

Méthode algorithmique de minimisation de fonctions d'écart entre champs de données.

Application à la reconstruction d'images astrophysiques.

Henri LANTÉRI¹, Céline THEYS¹, Federico BENVENUTO², David MARY¹

¹Laboratoire Hippolyte Fizeau, UMR 6525
CNRS/OCA/ Université de Nice Sophia-Antipolis
Parc Valrose, 06108 Nice Cedex 2, France

²DIMA

Università degli Studi di Genova, Via Dodecaneso 35, 16146 Genova, Italie

Henri.Lanteri@unice.fr, Celine.Theys@unice.fr, benvenut@dima.unige.it, David.Mary@unice.fr

Résumé – Cette communication traite le problème d'optimisation d'un critère convexe avec contrainte de non négativité et de conservation du flux dans le cas particulier d'un problème de convolution. L'algorithme proposé est un algorithme de descente et s'applique à toute fonction d'écart appliquée sur des variables normalisées. Il est démontré que le flux et la positivité sont maintenues au cours des itérations. La convergence de cet algorithme est assurée pour un pas de descente correctement calculé, par un algorithme d'optimisation unidimensionnel par exemple. De plus une version «optimale» au sens du nombre d'itérations minimal est proposée. Une application sur une image de type astrophysique est donnée dans le cas d'une mesure d'écart entre la «moyenne racine carrée» et la moyenne géométrique des champs de données. Les résultats sont comparés avec ceux obtenus avec une distance de Kullback-Leibler correspondant à un processus de Poisson.

Abstract – This communication deals with convex optimization under positivity and total flux preservation, in the special case of deconvolution. The proposed algorithm is a descent one and can be applied to any discrepancy function of normalized variables. The total flux preservation and the positivity constraints along the iterations are demonstrated. The algorithm convergence is ensured using a one-dimension optimization search on the descent step, for example. Moreover, we describe a version of the algorithm for which the total number of iterations is minimized. An application on an astrophysical image is given in the case of a discrepancy measure between the square root mean and the geometrical mean of the data. A comparison with a Kullback-Leibler distance, corresponding to a Poisson process, is performed.

1 Introduction

Le problème traité ici est celui de la déconvolution d'images ; il s'agit d'un problème inverse mal posé, qui est en général résolu par rapport à l'inconnue \mathbf{x} en minimisant une fonction d'écart entre des mesures bruitées \mathbf{y} et un modèle linéaire $\mathbf{H}\mathbf{x}$. Dans les cas classiques fondés sur la maximisation de la vraisemblance, on est conduit à minimiser une divergence de Kullback-Leibler entre mesures et modèle (dans le cas d'un processus de bruit de Poisson), ou bien un écart quadratique moyen (dans le cas d'un bruit additif Gaussien). Ceci étant, dans de nombreux cas, comme l'imagerie astrophysique par exemple, les données brutes sont pré-traitées afin d'éliminer les défauts connus du capteur ou du système instrumental, ainsi, les propriétés statistiques du bruit sont perdues. Cela permet de considérer les nombreuses distances ou divergences qui existent dans la littérature et qui peuvent être utilisées dans ce contexte. Le processus de minimisation doit

faire intervenir une contrainte de non négativité de la solution (ce problème a été résolu de plusieurs façons) et une contrainte plus difficile à respecter de conservation du flux total de l'image au cours du processus de reconstruction. Nous proposons une méthode algorithmique itérative permettant de minimiser toute divergence ou distance entre mesures et modèle en respectant l'ensemble des contraintes ci-dessus au cours du processus itératif. La méthode proposée est une extension de la Méthode du Gradient Séparé (Split Gradient Method, SGM), [5].

2 Fonctions d'écart

Les champs de données qui sont mis en jeu sont notés de façon générale \mathbf{p} et \mathbf{q} . La plupart des fonctions d'écart trouvées dans la littérature, [2], s'expriment sous la forme

$$D(P, Q) = w[\sum_i g(p_i, q_i)] = w[\sum_i g_i] = w[G] \quad (1)$$

où $w[\cdot]$ et $g[\cdot]$ sont deux fonctions. Dans ces expressions, i est l'indice des pixels. Ces fonctions d'écart comportent trois niveaux de simplification possibles, en général non précisés explicitement :

- Aucune relation entre $\sum_i p_i$ et $\sum_i q_i$ n'est prise en compte explicitement ; ce sont des distances ou des divergences qui sont des mesures générales de convexité : Jensen, [4], Bregman, [3], etc...
- Elles sont simplifiées en tenant compte de la relation $\sum_i p_i = \sum_i q_i$ et sont en général utilisées pour chiffrer un écart entre 2 champs de données qui sont nécessairement de même somme.
- Elles sont simplifiées à l'extrême en tenant compte du fait que $\sum_i p_i = \sum_i q_i = 1$; elles sont en général utilisées pour chiffrer un écart entre 2 lois de probabilités.

L'emploi de variables de somme égale à "1" permet d'utiliser indifféremment toutes les divergences sans se préoccuper de leur degré de simplification. Dans le cas de la déconvolution et en utilisant des variables normalisées, on aura :

$$p_i = \frac{y_i}{\sum_j y_j}, q_i = \frac{(\mathbf{H}\mathbf{x})_i}{\sum_j (\mathbf{H}\mathbf{x})_j} \quad (2)$$

On peut noter que ce choix n'implique aucune relation entre $\sum_i y_i$ et $\sum_i (\mathbf{H}\mathbf{x})_i$. Dans le problème considéré, la réponse impulsionnelle est normalisée à 1 ce qui entraîne

$$\sum_j h_{ji} = 1, \forall(i) \Rightarrow \sum_i (\mathbf{H}\mathbf{x})_i = \sum_i x_i. \quad (3)$$

Par ailleurs, on peut montrer facilement que si $D(P, Q)$ est convexe par rapport à \mathbf{q} et si les q_i sont liés aux inconnues \mathbf{x} par la relation (2), alors $D(P, Q)$ est convexe par rapport à \mathbf{x} . Cette remarque a son importance dans la mesure où la méthode algorithmique proposée s'appuie sur une méthode de descente de type gradient.

3 Méthode algorithmique

Cette méthode itérative de minimisation, par rapport à \mathbf{x} , de fonctionnelles convexes $D(P, Q)$, sous contrainte de non-négativité est appelée SGM. Elle s'appuie sur les conditions de Kuhn-Tucker et sur une décomposition du gradient de la fonctionnelle à minimiser sous la forme $-\nabla_{\mathbf{x}} D(P, Q) = (\mathbf{U} - \mathbf{V})$ où toutes les composantes des vecteurs \mathbf{U} et \mathbf{V} sont positives. Cette décomposition n'est pas unique, et ce point sera utilisé par la suite. L'algorithme de descente est écrit sous la forme :

$$\mathbf{x}^{k+1} = \mathbf{x}^k + \alpha^k \text{Diag}\left[\frac{\mathbf{x}^k}{\mathbf{V}^k}\right](\mathbf{U}^k - \mathbf{V}^k) \quad \alpha^k > 0 \quad (4)$$

Dans ces équations, le rapport de vecteurs est effectué au sens de Hadamard et $\text{Diag}\left[\frac{\mathbf{x}^k}{\mathbf{V}^k}\right]$ est une matrice diagonale de diagonale égale à $(\frac{x_1^k}{V_1^k} \dots \frac{x_N^k}{V_N^k})$. La direction de

descente $-\mathbf{x}^k \nabla_{\mathbf{x}} D(P, Q)$ vient des conditions de Kuhn-Tucker pour la contrainte de positivité, la multiplication par $\frac{1}{\sqrt{k}}$ permet d'avoir un algorithme multiplicatif pour un pas de descente $\alpha^k = 1$, lire [5] pour plus de détails.

La valeur maximale α_{\max}^k du pas assurant la positivité est d'abord déterminée, elle est toujours supérieure à 1 ; une recherche unidimensionnelle simplifiée, donc rapide, sur l'intervalle $[0, \alpha_{\max}^k]$ permet ensuite de calculer une valeur du pas assurant la convergence de l'algorithme.

En considérant que les divergences $D(P, Q)$ sont convexes par rapport à \mathbf{q} , on aura dans le cas général, éq. (1) :

$$\nabla_{\mathbf{x}} D(P, Q) = \frac{w'(G)}{\sum_j (\mathbf{H}\mathbf{x})_j} \mathbf{M} \left[\frac{\partial \mathbf{g}}{\partial \mathbf{q}} \right] \quad (5)$$

En tenant compte de l'éq. (2), on a

$$\mathbf{M} = \mathbf{H}^T - \mathbf{Q} \quad (6)$$

où \mathbf{Q} est une matrice dont toutes les lignes sont égales à $[q_1 \ q_2 \ \dots \ q_N]$, avec $\sum_i q_i = 1$, soit

$$-\nabla_{\mathbf{x}} D(P, Q) = \frac{w'(G)}{\sum_j (\mathbf{H}\mathbf{x})_j} (\mathbf{H}^T - \mathbf{Q}) \left[-\frac{\partial \mathbf{g}}{\partial \mathbf{q}} \right] \quad (7)$$

Compte tenu des propriétés de la matrice \mathbf{H} , éq. (3), on aura, pour tout vecteur constant \mathbf{r} :

$$[\mathbf{H}^T - \mathbf{Q}][\mathbf{r}] = [\mathbf{0}] \quad (8)$$

On peut donc ajouter au vecteur $\mathbf{t} = w'(G) \left[-\frac{\partial \mathbf{g}}{\partial \mathbf{q}} \right]$, un vecteur constant \mathbf{r} , libre en signe, sans changer l'expression (7) et donc remplacer \mathbf{t} par $\mathbf{a} = \mathbf{t} + \mathbf{r}$ tel que $\mathbf{a} \geq 0$. L'expression (7) est alors décomposée selon :

$$\mathbf{U} = \mathbf{H}^T \mathbf{a}, \mathbf{V} = \mathbf{Q}\mathbf{a} \quad (9)$$

et l'algorithme itératif est construit sous la forme (4). On peut remarquer que toutes les composantes du vecteur $\mathbf{V} = \mathbf{Q}\mathbf{a}$ sont égales à $\sum_j q_j a_j$, et que le vecteur des coefficients de correction $(\frac{\mathbf{U}}{\mathbf{V}} - 1)$ ne deviendra égal à $\mathbf{0}$ que si le vecteur \mathbf{a} , c.à.d. \mathbf{a}^k tend vers une constante au cours des itérations.

Nous pouvons en déduire deux résultats importants. Le premier est que l'algorithme (4) associé à la décomposition (9) permet de maintenir le flux, en plus de la non négativité, au cours du processus itératif. Le second est que la valeur $r^k = -\min(t^k)$ conduit au plus rapide des algorithmes de ce type même si le gain en rapidité dépend de l'image à traiter et de la divergence à minimiser.

3.1 Maintien du flux total

L'algorithme (3) associé à la décomposition (10) permet de maintenir le flux et la non négativité au cours du processus itératif, en effet on peut écrire :

$$\sum_i x_i^{k+1} = (1 - \alpha^k) \sum_i x_i^k + \alpha^k \sum_i x_i^k \frac{U_i^k}{V_i^k} \quad (10)$$

Compte tenu des expressions de U_i^k et de V_i^k , on aura

$$\sum_i x_i^{k+1} = (1 - \alpha^k) \sum_i x_i^k + \alpha^k \sum_i x_i^k \frac{(H^T \mathbf{a}^k)_i}{\sum_j q_j^k a_j^k} \quad (11)$$

ce qui peut s'écrire :

$$\sum_i x_i^{k+1} = (1 - \alpha^k) \sum_i x_i^k + \alpha^k \sum_i a_i^k \frac{(Hx^k)_i}{\sum_j q_j^k a_j^k} \quad (12)$$

En divisant les deux membres par $\sum_i (Hx^k)_i = \sum_i x_i^k$, on obtient :

$$\frac{\sum_i x_i^{k+1}}{\sum_i x_i^k} = (1 - \alpha^k) + \frac{\alpha^k}{\sum_j q_j^k a_j^k} \sum_i a_i^k \frac{(Hx^k)_i}{\sum_i (Hx^k)_i} = 1 \quad (13)$$

Ainsi, le flux total est maintenu au cours des itérations; il suffit donc de fixer la valeur de l'estimé initial tel que $\sum_i x_i^0 = \sum_i y_i$.

3.2 Choix optimal du décalage \mathbf{r}^k

La valeur $r^k = -\min(t^k)$ conduit au plus rapide des algorithmes de ce type; pour montrer cela, on considère l'algorithme (4) associé à la décomposition (9), avec un pas de descente constant, égal à 1, pour alléger l'écriture. Deux situations sont analysées :

1. $r^k = -\min(t^k)$, alors au moins l'une des composantes du vecteur \mathbf{a}^k est nulle et l'algorithme s'écrit :

$$\mathbf{x}^{k+1} = \mathbf{x}^k \frac{\mathbf{U}^k}{\mathbf{V}^k} \quad (14)$$

2. $r^k = -\min(t^k) + C$ avec $C > 0$, alors toutes les composantes du vecteur \mathbf{a}^k sont strictement positives, et l'algorithme s'écrit :

$$\mathbf{x}^{k+1} = \mathbf{x}^k \frac{\mathbf{U}^k + C}{\mathbf{V}^k + C} \quad (15)$$

On peut montrer sans difficulté que la variation entre x_i^{k+1} et x_i^k est toujours plus faible dans le cas 2 que dans le cas 1.

4 Exemple d'application à l'imagerie

On considère deux exemples de distances pour l'application à la déconvolution d'images astrophysiques : la distance de Kullback-Leibler qui correspond à une transformation de Poisson de l'objet convolué et une distance qui n'est pas directement attachée à la vraisemblance des données, il s'agit de la mesure d'écart entre la moyenne racine carrée et la moyenne géométrique des 2 champs de données.

4.1 Distance de Kullback-Leibler

La distance de Kullback-Leibler est la distance qui modélise classiquement les mesures astrophysiques, [1]. Dans ce cas, (1) s'exprime en fonction des champs de données :

$$D_{\text{KL}}(\mathbf{P}, \mathbf{Q}) = \sum_i \left(p_i \log\left(\frac{p_i}{q_i}\right) - p_i + q_i \right) \quad (16)$$

On a alors $w(u) = u$, et on obtient :

$$t_i^k = \frac{p_i}{q_i^k} \quad (17)$$

t_i^k est une fonction croissante de $p_i/q_i^k \geq 0$ et son minimum absolu est égal à 0. Deux cas sont analysés

1. $r_i^k = 0, \forall k$.
2. $r_i^k = -\min(\mathbf{t}^k)$.

4.2 Distance entre moyennes

On montre ici une application à la minimisation de la divergence M_{SG} [6], qui est égale à l'écart entre la moyenne "racine carrée" et la moyenne géométrique des 2 champs de données :

$$M_{\text{SG}}(\mathbf{P}, \mathbf{Q}) = \sum_i \left[\sqrt{\frac{p_i^2 + q_i^2}{2}} - \sqrt{p_i q_i} \right] \quad (18)$$

Dans ce cas, on a $w(u) = u$, et on obtient :

$$t_i^k = \sqrt{\frac{p_i}{q_i^k}} - \frac{\sqrt{2}}{\sqrt{1 + \left(\frac{p_i}{q_i^k}\right)^2}} \quad (19)$$

Cette expression est une fonction croissante de $p_i/q_i^k \geq 0$ et son minimum absolu est égal à $-\sqrt{2}$. Deux cas sont analysés :

1. $r_i^k = \sqrt{2}, \forall k$.
2. $r_i^k = -\min(\mathbf{t}^k)$.

On considère un objet synthétique typique des observations astronomiques fig. 1(a). Il a été convolué par une réponse impulsionnelle représentant de façon réaliste les effets de la turbulence atmosphérique et l'effet des dimensions finies du télescope, fig. 1(b); on a introduit un bruit de Poisson avec 10^6 photons dans l'image; les données ainsi obtenues sont représentées fig. 1(c). La meilleure reconstruction est montrée sur la figure 1(d) dans le cas de la distance entre moyennes; elle correspond au minimum de l'écart quadratique moyen relatif entre l'objet exact et l'objet reconstruit. L'erreur de reconstruction et le critère à minimiser en fonction du nombre d'itérations sont représentés sur les figures (2) et (3). Leurs allures sont typiques d'un problème mal posé où plusieurs solutions correspondent au minimum du critère à minimiser, on parle de semi-convergence, on cherche alors la meilleure solution au sens d'un critère choisi qui ici est l'écart quadratique moyen relatif entre l'objet exact et l'objet reconstruit. Les

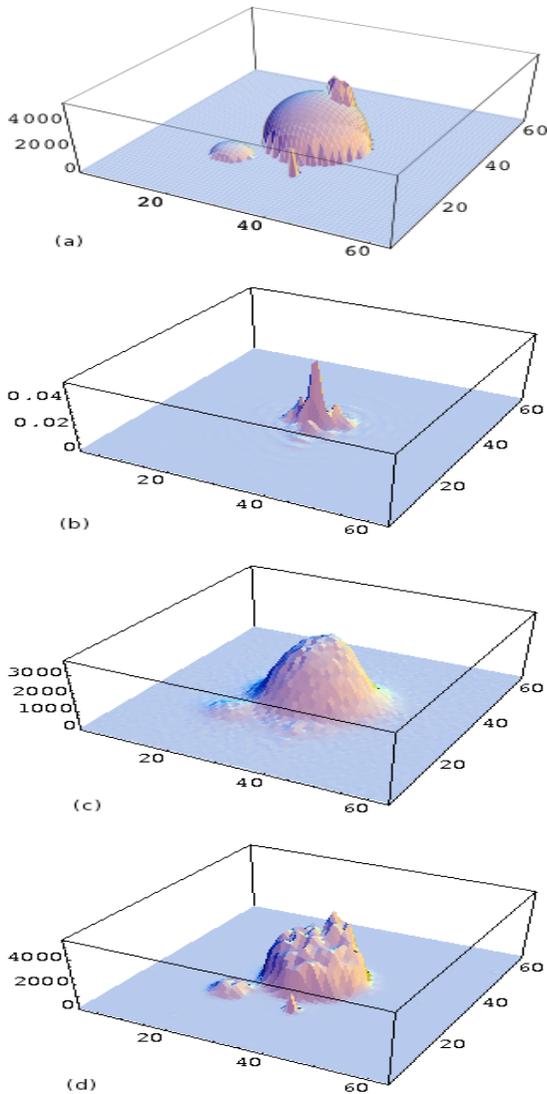


FIGURE 1 – (a) Objet, (b) Psf, (c) Image convoluée bruitée, (d) Meilleure reconstruction

valeurs numériques des minima des erreurs de reconstruction et du nombre d'itérations pour les deux distances et dans les deux cas sont données dans le tableau ci-dessous. On peut noter que, sur cet exemple (également vrai dans toutes nos simulations), les deux distances donnent sensiblement la même erreur de reconstruction mais que le respect d'un choix de distance correspondant à la vraisemblance des données, bruit de Poisson, permet d'obtenir un résultat plus rapide.

Distance	Kullback-Leibler		Moyennes	
	0	$-\min(\mathbf{t}^k)$	$\sqrt{2}$	$-\min(\mathbf{t}^k)$
r_k	0	$-\min(\mathbf{t}^k)$	$\sqrt{2}$	$-\min(\mathbf{t}^k)$
nbre iter.	86	81	126	98
erreur (%)	10.06	10.06	10.06	10.06

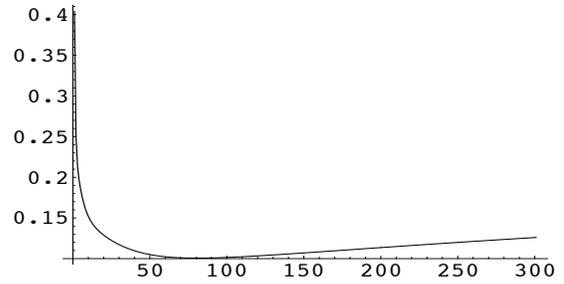


FIGURE 2 – Erreur de reconstruction en fonction du nombre d'itérations

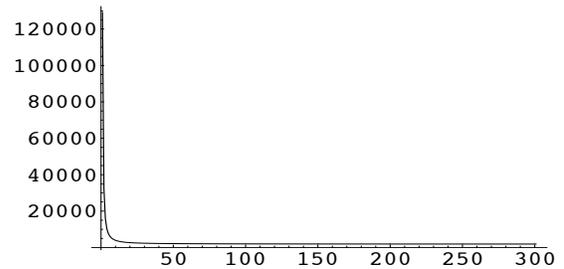


FIGURE 3 – Critère à minimiser en fonction du nombre d'itérations

5 Conclusion

Un algorithme itératif de minimisation d'un critère convexe sous contrainte de positivité de la solution et de maintien du flux a été proposé. La convergence (ou semi convergence pour un problème mal posé) est assurée avec une recherche sur le pas de descente. Des résultats numériques sont donnés dans le cas d'une déconvolution d'un objet de type astrophysique. On peut noter que sur cet exemple, et sur tous nos tests, l'erreur quantitative du résultat ne dépend pas du type de distance choisie, en accord ou pas avec la statistique des données.

Références

- [1] Lucy L B. An iterative technique for the rectification of observed distributions. *Astronomic Journal*, (79) :745–754, 1974.
- [2] Michèle Basseville. Information : entropies, divergences et moyennes. Technical report, IRISA, 1996.
- [3] 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*, 7 :200–217, 1967.
- [4] Jacob Burbea and C. Radhakrishna Rao. On the convexity of some divergence measures based on entropy functions. *IEEE Transactions on Information Theory*, 28(3) :489–495, 1982.
- [5] H. Lanteri, M. Roche, and C. Aime. Penalized maximum likelihood image restoration with positivity constraints- multiplicative algorithms. *Inverse problems*, 18 :1397–1419, 2002.
- [6] Inder Jeet Taneja. On mean divergences measures. *Math.ST*, 2005.