HAL open science

Echantillonnage de champs gaussiens de grande dimension

Olivier Feron, François Orieux, Jean-François Giovannelli

To cite this version:

Olivier Feron, François Orieux, Jean-François Giovannelli. Echantillonnage de champs gaussiens de grande dimension. 42èmes Journées de Statistique, 2010, Marseille, France, France. inria-00494771

HAL Id: inria-00494771

https://inria.hal.science/inria-00494771

Submitted on 24 Jun 2010

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Échantillonnage de champs gaussiens de grande dimension

Olivier Féron[1] & François Orieux[2] & Jean-François Giovannelli[3]

[1] EDF R&D et Laboratoire de Finance des Marchés de l'Énergie, Université Paris-Dauphine, Place du Maréchal de Lattre de Tassigny, 75775 Paris, olivier-2.feron@edf.fr.

[2] Laboratoire des Signaux et Systèmes (CNRS – SUPELEC – Univ. Paris-Sud 11) SUPELEC, Plateau de Moulon, 91192 Gif-sur-Yvette Cedex, orieux@lss.supelec.fr.

[3] Laboratoire d'Intégration du Matériau au Système, Équipe Signal-Image, Université de Bordeaux 1, 33405 Talence, France, Giova@IMS-Bordeaux.fr.

Résumé : Nous proposons une nouvelle approche pour l'échantillonnage de champs gaussiens corrélés dans le cas où les approches classiques ne sont pas utilisables : lorsque la dimension du problème est très grande et lorsque la matrice inverse de covariance (ou matrice de précision) n'est pas creuse. Cette approche est valide dans le cas où une structure particulière de la matrice de précision est disponible. Cette structure apparaît dans la résolution de problèmes inverses par des méthodes d'estimation bayésienne. L'algorithme proposé trouve une application directe pour les méthodes d'inversion myopes et/ou non supervisées, fondées sur des méthodes d'échantillonnage de type MCMC. L'efficacité de cette approche est illustrée sur l'inversion non supervisée d'un problème de super résolution.

Abstract : This work presents a novel approach for sampling Gaussian fields when classical approaches fail : in high-dimensional cases and when the precision matrix is not sparse. This approach is valid when a particular structure of the precision matrix is available. The main proposed application is inverse problems and their resolution in a Bayesian framework. It finds a direct application in myopic and/or unsupervised inversion methods based on MCMC sampling algorithms. An illustration in unsupervised super-resolution imaging assesses the effectiveness of the proposed approach.

Mots-clés : Échantillonnage, Problèmes inverses, Méthodes bayésiennes.

1 Introduction

Nous proposons une méthode d'échantillonnage de champs conditionnellement gaussiens de grande dimension, dont la matrice de précision n'est pas creuse. Ce problème

intervient notamment dans la résolution de problèmes inverses par des méthodes probabilistes. En effet, considérons le problème inverse général caractérisé par le modèle direct :

$$\boldsymbol{y} = \boldsymbol{H}\boldsymbol{x} + \boldsymbol{b}, \tag{1}$$

où $\boldsymbol{y}$ regroupe les observations, $\boldsymbol{x}$ est l'image inconnue, $\boldsymbol{H}$ est un opérateur linéaire et $\boldsymbol{b}$ le bruit. Nous nous plaçons dans un contexte d'estimation bayésienne et nous considérons des lois sur $\boldsymbol{b}$ et $\boldsymbol{x}$ gaussiennes conditionnellement à un ensemble de paramètres $\boldsymbol{\theta}$. Nous nous plaçons, de plus, dans le cas où l'objectif est d'estimer conjointement $\boldsymbol{\theta}$ et $\boldsymbol{x}$ par inférence sur la loi *a posteriori* jointe $p(\boldsymbol{x}, \boldsymbol{\theta}|\boldsymbol{y})$.

Ce cadre de travail est très général et peut être utilisé dans de nombreuses applications. En particulier, pour des problèmes inverses de reconstruction d'image, il recouvre une majeure partie des techniques actuelles comme l'inversion myope [5] ou non supervisée [3], dès lors que l'on considère dans $\boldsymbol{\theta}$ les paramètres instruments et les hyperparamètres. Ce cadre intègre aussi le cas de modélisations *a priori* non gaussiennes [2, 3, 1] par l'introduction de variables auxiliaires, dès lors qu'elles sont considérées dans $\boldsymbol{\theta}$.

La loi *a posteriori* jointe $p(\boldsymbol{x}, \boldsymbol{\theta}|\boldsymbol{y})$ étant généralement difficile à manipuler, on utilise des approches fondées sur l'échantillonnage de Gibbs qui amènent au problème d'échantillonnage suivant la loi *a posteriori* $p(\boldsymbol{x}|\boldsymbol{\theta}, \boldsymbol{y})$. Selon les hypothèses précédentes, cette loi conditionnelle est gaussienne, dont la matrice de précision dépend de $\boldsymbol{\theta}$ et de l'opérateur linéaire $\boldsymbol{H}$. Lorsque le problème est de grande dimension, la difficulté d'échantillonnage d'une telle loi réside dans la manipulation d'une matrice de covariance de très grande taille. De plus, la matrice de précision *a posteriori* n'est pas creuse car elle dépend de l'opérateur linéaire $\boldsymbol{H}$ qui, dans le cas général, n'a pas de propriété particulière. Dans ce cadre, les méthodes d'échantillonnage classiques [1, 4, 8], basées sur le caractère creux de la matrice de précision, ne sont pas utilisables en pratique. À notre connaissance, les seules autres solutions sont des méthodes d'échantillonnage pixel par pixel ou de type Metropolis-Hastings, très coûteuses en temps de calcul.

Dans ce papier, nous proposons une méthode d'échantillonnage adaptée à la résolution de problèmes inverses généraux, reliée à n'importe quelle forme de l'opérateur linéaire $\boldsymbol{H}$. Le principe général est fondé sur les travaux de [4] et [8] qui soulignent le fait que si une factorisation de la matrice de précision est disponible, alors un échantillon s'obtient par la génération d'un bruit blanc gaussien et la résolution d'un système linéaire. La généralisation de ce principe s'obtient en utilisant un algorithme d'optimisation pour résoudre le système linéaire sans nécessité de caractère creux. L'application d'origine étant la résolution de problèmes inverses en reconstruction d'images, ces travaux ont également été soumis [6] à la conférence ICIP 2010 (International Conference on Image Processing).

2 Algorithme de Perturbation-Optimisation

2.1 Simulation par optimisation

Nous considérons un problème inverse décrit par le modèle direct (1). L'objectif est de résoudre ce problème dans le cadre de l'estimation bayésienne à partir des hypothèses

suivantes :

- le bruit d'observation est gaussien $\boldsymbol{b} \sim \mathcal{N}(0, \boldsymbol{R}_b)$,
- l'opérateur direct $\boldsymbol{H}$ peut dépendre de paramètres instruments,
- la loi *a priori* de $\boldsymbol{x}$ est gaussienne $\boldsymbol{x} \sim \mathcal{N}(\boldsymbol{m}_x, \boldsymbol{R}_x)$ conditionnellement à $\boldsymbol{\theta}$.

Dans le but d'être le plus général possible, nous considérons que $\boldsymbol{\theta}$ regroupe les paramètres instruments, les hyperparamètres et les variables auxiliaires. L'inversion myope et non supervisée consiste alors à estimer $\boldsymbol{x}$ et $\boldsymbol{\theta}$ à partir de la loi $p(\boldsymbol{x}, \boldsymbol{\theta}|\boldsymbol{y})$. Nous supposons que la procédure d'estimation utilisée fait intervenir l'étape d'échantillonnage suivant la loi *a posteriori* conditionnelle $p(\boldsymbol{x}|\boldsymbol{y}, \boldsymbol{\theta})$ qui est un champ gaussien corrélé :

$$p(\boldsymbol{x}|\boldsymbol{y}, \boldsymbol{\theta}) = \mathcal{N}(\boldsymbol{m}_x^{post}, \boldsymbol{R}_x^{post}) \tag{2}$$

avec

$$\boldsymbol{R}_x^{post} = \left(\boldsymbol{H}^t \boldsymbol{R}_b^{-1} \boldsymbol{H} + \boldsymbol{R}_x^{-1}\right)^{-1} \quad \text{et} \quad \boldsymbol{m}_x^{post} = \boldsymbol{R}_x^{post} \left(\boldsymbol{H}^t \boldsymbol{R}_b^{-1} \boldsymbol{y} + \boldsymbol{R}_x^{-1} \boldsymbol{m}_x.\right)$$

Il est connu que $\boldsymbol{m}_x^{post}$ peut être obtenu par minimisation du critère

$$J(\boldsymbol{x}|\boldsymbol{y}, \boldsymbol{m}_x, \boldsymbol{R}_x, \boldsymbol{\theta}) = (\boldsymbol{y} - \boldsymbol{H}\boldsymbol{x})^t \boldsymbol{R}_b^{-1} (\boldsymbol{y} - \boldsymbol{H}\boldsymbol{x}) + (\boldsymbol{x} - \boldsymbol{m}_x)^t \boldsymbol{R}_x^{-1} (\boldsymbol{x} - \boldsymbol{m}_x). \tag{3}$$

Considérons alors des perturbations sur les données $\boldsymbol{y}$ et la moyenne *a priori* $\boldsymbol{m}_x$ à partir des distributions *a priori*

$$\widetilde{\boldsymbol{y}} \sim \mathcal{N}(\boldsymbol{y}, \boldsymbol{R}_b), \tag{4}$$

$$\widetilde{\boldsymbol{m}}_x \sim \mathcal{N}(\boldsymbol{m}_x, \boldsymbol{R}_x). \tag{5}$$

Définissons $\hat{\boldsymbol{x}}$ le minimiseur du critère perturbé $J(\boldsymbol{x}|\widetilde{\boldsymbol{y}}, \widetilde{\boldsymbol{m}}_x, \boldsymbol{R}_x)$ dont nous avons l'expression analytique :

$$\hat{\boldsymbol{x}} = \boldsymbol{R}_x^{post} \left(\boldsymbol{H}^{\mathrm{t}} \boldsymbol{R}_b^{-1} \widetilde{\boldsymbol{y}} + \boldsymbol{R}_x^{-1} \widetilde{\boldsymbol{m}}_x\right). \tag{6}$$

Proposition 2.1 – *Si $\widetilde{\boldsymbol{y}}$ et $\widetilde{\boldsymbol{m}}_x$ sont définis par les équations* (4) *et* (5)*, alors $\hat{\boldsymbol{x}}$ est gaussien de moyenne $\boldsymbol{m}_x^{post}$ et de covariance $\boldsymbol{R}_x^{post}$.*

Preuve – $\hat{\boldsymbol{x}}$ est gaussien comme combinaison linéaire de deux variables gaussiennes.
Moyenne de $\hat{\boldsymbol{x}}$

$$\mathbb{E}[\hat{\boldsymbol{x}}] = \boldsymbol{R}_x^{post} \left(\boldsymbol{H}^{\mathrm{t}} \boldsymbol{R}_b^{-1} \mathbb{E}[\widetilde{\boldsymbol{y}}] + \boldsymbol{R}_x^{-1} \mathbb{E}[\widetilde{\boldsymbol{m}}_x]\right) = \boldsymbol{m}_x^{post}. \tag{7}$$

Covariance de $\hat{\boldsymbol{x}}$

$$\begin{aligned}
\mathbb{E}[\hat{\boldsymbol{x}}\hat{\boldsymbol{x}}^{\mathrm{t}}] &= \boldsymbol{R}_x^{post} \mathbb{E}\Big[\left(\boldsymbol{H}^{\mathrm{t}} \boldsymbol{R}_b^{-1} \widetilde{\boldsymbol{y}} + \boldsymbol{R}_x^{-1} \widetilde{\boldsymbol{m}}_x\right) \left(\boldsymbol{H}^{\mathrm{t}} \boldsymbol{R}_b^{-1} \widetilde{\boldsymbol{y}} + \boldsymbol{R}_x^{-1} \widetilde{\boldsymbol{m}}_x\right)^{\mathrm{t}} \Big] \boldsymbol{R}_x^{post} \\
&= \boldsymbol{R}_x^{post} \Big[\boldsymbol{H}^{\mathrm{t}} \boldsymbol{R}_b^{-1} (\boldsymbol{y}\boldsymbol{y}^{\mathrm{t}} + \boldsymbol{R}_b) \boldsymbol{R}_b^{-1} \boldsymbol{H} + \boldsymbol{R}_x^{-1} (\boldsymbol{m}_x \boldsymbol{m}_x{}^{\mathrm{t}} + \boldsymbol{R}_x) \boldsymbol{R}_x^{-1} \Big] \boldsymbol{R}_x^{post} \\
&= \mathbb{E}\left[\hat{\boldsymbol{x}}\right] \mathbb{E}\left[\hat{\boldsymbol{x}}\right]^{\mathrm{t}} + \boldsymbol{R}_x^{post}.
\end{aligned}$$

On obtient :

$$\mathbb{V}[\hat{\boldsymbol{x}}] = \boldsymbol{R}_x^{post}.$$

2.2 Algorithme proposé

À partir du résultat précédent, on en déduit une méthode d'échantillonnage de champs gaussiens définis par (2) en deux étapes.

Étape P (perturbation) : perturber $\boldsymbol{y}$ et $\boldsymbol{m}_x$ suivant (4) et (5).

Étape O (optimisation) : minimiser le critère perturbé $J(\boldsymbol{x}|\widetilde{\boldsymbol{y}}, \widetilde{\boldsymbol{m}}_x, \boldsymbol{R}_x, \boldsymbol{\theta})$.

Le critère perturbé étant quadratique, on a accès à toute la littérature sur les algorithmes d'optimisation numérique. Notons qu'en théorie l'échantillon exact de la loi *a posteriori* est l'optimiseur du critère perturbé. Par conséquent, l'étape d'optimisation peut nécessiter autant d'étapes de descentes que la dimension du problème. Cependant, en pratique, la procédure d'optimisation peut être arrété plus rapidement sans conséquence visible.

Dans cette section, nous nous plaçons dans le cas de l'échantillonnage de champs gaussiens *a posteriori*. En fait, ce n'est pas l'unique application et nous pouvons effectuer quelques remarques. Premièrement, l'algorithme proposé est applicable pour tout champ *a priori* conditionnellement gaussien, dont la matrice de précision est de la forme $\boldsymbol{Q} = \sum_{k=1}^{m} \boldsymbol{M}_k \boldsymbol{R}_k^{-1} \boldsymbol{M}_k^{\mathrm{t}}$ et dès que des échantillons gaussiens de matrice de covariance $\boldsymbol{R}_k$ sont faciles à obtenir. De plus, dans un échantillonneur de Gibbs, le champ conditionnel peut ne pas être invariant puisque $\boldsymbol{H}$ peut dépendre de la valeur courante de $\boldsymbol{\theta}$, et la procédure d'optimisation reste valide. Par le même raisonnement, l'algorithme reste applicable dans le cas de champs gaussiens non stationnaires ou non homogènes.

3 Illustration

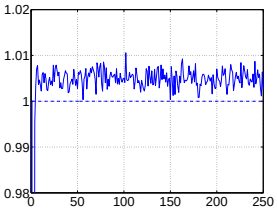
L'algorithme de perturbation-optimisation a déjà été utilisé [1, 4, 5] pour résoudre des problèmes inverses de reconstruction d'image, avec des modélisation *a priori* gaussiennes [5] ou gaussiennes conditionnellement à des variables auxiliaires [1]. Dans la suite, nous proposons une illustration sur un problème de super résolution. Dans ce problème plusieurs images floues et sous-échantillonnées, issues d'une même image originale, sont disponibles. L'idée principale repose sur l'utilisation des petits décalages entre les images à faible résolution pour reconstruire l'image d'origine avec une meilleure résolution. C'est un problème mal posé, l'approche générale consiste à introduire de la régularisation [7] avec des hyperparamètres fixés. Nous illustrons ici comment utiliser l'approche décrite précédemment pour estimer l'hyperparamètre dans un cadre bayésien.

Le modèle direct s'écrit $\boldsymbol{y} = \boldsymbol{P}\boldsymbol{H}\boldsymbol{x} + \boldsymbol{b}$, où $\boldsymbol{y} \in \mathbb{R}^M$ sont les images de basse résolution (128×128 pixels, 40 images sont utilisées), $\boldsymbol{H}$ est une matrice de convolution, $\boldsymbol{P}$ une matrice de sous-échantillonnage, $\boldsymbol{x} \in \mathbb{R}^N$ l'image originale (279×279 pixels) et $\boldsymbol{b}$ le bruit de loi $\mathcal{N}(\boldsymbol{0}, \gamma_n^{-1}\boldsymbol{I})$. La loi *a priori* de $\boldsymbol{x}$ est $\mathcal{N}(\boldsymbol{0}, \gamma_x^{-1}\boldsymbol{D}^t\boldsymbol{D})$ où $\boldsymbol{D}$ est le laplacien. Les deux paramètres γ_n et γ_x sont considérés *inconnus*. Ils sont estimés conjointement avec l'image $\boldsymbol{x}$ et leur loi *a priori* est une loi de Jeffreys. La loi *a posteriori* est alors :

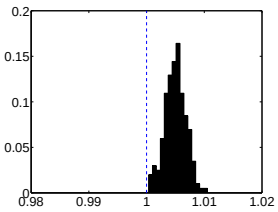
$$p(\boldsymbol{x}, \gamma_n, \gamma_x) \propto \gamma_n^{M/2-1} \gamma_x^{(N-1)/2-1} \exp\left[-\frac{\gamma_n}{2}\|\boldsymbol{y} - \boldsymbol{P}\boldsymbol{H}\boldsymbol{x}\|^2\right] \exp\left[-\frac{\gamma_x}{2}\|\boldsymbol{D}\boldsymbol{x}\|^2\right]. \quad (8)$$

Notons que la loi *a posteriori* conditionnelle de $\boldsymbol{x}$ est gaussienne dont la matrice de

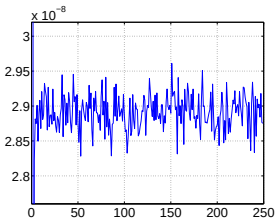
(a) Vraie image (b) Une image basse résolution (c) Image estimée



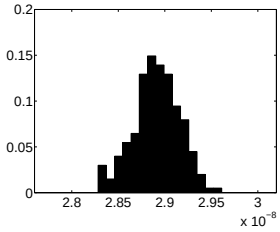
(d) γ_n



(e) γ_n



(f) γ_x



(g) γ_x

FIGURE 1 – Super-résolution non supervisée. Les figures (a), (b) et (c) sont respectivement la vraie image, une image basse résolution et l'image estimée. Les figuress (d) et (e) sont respectivement la chaîne de Markov et l'histogramme de γ_n. La ligne pointillée correspond à la vraie valeur. Les figures (f) et (g) sont respectivement la chaîne de Markov et l'histogramme de γ_x.

covariance n'est pas invariante à cause de $\boldsymbol{P}$.

La moyenne *a posteriori* est calculée par un échantillonneur de Gibbs pour $\boldsymbol{x}$, γ_n et γ_x récursivement dont le schéma est décrit par l'algorithme 1. Comme $\boldsymbol{P}$ n'est pas invariant et $\boldsymbol{H}$ non creux, la matrice $\boldsymbol{R}_x^{post}$ ne peut pas être calculée ou factorisée en un temps raisonnable. Par conséquent, l'échantillon pour $\boldsymbol{x}$ est donné par l'algorithme de perturbation-optimisation décrit en section 2 (avec un algorithme de gradient conjugué).

Les résultats sont présentés dans la figure 1. L'image 1(c) montre clairement que l'on obtient une image de plus haute résolution que la donnée initiale (Fig. 1(b)). Les paramètres de régularisation sont estimés automatiquement et les moyennes *a posteriori* obtenues sont $\widehat{\gamma}_n = 1.005$ avec une vraie valeur à $\gamma_n^* = 1$ et $\widehat{\gamma}_x = 2.8911 \times 10^{-8}$. Par conséquent, notre approche d'échantillonnage de champs gaussiens non stationnaire permet de proposer une méthode de super-résolution non supervisée.

4 Conclusion

Ce papier présente une nouvelle approche pour l'échantillonnage de champs gaussiens dans le cas où les approches classiques ne sont pas utilisables : lorsque la dimension du problème est très grande et lorsque la matrice de précision n'est pas creuse. Le champ d'application est très large, et recouvre en particulier les problèmes inverses de reconstruction d'image dont le modèle direct est linéaire et avec des lois *a priori* conditionnellement gaus-

Algorithme 1 –Échantillonneur de Gibbs pour l'inversion non supervisée.

1: Initialisation avec $k = 1$ et $\boldsymbol{x}^{(0)} = \boldsymbol{x}_0$.

2: Tirage de $\gamma_n^{(k)} \sim \mathcal{G}\left(1 + M/2, 2/\left\|\boldsymbol{y} - \boldsymbol{P}\boldsymbol{H}\boldsymbol{x}^{(k-1)}\right\|^2\right)$

3: Tirage de $\gamma_x^{(k)} \sim \mathcal{G}\left(1 + (N-1)/2, 2/\left\|\boldsymbol{D}\boldsymbol{x}^{(k-1)}\right\|^2\right)$

4: Tirage $\boldsymbol{x}^{(k)} \sim \mathcal{N}(\boldsymbol{m}_x^{post}, \boldsymbol{R}_x^{post})$ par optimisation où

$$\boldsymbol{R}_x^{post} = \left(\gamma_n^{(k)}\boldsymbol{H}^t\boldsymbol{P}^t\boldsymbol{P}\boldsymbol{H} + \gamma_x^{(k)}\boldsymbol{D}^t\boldsymbol{D}\right)^{-1} \quad \text{et} \quad \boldsymbol{m}_x^{post} = \gamma_n^{(k)}\boldsymbol{R}_x^{post}\boldsymbol{P}^t\boldsymbol{H}^t\boldsymbol{y}.$$

5: $k = k + 1$

6: Retour en 1 ou arrêt si respect d'un critère d'arrêt.

siennes. Nous illustrons l'efficacité de la méthode d'échantillonnage sur un problème de super résolution avec l'estimation des hyperparamètres par un échantillonneur de Gibbs.

5 Remerciements

Les auteurs voudraient remercier Jérôme Idier, Thomas Rodet et Ali Mohammad-Djafari pour toutes les fructueuses discussions autour de l'échantillonnage.

Références

[1] O. Féron, B. Duchêne, and A. Mohammad-Djafari. Microwave imaging of piecewise constant objects in a 2D-TE configuration. *Intern. Journ. App. Electr. Mec.*, 26(3-4), 2007.

[2] D. Geman and C. Yang. Nonlinear image recovery with half-quadratic regularization. *IEEE Trans. Image Proc.*, 4(7), July 1995.

[3] J.F. Giovannelli. Unsupervised bayesian convex deconvolution based on a field with an explicit partition function. *IEEE Trans. Im. Proc.*, 17(1), January 2008.

[4] P. Lalanne, D. Prévost, and P. Chavel. Stochastic artificial retinas : algorithm, optoelectronic circuits, and implementation. *Appl. Opt*, 40, 2001.

[5] F. Orieux. *Inversion bayésienne myope et non-supervisée pour l'imagerie sur-résolue. Application à l'instrument SPIRE de l'observatoire spatial Herschel.* PhD thesis, Université Paris-Sud 11, November 2009.

[6] F. Orieux, O. Féron, and J.-F. Giovannelli. Stochastic sampling of large dimension non-stationnary Gaussian field for image restoration. In *soumis à Int. Conf. Im. Proc.*, 2010.

[7] S. C. Park, M. K. Park, and M. G. Kang. Super-resolution image reconstruction : a technical overview. *IEEE Sig. Proc. Mag.*, May 2003.

[8] H. Rue. Fast sampling of Gaussian Markov random fields. *J. R. Statist. Soc B*, 63(2), 2001.