

Mélange de gaussiennes, Estimation de la loi de distribution et Espérance Bayésienne a posteriori

Albert BIJAOUÏ

Laboratoire Cassiopée
Observatoire de la Côte d'Azur, BP 4220, 06304 Nice Cedex 04, France
Albert.Bijaoui@oca.eu

Résumé – Le débruitage d'un signal ou d'une image à bruit gaussien blanc stationnaire est développé dans le cadre de l'estimation bayésienne associée à l'espérance a posteriori. Si on admet que les signaux sont identiquement distribués, la meilleure estimation au sens quadratique résulte de la relation de Miyasawa. Pour cela, il faut obtenir une estimation de la dérivée logarithmique de la loi de distribution a posteriori. Dans cette communication, une méthode d'estimation de cette loi, exprimée sous la forme d'un mélange de gaussiennes, est développée. Bien qu'il existe de nombreux travaux portant sur l'estimation des paramètres d'un mélange de gaussiennes, un nouvel algorithme est présenté. D'une part, il exploite les concepts de rejet itératif et d'espérance-maximisation, d'autre part, il identifie les gaussiennes par un algorithme glouton associé à une analyse de résidu dérivant de la loi de Poisson. De ce fait, l'analyse est exhaustive et aucun a priori n'est nécessaire sur les gaussiennes. L'application à la transformée en ondelettes d'une image permet d'obtenir la meilleure estimation pour chaque plan en ondelettes. La comparaison avec l'estimation du maximum a posteriori montre la supériorité de cette approche. Cette analyse conduit à une faible parcimonie, qui peut être augmentée, en relaxant un peu sur la précision de l'estimation, tout en satisfaisant l'attache aux données.

Abstract – The denoising of a signal or image in the case of a stationary white Gaussian noise is developed in the framework of a Bayesian estimation associated with the posterior expectation. Assuming that the signals are identically distributed, the Minimum Mean Square Estimate (MMSE) is given by the Miyasawa relation. Its application requires an estimate of the logarithmic derivative of the posterior distribution law. In this paper, an estimation method of this law, expressed as a Gaussian mixture, is developed. Although there are numerous works on estimating the parameters of a Gaussian mixture, a new algorithm is presented in this communication. On the one hand, he uses the concepts of iterative rejections and the expectation-maximization and, on the other hand, it identifies the Gaussian by a greedy algorithm associated to the analysis of the residuals derived from the Poisson statistics. Thus, the analysis is complete and no a priori is needed. Its application to the wavelet transform of an image provides the best estimate for each wavelet plane. The comparison with the maximum a posteriori estimate shows the superiority of this approach. This analysis leads to a low sparsity, that can be increased by relaxing slightly the accuracy of the estimate, whereas the data fidelity is kept.

1 L'estimation par l'espérance a posteriori.

Considérons une variable aléatoire x , de loi a priori $p(x)$ dont la mesure est y . Parmi toutes les estimations \hat{x} de x à partir de y , l'espérance bayésienne est celle qui minimise le carré de l'erreur moyenne $E[(\hat{x} - x)^2]$. Miyasawa [1] a montré que, dans le cas d'un bruit gaussien de variance N , \hat{x} dérive de la distribution des mesures $q(y)$ par la relation :

$$\hat{x} = y + N \frac{q'(y)}{q(y)} = y + N \log' q(y). \quad (1)$$

L'estimation nécessite de déterminer $\log' q(y)$ à partir des mesures. Plusieurs voies ont été explorées, dont la décomposition en somme de deux gaussiennes [2], la pyramide de lissages [3], la représentation sous la forme d'une distribution de Voigt [4] ou des lois exponentielles tronquées [5]. La décomposition en somme de gaussiennes est l'une des voies les plus élégantes

pour résoudre ce problème. L'algorithme proposé dans [2] étant apparu trop sommaire, une approche plus exhaustive a été développée.

2 Ajustement d'un histogramme par un mélange de gaussiennes.

L'algorithme Espérance-Maximisation (EM) [6] est utilisé de manière intensive pour l'estimation des paramètres d'un mélange de gaussiennes. Sa convergence vers un maximum local de la vraisemblance a été prouvée, mais il peut ne pas s'agir du maximum absolu. De nombreuses variantes, en particulier stochastiques, ont été proposées [7].

Dans cette communication nous proposons une démarche en deux étapes. On commence à déterminer une première représentation par un algorithme identifiant progressivement les gaussiennes par rejet. Cela permet d'avoir une représentation cou-

vant bien l'intervalle des données, mais pouvant ne pas être correcte localement. On introduit de nouvelles gaussiennes avec une identification sur des résidus.

2.1 Identification de gaussiennes par rejet.

Un algorithme de rejet itératif [8] permet identifier des valeurs nécessitant l'introduction d'une nouvelle gaussienne. L'algorithme EM permet de préciser les paramètres des gaussiennes identifiées par rejet. Pour accélérer l'algorithme, l'identification des gaussiennes s'effectue à partir de l'histogramme du signal à débruiter.

Cette phase est très rapide car, sur les images étudiées, le nombre de gaussiennes est faible, six au maximum. À son issue, tous les pixels ont été affectés à au moins une gaussienne identifiée, mais la représentation n'est pas parfaite, puisque localement des écarts significatifs peuvent exister. Néanmoins, pour nos expérimentations, le filtre issu de l'identification par rejet est très proche de celui que nous allons déterminer avec un algorithme d'identification plus complet des gaussiennes associées à l'histogramme.

2.2 Identification sur les résidus.

Considérons qu'à une itération n nous avons obtenu une approximation $f^{(n)}(l)$ de l'histogramme $H(l)$ du signal, $l \in (1, L)$. Nous allons ajouter une nouvelle gaussienne :

$$\phi_i(l) = \frac{1}{\sqrt{2\pi}\sigma_i} e^{-\frac{(x(l)-m_i)^2}{2\sigma_i^2}}. \quad (2)$$

Pour cela nous écrivons :

$$f^{(n+1)}(l) = f^{(n)}(l) + a_i \phi_i(l). \quad (3)$$

Pour un jeu défini de paramètres de la gaussienne (m_i, σ_i) , l'amplitude a_i est déterminée par la maximisation de la vraisemblance :

$$\log(L) = \sum_l H(l) \log f^{(n+1)}(l) - \sum_l f^{(n+1)}(l). \quad (4)$$

Pour obtenir cette relation, nous avons utilisé le fait que les valeurs de l'histogramme sont distribuées selon une loi de Poisson. En effet, si les valeurs du signal sont considérées comme des mesures indépendantes de la même distribution, l'affectation d'une valeur à un intervalle est régie selon un processus de Bernoulli. La loi de distribution d'une valeur de l'histogramme est donc binomiale. Cette loi est très bien approximée par une loi de Poisson, compte tenu de la faible probabilité d'un intervalle donnée.

Nous pouvons déterminer l'amplitude de la gaussienne par dérivation :

$$R(i) = \sum \phi_i(l) \left[\frac{H(l)}{f^{(n+1)}(l)} - 1 \right] = 0. \quad (5)$$

Si le modèle $f^{(n)}$ est correct, tous les R_i sont négligeables, quelle que soit la gaussienne introduite. Pour identifier les gaussiennes associées, nous allons considérer pour quel jeu de paramètres $R(i)$ est significativement différent de 0. Pour cela,

nous examinons la transformée suivante :

$$T^{(n)}(i) = \sum_l \phi_i(l) \left[\frac{H(l)}{f^{(n)}(l)} - 1 \right]. \quad (6)$$

Les paramètres m_i et σ_i^2 sont échantillonnés sur une pyramide. Nous choisissons une première valeur de l'écart-type s_1 , à partir de l'approximation obtenue par rejet. Nous échantillons les moyennes m_i pour cette écart-type avec un pas de $s_1/2$. Puis nous doublons l'écart-type et le pas d'échantillonnage en moyenne. Nous itérons en doublant à nouveau les échelles et les pas. Dans nos expérimentations, une douzaine d'échelles a été considérée.

Si l'approximation $f^{(n)}(l)$ est correcte, nous avons :

$$E[H(l)] = f^{(n)}(l) \quad \sigma^2[H(l)] = f^{(n)}(l), \quad (7)$$

compte tenu de la nature poissonnienne de la distribution des valeurs de l'histogramme. Ceci permet d'obtenir :

$$E[T^{(n)}(i)] = 0 \quad \sigma^2[T^{(n)}(i)] = \sum_l \frac{\phi_i^2(l)}{f^{(n)}(l)}. \quad (8)$$

La loi de la variable $T^{(n)}(i)$ n'est pas gaussienne, mais si le nombre d'événements qui ont contribué à sa détermination est élevé, cette approximation est suffisante pour la détection de la localisation de couples de paramètres gaussiennes pour lesquels l'hypothèse d'un modèle correct n'est pas acceptable. On considère donc le rapport :

$$r^{(n)}(i) = \frac{T^{(n)}(i)}{\sigma^{(n)}(i)}. \quad (9)$$

Si $r^{(n)}(i) > k$, où k a été choisi à 4 dans nos expériences, un écart significatif est détecté pour le jeu de paramètres gaussiens correspondant.

Dans le programme, seul le plus grand rapport est exploité à chaque échelle, ce qui rend l'algorithme lent, mais une version parallèle peut être possible, en détectant simultanément les gaussiennes si leur cohérence mutuelle est faible.

À l'étape n , nous identifions une nouvelle gaussienne, qui est insérée pour des itérations du type EM. Celles-ci peuvent conduire à une amplitude nulle. C'est le cas lorsque la valeur significative de $r^{(n)}(i)$ est associée à un faible nombre d'événements. La statistique réelle n'est plus gaussienne. L'algorithme EM joue ainsi un rôle de filtre.

L'algorithme s'arrête lorsque qu'il n'y a plus de valeurs $r^{(n)}(i)$ significatives, qui n'aient pas été prises en compte.

3 Application au débruitage d'image.

Nous avons examiné une image astronomique spécifique de 2000×2000 issue d'un relevé de l'ESO. Sur la figure 1 nous en avons extrait une petite zone. Une transformation en ondelettes avec l'algorithme à trous avec la fonction B-spline cubique [9] a été effectuée. L'image de la zone pour la première échelle a été tracée sur la figure 2.

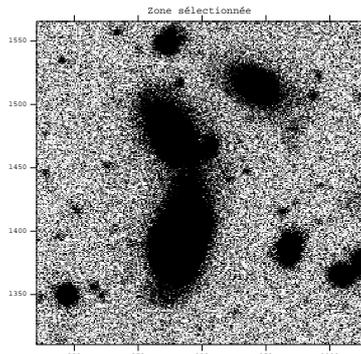


FIG. 1 – Zone sélectionnée de l'image. La représentation est linéaire, en négatif. Les valeurs extrêmes sont ajustées afin de bien mettre en évidence le bruit. La dynamique de l'image est environ mille fois plus grande.

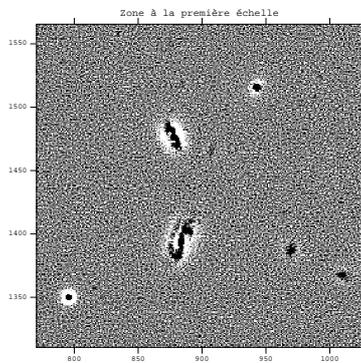


FIG. 2 – Premier plan en ondelettes de la zone sélectionnée. La représentation est la même que celle de l'image.

Sur la figure 3 nous avons porté en échelle logarithmique l'histogramme, l'ajustement obtenu par le mélange de gaussiennes et les 27 gaussiennes pour le premier plan en ondelettes de l'image. L'intervalle de tracé a été limité à la zone centrale pour mieux discerner l'ajustement du pic principal. Sur les ailes de la distribution, les valeurs sont réparties de manière éparse et on observe beaucoup de valeurs nulles pour l'histogramme. Dans la fenêtre tracée, on ne perçoit que 6 gaussiennes, dont la moitié d'entre elles ont une amplitude faible. Pour ce plan, l'information est répartie essentiellement dans le noyau central de la distribution. Au fur et à mesure que l'échelle augmente, la distribution devient de plus en plus complexe, en particulier dans les ailes de la distribution. Nous avons tracé, à droite, la relation entre la mesure et la valeur filtrée obtenue à partir de la relation de Miyasawa, toujours pour le premier plan en ondelettes. Pour cela l'écart-type du bruit a été choisi comme étant celui du pic principal de l'ajustement en somme de gaussiennes. Pour des valeurs proches de zéro, l'atténuation est importante. Par contre sur les ailes de la distribution, les valeurs sont conservées. Le filtre obtenu est, en quelque sorte, un fil-

trage dur atténué.

Sur la figure 4 nous avons représenté l'image filtrée du premier plan en ondelettes. En moyenne, l'écart-type du bruit pour ce plan a été réduit d'un facteur 3, 5.

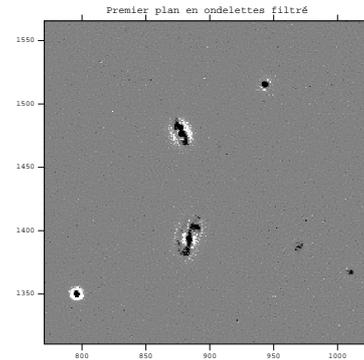


FIG. 4 – Zone du premier plan en ondelettes et de l'image après débruitage. La représentation est effectuée avec les mêmes paramètres que pour la figure 2.

Cette procédure est appliquée séparément échelle par échelle. Comme nous n'avons pas appliqué une transformation orthogonale, à une échelle donnée les valeurs de bruit sont corrélées. L'histogramme n'est pas en moyenne modifié, mais sa fluctuation est plus grande que celle induite par le simple bruit de Poisson. Cela ne met pas en cause l'application de la relation de Nadaraya, mais l'estimation de la loi a posteriori est sur-ajustée. En outre, les valeurs entre échelles sont aussi corrélées, ce qui fait que l'image de leur somme reste plus bruitée que si on avait appliqué une transformée orthogonale. La réduction du bruit est essentiellement liée à la première échelle, pour lesquelles l'effet de la corrélation du bruit est faible. Pour les plus grandes échelles, le filtrage est de plus en plus doux. Sur la figure 5 nous avons représentée l'image obtenue par cette procédure.

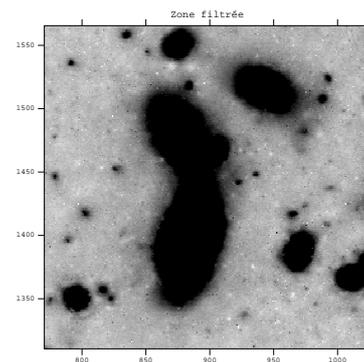


FIG. 5 – Zone sélectionnée de l'image après débruitage. La représentation est effectuée avec les mêmes paramètres que pour la figure 1.

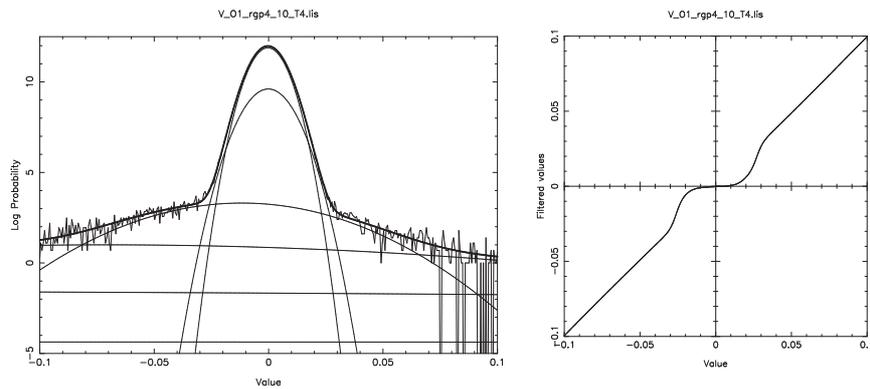


FIG. 3 – Ajustement de l’histogramme et fonction d’adoucissement pour les valeurs correspondant à la visualisation des images.

4 Conclusions.

La représentation obtenue après la première phase suffit pour le filtrage, même si l’ajustement des histogrammes n’est pas parfait. En fait, les différences apportées par une analyse exhaustive sont très faibles par rapport au bruit. L’algorithme EM avec rejet itératif est facile à mettre en œuvre et d’un coût de calcul faible.

Le débruitage résultant n’est pas parcimonieux. Même avec une loi a priori exponentielle, on n’obtient pas de parcimonie, contrairement à ce qui est obtenu avec le MAP. Pour obtenir de la parcimonie il faut augmenter la variance du bruit. Le rapport signal à bruit de l’image résultante est alors dégradée, mais l’attache aux données reste acceptable. On déduit aisément la loi a priori du signal $p(x)$ à partir de $q(y)$. L’approximation MAP conduit à une plus grande parcimonie, mais avec une réduction plus faible du bruit moyen (1, 94).

La méthode proposée permet de réduire au mieux l’écart-type entre l’espérance du signal et la mesure, sans hyperparamètre, dans le cas d’un bruit gaussien blanc. Pour l’appliquer, nous avons mis en œuvre une transformée, introduisant des corrélations du bruit. L’utilisation d’une transformation orthogonale permet de réduire ce défaut, mais en introduisant des artéfacts spécifiques. La relation de Miyasawa a été étendue au cas vectoriel [3], [5], mais la valeur du gradient du logarithme de la loi de probabilité a posteriori est très difficile à estimer correctement. Le procédé de Stein [10] permet d’éviter cette difficulté. Son application au débruitage a conduit à des résultats de très grand intérêt [11], [12]. Néanmoins, la détermination de la loi a posteriori permet d’obtenir le débruitage optimal.

Références

- [1] K. Miyasawa, *An empirical bayes estimator of the mean of a normal population*, Bull. Inst. Internat. Statist., 38, 181–188, 1961.
- [2] A. Bijaoui, *Wavelets, Gaussian Mixtures and the Wiener Filtering*, *Signal Processing*, 82, 709-712, 2002
- [3] A. Bijaoui, *A Universal Bayesian Estimate. Application to denoising*, *Astronomical Data Analysis IV*, ed. C.Surace, http://www.oamp.fr/conf/ada4/pub/19/DSD-ADAIV_Bijaoui.pdf, 2006.
- [4] A. Bijaoui. *Statistique empirique bayésienne. Estimation optimale et parcimonie*. XXIIème colloque du GRETSI, Dijon, <http://hdl.handle.net/2042/29031>, 2009.
- [5] M. Raphan, E.P. Simoncelli, *Empirical Bayes Least Squares Estimation without an Explicit Prior*, Computer Science Technical Report TR2007-900, Courant Institute of Mathematical Sciences, New York University, May 2007.
- [6] A. P. Dempster, N. M. Laird et D. B. Rubin. *Maximum likelihood from incomplete data via the EM algorithm*. *Journal of the Royal Statistical Society*, B 39 :1–38, 1977.
- [7] G. Celeux et G. Govaert (1991). *A classification EM algorithm for clustering and two stochastic versions*. Rapport de recherche RR-1364, Inria, Institut National de Recherche en Informatique et en Automatique, 1991.
- [8] J.L. Starck, F. Murtagh, A. Bijaoui, *Image processing and data analysis. The multiscale approach*, p.48, C.U.P., 1998.
- [9] J.L. Starck, F. Murtagh, A. Bijaoui, *Image processing and data analysis. The multiscale approach*, p.21, C.U.P., 1998.
- [10] C. Stein, *Estimation of the Mean of a Multivariate Normal Distribution*, *The Annals of Statistics*, 9, 1135–1151, 1981.
- [11] A. Benazza-Benyahia and J.-C. Pesquet, *Building Robust-Wavelet Estimators for Multicomponent Images Using Stein’s Principle*, *IEEE Trans. Image Processing*, 14, 1814–1830, (2005).
- [12] F. Luisier, T. Blu, M. Unser, *SURE-LET for Orthonormal Wavelet-Domain Video Denoising*, *IEEE Trans. Circuits and Systems for Video Technology*, 20, 913-919, 2010.