

# La $\gamma$ -divergence de Fujisawa-Eguchi - une alternative pour imposer la contrainte de somme. Application au démixage linéaire

Henri LANTÉRI, Céline THEYS

Laboratoire Lagrange, Université de Nice Sophia-Antipolis, CNRS, Observatoire de la Côte d'Azur

Henri.Lantéri@unice.fr, Céline.Theys@unice.fr

**Résumé** – Dans le problème de démixage linéaire, les données sont modélisées comme une somme pondérée de signaux unidimensionnels connus et positifs. Le problème est de déterminer les coefficients de pondération, les abondances, sous contrainte de positivité et de somme à un par minimisation d'une fonction d'écart entre mesures et modèle. La fonction d'écart que nous proposons d'utiliser est la  $\gamma$ -divergence de Fujisawa-Eguchi, typique d'une classe de divergences présentant la propriété d'invariance par changement d'échelle ; la contrainte de somme à un peut ainsi être facilement prise en compte. Nous développons deux formes d'algorithmes exploitant spécifiquement cette propriété et un algorithme général s'appliquant sur toute divergence, [9] ; dans tous les cas, la contrainte de positivité est prise en compte en s'appuyant sur les conditions de Karush-Kuhn-Tucker, selon la méthode proposée dans de précédents articles ([8], [9], ...).

Les algorithmes sont testés sur un jeu de données simulées et leurs performances sont comparées.

**Abstract** – This paper addresses the problem of linear unmixing for hyperspectral imagery. The data signal is modeled as a weighted sum of known positive signals. The problem is to recover the weights (abundances) by minimizing a discrepancy function between data and linear model subject to positivity and sum to one constraints of the unknowns.

We propose to use as a discrepancy function, the  $\gamma$ -divergence of Fujisawa and Eguchi which is typical of the class of scale invariant divergences. In so doing, the sum to one constraint is easily taken into account.

The scale invariance property of this specific class of divergences allows to obtain two particular forms of algorithms dealing with the sum to one constraint; moreover we propose a more general algorithm that can be used for any divergence, [9]. In all the algorithms, the positivity constraint is taken into account using a method founded on the Karush-Kuhn-Tucker conditions as previously proposed ([8], [9], ...).

The algorithms are checked on simulated data and their performances are compared.

## 1 Introduction

Le problème de démixage linéaire est généralement traité comme un problème de minimisation sous contraintes, d'une fonction d'écart (divergence) [2, 15] entre un vecteur  $y$  correspondant aux données et un modèle linéaire  $Hx$ , dans lequel  $H$  est une matrice connue à termes positifs et  $x$  est un vecteur inconnu à composantes positives et de somme égale à un (les abondances) ; ce sont les contraintes du problème.

Diverses fonctions d'écart ont été utilisées dans la littérature pour quantifier l'écart entre 2 champs de données dans les problèmes de démixage linéaire ou de NMF (Non negative Matrix Factorization) [3, 4, 5, 6]. Dans un article précédent traitant de la NMF [9], nous avons pris en compte la contrainte de somme égale à un en utilisant une méthode de changement de variables. Nous proposons une alternative qui consiste à utiliser la  $\gamma$ -divergence de Fujisawa et Eguchi [7] pour exprimer l'écart entre mesures et modèle. Cette divergence qui a montré des propriétés de robustesse et de faible biais pour l'estimation de paramètres, présente la particularité d'être invariante par changement d'échelle. On peut noter que cette même divergence apparaît également mais sous une autre forme dans [13]. En conséquence, au cours de la minimisation itérative de

cette divergence, on peut, à chaque itération, renormaliser les composantes du vecteur inconnu, sans modifier la valeur de la fonction objectif. Par ailleurs, cette divergence permet d'exhiber un algorithme de descente fondé sur les conditions de Karush Kuhn Tucker qui possède spontanément la propriété de contrainte de somme. L'utilisation de la divergence de Fujisawa et Eguchi est comparée à la méthode Fully Constrained Least Squares (FCLS), [3], du point de vue des performances en termes de temps de calcul et de moyenne et variance des paramètres estimés en fonction du rapport signal sur bruit.

## 2 Divergence de Fujisawa et Eguchi

La divergence proposée par Fujisawa et Eguchi, si  $i$  est l'indice du pixel, s'écrit :

$$F(p, q) = \frac{1}{\gamma(1 + \gamma)} \left\{ \log \sum_i p_i^{1+\gamma} + \gamma \log \sum_i q_i^{1+\gamma} - (1 + \gamma) \log \sum_i p_i q_i^\gamma \right\},$$

$\gamma > 0. \quad (1)$

Elle présente la particularité d'être invariante par changement d'échelle (divergence invariante par changement d'échelle) :

$$F(p, q) = F(p, aq), \quad a > 0. \quad (2)$$

Dans le cas d'une méthode de minimisation itérative, on peut donc renormaliser la solution  $\mathbf{x}$  à chaque itération sans modifier la valeur de la fonction objectif. Dans notre problème,  $p_i = y_i$ ,  $q_i = (H\mathbf{x})_i$ ,  $aq = H \times a\mathbf{x}$ . Nous considérons le problème à résoudre :

$$\min_{\mathbf{x}} F(\mathbf{y}, H\mathbf{x}) \quad \text{s.c.} \quad x_i \geq 0, \quad \sum_i x_i = 1. \quad (3)$$

Alors que la contrainte de non négativité est imposée en respectant les conditions de Karush Kuhn Tucker (KKT), nous proposons plusieurs alternatives pour contraindre la somme des paramètres constante, égale à un dans notre application.

## 2.1 Contrainte de non négativité - Conditions de KKT

Dans un premier temps, on considère uniquement le problème (3) sans contrainte d'égalité :

$$\min_{\mathbf{x}} F(\mathbf{y}, H\mathbf{x}) \quad \text{s.c.} \quad x_i \geq 0. \quad (4)$$

La fonction de Lagrange associée à (4) s'écrit :

$$L(\mathbf{y}, H\mathbf{x}; \boldsymbol{\lambda}) = F(\mathbf{y}, H\mathbf{x}) - \langle \boldsymbol{\lambda}, \mathbf{x} \rangle. \quad (5)$$

Où  $\boldsymbol{\lambda}$  est le vecteur des multiplicateurs positifs de Lagrange et  $\langle \cdot, \cdot \rangle$  est le produit scalaire. Les conditions de KKT doivent être satisfaites à l'optimum  $\boldsymbol{\lambda}^*$  et  $\mathbf{x}^*$ . La minimisation de (5) par rapport à  $\mathbf{x}$  conduit aux conditions de KKT, qui sont, pour le pixel  $i$  :

$$[\nabla_{\mathbf{x}} L(\mathbf{y}, H\mathbf{x}^*; \boldsymbol{\lambda}^*)]_i = 0, \quad (6)$$

$$\lambda_i^* \geq 0, \quad (7)$$

$$x_i^* \geq 0, \quad (8)$$

$$\langle \boldsymbol{\lambda}^*, \mathbf{x}^* \rangle = 0 \Leftrightarrow \lambda_i^* x_i^* = 0. \quad (9)$$

La condition (6) devient :

$$\lambda_i^* = [\nabla_{\mathbf{x}} F(\mathbf{y}, H\mathbf{x}^*)]_i \quad (10)$$

qui, combinée avec l'équation (9), donne la condition finale de KKT pour la non négativité :

$$x_i^* [\nabla_{\mathbf{x}} F(\mathbf{y}, H\mathbf{x}^*)]_i = 0.$$

que l'on peut écrire également sous la forme :

$$x_i^* [-\nabla_{\mathbf{x}} F(\mathbf{y}, H\mathbf{x}^*)]_i = 0. \quad (11)$$

Cette condition, qui doit être satisfaite à l'optimum, peut être utilisée comme une direction de descente, qui, plus généralement, peut s'écrire :

$$d_i = m_i x_i^* [-\nabla_{\mathbf{x}} F(\mathbf{y}, H\mathbf{x}^*)]_i, \quad (12)$$

où  $m_i > 0$ . Cette direction de descente permet de construire un algorithme itératif de type descente de gradient qui s'écrit pour le pixel  $i$  :

$$x_i^{k+1} = x_i^k + \alpha_i^k d_i. \quad (13)$$

Cet algorithme satisfait la condition de KKT et par conséquent garantit la non négativité de  $x_i$  uniquement à convergence. Afin de respecter les conditions de KKT et par conséquent la non négativité de la solution à chaque itération, on calcule d'abord la valeur du pas maximal assurant la non négativité de l'ensemble des composantes, soit  $\alpha_{\max}^k$ , puis le pas de descente assurant la convergence est calculé sur  $[0, \alpha_{\max}^k]$ , par une méthode de minimisation unidimensionnelle, par exemple, la méthode d'Armijo, [1].

Dans la suite, nous proposons trois algorithmes pour contraindre la somme à un et résoudre le problème complet, (3).

### 2.1.1 Divergence de Fujisawa et Eguchi

L'algorithme itératif, (13), avec  $m_i = 1$  après explicitation du gradient de la divergence de Fujisawa, (1), s'écrit :

$$x_i^{k+1} = x_i^k + \alpha^k x_i^k \left[ \frac{H^T p(q^k)^{\gamma-1}}{\sum_j p_j (q_j^k)^\gamma} - \frac{H^T (q^k)^\gamma}{\sum_j (q_j^k)^{1+\gamma}} \right]_i. \quad (14)$$

Cette forme a la particularité de conserver la somme, c.a.d  $\sum_i x_i^{k+1} = \sum_i x_i^k$ . En effet, cette condition est satisfaite si l'on a :

$$\sum_i x_i^k \left[ \frac{H^T p(q^k)^{\gamma-1}}{\sum_j p_j (q_j^k)^\gamma} - \frac{H^T (q^k)^\gamma}{\sum_j (q_j^k)^{1+\gamma}} \right]_i = 0. \quad (15)$$

Le second terme de l'expression ci-dessus est égal à :

$$\begin{aligned} \frac{1}{\sum_j (q_j^k)^{1+\gamma}} \mathbf{x}^T H^T (q^k)^\gamma &= \frac{1}{\sum_j (q_j^k)^{1+\gamma}} (q^k)^{\gamma T} H \mathbf{x} \\ &= \frac{\sum_j (q_j^k)^{1+\gamma}}{\sum_j (q_j^k)^{1+\gamma}}. \end{aligned}$$

Le second terme est donc égal à un, on trouve le même résultat pour le premier terme, (15) est donc satisfaite. La contrainte de somme égale à un est donc assurée en choisissant un estimé initial contraint à un.

### 2.1.2 Divergence de Fujisawa et Eguchi - SGM

Si l'on choisit une direction de descente  $d_i$ , (12), avec :

$$m_i = \left[ \frac{\sum_j (q_j^k)^{1+\gamma}}{H^T (q^k)^\gamma} \right]_i$$

l'algorithme (13) résultant est l'algorithme SGM (Split Gradient Method), [8] qui s'écrit alors :

$$x_i^{k+1} = x_i^k + \alpha^k x_i^k \left\{ \left[ \frac{\sum_j (q_j^k)^{1+\gamma}}{\sum_j p_j (q_j^k)^\gamma} \frac{H^T p(q^k)^{\gamma-1}}{H^T (q^k)^\gamma} \right]_i - 1 \right\}. \quad (16)$$

Le pas de descente est calculé de la même façon que pour (14); il permet d'assurer la convergence de l'algorithme. Compte tenu des propriétés de la divergence utilisée, la contrainte de flux peut être introduite par une renormalisation de la solution à chaque itération sans modification de la fonction objectif  $F$ .

Cette forme a la particularité de conduire à un algorithme «pu-  
rement multiplicatif» pour  $\alpha^k = 1, \forall k$  :

$$x_i^{k+1} = x_i^k \left[ \frac{\sum_j (q_j^k)^{1+\gamma} H^T p(q^k)^{\gamma-1}}{\sum_j p_j (q_j^k)^\gamma H^T (q^k)^\gamma} \right]_i$$

Le facteur multiplicatif correctif étant positif, la positivité de  $x_i$  est assurée à chaque itération à partir d'un estimé initial positif. Par contre, la convergence n'est pas démontrée formellement puisqu'il n'y a pas de contrôle de descente mais elle a été vérifiée numériquement dans tous les cas considérés.

### 2.1.3 Divergence quelconque, SGM Normalisé

Dans cette méthode, indépendante de la divergence utilisée, on adjoint à l'algorithme SGM précédent, la contrainte de somme en procédant au changement de variable  $x_i = u_i / \sum_j u_j$  comme il a été proposé dans [9].

On peut noter que si  $\sum_j u_j$  est constant au cours des itérations, alors la convexité de la divergence par rapport à  $x$  entraîne sa convexité par rapport à  $u$ . Conceptuellement, on minimise par rapport à  $u$  sous contrainte de non négativité, avec un algorithme qui maintient constante  $\sum_j u_j$  au cours des itérations, ce qui permet de revenir aux variables initiales,  $x$ . Le gradient par rapport à  $u$  est calculé sous la forme :

$$\frac{\partial F}{\partial u_r} = \sum_l \frac{\partial F}{\partial x_l} \frac{\partial x_l}{\partial u_r} \Rightarrow \frac{\partial F}{\partial u_r} = \frac{1}{\sum_j u_j} \left( \frac{\partial F}{\partial x_r} - \sum_l x_l \left( \frac{\partial F}{\partial x_l} \right) \right). \quad (17)$$

L'algorithme final, SGM normalisé, s'écrit sous forme relaxée :

$$x_i^{k+1} = x_i^k + \alpha^k x_i^k \left\{ \frac{\left[ -\frac{\partial F}{\partial x_i^k} - \min \left( -\frac{\partial F}{\partial x^k} \right) + \epsilon \right]_i}{\left[ \sum_l x_l^k \left( -\frac{\partial F}{\partial x_l^k} \right) - \min \left( -\frac{\partial F}{\partial x^k} \right) + \epsilon \right]_i} - 1 \right\}. \quad (18)$$

Le pas de descente est calculé de la même façon que pour (14). La contrainte de somme égale à 1 sur  $x$  est ainsi respectée si l'estimé initial est contraint. Le paramètre  $\epsilon$  est introduit pour éviter les problèmes numériques.

## 3 Résultats

L'imagerie hyperspectrale fait l'objet d'une attention toute particulière ces dernières années, voir par exemple [10], [11] et les références associées. Cela consiste en l'acquisition de données à haute résolution dans des centaines de bandes spectrales, géo référencées. Chaque pixel d'une image hyperspectrale caractérise une signature spectrale. Il est classique de considérer que chaque pixel étendu en longueur d'onde sur un vecteur peut être modélisé comme un mélange linéaire de différentes composantes spectrales élémentaires, appelés "endmembers".

Dans notre modèle,  $y$  peut être interprétée comme une signature spectrale obtenue par mélange linéaire des spectres élémentaires contenus dans  $H$  avec les poids à estimer  $x$ .

L'expérience présentée dans ce papier correspond à un mélange linéaire de 3 "endmembers", la longueur de chaque spectre est de 826. Les spectres élémentaires utilisés dans cet exemple proviennent de la bibliothèque ENVI, [12]. Le bruit utilisé est additif Gaussien de moyenne nulle.

Les algorithmes utilisés et dont quelques résultats sont montrés sont FUJI, eq. (14), NFUJI, eq. (18), FUGISGM, eq. (16), FUJIMULT, eq. (16) avec  $\alpha = 1$  et FCLS, un algorithme de référence dans le domaine. Il est à noter que les performances des algorithmes en terme de moyenne et de variance des estimés sont sensiblement identiques. L'algorithme FUJI et l'algorithme FUJIMULT sont respectivement le plus rapide et le plus lent en termes de nombre d'itérations et de temps de calcul. Il est à noter que tous les algorithmes proposés imposent la contrainte de non négativité et de somme à chaque itération alors que FCLS traite la somme comme un terme de régularisation, avec donc un coefficient à régler, la contrainte est donc satisfaite de façon approchée uniquement à convergence de l'algorithme.

## 4 Conclusion

Dans ce travail, nous avons proposé d'introduire dans le problème de démélange linéaire, la contrainte de somme à un en nous appuyant sur la propriété d'invariance par changement d'échelle de la  $\gamma$  divergence de Fujisawa et Eguchi. La contrainte de non négativité est prise en compte par explicitation des conditions de Karush Kuhn Tucker. On peut ainsi exhiber trois algorithmes de reconstruction respectant ces contraintes, l'un procède par changement de variables et s'applique quelle que soit la divergence à minimiser, deux autres font appel spécifiquement aux propriétés de la  $\gamma$ -divergence.

Ces algorithmes ont été testés sur des données perturbées par un bruit Gaussien additif et comparés à l'algorithme FCLS. Les vitesses de convergence sont analysées ainsi que les performances vis à vis du niveau de bruit.

Alors que les propriétés de cette divergence ne peuvent pas être totalement exploitées pour le problème de Non Negative Matrix Factorization (NMF), une application au problème de déconvolution aveugle peut en revanche être considéré.

Une forme plus générale de divergence présentant la propriété d'invariance par changement d'échelle est donnée dans [14] et fera l'objet d'un travail ultérieur.

## Références

- [1] L. Armijo. *Minimization of functions having continuous derivatives*. Pacific Journal of Mathematics 16 (1996) 1–3.

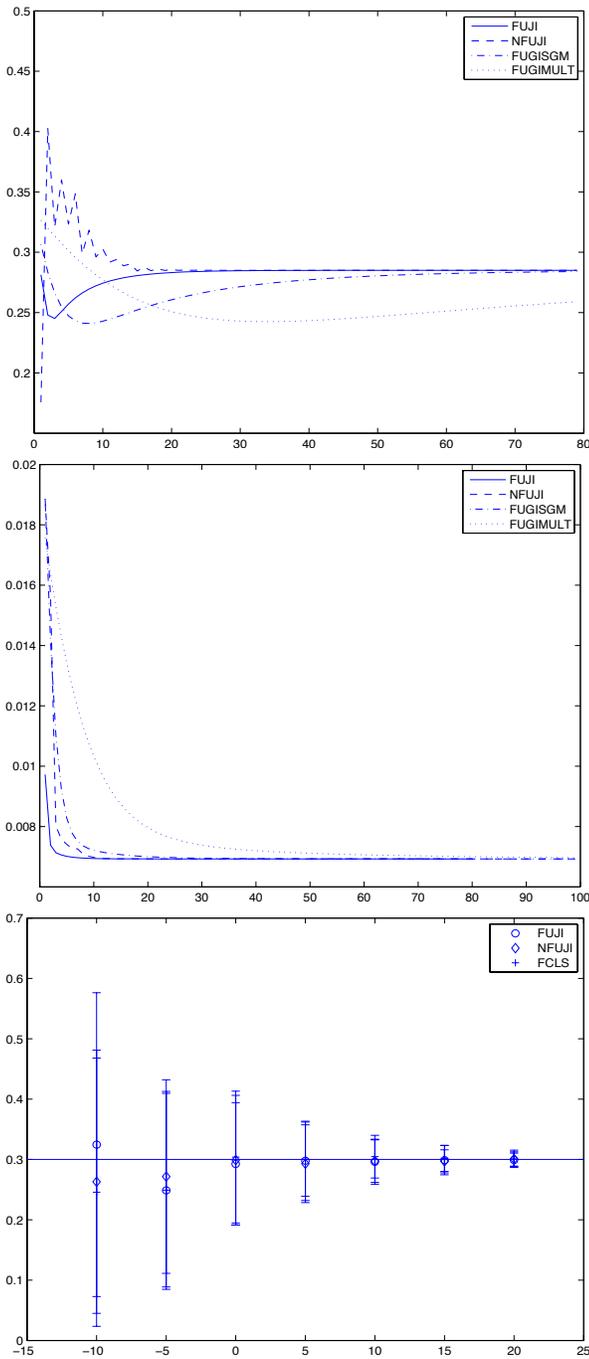


FIGURE 1 – De gauche à droite. Évolution du paramètre  $x_1$  en fonction des itérations, vraie valeur 0.3 pour un  $RSB = 10dB$ . Évolution de la divergence de Fujisawa en fonction des itérations. Moyenne et écart type de  $x_1$  estimé à convergence sur 100 réalisations de bruit en fonction du rapport signal sur bruit.

- [2] M. Basseville. *Information : entropies, divergences et moyennes*. Rapport technique, IRISA 1996.
- [3] D.C. Heinz, C.I. Chang. *Fully constrained least squares linear spectral mixture for material quantification in hyperspectral imagery*. IEEE Trans. Geosci. and Remote Sensing 39 (2001) 529-545.
- [4] D.D. Lee, H.S. Seung. *Algorithms for non negative matrix factorization*. In NIPS 2000.
- [5] C. Févotte, N. Bertin, J.L. Durrieu. *Non negative matrix factorization with the Itakura-Saito divergence, with application to music analysis*. Neural computations (2000).
- [6] A. Cichocki, R. Zdunek, S.I. Amari. *Csiszár's divergences for Non negative Matrix Factorization : Family of new algorithms*. V. 3889/2006 of Lecture Notes in Computer Science. Springer Berlin/Heidelberg, 2006.
- [7] H. Fujisawa, S. Eguchi. *Robust parameter estimation with a small bias against heavy contamination*. Multivariate analysis V.99, 9 (2008) 2053-2081.
- [8] H. Lantéri, M. Roche, C. Aime. *Penalized maximum likelihood image restoration with positivity constraint, multiplicative algorithms*. Inverse Problems, 18 (2002) 1397-1419.
- [9] H. Lantéri, C. Theys, C. Richard, C. Févotte. *Split Gradient Method for non negative matrix factorization*. EU-SIPCO (2010).
- [10] C. I. Chang. *Hyperspectral Imaging : Techniques for Spectral Detection and Classification*. Plenum Publishing Co., New York, 2003.
- [11] D. A. Landgrebe. *Signal Theory Methods in Multispectral Remote Sensing*. Wiley, New York, 2003.
- [12] RSI (Research Systems Inc.). *ENVI User's guide Version 4.0*. Boulder, CO 80301 USA, September 2003.
- [13] E. Lutwak, D. Yang and G. Zhang. *Cramèr-Rao and moment entropy inequalities for Renyi entropy and generalized Fisher information*. Information theory, IEEE Trans. on, vol. 51(2) (2005) 473-478.
- [14] J. F. Bercher. *Escort entropies and divergences and related canonical distributions*. Physics Letters A 375 (33) 2969-2973.
- [15] M. Basseville *Divergence measures for statistical data processing - An annotated bibliography*. Signal Processing, September 2012.