

# Reconstruction de signaux parcimonieux à l'aide d'un algorithme rapide d'échantillonnage stochastique

Yosra MARNISSI<sup>1</sup>, Emilie CHOUZENOUX<sup>1</sup>, Amel BENAZZA-BENYAHIA<sup>2</sup>, Jean-Christophe PESQUET<sup>1</sup>, Laurent DUVAL<sup>3</sup>

<sup>1</sup>LIGM, Université Paris-Est Marne-la-Vallée  
Cité Descartes, 5 boulevard Descartes, Champs-sur-Marne, 77454 Marne-la-Vallée Cedex 2, France

<sup>2</sup>COSIM SUP'COM, Université de Carthage,  
Cité Technologique des Communications, 2080, Tunisie

<sup>3</sup>IFP Energies nouvelles,  
1 et 4 avenue de Bois-Préau, 92852 Rueil-Malmaison, France

yosra.marnissi, emilie.chouzenoux, jean-christophe.pesquet@u-pem.fr  
benazza.amel@supcom.rnu.tn  
Laurent.duval@ifpen.fr

**Résumé** – Ce travail s'intéresse à la reconstruction d'un signal parcimonieux dégradé par un opérateur linéaire et perturbé par un bruit additif gaussien. Nous considérons une approche bayésienne où les coefficients du signal sont supposés mutuellement indépendants et identiquement distribués selon une loi de Student de paramètres inconnus. Nous nous proposons de calculer l'estimateur de la moyenne *a posteriori* du signal ainsi que des hyperparamètres de régularisation par un échantillonneur hybride de Gibbs. Pour échantillonner selon la loi *a posteriori* du signal, nous considérons une version accélérée de l'algorithme MALA (*Metropolis-Adjusted Langevin Algorithm*) reposant sur l'introduction d'une métrique variable. Cette métrique est choisie selon le principe de majoration-minimisation. Nos résultats de simulations portant sur la restauration d'un signal sismique illustrent l'accélération produite par cette approche sur la vitesse de convergence de l'algorithme d'échantillonnage.

**Abstract** – The aim of this work is to restore a sparse signal degraded by a linear operator and corrupted with an additive Gaussian noise. We consider a Bayesian framework where the coefficients of the signal are assumed to be mutually independent and identically distributed according to a Student prior. We propose to compute the posterior mean estimate of the signal and the regularization hyperparameters using a hybrid Gibbs sampler. In particular, we choose an accelerated version of the Metropolis-Adjusted Langevin Algorithm (MALA) based on the use of a variable metric. This metric is chosen according to the majorization-minimization principle. Our simulation results on a seismic signal restoration example illustrate the acceleration in terms of convergence speed of the sampling algorithm produced by the proposed approach.

## 1 Introduction

Dans ce travail, nous nous intéressons à l'estimation d'un signal parcimonieux correspondant à un vecteur  $\mathbf{x} \in \mathbb{R}^Q$ , à partir d'une observation  $\mathbf{z} \in \mathbb{R}^N$  de celui-ci, suivant le modèle :  $\mathbf{z} = \mathbf{H}\mathbf{x} + \mathbf{w}$  où  $\mathbf{H} \in \mathbb{R}^{N \times Q}$  et  $\mathbf{w} \in \mathbb{R}^N$  est un bruit additif. La matrice  $\mathbf{H}$  est souvent associée à une opération de convolution ou de projection. Ce problème inverse peut être résolu dans un cadre bayésien qui nécessite le calcul de la loi *a posteriori* de  $\mathbf{x}$  conditionnellement aux observations. Cette loi s'obtient par la formule de Bayes qui prend en compte à la fois les informations *a priori* sur le signal original  $p(\mathbf{x}) \propto \exp(-\Psi(\mathbf{x}))$  et les informations sur les données  $p(\mathbf{z}|\mathbf{x}) \propto \exp(-\Phi(\mathbf{H}\mathbf{x} - \mathbf{z}))$ . Ainsi, la vraisemblance *a posteriori* s'écrit  $\exp(-\mathcal{J}(\mathbf{x}))$  où  $\mathcal{J}$  est défini, à une constante additive près, par

$$\mathcal{J}(\mathbf{x}) = \Phi(\mathbf{H}\mathbf{x} - \mathbf{z}) + \Psi(\mathbf{x}). \quad (1)$$

Le calcul de l'estimateur du maximiseur *a posteriori* du signal  $\mathbf{x}$  est donc équivalent à résoudre le problème de minimisation

du critère pénalisé  $\mathcal{J}(\mathbf{x})$  dans un cadre déterministe. Afin de refléter la parcimonie du signal tout en préservant les détails les plus importants, l'idéal est de choisir comme pénalisation la pseudo-norme  $\ell^0$  (nombre d'éléments non nuls d'un vecteur). Cependant, cette pénalisation  $\ell^0$  n'est ni convexe, ni différentiable, ni même continue, et la minimisation devient alors un problème combinatoire NP difficile [12]. Cette difficulté est souvent contournée en remplaçant la mesure  $\ell^0$  par la norme  $\ell^1$  (cf. [2] pour des extensions bayésiennes plus sophistiquées). Celle-ci est associée à une loi *a priori* de Laplace [4]. D'autres lois à queues lourdes peuvent favoriser davantage la parcimonie que la loi de Laplace et se rapprocher plus du comportement de la pénalisation  $\ell^0$  comme par exemple la loi alpha stable [1], la loi gaussienne généralisée avec un paramètre de forme  $p < 1$  [11] et la loi de Student [16]. Néanmoins, ces pénalisations souvent non convexes peuvent conduire à des problèmes complexes d'optimisation, et les méthodes de résolution n'aboutissent souvent qu'à des minima locaux du critère. De plus, les résultats restent très sensibles à la qualité des estimateurs

des hyperparamètres du modèle. Ainsi, au lieu de chercher le mode de la loi *a posteriori*, nous nous intéressons à l'estimateur minimisant l'erreur quadratique, c'est-à-dire au calcul de la moyenne de la loi *a posteriori*. Comme le calcul direct de cet estimateur est souvent impossible, nous proposons de recourir à des méthodes d'échantillonnage de Monte-Carlo par chaînes de Markov (MCMC). Dans ce travail, nous considérons que le bruit  $\mathbf{w}$  est gaussien centré, de matrice de covariance  $\mathbf{\Lambda}$ , et que la loi *a priori* est une loi séparable de Student. Notre contribution réside en le développement d'une méthode rapide permettant de calculer à la fois l'estimateur de la moyenne *a posteriori* du signal ainsi que des hyperparamètres en utilisant un échantillonneur hybride de Gibbs. Sur le plan applicatif, nous montrons que l'algorithme proposé fournit des résultats intéressants pour la restauration de signaux sismiques. Cet article est organisé comme suit. La section 2 décrit la méthode proposée. Dans la section 3, nous illustrons les performances de cette méthode pour un exemple de déconvolution de signaux sismiques. Enfin, dans la section 4, nous tirons les conclusions de ce travail.

## 2 Méthode proposée

### 2.1 Modèles *a priori*

Afin de favoriser la parcimonie du signal restauré, nous proposons d'utiliser une loi *a priori* de Student dont la densité de probabilité est donnée par

$$(\forall u \in \mathbb{R}) \mathcal{T}_\nu(u; \mu, \gamma) = \frac{\Gamma(\frac{\nu+1}{2})\gamma^\nu}{\sqrt{\pi\nu} \Gamma(\frac{\nu}{2})} \left( \gamma^2 + \frac{(u - \mu)^2}{\nu} \right)^{-\frac{\nu+1}{2}}$$

où  $\nu > 0$  est le degré de liberté déterminant la forme de la distribution,  $\mu$  est le paramètre de position et  $\gamma > 0$  est le paramètre d'échelle. Notons que la loi de Cauchy est un cas particulier de cette distribution, obtenue pour  $\nu = 1$ . Les petites valeurs de  $\nu$  définissent des lois à queues lourdes tandis que la loi normale est retrouvée pour  $\nu$  très grand. La loi de Student est souvent utilisée en reconstruction d'images pour modéliser la distribution des coefficients d'ondelettes [3]. Nous supposons par la suite que les coefficients du signal à estimer sont mutuellement indépendants. Une loi de Student séparable est donc considérée et l'anti-log-vraisemblance *a priori* s'écrit

$$\Psi(\mathbf{x}) = \frac{\nu+1}{2} \sum_{q=1}^Q \log \left( \gamma^2 + \frac{(x_q - \mu)^2}{\nu} \right). \quad (2)$$

Notons que cette pénalisation a déjà été introduite dans [8] comme un compromis entre la norme  $\ell^2$  et l'approximation non convexe de la semi-norme  $\ell^0$  présentée dans [5] pour modéliser la parcimonie du signal et mieux préserver les discontinuités. On peut mentionner que la loi de Student peut s'exprimer sous une forme intégrale de mélange de gaussiennes où la variable cachée suit une loi Gamma inverse. Dans la plupart des méthodes bayésiennes, elle est généralement utilisée sous cette forme : le signal inconnu  $\mathbf{x}$  et la variable cachée sont estimés à partir de leur loi *a posteriori* conjointe. Dans ce travail, on se

propose d'utiliser directement l'expression (2).

En ce qui concerne les hyperparamètres de régularisation, on suppose  $\nu$  connu, mais on ne dispose d'aucune connaissance *a priori* sur le reste des hyperparamètres. Par conséquent, on choisit une loi *a priori* uniforme pour  $\mu$  définie sur  $[-\mu_m, \mu_M]$  et une loi *a priori* uniforme sur  $[\gamma_m, \gamma_M]$  pour  $\gamma$ , où  $\mu_m, \mu_M, \gamma_m$  et  $\gamma_M$  sont quatre valeurs positives fixées. Les lois *a posteriori* des hyperparamètres s'écrivent donc

$$p(\mu|\mathbf{x}, \mathbf{z}, \gamma) \propto \prod_{q=1}^Q \left( \gamma^2 + \frac{(x_q - \mu)^2}{\nu} \right)^{-\frac{\nu+1}{2}} \mathbb{1}_{[-\mu_m, \mu_M]},$$

$$p(\gamma|\mathbf{x}, \mathbf{z}, \mu) \propto \gamma^{Q\nu} \prod_{q=1}^Q \left( \gamma^2 + \frac{(x_q - \mu)^2}{\nu} \right)^{-\frac{\nu+1}{2}} \mathbb{1}_{[\gamma_m, \gamma_M]}(\gamma).$$

### 2.2 Algorithme utilisé

Dans le cadre bayésien, l'algorithme MCMC le plus utilisé est l'algorithme de Metropolis-Hastings (MH) [7]. À partir d'une loi de proposition  $q(\mathbf{x}, \cdot)$  préalablement choisie, l'algorithme construit une chaîne de Markov dont la loi stationnaire est la loi cible *a posteriori*  $\pi$  selon les étapes suivantes [7] :

Pour  $t = 0, 1, \dots$

Générer  $\mathbf{x}^* \sim q(\mathbf{x}^t, \cdot)$

Calculer la probabilité d'acceptation

$$\alpha(\mathbf{x}^t, \mathbf{x}^*) = \min \left( 1, \frac{\pi(\mathbf{x}^*)q(\mathbf{x}^t, \mathbf{x}^*)}{\pi(\mathbf{x}^t)q(\mathbf{x}^*, \mathbf{x}^t)} \right)$$

Générer  $u^t \sim \mathcal{U}(0, 1)$

Si  $u^t \leq \alpha(\mathbf{x}^t, \mathbf{x}^*)$  alors

$$\mathbf{x}^{t+1} = \mathbf{x}^*$$

sinon

$$\mathbf{x}^{t+1} = \mathbf{x}^t$$

L'algorithme de MH à marche aléatoire est associé au cas particulier d'une loi de proposition normale centrée sur l'état courant  $\mathbf{x}^t$ . Toutefois, la convergence de cet algorithme devient très lente quand la dimension du problème augmente [17]. Une option est d'utiliser l'algorithme MALA (*Metropolis-Adjusted Langevin Algorithm*) qui exploite le gradient du critère pour guider la chaîne vers les zones les plus probables afin d'accélérer la convergence [15]. Ainsi, la loi de proposition de MALA est construite selon l'équation de transition suivante :

$$\mathbf{x}^* = \mathbf{x}^t - \frac{\varepsilon^2}{2} \nabla \mathcal{J}(\mathbf{x}^t) + \varepsilon \mathcal{N}(0, 1) \quad (3)$$

où  $\varepsilon$  est une constante positive. L'ajout d'une matrice de préconditionnement adaptative (matrice hessienne ou matrice de Fisher) permet d'augmenter la vitesse de convergence de cet algorithme [6, 10, 17, 19]. Néanmoins, cela nécessite souvent l'inversion d'une matrice à chaque itération pouvant parfois conduire à un coût de calcul prohibitif dans des problèmes de grande dimension. L'originalité de notre travail est d'utiliser une matrice de préconditionnement définie positive construite par une approche de majoration-minimisation (MM) [9]. Ainsi, à chaque itération  $t$ , on cherche une matrice  $\mathbf{Q}(\mathbf{x}^t) \in \mathbb{R}^{Q \times Q}$

de façon à ce que la fonction quadratique :

$$f(\cdot, \mathbf{x}^t) = \mathcal{J}(\mathbf{x}^t) + (-\mathbf{x}^t)^\top \nabla \mathcal{J}(\mathbf{x}^t) + \frac{1}{2} (-\mathbf{x}^t)^\top \mathbf{Q}(\mathbf{x}^t) (-\mathbf{x}^t)$$

soit une fonction majorante de  $\mathcal{J}$  en  $\mathbf{x}^t$ , c'est-à-dire

$$\begin{cases} f(\mathbf{x}^t, \mathbf{x}^t) = \mathcal{J}(\mathbf{x}^t), \\ f(\mathbf{x}, \mathbf{x}^t) \geq \mathcal{J}(\mathbf{x}) \quad (\forall \mathbf{x} \in \mathbb{R}^Q). \end{cases} \quad (4)$$

Il en résulte qu'à chaque itération  $t$ , une réalisation  $\mathbf{x}^*$  de  $\mathbf{x}$  est proposée selon le schéma suivant, appelé algorithme 3MH :

$$\mathbf{x}^* = \mathbf{x}^t - \frac{\varepsilon^2}{2} \mathbf{Q}^{-1}(\mathbf{x}^t) \nabla \mathcal{J}(\mathbf{x}^t) + \varepsilon \mathbf{Q}^{-1/2}(\mathbf{x}^t) \mathcal{N}(0, 1). \quad (5)$$

Compte tenu de l'hypothèse gaussienne sur le bruit et puisque la vraisemblance *a posteriori* satisfait les hypothèses présentées dans [9], nous pouvons utiliser pour  $\mathbf{Q}$  la matrice de courbure définie positive  $\mathbf{Q}_1$  suivante :

$$(\forall \mathbf{x} \in \mathbb{R}^N) \quad \mathbf{Q}_1(\mathbf{x}) = \mathbf{H}^\top \mathbf{\Lambda}^{-1} \mathbf{H} + \text{diag}\{\omega(\mathbf{x})\}, \quad (6)$$

où  $\omega(\mathbf{x}) = (\omega_q(\mathbf{x}))_{1 \leq q \leq Q}$  avec, pour tout  $q \in \{1, \dots, Q\}$ ,

$$\omega_q(\mathbf{x}) = \frac{\nu + 1}{\nu \gamma^2 + (x_q - \mu)^2}.$$

L'inversion de la matrice  $\mathbf{Q}_1$  peut s'avérer coûteuse quand la dimension augmente dans la mesure où l'on doit la faire à chaque itération. Une alternative dans ce cas est de considérer une matrice de courbure  $\mathbf{Q}_2$  constante :

$$\mathbf{Q}_2 = \mathbf{H}^\top \mathbf{\Lambda}^{-1} \mathbf{H} + \frac{\nu + 1}{\nu \gamma^2} \text{Id}. \quad (7)$$

Notons que si le bruit est blanc gaussien de variance  $\sigma^2$  et  $\mathbf{H}$  est circulante, la matrice définie en (7) est facilement inversible dans le domaine de Fourier. Enfin, une dernière stratégie est de choisir une matrice de courbure  $\mathbf{Q}_3$  diagonale approchant  $\mathbf{Q}_1$  :

$$(\forall \mathbf{x} \in \mathbb{R}^N) \quad \mathbf{Q}_3(\mathbf{x}) = \text{Diag}(\mathbf{P}^\top \mathbf{1}_Q + \omega(\mathbf{x})), \quad (8)$$

où, pour tout  $q \in \{1, \dots, Q\}$  et  $n \in \{1, \dots, N\}$ , on a :

$$P_{q,n} = |(\mathbf{\Lambda}^{-1/2} \mathbf{H})_{q,n}| \sum_{k=1}^N |(\mathbf{\Lambda}^{-1/2} \mathbf{H})_{k,n}|.$$

En ce qui concerne les hyperparamètres, leurs lois *a posteriori* n'ayant pas une forme usuelle, il n'est pas facile de générer directement des échantillons de  $\mu$  et  $\gamma$ . Nous envisageons donc de les estimer à l'aide d'un algorithme de MH à marche aléatoire dont le pas est ajusté automatiquement pendant la phase transitoire de façon à atteindre une probabilité d'acceptation égale à 0.33.

### 3 Application à la déconvolution d'un signal sismique

On s'intéresse à la restauration d'un signal sismique  $\bar{\mathbf{x}}$  de longueur  $Q = 784$ . Ce signal parcimonieux présente une série de pics appelés coefficients de réflexion primaire [18, 14] qui

apportent des informations sur le temps mis par les ondes sismiques pour parcourir la distance entre deux réflecteurs, ainsi que l'amplitude des événements sismiques réfléchis vers le capteur. Le signal est dégradé artificiellement par un filtre passe-bande de type "chapeau mexicain" de taille 41 dont le spectre de fréquences est concentré entre 10 et 40 Hz et par un bruit gaussien additif de variance  $\sigma^2 = 2.5 \times 10^{-3}$  (figure 1). Le rapport signal sur bruit (RSB) initial est de  $-4.58$  dB.

Une loi *a priori* de Cauchy ( $\nu = 1$ ) est considérée. La figure 2 montre l'erreur entre le signal original et le signal dégradé ainsi que celle entre le signal original et l'estimateur de la moyenne *a posteriori* du signal (conduisant à un RSB de 8.24 dB).

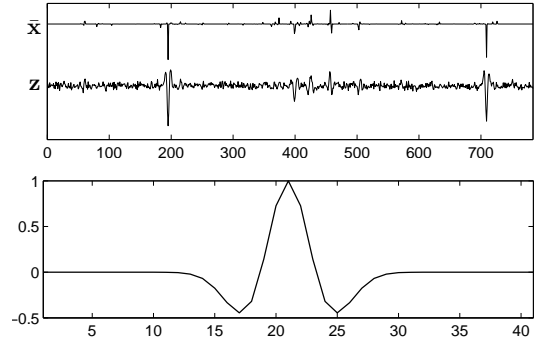


FIGURE 1 – Signal original et dégradé (haut) et noyau de convolution (bas).

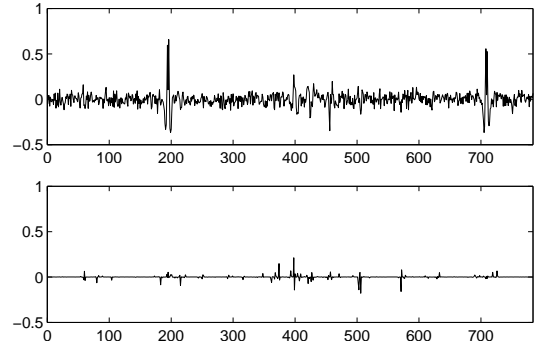


FIGURE 2 – Erreurs  $\bar{\mathbf{x}} - \mathbf{z}$  (haut) et  $\bar{\mathbf{x}} - \hat{\mathbf{x}}$  (bas).

On se propose de comparer la vitesse de convergence de l'algorithme 3MH en utilisant les matrices de courbures  $\mathbf{Q}_1$ ,  $\mathbf{Q}_2$ , et  $\mathbf{Q}_3$  avec l'algorithme standard MALA. Pour ce faire, on lance les différents algorithmes jusqu'à la convergence. Le pas de discrétisation est ajusté pour tous les algorithmes pendant la période transitoire afin de correspondre à une probabilité d'acceptation entre 0.3 et 0.6. Notons que pour réduire la complexité de chaque itération dans le cas où  $\mathbf{Q} = \mathbf{Q}_1$ , l'inversion de la matrice de courbure est remplacée par un algorithme de gradient conjugué et la génération des variables aléatoires selon la loi de proposition (5) est assurée grâce à la méthode rapide proposée dans [13].

La figure 3 montre l'évolution du  $\mathcal{J}$  au cours du temps pour

chacun d’entre eux. On peut remarquer que le comportement obtenu de la matrice de courbure  $\mathbf{Q}_2$  est proche de celui de MALA. La raison de ce phénomène tient au fait que, sur cet exemple, la dispersion des valeurs propres de  $\mathbf{Q}_2$  est faible. L’ajout d’une matrice de préconditionnement adaptative dans l’algorithme MALA permet de s’approcher des zones caractérisées par des valeurs élevées de la probabilité en moins d’itérations. Néanmoins, la manipulation de la matrice  $\mathbf{Q}_1$  à chaque itération est d’autant plus coûteuse que la dimension du problème augmente. Ainsi, le choix de la matrice adaptative diagonale  $\mathbf{Q}_3$  permet d’établir un bon compromis entre ces différents algorithmes en raison de la faible complexité opératoire qu’il induit à chaque itération.

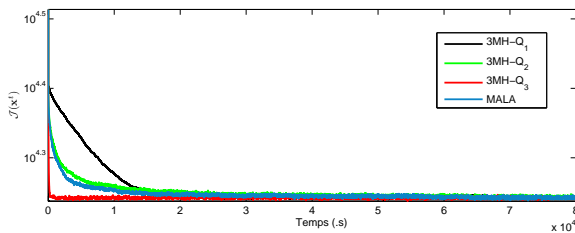


FIGURE 3 – Comparaison entre les vitesses de convergence des algorithmes MALA, 3MH -  $\mathbf{Q}_1$ , 3MH -  $\mathbf{Q}_2$  et 3MH -  $\mathbf{Q}_3$ .

## 4 Conclusion

Nous avons traité un problème de reconstruction de signaux parcimonieux en utilisant une loi *a priori* de Student. Nous avons calculé l’estimateur de la moyenne *a posteriori* du signal en utilisant un algorithme d’échantillonnage basé sur la diffusion de Langevin accélérée par l’introduction d’une matrice de préconditionnement selon le principe MM. Les résultats des simulations dans le cadre de la restauration d’un signal sismique montrent que notre algorithme se compare favorablement à l’algorithme standard MALA.

## References

- [1] A. Achim, P. Tsakalides, and A. Bezerianos. SAR image denoising via Bayesian wavelet shrinkage based on heavy-tailed modeling. *IEEE Trans. Geosci. Remote Sens.*, 41:1773–1784, 2003.
- [2] L. Chaari, H. Batatia, N. Dobigeon, and J.-Y. Tourneret. A hierarchical sparsity-smoothness Bayesian model for  $\ell_0 + \ell_1 + \ell_2$  regularization. In *Proc. IEEE Int. Conf. Acoust., Speech Signal Process.*, pages 1901–1905, Florence, Italy, 4-9 May 2014.
- [3] G. Chantas, N. Galatsanos, A. Likas, and M. Saunders. Variational Bayesian image restoration based on a product of  $t$ -distributions image prior. *IEEE Trans. Image Process.*, 17:1795–1805, 2008.
- [4] D. L. Donoho and Y. Tsaig. Fast solution of  $\ell^1$  norm minimization problems when the solution may be sparse. *IEEE Trans. Inform. Theory*, 54:4789–4812, 2008.

- [5] S. A. Geman and D. E. McClure. Statistical methods for tomographic image reconstruction. *MIT Industrial Liaison Program Report*, 8:5–21, 1987.
- [6] M. Girolami and B. Calderhead. Riemann manifold Langevin and Hamiltonian Monte Carlo methods. *J. R. Stat. Soc. Ser. B Stat. Methodol.*, 73(91):123–214, 2011.
- [7] W. K. Hastings. Monte Carlo sampling methods using Markov chains and their applications. *Biometrika*, 57(1):97–109, Apr. 1970.
- [8] T. Hebert and R. Leahy. A generalized EM algorithm for 3-D Bayesian reconstruction from Poisson data using Gibbs priors. *IEEE Trans. Med. Imag.*, 8:194–202, 1989.
- [9] Y. Marnissi, A. Benazza-Benyahia, E. Chouzenoux, and J.-C. Pesquet. Majorize-Minimize adapted Metropolis-Hastings algorithm. Application to multichannel image recovery. In *Proc. Eur. Sig. Image Proc. Conf.*, pages 1332–1336, Lisbon, Portugal, 1-5 Sept. 2014.
- [10] J. Martin, C. L. Wilcox, C. Burstedde, and O. Ghattas. A stochastic Newton MCMC method for large-scale statistical inverse problems with application to seismic inversion. *SIAM J. Sci. Comput.*, 34(3):A1460–A1487, 2012.
- [11] P. Moulin and J. Liu. Analysis of multiresolution image denoising schemes using Generalized Gaussian and complexity priors. *IEEE Trans. Inform. Theory*, 45:909–919, 1999.
- [12] B. K. Natarajan. Sparse approximate solutions to linear systems. *SIAM J. Comput.*, 24(2):227–234, 1995.
- [13] F. Orieux, O. Feron, and J. F. Giovannelli. Sampling high-dimensional Gaussian distributions for general linear inverse problems. *IEEE Signal Process. Lett.*, 19(5):251–254, May 2012.
- [14] A. Repetti, M. Q. Pham, L. Duval, E. Chouzenoux, and J.-C. Pesquet. Euclid in a taxicab: Sparse blind deconvolution with smoothed  $\ell_1/\ell_2$  regularization. *IEEE Signal Process. Lett.*, 22(5):539–543, May 2015.
- [15] O. G. Roberts and L. R. Tweedie. Exponential convergence of Langevin distributions and their discrete approximations. *Bernoulli*, 2(4):341–363, 1996.
- [16] M. W. Seeger and H. Nickisch. Large scale Bayesian inference and experimental design for sparse linear models. *SIAM J. Imaging Sci.*, 4:166–199, 2011.
- [17] C. Vacar, J. F. Giovannelli, and Y. Berthoumieu. Langevin and Hessian with Fisher approximation stochastic sampling for parameter estimation of structured covariance. In *Proc. IEEE Int. Conf. Acoust., Speech Signal Process.*, pages 3964–3967, Prague, Czech Republic, 22-27 May 2011.
- [18] A. T. Walden and J. W. J. Hosken. The nature of the non-Gaussianity of primary reflection coefficients and its significance for deconvolution. *Geophys. Prospect.*, 34(7):1038–1066, 1986.
- [19] Y. Zhang and C. A. Sutton. Quasi-Newton methods for Markov chain Monte Carlo. In J. Shawe-Taylor, R.S. Zemel, P.L. Bartlett, F. Pereira, and K.Q. Weinberger, editors, *Adv. Neural Inf. Process. Syst.*, pages 2393–2401. Curran Associates, Inc., 2011.