

# Divergences invariantes par changement d'échelle pour la reconstruction de signaux ou d'images

## Application à l'interférométrie optique

Henri LANTÉRI<sup>1</sup>, Céline THEYS<sup>1</sup>, Claude AIME<sup>1</sup>

<sup>1</sup>Laboratoire Lagrange

UMR CNRS 7293

Université de Nice Sophia-Antipolis

Observatoire de la Côte d'Azur

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

Claude.Aime@unice.fr

**Résumé** – Le problème considéré est la reconstruction d'un signal ou d'une image sous contraintes de non-négativité et de somme constante. La contrainte de somme est imposée par l'utilisation de divergences invariantes par changement d'échelle, ce qui conduit à des algorithmes itératifs de forme très simple. Deux familles de divergences entre deux champs de données  $p$  et  $q$  sont considérées, les  $\alpha$ -divergences et les  $\beta$ -divergences. Une procédure permet de rendre ces divergences invariantes. Un algorithme de point intérieur est alors développé et utilisé dans le contexte des problèmes mal posés. Des illustrations numériques sont proposées pour la déconvolution d'images interférométriques.

**Abstract** – The subject of this paper is the reconstruction of a signal or an image under constraints of non negativity and of constant sum. The sum constraint is imposed by the use of scale invariant divergences, which allows the development of simple iterative algorithms. Two families of divergences between two data fields  $p$  and  $q$  are considered, the  $\alpha$ -divergence and the  $\beta$ -divergence. A procedure is applied to make them scale-invariant w.r.t.  $p$  and  $q$ . The resulting method is an interior point type algorithm useful in the context of ill-posed problems. Numerical illustrations are given for the deconvolution of interferometric images.

## 1 Introduction

Le problème considéré dans ce travail est rencontré dans de nombreuses applications physiques, parmi lesquelles on peut citer la déconvolution, le démélange linéaire pour l'imagerie hyperspectrale ou la factorisation par matrices non négatives (NMF). Il s'agit de minimiser une divergence par rapport à  $x$  entre des données mesurées  $p$  et un modèle  $q = Hx$  avec  $h_{ij} \geq 0$ , en respectant les contraintes  $x_i \geq 0 \forall i$  et  $\sum_i x_i = C$ ,  $C > 0$ , de plus  $\sum_i h_{ij} = 1$  et on fixera  $C = \sum_i p_i$ .

Ce problème de minimisation sous contraintes a été très étudié ces dernières années et des solutions ont été proposées principalement dans le cas d'une fonction de coût quadratique. Nous proposons ici une méthode applicable à toute divergence construite à partir d'une fonction convexe. La contrainte de non-négativité est imposée en appliquant les conditions de Karush Kuhn Tucker (KKT), [1] et nous avons proposé dans des travaux précédents d'imposer la contrainte de somme par changement de variables avec application à la NMF, [2] et au démélange linéaire pour l'imagerie hyperspectrale, [3, 4]. Nous proposons ici une procédure systématique pour obtenir une divergence invariante par changement d'échelle afin que la contrainte de somme soit vérifiée très simplement en raison des propriétés des divergences obtenues. L'algorithme proposé, de

type gradient, vérifie les contraintes à chaque itération, ce qui en fait un algorithme de point intérieur.

La méthode est appliquée à un problème de reconstruction d'un objet astrophysique. Il s'agit de la déconvolution d'une éventuelle exoplanète de type Saturne à partir d'observations interférométriques dans l'espace.

## 2 Divergences invariantes par changement d'échelle

On considère les divergences construites sur la base de la fonction convexe :

$$f(x) = \frac{1}{\gamma(\gamma-1)} [x^\gamma - \gamma x - (1-\gamma)]. \quad (1)$$

Pour cette fonction, on a  $f(1) = 0$ ,  $f'(1) = 0$  et  $f''(1) = 1$ . Une divergence de Csizsár, [5], entre deux champs de données  $p$  et  $q$ , construite sur une fonction convexe  $f$  est définie par :

$$A(p||q) = \sum_i q_i f\left(\frac{p_i}{q_i}\right) \quad (2)$$

Dans le cas particulier de la fonction 1, on a :

$$A_\gamma(p||q) = \frac{1}{\gamma(\gamma-1)} \left[ \sum_i \left( p_i^\gamma q_i^{1-\gamma} - \gamma p_i + (\gamma-1) q_i \right) \right]. \quad (3)$$

La forme obtenue est appelée la  $\alpha$ -divergence, étudiée en détails par Amari et al., [6, 7]. Parallèlement, la divergence de Bregman, [8], construite sur une fonction convexe  $f$ , est définie comme :

$$B(p||q) = \sum_i \left( f(p_i) - f(q_i) - (p_i - q_i) f'(q_i) \right), \quad (4)$$

et cette définition appliquée à la fonction 1 conduit à :

$$B_\gamma(p||q) = \frac{1}{\gamma(\gamma-1)} \left[ \sum_i \left( p_i^\gamma + (\gamma-1) q_i^\gamma - \gamma p_i q_i^{\gamma-1} \right) \right]. \quad (5)$$

C'est la  $\beta$ -divergence proposée par Basu et al., [9] et par Eguchi et Kano, [10].

La première étape de notre analyse consiste à appliquer sur les divergences 3 et 5 une transformation qui permet d'obtenir de nouvelles divergences invariantes par rapport à un facteur multiplicatif positif sur  $q$ , c'est à dire invariantes par changement d'échelle. L'objectif est d'obtenir des divergences dont les valeurs ne dépendront que de la forme de  $p$  et de  $q$  et non de leurs amplitudes relatives.

La procédure permettant d'atteindre ce résultat a été suggérée par Eguchi et Kato, [11] et peut être résumée comme suit, pour une divergence  $D(p||q)$  :

1. Considérant le facteur  $T(p, q) > 0$ , exprimer  $D(p||Tq)$ .
2. Résoudre, pour  $T > 0$  l'équation  $\frac{\partial D(p||Tq)}{\partial T} = 0$ .
3. Remplacer  $T$  par l'expression obtenue dans  $D(p||Tq)$ .

Les facteurs  $TA_\gamma(p, q)$  et  $TB_\gamma(p, q)$  obtenus pour les  $\alpha$ -divergences et les  $\beta$ -divergences sont :

$$TA_\gamma(p, q) = \left[ \frac{\sum_j p_j^\gamma q_j^{1-\gamma}}{\sum_j q_j} \right]^{\frac{1}{\gamma}}, \quad TB_\gamma(p, q) = \frac{\sum_j p_j q_j^{\gamma-1}}{\sum_j q_j^\gamma}. \quad (6)$$

On peut remarquer que lorsque  $q \rightarrow p$ ,  $TA_\gamma(p, q)$  et  $TB_\gamma(p, q) \rightarrow 1$ . Les divergences invariantes ainsi obtenues à partir de 3 et 5 sont données par :

$$AI_\gamma(p||q) = \frac{1}{\gamma-1} \left[ \left( \sum_i q_i \right)^{\frac{\gamma-1}{\gamma}} \left( \sum_i p_i^\gamma q_i^{1-\gamma} \right)^{\frac{1}{\gamma}} - \sum_i p_i \right] \quad (7)$$

et

$$BI_\gamma(p||q) = \frac{1}{\gamma(\gamma-1)} \left[ \sum_i p_i^\gamma - \left( \sum_i q_i \right)^{1-\gamma} \left( \sum_i p_i q_i^{\gamma-1} \right)^\gamma \right]. \quad (8)$$

Ces nouvelles divergences sont insensibles à un changement d'échelle sur  $q$  et apparaissent comme la différence de deux termes. Nous pouvons appliquer sur chacun de ces termes une fonction croissante sans modifier le signe de la divergence. On effectue cette opération en utilisant la fonction « logarithme déformé ou généralisé », définie pour  $u > 0$  par :

$$L_d(u) = \frac{u^d - 1}{d} \quad (9)$$

dont les limites sont respectivement  $(u-1)$  si  $d \rightarrow 1$  et  $\ln(u)$  si  $d \rightarrow 0$ . Dans ce dernier cas, nous obtenons à partir de (7) et (8) respectivement :

$$LAI_\gamma(p||q) = \frac{1}{\gamma} \ln \sum_i q_i - \frac{1}{(\gamma-1)} \ln \sum_i p_i + \frac{1}{\gamma(\gamma-1)} \ln \sum_i p_i^\gamma q_i^{1-\gamma} \quad (10)$$

qui a été obtenue par Cichocki et al., [12, 13], et

$$LBI_\gamma(p||q) = \frac{1}{\gamma(\gamma-1)} \ln \sum_i p_i^\gamma + \frac{1}{\gamma} \ln \sum_i q_i^\gamma - \frac{1}{\gamma-1} \ln \sum_i p_i q_i^{\gamma-1}. \quad (11)$$

Cette dernière forme est la divergence de Fujisawa, [14] utilisée précédemment dans [4]. On peut remarquer que les formes logarithmiques des divergences sont également invariantes à un changement d'échelle sur  $p$ .

### 3 Algorithmes de minimisation

Dans cette partie, nous développons des algorithmes itératifs basés sur les divergences invariantes  $LAI_\gamma$  et  $LBI_\gamma$ . On note  $\mathbf{p}$  le vecteur des données bruitées et  $\mathbf{q}$  le vecteur décrivant le modèle linéaire physique :  $\mathbf{q} = H\mathbf{x}$ . Le problème considéré est alors :

$$\min_{\mathbf{x}} D(\mathbf{p}||H\mathbf{x}) \quad s.c \quad x_i \geq 0, \quad \sum_i x_i = C. \quad (12)$$

où  $D$  est une divergence. La contrainte de non-négativité est prise en compte en s'appuyant sur les conditions de KKT. Comme il a été développé dans [1, 4], ceci se traduit par la condition suivante qui doit être satisfaite à l'optimum  $\mathbf{x}^*$  :

$$x_i^* [\nabla_{\mathbf{x}} D(\mathbf{p}||H\mathbf{x}^*)]_i = 0. \quad (13)$$

Cette expression permet de proposer un algorithme itératif de minimisation dont la direction de descente est reliée au gradient, de la forme :

$$x_i^{k+1} = x_i^k + \alpha_i^k x_i^k [-\nabla_{\mathbf{x}} D(\mathbf{p}||H\mathbf{x}^k)]_i. \quad (14)$$

A chaque itération, le pas maximal  $\alpha_{\max}^k$  assurant la non-négativité de l'estimé suivant doit d'abord être déterminé, comme décrit dans [1], puis le pas  $\alpha^k$  assurant la convergence est calculé sur l'intervalle  $[0, \alpha_{\max}^k]$  en utilisant une méthode de recherche de pas telle que la règle d'Armijo, [15], par exemple.

En ce qui concerne la contrainte de somme, on peut opérer de façon générale, par un changement de variables, comme cela a été proposé dans [4]. Lorsque la fonctionnelle à minimiser est invariante par changement d'échelle, deux solutions simples sont possibles, soit on normalise le résultat à chaque itération et ceci ne modifie pas la valeur de la fonctionnelle, soit on utilise les propriétés spécifiques du gradient de ces divergences. En effet, l'algorithme itératif, (14), développé pour la divergence  $LAI_\gamma$  conduit à :

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

et pour la divergence  $LBI_\gamma$  :

$$x_i^{k+1} = x_i^k + \alpha^k x_i^k \left[ H^T \frac{\mathbf{p} \circ (\mathbf{q}^k)^{\gamma-2}}{\sum_j p_j (q_j^k)^{\gamma-1}} - H^T \frac{(\mathbf{q}^k)^{\gamma-1}}{\sum_j (q_j^k)^\gamma} \right]_i \quad (16)$$

Dans ces expressions, le symbole "o" est utilisé pour le produit de Hadamard (c.a.d. point à point).

On peut vérifier facilement, que pour les algorithmes 15 et 16, nous avons  $\sum_i x_i^{k+1} = \sum_i x_i^k$ , de sorte que la contrainte de somme sera satisfaite pour tous les estimés successifs si elle est satisfaite pour l'estimé initial. Cette propriété, liée à la forme spécifique du gradient et à la contrainte de non-négativité, est la principale particularité de ces divergences invariantes utilisée dans notre application, ce qui nous conduit naturellement à une procédure de type point intérieur, ce qui signifie que tous les estimés successifs respectent les contraintes.

Les algorithmes correspondant aux divergences  $AI_\gamma$  et  $BI_\gamma$ , 7 et 8, sont très similaires à ceux des équations 15 et 16. Ils diffèrent par un coefficient dépendant de l'itération  $k$ , apparaissant comme un terme multiplicatif correctif sur  $\alpha^k$ . On peut noter que ce facteur est pris implicitement en compte dans le calcul du pas de descente.

## 4 Interférométrie optique

Le problème de reconstruction d'image que nous considérons dans cette section est celui des futurs grands interféromètres optiques spatiaux, tels que ceux projetés dans le projet *Luciola* de Labeyrie et al. [17]. L'ultime objectif de ces interféromètres sera l'imagerie d'une exoterre, avec probablement comme étape intermédiaire l'imagerie des surfaces d'étoiles géantes comme Bételgeuse. Le coût de la mise en place d'un tel système dépassera de beaucoup les réalisations actuelles les plus ambitieuses. La particularité des images obtenues avec ces pupilles diluées (ou parcimonieuses) est que les fréquences spatiales ne sont transmises que dans des régions non contiguës du plan des fréquences, ce qui correspond à des réponses impulsionnelles (ou PSF) tout à fait particulières.

Dans l'étude présentée ici, nous nous limitons à l'obtention d'images instantanées pour une configuration non-redondante de 25 télescopes, optimisée sur une grille hexagonale par Kopilovich et Sodin [16], représentée dans la figure 1. Elle s'avère

particulièrement efficace en couverture du plan de Fourier. La réponse impulsionnelle (PSF) correspondante idéale est donnée dans la figure 2, en haut à gauche. Elle présente des répliques d'une tache d'Airy sur une grille hexagonale, approchant la résolution du meta-télescope englobant les pupilles diluées. De plus, une structure de type bruit de speckles, due à la forte dilution des ouvertures, s'ajoute à ces répliques.

L'objet choisi pour l'exoplanète  $\mathbf{x}$  est une image basse résolution de Saturne, donnée dans la figure 3. En raison de la structure de la PSF, l'image  $H\mathbf{x}$  présente des répliques de  $\mathbf{x}$  sur le maillage hexagonal, figure 2. Un capteur réalise une photo conversion, transformation de Poisson de moyenne  $H\mathbf{x}$ , puis un bruit de lecture Gaussien s'ajoute (*read out noise*), indépendant du processus de Poisson. Les données bruitées  $\mathbf{p}$  s'écrivent alors suivant le modèle :

$$\mathbf{p} = \mathcal{P}(H\mathbf{x}) + \mathbf{n} \quad (17)$$

où  $\mathcal{P}$  est la transformation de Poisson et  $\mathbf{n} \sim \mathcal{N}(0, \sigma^2)$ . L'image résultante est représentée dans la figure 2 pour  $3 \times 10^7$  photons et  $\sigma^2 = 29$ . Toutes les images sont de taille  $1024 \times 1024$  points.

Les algorithmes 15 et 16 ont été implémentés sur les données précédentes. Les résultats sont donnés pour  $\gamma = 2$ , valeur choisie car une première étude dans [19] montre un minimum de l'erreur quadratique pour cette valeur de  $\gamma$ . Au cours des itérations, l'effet des algorithmes est de réduire progressivement les répliques de l'objet jusqu'à les supprimer dans la meilleure reconstruction au sens de l'erreur quadratique moyenne  $\epsilon^k = \|\mathbf{x}^k - \mathbf{x}^*\| / \|\mathbf{x}^*\|$ . Les meilleures reconstructions de l'objet sont données dans la figure 3. La comparaison de l'image convoluée bruitée et des reconstructions, figure 3 du bas montre les excellents résultats obtenus par application de ces algorithmes. Sur cet exemple, les meilleurs résultats en terme de rapidité et de minimum de l'erreur de reconstruction sont obtenus avec LBI.

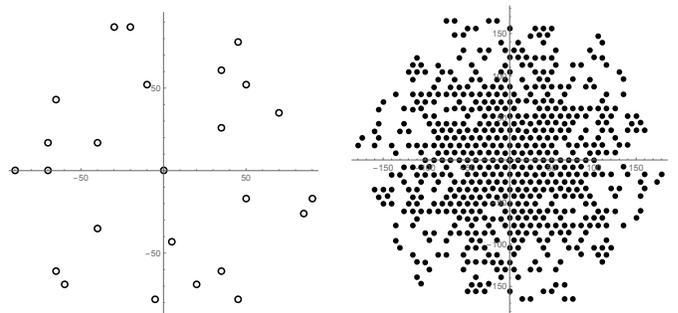


FIGURE 1 – Configuration non-redondante de 25 télescopes d'après Kopilovich et Sodin [16] (à gauche) et couverture correspondante du plan des fréquences spatiales (à droite).

## 5 Conclusion

Nous avons proposé des algorithmes de minimisations sous contraintes de non-négativité et de somme constante. L'appli-

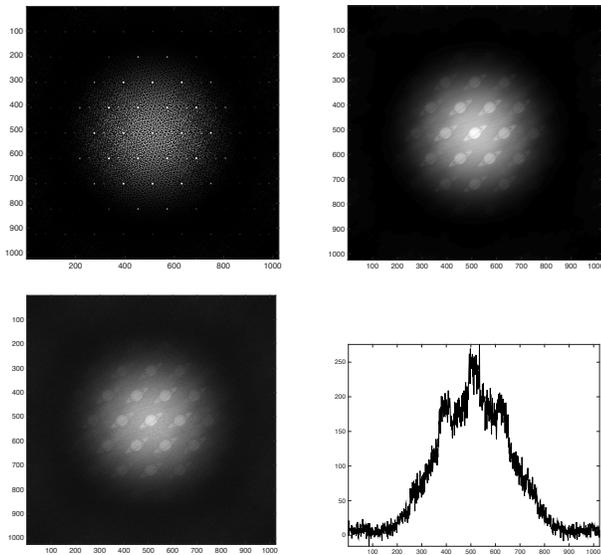


FIGURE 2 – Haut : PSF  $H$  et image convoluée  $Hx$ . Bas : image bruitée  $p$  et coupe au centre de l’image bruitée.

cation à la reconstruction d’images pour l’interférométrie optique donne d’excellents résultats. La contrainte de somme est prise en compte par l’utilisation, comme fonctions objectifs, de divergences invariantes par changement d’échelle. Ceci conduit à des algorithmes de type point intérieur, ce qui permet l’arrêt des itérations avant convergence, essentiel dans le contexte de problèmes mal posés. Des études complémentaires devront être menées, comme par exemple, la différence de comportement entre les algorithmes LAI et LBI et l’influence du paramètre  $\gamma$ . Un point intéressant et qui fait l’objet d’un travail en cours est la régularisation du problème par ajout d’un terme de pénalité basé sur ces mêmes divergences, ce qui permet de conserver la propriété de maintien du flux des algorithmes.

## Références

- [1] H. Lantéri, M. Roche, and C. Aime, “Penalized maximum likelihood image restoration with positivity constraints- multiplicative algorithms,” *Inverse problems*, vol. 18, pp. 1397–1419, 2002.
- [2] H. Lantéri, C. Theys, C. Richard, and C. Févotte, “Split gradient method for nonnegative matrix factorization,” in *Proc. 18th European Signal Processing Conference (EUSIPCO’10)*, Aalborg, Denmark, 2010.
- [3] C. Theys, N. Dobigeon, J.Y. Tourneret, and H. Lantéri, “Linear unmixing of hyperspectral images using a scaled gradient method,” in *Proc. IEEE Workshop Stat. Signal Process. (SSP)*, Cardiff, Wales, UK, Aug. 2009, pp. 729–732.
- [4] H. Lantéri and C. Theys, “La gamma divergence de Fujisawa-Eguchi, une alternative pour imposer la contrainte de somme. application au mélange linéaire.,” in *GRETSI*, 2013.
- [5] I. Csiszár, “Information measures : a critical survey,” in *Transactions of the 7<sup>th</sup> Prague conference*, 1974.
- [6] S. I. Amari, “Alpha divergence is unique belonging to the f-divergence and Bregman divergence classes,” *IEEE Transactions on Information Theory*, 2009.

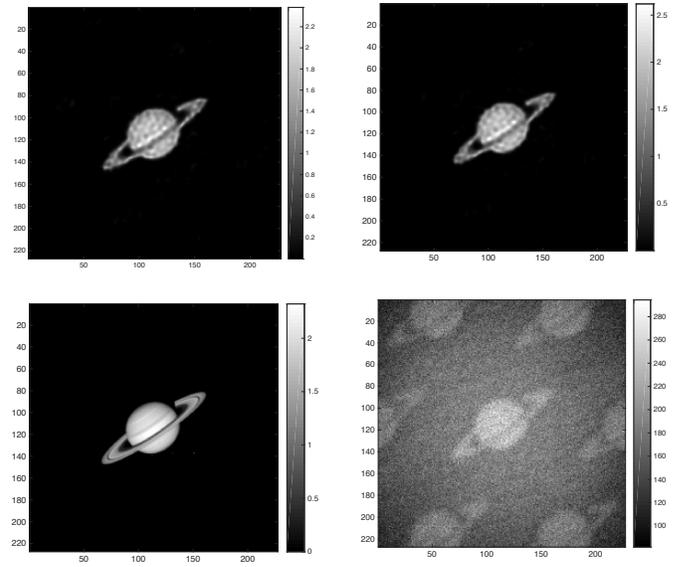


FIGURE 3 – Haut : Reconstructions avec  $LAI_2$  ( $k = 496$ ,  $\epsilon_{\min} = 0.19$ ) et  $LBI_2$  ( $k = 49$ ,  $\epsilon_{\min} = 0.14$ ). Bas : objet  $x$  et image convoluée bruitée  $p$ . (Parties centrales des images)

- [7] S. I. Amari and H. Nagaoka, *Methods of information geometry*, Oxford University Press, New York, 2000.
- [8] L. M. Bregman, “The relaxation method of finding the common point of convex sets and its application to the solution of problems in convex programming,” *USSR Computational Mathematics and Mathematical Physics*, vol. 7, pp. 200–217, 1967.
- [9] A. Basu, I. R. Harris, N. L. Hjort, and M. C. Jones, “Robust and efficient estimation by minimizing a density power divergence,” *Biometrika*, vol. 85, no. 3, pp. 549–559, 1998.
- [10] S. Eguchi and Y. Kano, “Robustifying maximum likelihood estimation,” Tech. Rep., Institute of statistical mathematics, 2001.
- [11] S. Eguchi and S. Kato, “Entropy and divergences associated with power function and the statistical application,” *Entropy*, 2010.
- [12] A. Cichocki and S. Amari, “Families of alpha-beta and gamma divergences : flexible and robust measures of similarities,” *Entropy*, vol. 12, 2010.
- [13] A. H. Phan A. Cichocki, R. Zdunek and S.I. Amari, *Non negative matrix and tensor factorization*, J. Wiley, 2010.
- [14] H. Fujisawa and S. Eguchi, “Robust parameter estimation with small bias against heavy contamination,” *Multivariate Analysis*, vol. 99, no. 9, 2008.
- [15] L. Armijo, “Minimization of functions having continuous derivatives,” *Pacific Journal of Mathematics*, , no. 16, pp. 1–3, 1966.
- [16] L. E. Kopilovich and L. G. Sodin, Eds., *Multielement system design in astronomy and radio science*, vol. 268 of *Astrophysics and Space Science Library*, 2001.
- [17] A. Labeyrie, H. Le Coroller, J. Dejonghe, O. Lardière, C. Aime, K. Dohlen, D. Mourard, R. Lyon, and K. G. Carpenter, “Luciola hypertelescope space observatory : versatile, upgradable high-resolution imaging, from stars to deep-field cosmology,” *Experimental Astronomy*, vol. 23, pp. 463–490, Mar. 2009.
- [18] C. Aime, H. Lantéri, M. Diet, and A. Carlotti, “Strategies for the deconvolution of hypertelescope images,” *Astronomy and Astrophysics*, vol. 543, pp. A42, July 2012.
- [19] H. Lantéri, C. Theys and C. Aime “Scale invariant divergences for signal and image reconstruction,” *EUSIPCO 2015*, to be published.