x)$ est strictement convexe sur un voisinage de $x = 1$, on a la propriété fondamentale suivante

$$I_\varphi(Q, P) = 0 \text{ si et seulement si } Q = P. \quad (2.2)$$

Toutes ces propriétés sont présentées et démontrées dans Csiszár (1967) et Liese et Vajda (1987). Notons que si Q et P admettent des densités q et p par rapport à une même mesure dominante, la divergence (2.1) s’écrit

$$I_\varphi(Q, P) := \int \varphi\left(\frac{q}{p}\right) dP. \quad (2.3)$$

Largement utilisée en théorie de l’information, la divergence de Kullback-Leibler, notée KL -divergence, est associée à la fonction convexe réelle $\varphi(x) = x \log x - x + 1$; elle est définie comme suit :

$$KL(Q, P) := \int \log\left(\frac{dQ}{dP}\right) dQ. \quad (2.4)$$

La divergence de Kullback-Leibler modifiée, notée KL_m divergence, est associée à la fonction convexe $\varphi(x) = -\log x + x - 1$, i.e.,

$$KL_m(Q, P) := \int -\log\left(\frac{dQ}{dP}\right) dP. \quad (2.5)$$

Elle est appelée aussi Information mutuelle (IM). D'autres divergences, largement utilisées sont les divergences de χ^2 et χ^2 -modifiée (χ_m^2) :

$$\chi^2(Q, P) := \frac{1}{2} \int \left(\frac{dQ}{dP} - 1\right)^2 dP \quad (2.6)$$

et

$$\chi_m^2(Q, P) := \frac{1}{2} \int \frac{\left(\frac{dQ}{dP} - 1\right)^2}{\frac{dQ}{dP}} dP \quad (2.7)$$

qui sont associées aux fonctions convexes $\varphi(x) = \frac{1}{2}(x-1)^2$ et $\varphi(x) = \frac{1}{2}(x-1)^2/x$, respectivement. La distance de Hellinger (H) est également une I_φ -divergence; elle est associée à la fonction convexe $\varphi(x) = 2(\sqrt{x}-1)^2$,

$$H(Q, P) := \int 2 \left(\sqrt{\frac{dQ}{dP}} - 1\right)^2 dP. \quad (2.8)$$

Tous les exemples précédents des I_φ -divergences font partie de la classe des "divergences de puissance" introduite par Cresie et Read (1984) qui est définie par la classe des fonctions convexes

$$x \in \mathbb{R}_+^* \mapsto \varphi_\alpha(x) := \frac{x^\alpha - \alpha x + \alpha - 1}{\alpha(\alpha - 1)} \quad (2.9)$$

pour tout $\alpha \in \mathbb{R} \setminus \{0, 1\}$,

$$\begin{aligned} \varphi_0(x) &= -\log x + x - 1, \\ \varphi_1(x) &= x \log x - x + 1, \end{aligned}$$

et

$$\varphi_\alpha(0) = \lim_{x \rightarrow 0} \varphi_\alpha(x), \quad \varphi_\alpha(+\infty) = \lim_{x \rightarrow +\infty} \varphi_\alpha(x),$$

pour tout $\alpha \in \mathbb{R}$. Le tableau suivant présente, suivant le choix de la fonction φ_α , la divergence associée.

TAB. 1: Exemples de fonctions convexes et leurs divergences associées

la fonction φ_α	φ_1	φ_0	φ_2	φ_{-1}	$\varphi_{\frac{1}{2}}$
la divergence associée	KL	KL_m	χ^2	χ_m^2	H

Ces divergences ont été utilisées en estimation paramétriques et semiparamétriques par Keziou (2003), Broniatowski et Keziou (2006) et Broniatowski et Keziou (2009). Il a été montré que leur utilisation généralise la méthode classique du maximum de vraisemblance, et conduit à des estimateurs ayant des propriétés similaires, voire meilleures dans certains cas; l'estimateur utilisant la divergence de Hellinger améliore celui du maximum de vraisemblance en terme d'efficacité-robustesse pour des données bruitées.

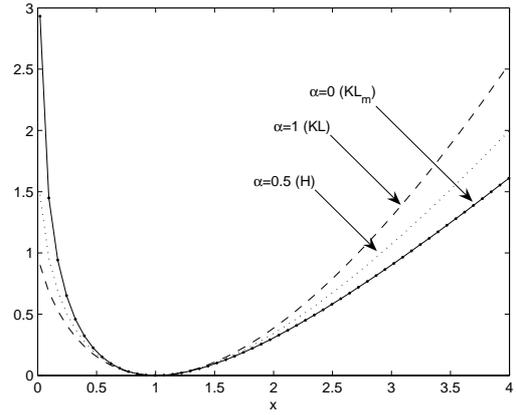


FIG. 1: Les fonctions convexes φ_α des divergences KL , KL_m et H

3 Principe de la séparation de sources dans le cas de mélange convolutif

Le mélange s'écrit sous la forme

$$\mathbf{x}(t) = \mathbf{A} * \mathbf{s}(t) + \mathbf{b}(t), \quad (3.1)$$

où $*$ est le produit de convolution, \mathbf{s} est le vecteur des signaux sources et \mathbf{A} est la matrice des filtres de mélange. \mathbf{b} est le vecteur des signaux bruits, dont les éléments sont supposés centrés, mutuellement indépendants, et indépendants des signaux sources. La présence du bruit additif dans le modèle de mélange complique le problème de la séparation de sources. Pour simplifier le problème, le bruit est supposé négligeable (après prétraitement). Le but est d'estimer les sources \mathbf{s} à partir des observations \mathbf{x} . L'estimateur, dans ce cas, est de la forme

$$\mathbf{y}(t) = \mathbf{B} * \mathbf{x}(t), \quad (3.2)$$

où \mathbf{B} est la matrice des filtres séparante. Le problème consiste donc à chercher un estimateur $\hat{\mathbf{B}}$ de \mathbf{B} , à partir des observations \mathbf{x} , qui conduit à une estimation

$$\hat{\mathbf{y}}(t) = \hat{\mathbf{B}} * \mathbf{x}(t),$$

des sources \mathbf{s} . La transposition dans le domaine discret des modèles (3.1) et (3.2) conduit à

$$\mathbf{x}(n) = [\mathbf{A}(z)]\mathbf{s}(n) + \mathbf{b}(n) = \sum_k \mathbf{A}_k \mathbf{s}(n-k), \quad (3.3)$$

$$\mathbf{y}(n) = [\mathbf{B}(z)]\mathbf{x}(n) = \sum_k \mathbf{B}_k \mathbf{x}(n-k), \quad (3.4)$$

où \mathbf{A}_k et \mathbf{B}_k correspondent respectivement à la transformée en z des matrices des filtres \mathbf{A} et \mathbf{B} . Si \mathbf{A} est une matrice inversible à gauche et les sources sont statistiquement indépendantes, la solution du problème de séparation est possible à une permutation et un filtrage près :

$$\mathbf{y}(n) = [\hat{\mathbf{B}}(z)]\mathbf{x}(n) = [\hat{\mathbf{B}}(z)\mathbf{A}(z)]\mathbf{s}(n), \quad (3.5)$$

où la matrice des filtres $\hat{\mathbf{B}}(z)$ vérifie $[\hat{\mathbf{B}}(z)\mathbf{A}(z)] = [\mathcal{P}\mathcal{H}(z)]$, avec \mathcal{P} une permutation et \mathcal{H} un opérateur de filtrage.

4 Nouveaux critères de séparation via les divergences

Soit $\mathbf{y} := (y_1, \dots, y_d)^T$ un vecteur aléatoire, et notons f la densité jointe du vecteur \mathbf{y} et f_i la densité marginale de y_i , la i ème composante de \mathbf{y} . La I_{φ_α} -divergence entre le produit $f_1 \cdots f_d$ des densités marginales et la densité jointe f s'écrit (voir 2.1)

$$I_{\varphi_\alpha}(f_1 \cdots f_d, f) = \int_{\mathbb{R}^d} \varphi_\alpha \left(\frac{f_1(t_1) \cdots f_d(t_d)}{f(t_1, \dots, t_d)} \right) f(t_1, \dots, t_d) dt_1 \cdots dt_d = \mathbb{E} \left[\varphi_\alpha \left(\frac{f_1(y_1) \cdots f_d(y_d)}{f(y_1, \dots, y_d)} \right) \right]. \quad (4.1)$$

D'après la propriété (2.2), le critère (4.1) est positive, et il s'annule si est seulement si $f_1 \cdots f_d = f$, i.e., ssi les composantes y_1, \dots, y_d sont mutuellement indépendantes. Ce critère peut être estimé par

$$I_{\varphi_\alpha}(\mathbf{y}) := \frac{1}{N} \sum_{n=1}^N \varphi_\alpha \left(\frac{\hat{f}_1(y_1(n)) \cdots \hat{f}_d(y_d(n))}{\hat{f}(y_1(n), \dots, y_d(n))} \right) \quad (4.2)$$

où $\hat{f}_1, \dots, \hat{f}_d, \hat{f}$ sont respectivement les estimateurs à noyau des densités f_1, \dots, f_d, f . Dans le cas des mélanges convolutifs, l'indépendance des signaux $y_1(n), y_2(n), \dots, y_d(n)$ (pour tout n) ne suffit pas pour séparer les sources. En d'autres termes, dans le cas de 2 sources, pour séparer les sources nous avons besoin de rendre indépendants les signaux $y_1(n)$ et $y_2(n-m)$ pour tout n et tout retard m ; (voir Babaie-Zadeh *et al.* (2001); El Rhabi *et al.* (2005)). Par conséquent, nous remplaçons le critère (4.2) par

$$J_\alpha := \sum_m I_{\varphi_\alpha}(\mathbf{y}^m) = \sum_m \frac{1}{N} \sum_{n=1}^N \varphi_\alpha \left(\frac{\hat{f}_1(y_1(n)) \cdots \hat{f}_d(y_d(n-m_d))}{\hat{f}(y_1(n), \dots, y_d(n-m_d))} \right) \quad (4.3)$$

où $\mathbf{y}^m(n) := (y_1(n-m_1), y_2(n-m_2), \dots, y_d(n-m_d))^T$ et $m_1 = 0$.

La séparation sera obtenue par minimisation de celui-ci sur la matrice des filtres \mathbf{B} :

$$\min_{\mathbf{B}} J_\alpha = \min_{\mathbf{B}} \sum_m I_{\varphi_\alpha}(\mathbf{y}^m). \quad (4.4)$$

Nous allons utiliser la méthode de descente du gradient pour calculer ce minimum. Le calcul explicite du gradient est donné dans Ould Mohamed *et al.* (2009). Dans le cas de deux sources - deux observations, après calcul, nous obtenons

$$\frac{\partial J_\alpha}{\partial \mathbf{B}_k(p, q)} = \sum_m \frac{1}{N} \sum_{n=1}^N \varphi'_\alpha \left(\frac{\hat{f}_1(y_1(n)) \hat{f}_1(y_2(n-m))}{\hat{f}(y_1(n), y_2(n-m))} \right) \frac{\partial \left(\frac{\hat{f}_1(y_1(n)) \hat{f}_1(y_2(n-m))}{\hat{f}(y_1(n), y_2(n-m))} \right)}{\partial \mathbf{B}_k(p, q)},$$

où $\varphi'_\alpha(\cdot)$ est la dérivée de la fonction convexe $\varphi_\alpha(\cdot)$ introduit dans (2.9), et $\mathbf{B}_k(p, q)$, $p, q = 1, 2$, sont les éléments de la matrice \mathbf{B}_k pour tout retard k . Ainsi, pour $\alpha = 0$, i.e., dans le cas de l'information mutuelle, nous obtenons

$$\frac{\partial J_0}{\partial \mathbf{B}_k(p, q)} = \sum_m \frac{1}{N} \sum_{n=1}^N \left(1 - \frac{\hat{f}(y_1(n), y_2(n-m))}{\hat{f}_1(y_1(n)) \hat{f}_2(y_2(n-m))} \right) \frac{\partial \left(\frac{\hat{f}_1(y_1(n)) \hat{f}_2(y_2(n-m))}{\hat{f}(y_1(n), y_2(n-m))} \right)}{\partial \mathbf{B}_k(p, q)}.$$

Pour $\alpha = 1$, i.e., le cas de la divergence de Kullback-Leibler,

$$\frac{\partial J_1}{\partial \mathbf{B}_k(p, q)} = \sum_m \frac{1}{N} \sum_{n=1}^N \ln \left(\frac{\hat{f}_1(y_1(n)) \hat{f}_2(y_2(n-m))}{\hat{f}(y_1(n), y_2(n-m))} \right) \frac{\partial \left(\frac{\hat{f}_1(y_1(n)) \hat{f}_2(y_2(n-m))}{\hat{f}(y_1(n), y_2(n-m))} \right)}{\partial \mathbf{B}_k(p, q)}.$$

Enfin, pour $\alpha \neq 0, 1$, nous obtenons $\frac{\partial J_\alpha}{\partial \mathbf{B}_k(p, q)} =$

$$\sum_m \frac{1}{N(\alpha-1)} \sum_{n=1}^N \left(\left(\frac{\hat{f}_1(y_1(n)) \hat{f}_2(y_2(n-m))}{\hat{f}(y_1(n), y_2(n-m))} \right)^{\alpha-1} - 1 \right) \frac{\partial \left(\frac{\hat{f}_1(y_1(n)) \hat{f}_2(y_2(n-m))}{\hat{f}(y_1(n), y_2(n-m))} \right)}{\partial \mathbf{B}_k(p, q)}.$$

5 Résultats de simulations

Les deux signaux sources $\mathbf{s}_1(t)$ et $\mathbf{s}_2(t)$ sont des variables aléatoires indépendantes et de même loi uniforme (sur $[0,1]$) centrées et réduites. La matrice des filtres de mélange considérée est

$$\mathbf{A}(\mathbf{z}) = \begin{bmatrix} 1 + 0.2z^{-1} + 0.1z^{-2} & 0.5 + 0.3z^{-1} + 0.1z^{-2} \\ 0.5 + 0.3z^{-1} + 0.1z^{-2} & 1 + 0.2z^{-1} + 0.1z^{-2} \end{bmatrix}$$

Nous comparons la qualité de séparation des trois critères KL , KL_m et Hellinger, en fonction du rapport signal sur bruit (RSB). La somme dans (4.3) est calculée sur $m \in \{-M, M\}$ avec $M = 2p$ où p est l'ordre du filtre ($p=2$). Nous utilisons le critère de performance suivant

$$SNR_i = 10 \log_{10} \left(\frac{\mathbb{E}[y_i^2]}{\mathbb{E}[y_i^2 | s_i=0]} \right), \quad i = 1, 2. \quad (5.1)$$

Pour estimer les densités et ses dérivées, nous avons utilisé un noyau gaussien standard avec la fenêtre $h = N^{-1/5}$ où N est la taille des signaux. Sur la figure 2, nous présentons les SNR moyens (des deux sources) en fonction du RSB pour les trois critères, KL_m , KL et Hellinger. Le RSB varie entre $-20(dB)$ et $0(dB)$. La taille des signaux est $N = 500$, et les simulations sont répétées 100 fois. Les intervalles de confiance (au niveau 95%) des différentes valeurs des SNRs sont représentés par des segments. Nous constatons que KL_m est la plus robuste pour certaines valeurs du RSB (entre -20 et -12 dB, et entre -8 et 0 dB), par contre entre -12 et -8 dB, KL_m est la moins robuste. On observe la propriété inverse pour la KL . Le critère

basé sur Hellinger présente un bon compromis entre les critères KL et KL_m et ce quelque soit le degré de contamination des données; voir figure 2. Ceci peut s'expliquer de la façon suivante. Lorsque les données sont bruitées, deux cas peuvent se produire.

cas 1: $\frac{\hat{f}_1(y_1(n))\hat{f}_d(y_2(n))}{\hat{f}(y_1(n),y_2(n))} \ll 1$, on l'appelle "outlier".

cas 2: $\frac{\hat{f}_1(y_1(n))\hat{f}_d(y_2(n))}{\hat{f}(y_1(n),y_2(n))} \gg 1$, on l'appelle "inlier".

D'après la figure 1, il est clair que KL_m est la plus robuste par rapport aux outliers, et la moins robuste par rapport aux inliers. A contrario, KL est la moins robuste par rapport aux outliers et la plus robuste par rapport aux inliers. Comme la fonction convexe associée à Hellinger (figure 1) est comprise entre celles associées à KL et KL_m , le critère basé sur Hellinger présente un compromis en terme de robustesse dans le cas des outliers et inliers. Comme dans la pratique, nous ne connaissons pas si le bruit conduit à un outlier ou un inlier, il convient donc d'utiliser le critère basé sur Hellinger.

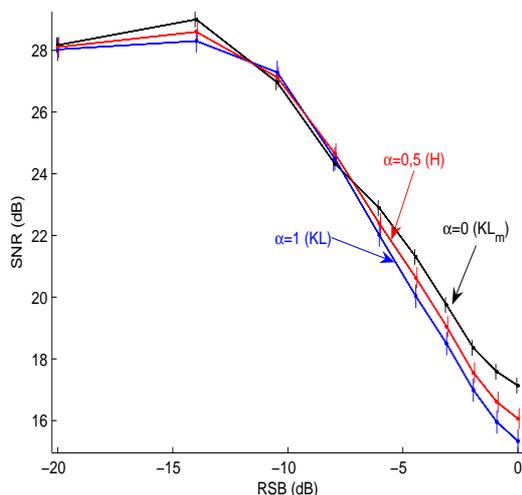


FIG. 2: Qualité de la séparation des trois critères: IM ($\alpha = 0$), Hellinger ($\alpha = 0.5$) et KL ($\alpha = 1$), en fonction du RSB

6 Conclusion

Nous avons présenté dans ce papier un nouvel algorithme de séparation aveugle de signaux basé sur la minimisation des α -divergences. La méthode proposée généralise l'approche de l'information mutuelle. Les méthodes proposées s'appliquent dans le cas des mélanges instantanés ou convolutifs. Nous avons montré par simulation que l'algorithme utilisant la divergence de Hellinger a de bonnes propriétés en terme d'efficacité et robustesse.

Références

- Babaie-Zadeh, M., Jutten, C., et Nayebi, K. (2001). Separating convolutive mixtures by mutual information minimization. *Proc. of IWANN, Granada, Spain*, pages 834–842.
- Beran, R. (1977). Minimum Hellinger distance estimates for parametric models. *Ann. Statist.*, **5**(3), 445–463.
- Broniatowski, M. et Keziou, A. (2006). Minimization of ϕ -divergences on sets of signed measures. *Studia Sci. Math. Hungar.*, **43**(4), 403–442.
- Broniatowski, M. et Keziou, A. (2009). Parametric estimation and tests through divergences and the duality technique. *J. Multivariate Anal.*, **100**(1), 16–36.
- Cressie, N. et Read, T. R. C. (1984). Multinomial goodness-of-fit tests. *J. Roy. Statist. Soc. Ser. B*, **46**(3), 440–464.
- Csiszár, I. (1967). Information-type measures of difference of probability distributions and indirect observations. *Studia Scientiarum Mathematicarum Hungarica*, **2**, 299–318.
- El Rhabi, M., Fenniri, H., Gelle, G., et Delaunay, G. (2005). Blind separation of rotating machine signals using penalized mutual information criterion and minimal distortion principle. *Mechanical Systems and Signal Processing*, **19**, 1282–1292.
- Erdogmus, D., Hild, K., et Principe, J. (2001). Blind source separation using renyi's mutual information. *IEEE Signal Processing Letters*, **8**(6), 174–176.
- Jiménez, R. et Shao, Y. (2001). On robustness and efficiency of minimum divergence estimators. *Test*, **10**(2), 241–248.
- Keziou, A. (2003). Dual representation of ϕ -divergences and applications. *C. R. Math. Acad. Sci. Paris*, **336**(10), 857–862.
- Liese, F. et Vajda, I. (1987). *Convex statistical distances*, volume 95 of *Teubner-Texte zur Mathematik [Teubner Texts in Mathematics]*. BSB B. G. Teubner Verlagsgesellschaft, Leipzig.
- Lindsay, B. G. (1994). Efficiency versus robustness: the case for minimum Hellinger distance and related methods. *Ann. Statist.*, **22**(2), 1081–1114.
- Ould Mohamed, M., Keziou, A., et Fenniri, H. (2009). Séparation aveugle de sources par minimisation des alpha-divergences: application au diagnostic et à la surveillance des machines tournantes. *Second International Conference on Industrial Risk Engineering, Reims, France*.
- Pham, D.-T. (2002). Mutual information approach to blind separation of stationary sources. *IEEE Trans. Information Theory*, **48**(7).
- Vrins, F., Pham, D.-T., et Verleysen, M. (2007). *Is the general form of Renyi's entropy a contrast for source separation?*, volume 4666/2007 of *Lecture Notes in Computer Science*. Springer, Berlin.