

Rééchantillonnage gaussien en grande dimension pour les problèmes inverses

Gilavert Clément

Mél : clement.gilavert@irccyn.ec-nantes.fr

Résumé : Nous nous intéressons aux algorithmes MCMC dans le cas où une des composantes à rééchantillonner est un vecteur gaussien de grande dimension dont la covariance varie au cours des itérations (du fait de sa dépendance à des hyperparamètres eux-même rééchantillonnés par exemple). Pour éviter le coût de calcul d'une factorisation de type Cholesky à chaque itération, une solution récente propose de passer par la résolution d'un système linéaire de grande dimension, effectuée de façon approchée par gradient conjugué tronqué. Notre contribution consiste (1) à mettre en évidence que cette troncature empêche la convergence vers la loi cible ; (2) à proposer l'incorporation d'une étape d'acceptation-rejet peu coûteuse rétablissant la convergence. Nous illustrons nos résultats sur un problème de super-résolution en imagerie. Ils mettent en avant de façon concrète le biais engendré par la version sans étape d'acceptation-rejet, ainsi que le degré de troncature possible grâce à l'algorithme proposé.

Mots clés : *MCMC, échantillonnage gaussien en grande dimension, problème inverse, déconvolution.*

1 Introduction

En inférence bayésienne, il est fréquent qu'une partie des composantes à rééchantillonner correspondent à un vecteur ou à une image de grande dimension et de loi conditionnelle gaussienne. C'est souvent le cas par exemple en restauration et en reconstruction de signaux et d'images. Dans la suite, nous considérons en particulier une structure de problème inverse linéaire donné par le modèle d'observation $\mathbf{y} = \mathbf{M}\mathbf{x} + \mathbf{b}$ avec \mathbf{y} la mesure, \mathbf{M} la matrice d'observation, \mathbf{b} le bruit additif et $\mathbf{x} \in \mathbb{R}^N$ le vecteur d'intérêt. Supposons des modèles gaussiens $\mathcal{N}(\boldsymbol{\mu}_b, \mathbf{R}_b)$ et $\mathcal{N}(\boldsymbol{\mu}_x, \mathbf{R}_x)$ pour le bruit et la variable d'intérêt. Notons $\boldsymbol{\Theta}$ l'ensemble des hyperparamètres dont dépendent $\{\boldsymbol{\mu}_n, \mathbf{R}_n, \boldsymbol{\mu}_x, \mathbf{R}_x\}$. Le problème d'inférence bayésienne considéré ici consiste à estimer à la fois \mathbf{x} et $\boldsymbol{\Theta}$. Pour cela une approche de type MCMC est souvent utilisée pour tirer des échantillons de la loi *a posteriori* $P(\mathbf{x}, \boldsymbol{\Theta} | \mathbf{y})$ afin d'en extraire des estimateurs. Une structure de Gibbs par bloc est souvent utilisée pour le rééchantillonnage. A chaque itération k :

1. Tirer $\boldsymbol{\Theta}^{(k)}$ suivant la loi conditionnelle $P(\boldsymbol{\Theta} | \mathbf{x}^{(k-1)}, \mathbf{y})$.
2. Tirer $\mathbf{x}^{(k)}$ suivant la loi conditionnelle $P(\mathbf{x} | \boldsymbol{\Theta}^{(k)}, \mathbf{y})$.

On montre facilement que la loi conditionnelle $P(\mathbf{x} | \boldsymbol{\Theta}, \mathbf{y})$ est une loi gaussienne dont la matrice de précision (inverse de la matrice de covariance) et la moyenne sont données par

$$\mathbf{Q} = \mathbf{M}^t \mathbf{R}_b^{-1} \mathbf{M} + \mathbf{R}_x^{-1} \quad \text{et} \quad \mathbf{Q}\boldsymbol{\mu} = \mathbf{M}^t \mathbf{R}_b^{-1} (\mathbf{y} - \boldsymbol{\mu}_b) + \mathbf{R}_x^{-1} \boldsymbol{\mu}_x \quad (1)$$

En grande dimension, deux difficultés apparaissent dans l'étape 2 du rééchantillonnage : la résolution du système pour calculer $\boldsymbol{\mu}$, puis l'échantillonnage de la gaussienne de matrice de précision \mathbf{Q} .

L'approche la plus classique consiste à inverser la matrice de précision $\mathbf{R} = \mathbf{Q}^{-1}$, à calculer sa factorisation de Cholesky $\mathbf{R} = \mathbf{L}_r \mathbf{L}_r^t$, puis à l'utiliser pour générer le nouvel échantillon $\mathbf{x} = \boldsymbol{\mu} + \mathbf{L}_r \boldsymbol{\omega}$, avec $\boldsymbol{\omega} \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$. Mais cette approche nécessite $O(N^3)$ opérations si \mathbf{R} n'a pas de structure spécifique, ce qui la rend infaisable en grande dimension. Des solutions exactes existent si la matrice \mathbf{Q} est creuse [1, 2] ou circulante [3]. Dans le premier cas, la décomposition de Cholesky devient possible et s'applique directement à la matrice \mathbf{Q} . Dans le second, les opérations se passent dans le domaine fréquentiel à l'aide de transformées de Fourier.

Une autre approche appelée "perturbation gaussienne" [4] consiste à simuler un échantillon $\boldsymbol{\eta} \sim \mathcal{N}(\mathbf{Q}\boldsymbol{\mu}, \mathbf{Q})$, puis à résoudre le système $\mathbf{Q}\mathbf{x} = \boldsymbol{\eta}$. Il est facile de vérifier que la loi de \mathbf{x} est alors bien $\mathcal{N}(\boldsymbol{\mu}, \mathbf{Q}^{-1})$. Dans le cadre étudié ici, le tirage de $\boldsymbol{\eta}$ peut se faire rapidement de la façon suivante :

$$\boldsymbol{\eta} = \mathbf{M}^t \mathbf{R}_b^{-1} \boldsymbol{\eta}_b + \mathbf{R}_x^{-1} \boldsymbol{\eta}_x, \quad \text{avec} \quad \boldsymbol{\eta}_b \sim \mathcal{N}(\mathbf{y} - \boldsymbol{\mu}_b, \mathbf{R}_b) \quad \text{et} \quad \boldsymbol{\eta}_x \sim \mathcal{N}(\boldsymbol{\mu}_x, \mathbf{R}_x) \quad (2)$$

Notons que (2) est peu coûteuse si \mathbf{R}_b et \mathbf{R}_x ont des structures simples, en revanche le coût de l'inversion du système linéaire $\mathbf{Q}\mathbf{x} = \boldsymbol{\eta}$ reste $O(N^3)$ opérations. C'est pourquoi plusieurs travaux [4, 5, 6] préconisent de résoudre ce système par gradient conjugué en effectuant un nombre réduit $K \ll N$ d'itérations. Cependant, l'effet de cette troncature en termes de convergence de la chaîne de Markov n'a pas été étudié pour l'instant.

Notre principale contribution est une modification de l'algorithme de perturbation gaussienne tronquée : nous proposons d'utiliser cet algorithme pour proposer un candidat, puis d'accepter cette proposition selon une étape d'acceptation-rejet dont le ratio est calculé d'après la théorie des sauts réversibles [7]. Cette étape permet de garantir la convergence vers la loi cible, quel que soit le degré de troncature dans la résolution du système. De plus, le calcul du taux d'acceptation étant peu coûteux, celui-ci peut être utilisé pour contrôler le niveau de troncature, c'est-à-dire arrêter la résolution du système quand le taux d'acceptation atteint un seuil prédéfini.

La méthode proposée est illustrée sur un problème de super-résolution en imagerie. Les résultats mettent en avant le biais engendré par la version sans étape d'acceptation-rejet, ainsi que le degré de troncature possible grâce à l'algorithme proposé.

2 Formulation à l'aide des MCMC à sauts réversibles

2.1 Cadre général et cas particulier de la *perturbation gaussienne*

Afin de définir un schéma convergent, nous nous plaçons dans le cadre des MCMC à sauts réversibles [7], en dimension constante. Cette approche s'appuie sur la simulation d'une variable auxiliaire $\mathbf{z} \sim P_{\mathbf{Z}}(\mathbf{z}|\mathbf{x}^-)$, où \mathbf{x}^- désigne l'échantillon précédent, et l'application d'une transformation du couple $(\mathbf{x}^-, \mathbf{z})$ pour produire un couple de candidats $(\mathbf{x}, \mathbf{s}) = \phi(\mathbf{x}^-, \mathbf{z})$ ². Cette transformation déterministe doit également être réversible, c'est-à-dire $\phi(\mathbf{x}, \mathbf{s}) = (\mathbf{x}^-, \mathbf{z})$. Ainsi, le candidat \mathbf{x} issu de la transformation est soumis à une règle d'acceptation-rejet, pour définir le nouvel échantillon \mathbf{x}^+ , avec un taux d'acceptation donné par :

$$\alpha = \min \left(1, \frac{P_{\mathbf{X}}(\mathbf{x})P_{\mathbf{Z}}(\mathbf{s}|\mathbf{x})}{P_{\mathbf{X}}(\mathbf{x}^-)P_{\mathbf{Z}}(\mathbf{z}|\mathbf{x}^-)} |\Phi| \right) \quad (3)$$

avec Φ le jacobien de la transformation ϕ . Ce cadre général des MCMC à sauts réversibles laisse donc comme degrés de liberté le choix de la distribution $P_{\mathbf{Z}}$ de la variable auxiliaire ainsi que la transformation déterministe ϕ .

Pour l'échantillonnage d'un vecteur gaussien $\mathbf{x} \sim \mathcal{N}(\boldsymbol{\mu}, \mathbf{Q}^{-1})$, nous proposons une généralisation de l'approche introduite dans [8], en simulant \mathbf{z} suivant une loi normale $\mathcal{N}(\mathbf{A}\mathbf{x}^- + \mathbf{c}, \mathbf{B})$, et nous optons pour une transformation déterministe ϕ définie par

$$\mathbf{x} = \boldsymbol{\varphi}_1(\mathbf{x}^-, \mathbf{z}) = -\mathbf{x}^- + \mathbf{f}(\mathbf{z}) \quad \text{et} \quad \mathbf{s} = \boldsymbol{\varphi}_2(\mathbf{x}^-, \mathbf{z}) = \mathbf{z}. \quad (4)$$

Le jacobien de cette transformation valant toujours -1, le taux d'acceptation est donné par

$$\alpha = \min \left(1, e^{-\frac{1}{2}\Delta S} \right)$$

avec

$$\Delta S = (\mathbf{x} - \mathbf{x}^-)^\top \left[(\mathbf{Q} + \mathbf{A}^\top \mathbf{B}^{-1} \mathbf{A}) \mathbf{f}(\mathbf{z}) - 2(\mathbf{Q}\boldsymbol{\mu} + \mathbf{A}^\top \mathbf{B}^{-1}(\mathbf{z} - \mathbf{c})) \right] \quad (5)$$

Le choix des paramètres \mathbf{A} , \mathbf{B} , \mathbf{c} et de la fonction $\mathbf{f}(\cdot)$ dépend des caractéristiques de la loi gaussienne d'intérêt et des performances attendues en termes de taux d'acceptation et de corrélation entre échantillons successifs. On montre que :

C1. Si $\mathbf{f}(\mathbf{z}) = 2(\mathbf{Q} + \mathbf{A}^\top \mathbf{B}^{-1} \mathbf{A})^{-1}(\mathbf{Q}\boldsymbol{\mu} + \mathbf{A}^\top \mathbf{B}^{-1}(\mathbf{z} - \mathbf{c}))$ alors le taux d'acceptation est unitaire.

C2. Si, de plus, $\mathbf{A}^\top \mathbf{B}^{-1} \mathbf{A} = \mathbf{Q}$, alors les deux échantillons successifs seront totalement indépendants.

Le choix $\mathbf{A} = \mathbf{B} = \mathbf{Q}$, $\mathbf{c} = \mathbf{Q}\boldsymbol{\mu}$ permet de respecter ces deux propriétés et conduit à une fonction $\mathbf{f}(\mathbf{z}) = \mathbf{Q}^{-1}\mathbf{z}$ qui correspond à la résolution d'un système linéaire $\mathbf{Q}\mathbf{u} = \mathbf{z}$. Le schéma de simulation résultant est :

1. Tirer la variable auxiliaire $\mathbf{z} \sim \mathcal{N}(\mathbf{Q}\mathbf{x}^- + \mathbf{Q}\boldsymbol{\mu}, \mathbf{Q})$.
 2. Calculer $\mathbf{u}^* = \mathbf{f}(\mathbf{z})$ par résolution du système $\mathbf{Q}\mathbf{u} = \mathbf{z}$.
 3. Mettre à jour l'échantillon selon $\mathbf{x}^+ = -\mathbf{x}^- + \mathbf{u}^*$.
- (6)

2. Notons que dans notre cas, la variable \mathbf{s} ne sera pas utilisée.

Notons que ce schéma est strictement équivalent à l'algorithme de *perturbation gaussienne* introduit en section 1. En effet, comme $\mathbf{z} = \mathbf{Q}\mathbf{x}^- + \boldsymbol{\eta}$, avec $\boldsymbol{\eta} \sim \mathcal{N}(\mathbf{Q}\boldsymbol{\mu}, \mathbf{Q})$, alors $\mathbf{x}^+ = -\mathbf{x}^- + \mathbf{Q}^{-1}(\mathbf{Q}\mathbf{x}^- + \boldsymbol{\eta}) = \mathbf{Q}^{-1}\boldsymbol{\eta}$. Ce schéma se résume à simuler la variable $\boldsymbol{\eta}$ et à retenir $\mathbf{x}^+ = \mathbf{x}^*$, solution exacte de $\mathbf{Q}\mathbf{x} = \boldsymbol{\eta}$.

2.2 Algorithme de perturbation gaussienne tronquée

Retenir une solution approchée $\tilde{\mathbf{u}}$ du système linéaire qui apparaît dans l'étape 2 de (6) correspond à l'utilisation d'une fonction $\mathbf{f}(\mathbf{z})$ qui ne respecte pas la condition C1 énoncée précédemment. On montre que le taux d'acceptation dépend du résidu $\mathbf{r} = \mathbf{z} - \mathbf{Q}\tilde{\mathbf{u}}$ de la résolution du système selon $\alpha = \min(1, e^{-\mathbf{r}^\top(2\mathbf{x}^- - \tilde{\mathbf{u}})})$.

Finalement, en utilisant un gradient conjugué tronqué initialisé à $\mathbf{u}_0 = -\mathbf{x}^-$ pour résoudre ce système, le schéma de simulation que nous proposons se réduit à :

1. Tirer la variable auxiliaire $\boldsymbol{\eta} \sim \mathcal{N}(\mathbf{Q}\boldsymbol{\mu}, \mathbf{Q})$.
2. Calculer $\tilde{\mathbf{x}}$, solution approchée du système $\mathbf{Q}\mathbf{x} = \boldsymbol{\eta}$.
3. Calculer le taux d'acceptation $\alpha = \min(1, e^{-\mathbf{r}^\top(\mathbf{x}^- - \tilde{\mathbf{x}})})$.
4. Poser $\mathbf{x}^+ = \tilde{\mathbf{x}}$ avec la probabilité α , sinon rejeter $\tilde{\mathbf{x}}$ en posant $\mathbf{x}^+ = \mathbf{x}^-$.

De plus, le calcul du taux d'acceptation n'étant pas coûteux, le niveau de troncature est ajusté directement en fonction de celui-ci et non en fonction du résidu relatif.

3 Application : super-résolution non supervisée

Nous reprenons pour illustration le problème de la super-résolution non supervisée en imagerie, traité dans [6]. Le but est d'obtenir une image nette de résolution supérieure à partir de plusieurs images floues de basse résolution. Le modèle d'observation de ce problème inverse linéaire est donné par $\mathbf{y} = \mathbf{M}\mathbf{x} + \mathbf{b} = \mathbf{P}\mathbf{H}\mathbf{x} + \mathbf{b}$, avec $\mathbf{y} \in \mathbb{R}^M$ le vecteur des images observées, $\mathbf{x} \in \mathbb{R}^N$ l'image nette haute résolution à récupérer, \mathbf{H} la matrice $N \times N$ de convolution associée à une réponse impulsionnelle laplacienne dont la largeur à mi-hauteur vaut 4 pixels, \mathbf{P} la matrice $M \times N$ de décimation et \mathbf{b} le bruit additif. Dans notre cas, nous observons cinq images de 128×128 pixels ($M = 81920$) et nous recherchons une image de 256×256 pixels ($N = 65536$). Le bruit est modélisé par une loi $\mathcal{N}(\mathbf{0}, \gamma_b^{-1}\mathbf{I})$ et l'image \mathbf{x} par une distribution $\mathcal{N}(\mathbf{0}, \gamma_x^{-1}\mathbf{D}^\top\mathbf{D})$, avec \mathbf{D} la matrice de convolution associée au filtre laplacien. Pour modéliser les deux hyper-paramètres, on utilise des distributions de Jeffreys, non informatives. On cherche à estimer l'image \mathbf{x} et les deux hyperparamètres γ_b et γ_x à partir de la distribution à *postérieure* $P(\mathbf{x}, \gamma_x, \gamma_n | \mathbf{y})$ obtenue grâce à la formule de Bayes

$$P(\mathbf{x}, \gamma_x, \gamma_n | \mathbf{y}) \propto \gamma_x^{(N-1)/2-1} \gamma_n^{M/2-1} e^{-\frac{1}{2}\gamma_n(\mathbf{y}-\mathbf{P}\mathbf{H}\mathbf{x})^\top(\mathbf{y}-\mathbf{P}\mathbf{H}\mathbf{x}) - \frac{1}{2}\gamma_x\mathbf{x}^\top\mathbf{D}^\top\mathbf{D}\mathbf{x}} \quad (8)$$

Pour cela, un échantillonneur de gibbs est utilisé pour explorer cette distribution, il consiste à rééchantillonner successivement les distributions conditionnelles, à chaque itération k :

- Tirer $\gamma_x^{(k)} \sim P(\gamma_x^{(k)} | \gamma_n^{(k-1)}, \mathbf{x}^{(k-1)}, \mathbf{y}) = \mathcal{G}(1 + \frac{N-1}{2}, 2/\|\mathbf{D}\mathbf{x}^{(k-1)}\|^2)$
- Tirer $\gamma_n^{(k)} \sim P(\gamma_n^{(k)} | \gamma_x^{(k)}, \mathbf{x}^{(k-1)}, \mathbf{y}) = \mathcal{G}(1 + \frac{M}{2}, 2/\|\mathbf{y} - \mathbf{P}\mathbf{H}\mathbf{x}^{(k-1)}\|^2)$
- Tirer $\mathbf{x}^{(k)} \sim P(\mathbf{x}^{(k)} | \gamma_x^{(k)}, \gamma_n^{(k)}, \mathbf{y}) = \mathcal{N}(\boldsymbol{\mu}^{(k)}, [\mathbf{Q}^{(k)}]^{-1})$

avec

$$\mathbf{Q}^{(k)} = \gamma_n^{(k)} \mathbf{H}^\top \mathbf{P}^\top \mathbf{P} \mathbf{H} + \gamma_x^{(k)} \mathbf{D}^\top \mathbf{D} \quad \text{et} \quad \mathbf{Q}^{(k)} \boldsymbol{\mu}^{(k)} = \gamma_n^{(k)} \mathbf{P}^\top \mathbf{H}^\top \mathbf{y} \quad (9)$$

Cette dernière étape est réalisée par la *perturbation gaussienne* (PG) contrôlée classiquement par le résidu relatif, par l'algorithme proposé ici (PGT) directement contrôlé par le taux d'acceptation et par la factorisation de Cholesky comme référence. 1000 échantillons sont générés avec un temps de chauffe de 100 ; on observe les moyennes et écarts-types, obtenus par les approximations classiques de Monte-Carlo, des deux hyperparamètres et d'un pixel de l'image reconstruite, présents dans le tableau 1. La figure 2 montre l'image d'origine, un exemple d'observation basse résolution et l'image reconstruite par PGT avec un taux d'acceptation minimum de 0.1. Les résultats montrent le biais induit par la méthode PG même en réduisant l'erreur relative de résolution du système jusqu'à 10^{-8} . Les estimateurs obtenus par PGT sont, quant à eux, très fidèles à ceux obtenus par une approche de type Cholesky, que ce soit en terme de moyenne ou d'écart-type. Par ailleurs, la figure 1 montre le taux d'acceptation moyen en fonction du nombre d'itérations du gradient conjugué. On voit qu'il est possible d'échantillonner parfaitement avec le PGT en effectuant moins de 700 itérations du gradient conjugué, ce qui est très inférieur à $N = 65536$.

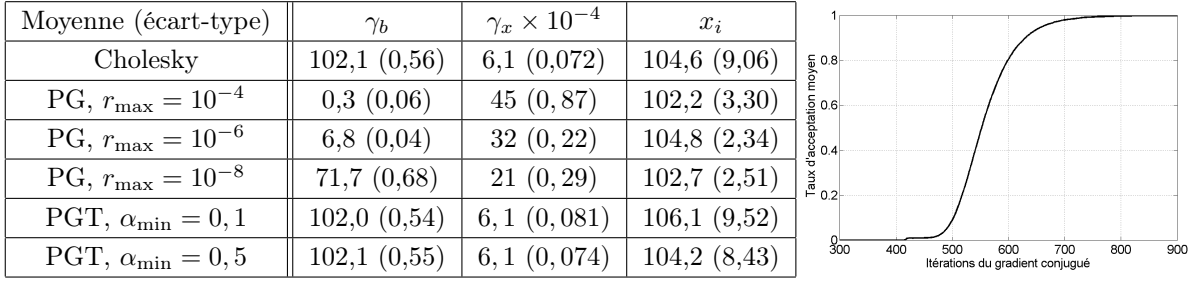


FIGURE 1 – (Gauche) comparaison des méthodes en terme de moyenne et d'écart-type des paramètres estimés ; (Droite) taux d'acceptation moyen en fonction du nombre d'itérations du gradient conjugué.

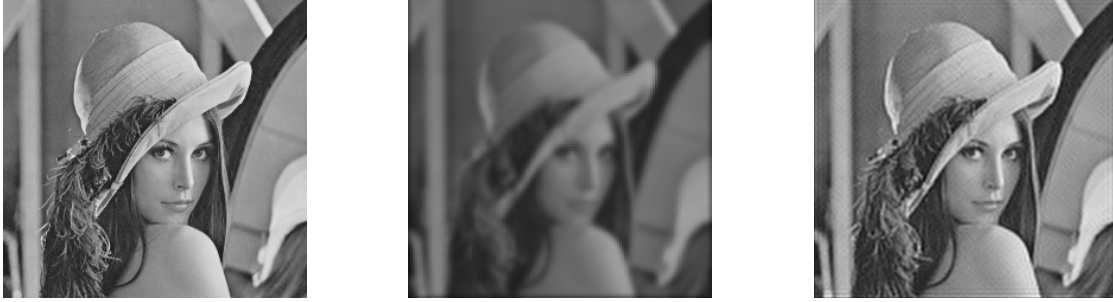


FIGURE 2 – (Gauche) image d'origine recherchée ; (Milieu) une image basse résolution observée ; (Droite) image reconstruite par PGT avec $\alpha_{\min} = 0.1$.

4 Conclusion

Nous avons présenté un algorithme d'échantillonnage de distribution gaussienne dans un contexte de résolution de problèmes inverses de grande taille par des méthodes d'inférence, jusqu'alors contraints à des temps de calcul rédhitoire. Il repose sur la théorie des sauts réversibles et peut être vu comme une version tronquée de l'algorithme de *perturbation gaussienne* dont la convergence est assurée par une étape d'acceptation-rejet.

Il peut être appliqué à une grande variété de problèmes inverses où le bruit et l'objet sont modélisés soit directement par des distributions gaussiennes, soit par des modèles plus complexes tel que les mélanges d'échelle gaussien [9].

Références

- [1] H. Rue. Fast sampling of Gaussian Markov random fields. *J. R. Statist. Soc. B*, 63(2) :325–338, 2001.
- [2] P. Lalanne, D. Prévost, and P. Chavel. Stochastic artificial retinas : algorithm, optoelectronic circuits, and implementation. *Appl. Opt.*, 40(23) :3861–3876, 2001.
- [3] D. Geman and C. Yang. Nonlinear image recovery with half-quadratic regularization. *IEEE Trans. Image Processing*, 4(7) :932–946, 1995.
- [4] G. Papandreou and A. Yuille. Gaussian sampling by local perturbations. In *Proc. NIPS*, 2010.
- [5] X. Tan, J. Li, and P. Stoica. Efficient sparse Bayesian learning via Gibbs sampling. In *Proc. IEEE ICASSP*, pages 3634–3637, 2010.
- [6] F. Orieux, O. Féron, and JF Giovannelli. Sampling high-dimensional Gaussian distributions for general linear inverse problems. *IEEE Signal Processing Lett.*, 19(5) :251, 2012.
- [7] R. Waagepetersen and D. Sorensen. A tutorial on reversible jump MCMC with a view toward applications in qtl-mapping. *Int. Statist. Rev.*, 69(1) :49–61, 2001.
- [8] P. De Forcrand. Monte Carlo quasi-heatbath by approximate inversion. *Phys. Rev. E*, 59(3) :3698–3701, 1999.
- [9] D.F. Andrews and C.L. Mallows. Scale mixtures of normal distributions. *J. R. Statist. Soc. B*, pages 99–102, 1974.