

# Optimisation itérative par majoration-minimisation pour l'estimation statistique sous l'hypothèse d'un bruit ricien

Qianyi JIANG<sup>1</sup>, Saïd MOUSSAOUI<sup>1</sup>, Jérôme IDIER<sup>1</sup>, Guylaine COLLEWET<sup>2</sup>

<sup>1</sup>Laboratoire des Sciences du Numérique de Nantes (LS2N, CNRS UMR 6004), 1 rue de la Noë, 44321 Nantes Cedex 3, France

<sup>2</sup>IRSTEA, UR OPAALE CS 64427, 35044 Rennes, France

qianyi.jiang@eleves.ec-nantes.fr

**Résumé** – Cette communication s'intéresse à l'estimation statistique d'un signal perturbé par un bruit ricien avec comme objectif de proposer une méthode d'optimisation efficace. L'application ciblée concerne le traitement d'images biologiques obtenues par résonance magnétique. Partant du constat que le critère à minimiser est unimodal mais non-convexe, la principale contribution de cette communication consiste en la proposition d'une stratégie d'optimisation fondée sur les techniques de majoration-minimisation quadratique après introduction d'une transformation permettant la convexification du critère à minimiser. Le schéma de descente itérative proposé sera comparé à l'approche de descente MM de référence. Les performances sont analysées en utilisant des données réelles issues de l'analyse d'échantillons agroalimentaires.

**Abstract** – This paper addresses statistical estimation of signals corrupted by a Rician noise with the aim to propose an efficient optimization method. The aimed application is the restoration of magnetic resonance images. Starting from the fact that the criterion to minimize is non-convex but unimodal, the main contribution of this work is to propose an optimization scheme based on the majoration-minimization principle after introducing a variable change allowing to get a strictly convex criterion. The resulting descent algorithm is compared to the classical MM descent algorithm and its performances are analyzed using real data sets of food sample analysis.

## 1 Introduction

Le bruit en imagerie par résonance magnétique (IRM) est la principale source de difficulté pour l'interprétation des images. Des temps d'acquisition doivent être suffisamment longs pour obtenir des images ayant un rapport sur bruit (RSB) suffisant. Dans certaines applications, il est nécessaire de réduire les temps d'acquisition des images et d'associer au système d'imagerie des routines de réduction de bruit [1]. Cependant, l'efficacité des méthodes de débruitage doit être éprouvée et notamment par la prise en compte de la statistique du bruit de mesure. En effet, comme le bruit IRM est supposé gaussien indépendant et identiquement distribué dans l'espace complexe des mesures (partie réelle et partie imaginaire), les images en module suivent une distribution de Rice [2]. Or, dans plusieurs applications de l'IRM, c'est le module du signal qui est traité du fait de la sensibilité de la phase à plusieurs effets instrumentaux tels que l'hétérogénéité du champ magnétique, la température de l'échantillon et le mouvement de l'objet observé. Ceci rend inefficace l'utilisation des techniques de débruitage fondées sur l'hypothèse d'un bruit gaussien.

Une méthode dite conventionnelle (CA), fondée sur le moment d'ordre deux de la distribution de Rice, a été proposée dans [3]. Quelques années plus tard, Sijber et al. [4] ont proposé une autre méthode d'estimation basée sur le principe de maximum vraisemblance (MV). L'estimateur MV est un compromis entre précision et efficacité statistique mais son application à des images réelles, tel que mentionné par Santiago [5],

ne permet pas d'avoir une performance satisfaisante à cause de l'apparition d'un flou. Ce phénomène n'est pas dû à la méthode MV elle-même, mais au fait qu'elle était initialement conçue pour plusieurs images. En fait, il y a deux approches possibles pour son application dans le cas d'une seule image : utiliser les voisinages de chaque pixel pour estimer la vraisemblance ou prendre en compte simplement le pixel unique. Dans le premier cas, l'estimation locale provoque le flou. Dans l'autre cas, le manque d'information entraîne une faible réduction de bruit. La solution dans ce cas est d'ajouter de l'information a priori à travers une régularisation.

Dans cette communication, nous proposons une méthode de maximum de vraisemblance régularisé pour améliorer la performance de restauration d'images IRM à partir d'une seule mesure, à la fois en temps de calcul et en qualité de restauration. Nous introduisons une régularisation entropique et nous proposons un algorithme d'optimisation itérative par majoration-minimisation (MM) adapté à la structure du critère à minimiser. La validité de la méthode est illustrée sur la base d'images réelles issues de mesures sur des échantillons agroalimentaires.

## 2 Formulation du problème

Le module du signal mesuré suit une distribution de Rice :

$$p(M_i | A) = \frac{M_i}{\sigma^2} e^{-\frac{M_i^2 + A^2}{2\sigma^2}} I_0\left(\frac{AM_i}{\sigma^2}\right) u(M_i), \quad (1)$$

où  $I_0(\cdot)$  est la fonction de Bessel modifiée de premier espèce d'ordre zéro,  $A$  représente l'amplitude du signal à estimer<sup>1</sup> et  $M_i$  correspond à la  $i$ -ème mesure,  $\sigma^2$  est la variance du bruit gaussien,  $u$  désigne la fonction indicatrice.

Sous l'hypothèse d'indépendance des  $N$  mesures  $M_i$ , la fonction objectif à minimiser pour l'estimation de l'intensité  $A$  par maximum de vraisemblance s'exprime par :

$$L(A) = -\log \prod_{i=1}^N p(M_i|A) = -\sum_{i=1}^N \log I_0\left(\frac{AM_i}{\sigma^2}\right) + \sum_{i=1}^N \frac{A^2 + M_i^2}{2\sigma^2} - \sum_{i=1}^N \log\left(\frac{M_i}{\sigma^2}\right). \quad (2)$$

D'après le critère (2), la méthode classique du MV est formulée pour plusieurs acquisitions. En présence d'une seule mesure ( $N = 1$ ), une approche alternative pour compléter l'information statistique pour reconstruire l'image est d'introduire un terme de régularisation issu d'informations *a priori* sur l'image.

Nous proposons d'appliquer une régularisation de type entropique [6], qui du fait que son gradient soit infini en 0 permet d'assurer la positivité des intensités des pixels de l'image. Considérons le cas de l'entropie de Burg qui correspond à :

$$R(A) = -\log A. \quad (3)$$

Signalons que d'autres formes d'entropies peuvent être utilisées et les développements qui vont suivre resteront valables, moyennant quelques adaptations techniques. L'expression finale du critère à minimiser et l'estimateur qui en découle sont donnés par :

$$F(A) = -\log p(M|A) + \beta R(A) \quad (4)$$

$$\hat{A} = \arg \min_A F(A) \quad (5)$$

où  $\beta$  est le paramètre qui permet d'ajuster le poids du terme de régularisation.

Du point de vue de l'optimisation, en l'absence de régularisation, la fonction objectif possède un minimum local et  $A = 0$  est toujours une solution triviale [7]. Dans [8], les auteurs ont proposé une estimation de maximum vraisemblance pénalisé fondée sur une majoration quadratique et une régularisation de type variation totale. Cependant, à notre connaissance, la question des performances des algorithmes d'optimisation dans ce cadre non-convexe n'est pas encore complètement traitée.

## 2.1 Critère initial, non convexe et unimodal

La fonction objectif ne peut pas être minimisée explicitement et elle n'est pas convexe mais reste unimodale, donc un algorithme d'optimisation itérative peut être appliqué. Nous optons pour un algorithme de descente de type MM. Comme  $f(\cdot) = -\log I_0(\cdot)$  est une fonction concave, celle-ci peut être majorée en tout point  $A_k$  par une fonction affine :

$$f(A) = -\log I_0\left(\frac{AM}{\sigma^2}\right) \leq g_k(A - A_k) + f(A_k), \quad (6)$$

avec  $g_k$  le gradient de  $f(\cdot)$  en  $A_k$  :

$$g_k = -\frac{M}{\sigma^2} \frac{I_1\left(\frac{A_k M}{\sigma^2}\right)}{I_0\left(\frac{A_k M}{\sigma^2}\right)}, \quad (7)$$

où  $A_k$  la valeur estimée à la  $k$ ème itération d'un algorithme de descente et  $I_1(\cdot)$  la fonction de Bessel modifiée du premier espèce et d'ordre un. Par conséquent, la fonction tangente majorante  $H_L(\cdot, A_k)$  de  $L(\cdot)$  et le gradient du critère régularisé  $\nabla H_F(\cdot, A_k)$ , sont sous la forme :

$$H_L(A, A_k) = \frac{A^2}{2\sigma^2} + g_k(A - A_k) - \log I_0\left(\frac{A_k M}{\sigma^2}\right), \quad (8)$$

$$\nabla H_F(A, A_k) = \frac{A}{\sigma^2} - \frac{\beta}{A} + g_k. \quad (9)$$

La racine positive qui annule ce gradient est exprimée par :

$$A_{k+1} = \frac{\sigma^2}{2} \left( -g_k + \sqrt{g_k^2 + \frac{4\beta}{\sigma^2}} \right). \quad (10)$$

Notons aussi que pour  $\beta = 0$ , cette solution rejoint celle du MV classique [9].

## 2.2 Critère transformé convexe

D'après la proposition 3 de [9], en introduisant le changement de variable  $y = A^2$  la fonction  $L$  est une fonction convexe et uni-modale par rapport à la variable  $y$  (voir l'illustration de la Figure 1). Ainsi, nous proposons de construire une nouvelle fonction majorante du critère convexe et nous appliquerons le schéma de descente MM pour l'optimisation par rapport à  $y$  et en déduire la valeur de  $A$ .

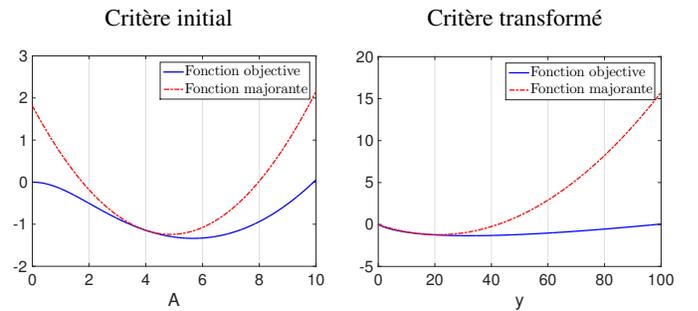


FIGURE 1 – Illustration du modèle initial non convexe unimodal, le critère transformé convexe et les approximations majorantes quadratiques

En notant  $y = A^2$ , la fonction convexe  $\tilde{L}$  et la fonction objectif  $\tilde{F}$  s'écrivent :

$$\tilde{L}(y) = \frac{y}{2\sigma^2} + f(\sqrt{y}) \quad (11)$$

$$\tilde{F}(y) = \tilde{L}(y) - \frac{\beta}{2} \log y \quad (12)$$

Nous proposons la fonction tangente majorante quadratique

$$\tilde{H}_L(y, y_k) = \frac{1}{2} B_k (y - y_k)^2 + \nabla \tilde{L}(y_k) (y - y_k) + \tilde{L}(y_k), \quad (13)$$

dont la courbure est définie par  $B_k = \frac{2 \left( \nabla \tilde{L}(y_k) y_k - \tilde{L}(y_k) \right)}{y_k^2}$ .

1. L'indice de pixels correspondant est omis pour simplifier les notations

**Discussion.** Cette courbure résulte du choix que la fonction majorante passe par le même point que  $\tilde{L}$  en  $y = 0$ . Cette construction permet d’obtenir l’approximation majorante quadratique de plus faible courbure. La propriété de majoration se déduit du fait que la dérivée troisième de la  $\tilde{F}$  est négative et donc sa courbure est décroissante. Les démonstrations sont omises ici par manque de place.

**Proposition 1.** La courbure  $B_k$  de la fonction tangente majorante  $\tilde{H}(\cdot, y_k)$  définie par l’équation (13) est définie positive pour toute valeur de  $y_k$  non-négative.

*Démonstration.* Notons que  $B_k(y)$  peut être formulée comme étant la dérivée de  $b(y) = 2\frac{\tilde{L}(y)}{y} = -2\frac{\log I_0(\sqrt{y})}{y} + \frac{1}{\sigma^2}$ . La preuve consiste à montrer que cette fonction est croissante ou encore à montrer que  $\exp(-P(y))$  est décroissante. Or, la dérivée de  $\exp(-P(y))$  est sous la forme  $I_1(y) - 2yI_0(y)$ , et d’après les théorème (15) et (16) de [10], il en résulte  $\frac{I_1(y)}{I_0(y)} \leq 2y$ . Ainsi, la positivité de la courbure est vérifiée.  $\square$

La valeur estimée de  $A$  à l’itération  $k + 1$  est obtenue avec la racine positive du gradient de  $\tilde{H}_L(\cdot, y_k) - \frac{\beta}{2} \log(\cdot)$  :

$$y_{k+1} = \frac{(y_k B_k - \nabla \tilde{F}(y_k)) + \sqrt{(y_k B_k - \nabla \tilde{F}(y_k))^2 + \beta B_k}}{2B_k}.$$

## 3 Analyse des performances

### 3.1 Intérêt du changement de variable

Comme les critères définis en fonction de  $A$  et  $y$  sont identiques, nous pouvons comparer les vitesses de convergence des deux algorithmes de descente MM. Pour cela, nous considérons une image IRM réelle en module. La variance du bruit est estimée à partir des pixels du fond de l’image en utilisant la méthode proposée dans [5]. L’optimisation est réalisée de façon indépendante pour chaque pixel de l’image. L’image mesurée est prise comme valeur initiale et le test d’arrêt des algorithmes MM se base sur un seuil sur la norme du gradient fixé à  $10^{-8}$ .

La figure 2 résume le nombre d’itérations nécessaires pour les deux algorithmes MM (critère initial et critère transformé) ainsi que pour une version optimisée mixant les deux algorithmes. Cette approche optimisée consiste à retenir à chaque itération la solution MM qui induit la plus forte descente du critère. Nous pouvons constater que le nombre d’itérations nécessaires pour la convergence de l’algorithme de descente en  $y$  est beaucoup plus faible quand  $\frac{M_i}{\sigma}$  est inférieur à une valeur de l’ordre de 1.8 environ, ce qui montre l’intérêt de l’approche proposée dans le cadre d’un faible rapport signal sur bruit (RSB). Cependant, la descente en  $A$  nécessite un nombre d’itérations moins important quand le RSB augmente. Ceci peut s’expliquer par la proximité de la distribution de Rice d’une distribution gaussienne lorsque le RSB est très fort. Néanmoins, l’approche optimisée permet de tirer profit des deux algorithmes en réduisant significativement le nombre d’itérations quel que soit la valeur du RSB. Une analyse théorique de ces résultats reste à réaliser.

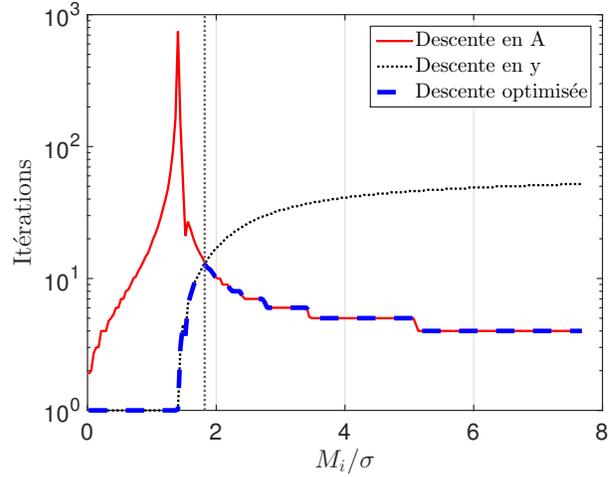


FIGURE 2 – Nombre d’itération du modèle initial, modèle transformé et modèle mixé pour estimation du MV

### 3.2 Débruitage avec régularisation entropique

Sachant que le RSB en IRM est proportionnel au nombre d’acquisitions, nous prenons une image IRM d’une coupe de fromage avec 32 acquisitions comme image de référence et la même coupe avec 8 acquisitions pour tester la méthode proposée. La variance du bruit est calculée selon une méthode d’estimation basée sur la distribution locale de la variance du signal [5]. La version optimisée de l’algorithme MM est employée pour l’optimisation du critère. Pour analyser la qualité du débruitage, l’erreur quadratique moyenne normalisée (EQMN) et la déviation absolue moyennée (DAM) sont calculées pour différentes valeurs du paramètre  $\beta$  (voir Figure.3).

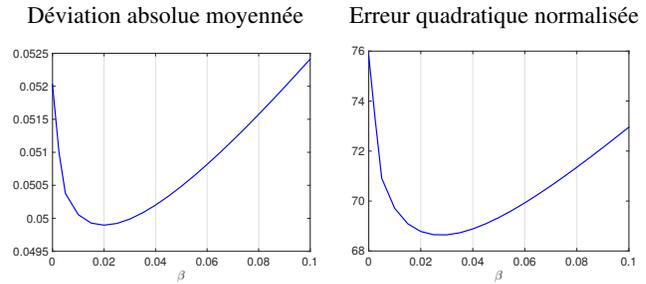


FIGURE 3 – Déviation absolue moyennée (Gauche) et erreur quadratique moyenne normalisée (Droite) pour différentes valeurs de  $\beta$  pour une régularisation entropique.

D’après les résultats de la Figure 3, et d’autres expérimentations avec des images réelles, nous pouvons affirmer que l’estimation par maximum de vraisemblance pénalisé tend à améliorer la qualité de l’image si le paramètre  $\beta$  est bien choisi. Cependant la gain apporté par la régularisation entropique doit être comparé à celui qui sera apporté par d’autres formes de régularisations.

Nous réalisons également une comparaison avec les méthodes classiques de maximum vraisemblance. L’indice de qualité basé

sur la variance locale (QIVL) [11] et l'erreur absolue normalisée (NAE) sont calculés. Le premier donne une mesure du niveau de flou ; plus l'indice est proche de 1 meilleure est l'image. Le deuxième évalue la qualité de la réduction du bruit. À partir des résultats présentés dans le tableau 1, on constate que l'estimation de MV pénalisé par l'entropie de Burg permet d'améliorer la qualité de l'image sans introduction de flou. La figure 4 illustre l'image originale bruitée et les résultats avec différents estimateurs classiques de MV. Nous remarquons que visuellement l'amélioration n'est pas évidente à percevoir en comparant avec la méthode de MV.

	MV pixel	MV entropie	MV voisinage
QIVL	0.8719	<b>0.8980</b>	0.4162
NAE	0.0431	<b>0.0421</b>	0.0423

TABLE 1 – Mesures de la qualité pour l'estimation par MV par pixel étant donné une seule image, MV pénalisé avec l'entropie de Burg et MV conventionnel utilisant un voisinage  $5 \times 5$

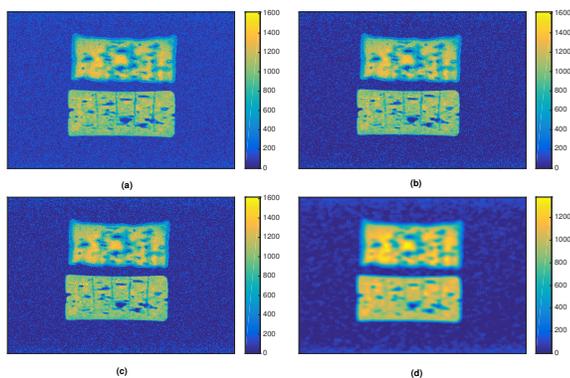


FIGURE 4 – Images IRM estimées d'une coupe de fromage : (a) Image originale. (b) MV pixel par pixel. (c) MV pénalisé,  $\beta = 0.03$ . (d) MV conventionnel (voisinage de  $5 \times 5$  pixels)

## 4 Conclusion

Dans cet article, nous avons étudié le problème de l'optimisation dans le cas de l'estimation en présence d'un bruit de Rice. Le schéma d'optimisation classique repose sur une structure de majoration quadratique d'un critère non-convexe mais unimodal. Nous avons proposé une approche originale fondée sur un changement de variable pour assurer la convexité du critère et une majoration quadratique efficace. Le résultat majeur est d'obtenir une accélération notable de la convergence de l'algorithme du maximum de vraisemblance en mixant ces deux solutions algorithmiques. Des expériences réalisées avec des images réelles ont montré que la régularisation conduit à une amélioration des estimations. En plus, en utilisant des indices d'évaluation de la qualité des images, nous avons démontré

que la méthode proposée permet un compromis entre réduction du bruit et préservation des bords quand une seule image est disponible. Les travaux futurs consistent à analyser les performances théoriques de l'algorithme proposé et à l'adapter à d'autres critères de régularisation entropique (Shannon, l'entropie croisée) ou de variation totale.

## 5 Remerciements

Ce travail a été financé par le GDR 720 ISIS dans le cadre de l'appel à projets exploratoires 2016.

## Références

- [1] J. Mohan, V. Krishnaveni et Y. Guo, « A survey on the magnetic resonance image denoising methods », *Biomedical Signal Processing and Control*, vol. 9, pp. 56–69, 2014.
- [2] H. Gudbjartsson et S. Patz, « The Rician distribution of noisy MRI data », *Magnetic Resonance in Medicine*, vol. 34, 6, pp. 910–914, 1995.
- [3] G. McGibney et M. Smith, « An unbiased signal-to-noise ratio measure for magnetic resonance images », *Medical Physics*, vol. 20, 4, pp. 1077–1078, 1993.
- [4] J. Sijbers et A. Den Dekker, « Maximum likelihood estimation of signal amplitude and noise variance from MR data », *Magnetic Resonance in Medicine*, vol. 51, 3, pp. 586–594, 2004.
- [5] S. Aja-Fernández, C. Alberola-López et C.-F. Westin, « Noise and signal estimation in magnitude MRI and Rician distributed images : a LMMSE approach », *IEEE Transactions on Image Processing*, vol. 17, 8, pp. 1383–1398, 2008.
- [6] C. A. Johnson, D. McGarry, J. A. Cook, N. Devasahayam, J. B. Mitchell, S. Subramanian et M. C. Krishna, « Maximum entropy reconstruction methods in electron paramagnetic resonance imaging », *Annals of Operations Research*, vol. 119, 1, pp. 101–118, 2003.
- [7] T. V. Yakovleva et N. S. Kulberg, « Noise and signal estimation in MRI : two-parametric analysis of Rice-distributed data by means of the maximum likelihood approach », *American Journal of Theoretical and Applied Statistics*, vol. 2, 3, pp. 67–79, 2013.
- [8] D. Varadarajan et J. P. Haldar, « A majorize-minimize framework for Rician and non-central Chi MR images », *IEEE Transactions on Medical Imaging*, vol. 34, 10, pp. 2191–2202, 2015.
- [9] J. Idier et G. Collewet, « Properties of Fisher information for Rician distributions and consequences in MRI », Rapport interne, IRCCyN/IRSTEA, 2014.
- [10] D. E. Amos, « Computation of modified Bessel functions and their ratios », *Mathematics of Computation*, vol. 28, 125, pp. 239–251, 1974.
- [11] Y.-q. Wang, W.-y. Liu et Y. Wang, « Image quality assessment based on local variance and structure similarity », *Journal of Optoelectronics-Laser*, vol. 19, 11, pp. 1546–1553, 2008.