Une formulation bayésienne du codage antiparcimonieux

Clément Elvira, Pierre Chainais, Nicolas Dobigeon

Une formulation bayésienne du codage antiparcimonieux

Clément ELVIRA[1], Pierre CHAINAIS[1] et Nicolas DOBIGEON[2]

[1]Université de Lille, CNRS, Centrale Lille, UMR 9189 - CRIStAL - Centre de Recherche en Informatique
Signal et Automatique de Lille, F-59000 Lille, France

[2]Université of Toulouse, IRIT/INP-ENSEEIHT, CNRS, 2 rue Charles Camichel,
BP 7122, 31071 Toulouse cedex 7, France

{clement.elvira, pierre.chainais}@centralelille.fr, nicolas.dobigeon@enseeiht.fr

Résumé – Dans un but de robustesse, un codage antiparcimonieux répartit uniformément l'information d'un signal sur toutes les composantes de sa représentation. La recherche d'un tel codage s'exprime naturellement sous la forme d'un problème variationnel impliquant une régularisation de type ℓ_∞. Dans cet article une formulation bayésienne du problème est proposée, impliquant une nouvelle loi de probabilité, la *loi démocratique*, qui pénalise les fortes amplitudes. Cette distribution est choisie comme loi *a priori* sur les coefficients de représentation, couplée avec une vraisemblance gaussienne. Les estimateurs bayésiens des coefficients de représentation peuvent être approchés à l'aide d'un échantillonneur de Gibbs. Cette méthode passe cependant difficilement à l'échelle et un algorithme de Monte Carlo proximal a été proposé. On discute une nouvelle façon de choisir et régler la loi *a priori* sur les paramètres de nuisance. Deux simulations numériques permettent de valider le réglage des hyperparamètres et la recherche du paramètre de régularisation.

Abstract – For sake of robustness, anti-sparse coding aims at spreading the information uniformly over representation coefficients and can be naturally expressed by an ℓ_∞-norm regularization. This article discusses a Bayesian formulation of the anti-sparse coding problem, relying on a new probability distribution called *democratic distribution*. Once elected as a prior distribution in a linear Gaussian inverse problem, inference can be conducted using a Gibbs sampling scheme. This paper discusses an alternative choice of prior for the nuisance parameters and proposes a comparison between two scalable inference schemes based on recent developements in Monte Carlo methods. These findings are illustrated by simulations on synthetic data, and compared with the recent deterministic variational FITRA algorithm.

1 Introduction

La régularisation de problèmes inverses mal posés par des contraintes de parcimonie a donné lieu à de nombreux travaux dans la littérature, notamment motivés par le paradigme de l'échantillonnage compressé. À l'inverse, répartir uniformément l'information d'un signal sur toutes les composantes d'une représentation est également recherché dans certaines situations telles que la conception de convertisseurs analogique/numérique [1] ou la recherche approchée de plus proches voisins [2]. Ces représentations peuvent s'obtenir en minimisant l'amplitude maximale des coefficients sous contrainte d'une erreur de reconstruction quadratique. Ces représentations antiparcimonieuses, dites de Kashin [3] ou encore *démocratiques* [4], correspondent à la décomposition d'un signal sur une famille donnée ayant l'amplitude la plus faible. Une autre manière d'obtenir la représentation antiparcimonieuse d'un signal consiste à résoudre un problème variationnel [5] où l'amplitude des coefficients est pénalisée via la norme ℓ_∞

$$\min_{\mathbf{x}\in\mathbb{R}^N} \frac{1}{2\sigma^2}\left\|\mathbf{y}-\mathbf{H}\mathbf{x}\right\|_2^2 + \lambda\left\|\mathbf{x}\right\|_\infty . \qquad (1)$$

Dans l'équation (1), $\mathbf{H}$ est une matrice de codage de taille $M \times N$ et σ^2 est la variance de l'erreur résiduelle résultant de l'approximation. En outre, on note $J(\mathbf{x},\beta)$ la fonction de coût implicitement définie par (1), i.e.,

$$J(\mathbf{x},\beta) \triangleq \left\|\mathbf{y}-\mathbf{H}\mathbf{x}\right\|_2^2 + \beta\left\|\mathbf{x}\right\|_\infty \text{ avec } \beta \triangleq 2\sigma^2\lambda.$$

Ce problème a récemment été considéré pour des applications en réduction de facteur de crête [6]. Cet article propose une formulation bayésienne du problème de reconstruction antiparcimonieuse sous jacent à (1). Les paramètres de nuisance λ et σ^2 sont intégrés au modèle via le choix de lois *a priori*. Cette formulation repose sur l'utilisation de la *loi démocratique* introduite dans [7]. L'inférence des paramètres inconnus du modèle bayésien proposé est réalisée à l'aide d'un algorithme de Gibbs. Les échantillons générés sont ensuite utilisés pour approximer deux estimateurs Bayésiens. On se propose également d'étudier le passage à l'échelle de la méthode. L'échantillonneur de Gibbs étant peu efficace en grande dimension, deux alternatives intégrant des méthodes proximales sont étudiées.

Cet article est organisé comme suit. La *loi démocratique* est introduite dans la partie 2. Une formulation bayésienne du problème de codage antiparcimonieux est proposée dans la partie 3. Cette partie décrit également deux solutions algorithmiques permettant de générer des échantillons asymptotiquement distribués suivant la loi *a posteriori* des paramètres. Les performances de ces méthodes sont comparées dans la partie 4. La partie 5 conclut l'article.

2 La loi démocratique

On s'appuie sur la pénalité en norme ℓ_∞ du problème variationnel (1) pour définir une nouvelle loi de probabilité appelée *loi démocratique*. La densité de la loi démocratique $\mathcal{D}_N(\lambda)$ est donnée pour tout vecteur $\mathbf{x}$ de $\mathbb{R}^N$ par

$$f(\mathbf{x}|\lambda) = \frac{\lambda^N}{2^N N!} \exp\left(-\lambda \|\mathbf{x}\|_\infty\right). \quad (2)$$

La Fig. 1 représente la fonction densité de probabilité associée à la loi $\mathcal{D}_2(3)$.

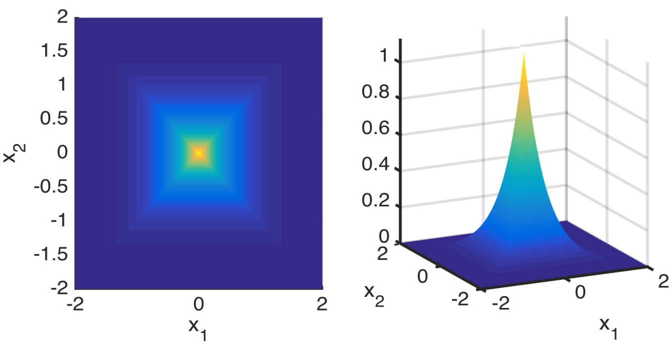

FIGURE 1 – Densité de la loi démocratique $\mathcal{D}_2(3)$.

2.1 Lois conditionnelles

La densité (2) n'est pas pratique à manipuler, principalement en raison de la norme ℓ_∞. On introduit les cônes $\mathcal{C}_n$ de $\mathbb{R}^N$

$$\mathcal{C}_n \triangleq \left\{\mathbf{x} = [x_1, \ldots, x_N]^T \in \mathbb{R}^N : \forall j \neq n,\ |x_j| < |x_n|\right\}. \quad (3)$$

induits par les symétries de la norme. Pour tout $\mathbf{x} \in \mathcal{C}_n$, on a $\|\mathbf{x_n}\|_\infty = |x_n|$. Les cônes $(\mathcal{C}_n)$ forment une partition [1] de $\mathbb{R}^N$, et l'on a directement par symétrie

$$\mathrm{P}\left[\mathbf{x} \in \mathcal{C}_n\right] = \frac{1}{N},\ \forall n \in \{1, \ldots, N\}. \quad (4)$$

En conditionnant aux cônes, on obtient les lois conditionnelles explicites des composantes dominées et dominantes

$$x_n|\mathbf{x}_{\setminus n}, \mathbf{x} \in \mathcal{C}_n \sim \frac{\lambda}{2} e^{-\lambda\left(|x_n| - \|\mathbf{x}_{\setminus n}\|_\infty\right)} \mathbf{1}_{\mathbb{R}\setminus\mathcal{I}_n}(x_n) \quad (5)$$

$$\mathbf{x}_{\setminus n}|x_n, \mathbf{x} \in \mathcal{C}_n \sim \prod_{l \neq n} \mathcal{U}\left(\mathcal{I}_n\right), \quad (6)$$

où $\mathbf{x}_{\setminus n}$ désigne le vecteur $\mathbf{x}$ privé de sa n-ième composante et $\mathcal{I}_n \triangleq \left(-\|\mathbf{x}_{\setminus n}\|_\infty, \|\mathbf{x}_{\setminus n}\|_\infty\right)$. En combinant les équations (4), (5) et (6), et en marginalisant sur les cônes, on obtient la loi conditionnelle

$$\begin{aligned} \mathrm{p}\left(x_n|\mathbf{x}_{\setminus n}\right) &= (1 - c_n) \frac{1}{2\left\|\mathbf{x}_{\setminus n}\right\|_\infty} \mathbf{1}_{\mathcal{I}_n}(x_n) \\ &\quad + c_n \frac{\lambda}{2} e^{-\lambda\left(|x_n| - \|\mathbf{x}_{\setminus n}\|_\infty\right)} \mathbf{1}_{\mathbb{R}\setminus\mathcal{I}_n}(x_n) \end{aligned} \quad (7)$$

où

$$c_n \triangleq \mathrm{P}\left[\mathbf{x} \in \mathcal{C}_n|\mathbf{x}_{\setminus n}\right] = \frac{1}{1 + \lambda\left\|\mathbf{x}_{\setminus n}\right\|_\infty}. \quad (8)$$

La loi conditionnelle est donc un mélange à deux composantes définies sur des intervalles disjoints. L'échantillonnage de la loi *démocratique* exploitant ces propriétés est discutée dans [7].

2.2 Opérateur proximal de la neg-log densité

La densité de la loi démocratique peut se réécrire $f(\mathbf{x}) \propto \exp\left(-g(\mathbf{x})\right)$ avec $g(\mathbf{x}) = \lambda\|\mathbf{x}\|_\infty$. L'opérateur proximal de g de paramètre δ, est donné par

$$\mathrm{prox}_g^\delta(\mathbf{x}) = \underset{\mathbf{u} \in \mathbb{R}^N}{\mathrm{argmin}}\ \lambda\|\mathbf{u}\|_\infty + \frac{1}{2\delta}\|\mathbf{x} - \mathbf{u}\|_2^2. \quad (9)$$

Le problème de minimisation (9) n'a pas de solution explicite. On peut toutefois approcher des solutions pour un faible coût numérique en s'appuyant sur les propriétés de l'opérateur proximal et sur la dualité des normes ℓ_1 et ℓ_∞ [7]

$$\mathrm{prox}_{\lambda\|\cdot\|_\infty}(\mathbf{x}) = \mathbf{x} - \lambda\Pi_{\{\boldsymbol{u}, \|\|\boldsymbol{u}/\lambda \leq 1\}}(\mathbf{x}), \quad (10)$$

où Π est le projecteur sur la boule ℓ_1. Cet opérateur proximal servira dans la partie 3 à échantillonner la loi *a posteriori* résultant de la formulation bayésienne du codage antiparcimonieux.

3 Codage anti-parcimonieux bayésien

Cette partie décrit la formulation bayésienne du problème de codage antiparcimonieux associé à l'équation (1).

3.1 Modèle bayésien hiérarchique

Soit $\mathbf{y}$ un vecteur d'observation de $\mathbb{R}^M$, et $\mathbf{H}$ une matrice de codage de taille $M \times N$ connue. On souhaite reconstruire ce vecteur d'observation à travers le modèle linéaire $\mathbf{y} = \mathbf{H}\mathbf{x} + \mathbf{e}$ où $\mathbf{e} = [e_1 \ldots e_N]^T$ est le vecteur des erreurs résiduelles, supposé gaussien centré et de matrice de covariance $\sigma^2\boldsymbol{I}_M$. Ce choix d'erreur conduit à la fonction de vraisemblance

$$f(\mathbf{y}|\mathbf{x}, \sigma^2) = \left(\frac{1}{2\pi\sigma^2}\right)^{\frac{M}{2}} \exp\left[-\frac{1}{2\sigma^2}\|\mathbf{y} - \mathbf{H}\mathbf{x}\|_2^2\right]. \quad (11)$$

On définit maintenant les lois *a priori* pour les paramètres inconnus liés au modèle d'observation.

Vecteur de représentation. On choisit la loi démocratique introduite dans la partie 2 comme loi *a priori* sur le vecteur de représentation $\mathbf{x}$ afin de favoriser l'antiparcimonie, *i.e.*, l'utilisation la plus uniforme possible de toutes les composantes

$$\mathbf{x} \mid \lambda \sim \mathcal{D}_N(\lambda). \quad (12)$$

Notons que c'est le produit $\beta = 2\sigma^2\lambda$ qui contrôle effectivement l'estimateur du maximum *a posteriori* (MAP) de $\mathbf{x}$, ou, de manière équivalente, l'estimateur associé au problème (1). Le choix de la valeur du paramètre λ est discuté au paragraphe 3.2.3.

Hyperparamètre. Une loi conjuguée inverse Gamma est choisie comme loi *a priori* pour l'hyperparamètre β

$$f(\beta) \propto \left(\frac{1}{\beta}\right)^{a_\beta + 1} \exp\left(-\frac{b_\beta}{\beta}\right). \quad (13)$$

Le choix des hyperparamètres a_β et b_β est détaillé à la partie 4.

1. Les frontières, qui sont des droites de mesure nulle, doivent être exclues.

Loi jointe *a posteriori*. Le modèle hiérarchique défini précédemment conduit à la loi jointe *a posteriori* du couple $(\mathbf{x}, \beta)$ plutôt que $(\mathbf{x}, \sigma^2)$

$$f(\mathbf{x}, \beta|\mathbf{y}, \lambda) \propto \mu^N \exp\left(-\frac{\lambda}{\beta}\|\mathbf{y}-\mathbf{H}\mathbf{x}\|_2^2 - \lambda\|\mathbf{x}\|_\infty\right) \times \left(\frac{1}{\beta}\right)^{a_\beta+\frac{M}{2}+1} \exp\left(-\frac{b_\beta}{\beta}\right). \quad (14)$$

On souligne le fait que, pour une valeur de β fixée, l'estimateur MAP du vecteur de représentation $\mathbf{x}$ associé à (14) est exactement la solution du problème défini en (1). Dans un paradigme non supervisé, ce paramètre dit de nuisance est estimé conjointement au vecteur de reconstruction. La prochaine partie présente un échantillonneur de Gibbs permettant de générer des échantillons $\left\{\beta^{(t)}, \mathbf{x}^{(t)}\right\}$ asymptotiquement distribués suivant (14). Ces échantillons sont utilisés pour approcher deux estimateurs bayésiens, à savoir l'estimateur qui minimise l'erreur quadratique moyenne (MMSE) $\hat{\mathbf{x}}_{\mathrm{MMSE}} = \mathbb{E}[\mathbf{x}|\mathbf{y}]$ et l'estimateur MAP marginal $\hat{\mathbf{x}}_{\mathrm{MAPm}} = \underset{\mathbf{x}}{\mathrm{argmax}} f(\mathbf{x}|\mathbf{y})$.

3.2 Échantillonneur de Gibbs

3.2.1 Vecteur de reconstruction

La loi *a posteriori* de chaque composante x_n est disponible [7], mais un échantillonneur s'appuyant sur ce résultat aboutit à un mélange peu efficace de la chaine et passe difficilement à l'échelle. On s'appuie sur P-MYULA [8], une version améliorée de P-MALA [9] pour simuler efficacement des échantillons suivant une distribution multidimensionnelle en s'appuyant sur les méthodes proximales pour l'optimisation convexe. P-MYULA exploite le fait que la fonction d'attache aux données liée à la vraisemblance est différentiable pour proposer des mouvements plus efficaces par rapport à P-MALA.

Notons h l'opposé du logarithme de la loi *a posteriori* de $\mathbf{x}$, $h(\mathbf{x}) \triangleq J(\mathbf{x}, 2\lambda\sigma^2)/(2\sigma^2)$. La fonction h est la somme d'une fonction différentiable et d'une fonction convexe. P-MYULA consiste à échantillonner suivant la distribution régularisée $\exp(-h^\nu)$, où h^ν s'obtient en remplaçant la partie non différentiable de h par son enveloppe de Moreau-Yoshida [10]. Un algorithme de type Metropolis-Hastings est utilisé pour échantillonner suivant h^ν, de loi de proposition

$$\mathbf{x}^*|\mathbf{x}^{(t-1)} \sim \mathcal{N}\left(\bar{\mathbf{x}}^{(t-1)}, 2\gamma\right) \quad (15)$$

avec

$$\bar{\mathbf{x}}^{(t-1)} = \left(1-\frac{\gamma}{\nu}\right)\mathbf{x}^{(t-1)} - \gamma\nabla f\left(\mathbf{x}^{(t-1)}\right) + \frac{\gamma}{\nu}\mathrm{prox}_h^\nu\left(\mathbf{x}^{(t-1)}\right).$$

En choisissant $\nu = L_{h_1}^{-1}$ où L_{h_1} est la constante de Lipschitz de h_1 et $\gamma = \nu/4$, l'erreur relative entre la distribution cible et la posterior est estimée à 1% dans [8]. Cette erreur peut également être corrigée par une étape de Metropolis-Hastings.

3.2.2 Paramètres de nuisance

Par conjugaison, la loi *a posteriori* du paramètre de nuisance est donnée par

$$\beta|\mathbf{y}, \mathbf{x} \sim \mathcal{IG}\left(a_\beta + \frac{M}{2}, b_\beta + \lambda\|\mathbf{y}-\mathbf{H}\mathbf{x}\|_2^2\right). \quad (16)$$

3.2.3 Réglage des hyper-paramètres

Les paramètres à régler sont λ, a_β et b_β. On rappelle que les paramètres ν et γ de MYULA sont réglés par la géométrie du problème. Des lois *a priori* vagues ont précédemment été proposées pour les paramètres de nuisance dans [7]. Cependant, dans le cas de la recherche d'une représentation antiparcimonieuse qui répartit uniformément l'énergie sur toutes les composantes, il est légitime de chercher $|x_n| \simeq \frac{\|\mathbf{y}\|_2}{\sqrt{N}}$. Or, on peut calculer

$$\mathbb{E}\left[|x_n| \big| \mathbf{x} \sim \mathcal{D}_N(\lambda)\right] = \frac{N+1}{2\lambda}. \quad (17)$$

Une équipartition de l'énergie étant visée pour une représentation démocratique, on espère $|x_n| \simeq \|\boldsymbol{y}\|_2/\sqrt{N}$. Un choix pertinent serait donc de choisir *a priori* $\lambda \propto \frac{(N+1)\sqrt{N}}{2\|\boldsymbol{y}\|_2}$. Notons que cette compréhension du paramètre a été rendue possible grâce à l'interprétation bayésienne du problème.

4 Résultats expérimentaux

Cette partie présente les résultats de deux expériences où la paramétrisation est différente en dimensions $M = 30$ et $N = 50$. Les matrices d'encodage sont générées en sous échantillonnant les lignes d'une matrice de transformée en cosinus discrets (DCT), et les observations $\boldsymbol{y}$ de $\mathbb{R}^M$ sont des tirages d'une loi normale [3]. On cherche un vecteur de représentation $\hat{\boldsymbol{x}}$ telle que $\boldsymbol{y} \approx \mathbf{H}\hat{\boldsymbol{x}}$ tout en minimisant $\|\boldsymbol{x}\|_\infty$. Les estimateurs bayésiens obtenus sont comparés à l'algorithme FITRA [6] qui résout le problème (1) par une méthode itérative de *forward-backward*. Les métriques utilisées pour évaluer la qualité des résultats est une mesure de reconstruction $\mathrm{SNR}_\mathbf{y} = 20\ln\frac{\|\mathbf{y}\|_2}{\|\mathbf{y}-\mathbf{H}\hat{\mathbf{x}}\|_2}$ et une mesure d'antiparcimonie définie par $\mathrm{PAPR} = \frac{N\|\hat{\mathbf{x}}\|_\infty^2}{\|\hat{\mathbf{x}}\|_2^2}$. L'objectif est de trouver des solutions avec un $\mathrm{SNR}_\mathbf{y}$ élevé et un faible PAPR traduisant une représentation démocratique. On affichera souvent les solutions dans le plan PAPR/$\mathrm{SNR}_\mathbf{y}$.

4.1 Comportement des estimateurs

Dans cette partie, le paramètre β n'est pas estimé, il est arbitrairement fixé à $\beta = 0.01$. Notons que le comportement des estimateurs est qualitativement indépendant de ce choix. On a montré dans la partie 3 que pour des valeurs fixées des paramètres λ et β, l'estimateur MAP $\arg\max f(\mathbf{x}|\mathbf{y}, \beta, \lambda)$ correspond à la solution du problème (1). On étudie dans cette partie la robustesse de la méthode en faisant varier la valeur de λ. La valeur de σ^2 s'obtient alors via $\sigma^2 = \beta/2\lambda$. Les estimateurs sont calculés sur 10^5 itérations de l'algorithme MCMC. La Figure 2-a affiche la valeur de $J(\hat{\mathbf{x}}, 0.01)$ pour les estimateurs MAP et MMSE. La droite horizontale est la valeur de $J(\hat{\mathbf{x}}, 0.01)$ obtenue avec FITRA, qui joue ici le rôle de solution de référence. Comme l'estimateur MAP ne dépend que de β, les performances devraient être les mêmes. On constate que les estimateurs sont d'autant plus proches de FITRA que le paramètre λ est grand. On note l'existence d'une valeur seuil,

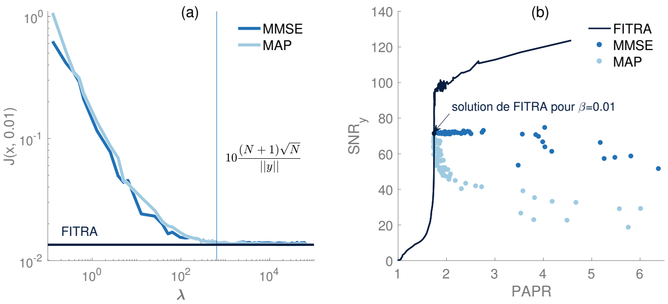

FIGURE 2 – (a) Valeurs de $J(\hat{\mathbf{x}}, 0.01)$ pour les estimateurs MAP et MMSE en fonction de λ. (b) Position des estimateurs MAP et MMSE dans le plan $\mathrm{PAPR}/\mathrm{SNR}_{\mathbf{y}}$ en fonction de λ.

environ égale à $10\frac{(N+1)\sqrt{N}}{\|\boldsymbol{y}\|_2}$, à partir de laquelle on peut numériquement considérer que les estimateurs sont les mêmes.

La Figure 2-b représente les estimateurs MAP (points bleu clair) et MMSE (points bleu foncé) dans le plan $\mathrm{PAPR}/\mathrm{SNR}_{\mathbf{y}}$ lorsque λ varie. La courbe noire correspond aux solutions proposées par FITRA pour toutes les valeurs du paramètre de régularisation β. On constate que les chemins des deux estimateurs obtenus en faisant varier λ convergent vers le point correspondant à la solution proposée par FITRA pour $\beta = 0.01$.

4.2 Inférence du paramètre de régularisation

Dans cette partie, le paramètre β n'est plus fixé, mais est estimé suivant le modèle hiérarchique présenté dans la partie 3. L'objectif est d'inférer la valeur de β qui correspond au meilleur compromis entre $\mathrm{SNR}_{\mathbf{y}}$ et PAPR. On s'intéresse en particulier à la zone critique mise en évidence par FITRA. Cette zone critique correspond à la transition de phase des solutions dans le plan $\mathrm{PAPR}/\mathrm{SNR}_{\mathbf{y}}$ obtenue avec les solutions de FITRA en faisant varier le paramètre de régularisation. Sur la Figure 3, cela correspond à un PAPR entre 1.5 et 1.8. On souhaite *a priori* que β explore cette région des solutions. On choisit pour cela $a_\delta = 2000$ et $b_\delta = 200$.
La Figure 3 montre les 200 dernières itérations de l'algorithme MCMC. On observe que les itérations sont proches de la courbe obtenue par l'algorithme FITRA et de la transition de phase. Cependant, les points correspondants restent tous à une certaine distance de la courbe FITRA, ce qui n'était pas le cas estimateurs présentés dans la partie 4.1. Par conséquent, l'estimateur MAPm défini par $\hat{\mathbf{x}}_{\mathrm{MAPm}} = \operatorname*{argmax}_{\mathbf{x}} f(\mathbf{x}|\mathbf{y})$ et l'estimateur MMSE fourniront tous les deux des estimateurs biaisés relativement à la référence FITRA. On en déduit que dans le cas où β est conjointement estimé, le modèle incluant une loi *a priori* conjuguée sur β induit un biais faible mais sensible dans les estimateurs des coefficients $\mathbf{x}$ considérés.

5 Conclusion

Cet article a introduit une nouvelle loi de probabilité, la loi *démocratique*, qui est ensuite utilisée comme loi *a priori* sur des coefficients de représentation dans un problème de codage

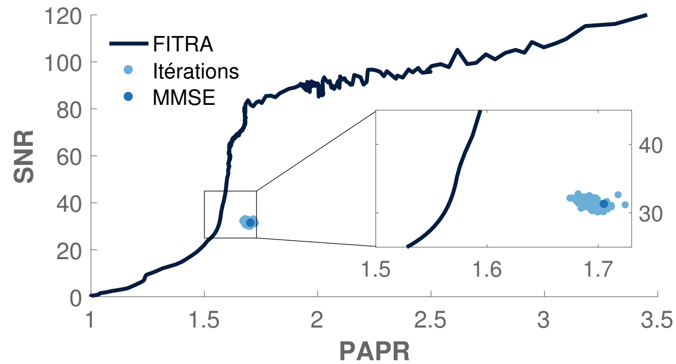

FIGURE 3 – 200 dernières itérations de l'algorithme MCMC lorsque β est également échantillonné.

antiparcimonieux. Un algorithme MCMC basé sur P-MYULA a été utilisé pour générer des échantillons asymptotiquement distribués suivant la loi *a posteriori*. Cet échantillonneur s'est avéré plus efficace qu'un échantillonneur de Gibbs classique. Si l'on fixe les paramètres du modèle, les résultats expérimentaux ont montré que les estimateurs MAP et MMSE convergent vers les solutions proposées par l'algorithme d'optimisation FITRA. En revanche, l'estimation du paramètre de régularisation est une question plus difficile. Le choix d'une loi *a priori* conjuguée pour le paramètre de régularisation semble induire un biais dans les estimateurs usuels MAPm et MMSE.

Références

[1] A. R. Calderbank and I. Daubechies, "The pros and cons of democracy," *IEEE Trans. Inf. Theory*, 2002.

[2] H. Jégou, T. Furon, and J. J. Fuchs, "Anti-sparse coding for approximate nearest neighbor search," in *ICASSP*, 2012.

[3] Y. Lyubarskii and R. Vershynin, "Uncertainty principles and vector quantization," *IEEE Trans. Inf. Theory*, 2010.

[4] C. Studer, Y. Wotao, and R. G. Baraniuk, "Signal representations with minimum ℓ_∞-norm," in *Proc. Ann. Allerton Conf. Comm. Control Comput.*, 2012.

[5] J.-J. Fuchs, "Spread representations," in *Proc. IEEE Asilomar Conf. Signals, Systems, Computers*, 2011.

[6] C. Studer and E. G. Larsson, "PAR-aware large-scale multi-user MIMO-OFDM downlink," *IEEE J. Sel. Areas Comm.*, 2013.

[7] C. Elvira, P. Chainais, and N. Dobigeon, "Bayesian antisparse coding," *IEEE Trans. Signal Process.*, April 2017.

[8] A. Durmus, E. Moulines, and M. Pereyra, "Efficient Bayesian computation by proximal Markov Chain Monte Carlo : when Langevin meets Moreau," 2016.

[9] M. Pereyra, "Proximal Markov chain Monte Carlo algorithms," *Stat. Comput.*, 2015.

[10] J.-J. Moreau, "Fonctions convexes duales et points proximaux dans un espace hilbertien," *CR Acad. Sci. Paris Sér. A Math*, 1962.