

Mélange de bruits et échantillonnage de posterior non log-concave

Pierre PALUD¹, Pierre CHAINAIS¹, Franck Le PETIT², Emeric BRON², Pierre-Antoine THOUVENIN¹, Maxime VONO³, L. EINIG⁴, M. GARCIA SANTA-MARIA⁴, M. GAUDEL⁴, J. H. ORKISZ⁴, V. DE SOUZA MAGALHAES⁴, S. BARDEAU⁴, M. GERIN⁴, J. R. GOICOECHEA⁴, P. GRATIER⁴, V. V. GUZMAN⁴, J. KAINULAINEN⁴, F. LEVRIER⁴, N. PERETTO⁴, J. PETY⁴, A. ROUEFF⁴, A. SIEVERS⁴

¹Univ. Lille, CNRS, Centrale Lille, UMR 9189 CRISTAL, F-59000 Lille, France

²LERMA, Observatoire de Paris, PSL Research University, CNRS, Sorbonne Universités, 92190 Meudon, France

³Criteo AI Lab, Paris, France

⁴Consortium ORION-B (<https://www.iram.fr/~pety/ORION-B/team.html>)

Résumé – Ce travail considère un problème inverse en astrophysique, qui consiste à estimer un ensemble de paramètres physiques à partir d’images multispectrales. Le modèle direct sous-jacent est une simulation numérique et le modèle d’observation mélange différentes sources de bruit. Ces caractéristiques donnent lieu à une fonction de log-vraisemblance non explicite et non concave. Nous en définissons une approximation qui permet de concevoir une méthode de Monte Carlo par chaîne de Markov (MCMC). L’échantillonneur obtenu fournit estimations ponctuelles et intervalles de crédibilité associés. L’approche proposée est suffisamment générale pour être appliquée à des problèmes inverses similaires.

Abstract – This work considers a radio-astronomy inverse problem of physical parameters inference from multispectral images. The forward model is a numerical simulation, and the observation model mixes different sources of noise. This results in a non-explicit non-log-concave likelihood function. We introduce a likelihood approximation with controlled error that allow the conception of a Monte Carlo Markov Chain (MCMC) method. The obtained sampler provides credibility intervals along with point estimates. We believe that the proposed approach is sufficiently generic to be applied to similar inverse problems.

1 Introduction

Notre motivation de départ est un problème inverse de radio-astronomie, ce sera notre « fil rouge », qui consiste à estimer des cartes de paramètres physiques à partir d’images multispectrales. Dans la plupart des problèmes inverses classiques, la fonction de vraisemblance est explicite, les observations ne sont affectées que par une unique source de bruit (par exemple, un bruit additif ou multiplicatif), et admettent une dynamique de variations limitée en amplitude. Nous supposons ici qu’aucune de ces trois conditions n’est vérifiée, comme c’est le cas dans notre exemple en astrophysique. Le modèle direct est une simulation numérique, ce qui amène à une fonction de vraisemblance qui n’est ni explicite, ni log-concave. Les observables peuvent typiquement couvrir plus de dix décades, ce qui provoque des problèmes de régularité et l’absence de constante de Lipschitz globale exploitable. La complexité du processus de mesure implique que les observations sont dégradées par des bruits de différentes natures (additif et multiplicatif) et peuvent être censurées. Puisque les intensités couvrent de nombreux ordres de grandeurs, il n’est pas possible de négliger un bruit sur l’ensemble de la carte : il est donc nécessaire de traiter les deux simultanément. De plus, pour pallier l’absence de vérité terrain, systématique en radio-astronomie, nous utilisons un cadre bayésien et une méthode d’échantillonnage pour fournir estimateurs ponctuels et intervalles de crédibilité.

Un modèle réduit facilite l’évaluation de la vraisemblance et une approximation de la vraisemblance permet de traiter si-

multanément les deux types de bruits. La distribution a priori pourra inclure lissage spatial, intervalles de validité, etc. selon le contexte applicatif. À cause des grandes variations d’amplitudes du modèle direct, la constante de Lipschitz du gradient de la log-posterior est très grande. Une version préconditionnée du Metropolis Adjusted Langevin Algorithm (P-MALA) [1], qui requiert une log-posterior \mathcal{C}^2 , c’est-à-dire (c.-à-d.) deux fois continûment différentiable, permet de prendre en compte cette difficulté. De plus, la non-concavité de la log-posterior est surmontée par la combinaison de P-MALA avec un noyau d’échantillonnage Multiple-Try Metropolis (MTM) [2]. L’approche proposée paraît suffisamment générale pour rester pertinente pour des problèmes inverses similaires.

La partie 2 présente le modèle bayésien employé. La partie 3 présente l’échantillonneur utilisé pour résoudre le problème inverse. La partie 4 illustre les résultats de l’approche proposée sur notre exemple « fil rouge » en radio-astronomie.

2 Modèle bayésien

2.1 Description du problème

Nous considérons des observations $Y \in \mathbb{R}^{N \times L}$, où N est le nombre de composantes et L le nombre d’observables par composante. La carte des paramètres à estimer est notée $\Theta = (\theta_n)_{n \in [1, N]} \in \mathbb{R}^{N \times D}$. Soit $f : \mathbb{R}^D \rightarrow \mathbb{R}^L$ un modèle direct « boîte noire ». Chaque observation $y_{n, \ell}$ est affectée par un bruit additif $\epsilon_{n, \ell}^{(a)} \sim \mathcal{N}(0, \sigma_a^2)$ et par un bruit multiplicatif

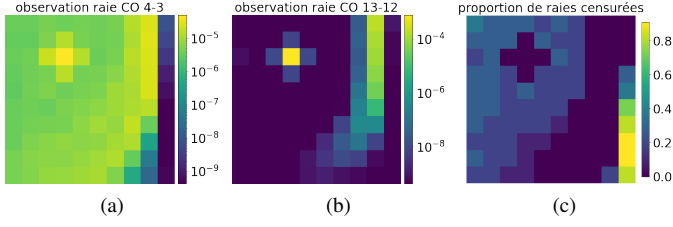


FIGURE 1 – (a), (b) Cartes synthétiques de deux raies de ^{12}CO . (c) proportion des raies censurées parmi les L considérées.

$\epsilon_{n,\ell}^{(m)} \sim \log \mathcal{N}(0, \sigma_m^2)$. En outre, la sensibilité limitée du télescope implique que les observations sont censurées en deçà d'un seuil ω , c.-à-d., pour tout $n \in \llbracket 1, N \rrbracket$, $\ell \in \llbracket 1, L \rrbracket$:

$$y_{n,\ell} = \max \left\{ \omega, \epsilon_{n,\ell}^{(m)} f_\ell(\theta_n) + \epsilon_{n,\ell}^{(a)} \right\}. \quad (1)$$

Les bruits $\epsilon_{n,\ell}^{(a)}$ et $\epsilon_{n,\ell}^{(m)}$ sont supposés indépendants et de variances respectives σ_a^2 and σ_m^2 connus. Comparé à d'autres travaux de la littérature [3], la contrainte $\mathbb{E}[\epsilon_{n,\ell}^{(m)}] = 1$ n'est pas considérée dans ce travail. Une telle approche permet de définir les vraisemblances approchées de la partie 2.3 à partir des deux premiers moments de (1).

Dans notre application « fil rouge », Y est une image multispectrale dont chaque pixel contient l'intensité de L raies spectrales de ^{12}CO , voir Fig. 1. Les écarts-type des bruits valent $\sigma_a = 1,4 \times 10^{-10}$ erg cm $^{-2}$ s $^{-1}$ sr $^{-1}$ et $\sigma_m = \log(1.1)$ pour des intensités variant typiquement entre 10^{-18} et 10^{-3} . Le modèle direct utilisé est encodé par le code Meudon PDR [4], qui calcule des intensités de raies d'émission à partir de paramètres physiques. En particulier, la Fig. 1c montre la proportion de raies censurées pour chacun des pixels.

2.2 Réduction de modèle

Dans l'application astrophysique considérée, pour un vecteur de paramètres physiques $\theta \in \mathbb{R}^D$, une évaluation du code Meudon PDR [4] prend quelques heures. Pour éviter de recourir à des approches telles que les algorithmes Approximate Bayesian Computation (ABC) [5] qui nécessiteraient de nombreuses évaluations du code Meudon PDR, nous l'approchons par une fonction explicite facile à évaluer numériquement. Des considérations physiques nous amène à supposer f_ℓ strictement positive et \mathcal{C}^∞ pour tout $\ell \in \llbracket 1, L \rrbracket$. Pour calculer efficacement les gradients du modèle approché, nous proposons une approximation polynomiale sur le log des intensités :

$$\forall \ell, \quad \tilde{f}_\ell(\theta) = \exp \left[\tilde{P}_\ell(\theta) \right]. \quad (2)$$

Les coefficients des polynômes \tilde{P}_ℓ sont estimés avec une régression avec régularisation de Tikhonov sur une grille de 1362 simulations. Avec une validation croisée *leave one out*, nous estimons une erreur moyenne d'un facteur 1.09 (en échelle linéaire). Guidés par l'application, nous définissons un ensemble de validité pour $\theta : \mathcal{C} = [l_1, u_1] \times \dots \times [l_D, u_D] \subset \mathbb{R}^D$, où $l_d, u_d \in \mathbb{R}$ sont des bornes sur θ_d .

2.3 Vraisemblance avec mélange de bruits

Pour simplifier la présentation, seul le cas non censuré est détaillé. À notre connaissance, aucune approche MCMC de la littérature ne permet de prendre en compte les deux bruits simultanément et de façon exacte. En général, une vraisemblance approchée purement additive ou multiplicative est considérée [3]. Le bruit additif (resp. multiplicatif) dans (1) peut être négligé lorsque $f_\ell(\theta) \rightarrow +\infty$ (resp. lorsque $f_\ell(\theta) \rightarrow 0$). Nous considérons donc une combinaison de deux approximations de la vraisemblance : l'une additive et l'autre multiplicative. En procédant de manière similaire à [3], la première s'écrit

$$y_{n,\ell} \simeq \tilde{f}_\ell(\theta_n) + e_{n,\ell}^{(a)}, \quad e_{n,\ell}^{(a)} \sim \mathcal{N}(m_{a,n,\ell}, s_{a,n,\ell}^2), \quad (3)$$

où $m_{a,n,\ell}$ et $s_{a,n,\ell}^2$ correspondent aux deux premiers moments du mélange de bruits de (1). La seconde est donnée par

$$y_{n,\ell} \simeq e_{n,\ell}^{(m)} \tilde{f}_\ell(\theta_n), \quad e_{n,\ell}^{(m)} \sim \log \mathcal{N}(m_{m,n,\ell}, s_{m,n,\ell}^2). \quad (4)$$

Ces approximations additive et multiplicative définissent deux vraisemblances, notées $\pi^{(a)}(y_{n,\ell}|\cdot)$ et $\pi^{(m)}(y_{n,\ell}|\cdot)$. La vraisemblance d'origine $\pi(y_{n,\ell}|\cdot)$ est approchée par

$$\tilde{\pi}(y_{n,\ell}|\theta_n) \propto \pi^{(a)}(y_{n,\ell}|\theta_n)^{\lambda(\theta_n)} \pi^{(m)}(y_{n,\ell}|\theta_n)^{1-\lambda(\theta_n)}, \quad (5)$$

où $\lambda \in \mathcal{C}^2(\mathbb{R}^D, [0, 1])$ est une sigmoïde qui contrôle la combinaison des deux approximations et vérifie : $\lambda(\theta) \rightarrow 1$ quand $\tilde{f}_\ell(\theta) \rightarrow 0$, et $\lambda(\theta) \rightarrow 0$ quand $\tilde{f}_\ell(\theta) \rightarrow +\infty$. Cette sigmoïde est paramétrée par deux scalaires $a > 0$ et $\alpha_f > 1$, qui déterminent respectivement la position et la vitesse de transition. Lorsque $a \rightarrow 0$ (resp. quand $a \rightarrow +\infty$), seule l'approximation multiplicative (resp. additive) est considérée. De même, plus α_f est grand, plus la transition est rapide.

2.4 Minimisation de l'erreur d'approximation

Pour évaluer la pertinence de l'approximation proposée à (σ_a, σ_m) et (a, α_f) fixés et la comparer à des approximations purement additive ou multiplicative, nous proposons le critère quantitatif suivant. Pour chaque raie ℓ , nous utilisons la distance de Kolmogorov-Smirnov $D_{\text{K-S}}^{(\ell)}$:

$$D_{\text{K-S}}^{(\ell)}(\pi, \tilde{\pi}_{(a,\alpha_f)}) : \theta \mapsto \sup_{y \in \mathbb{R}} \left| F^{(\ell)}(y|\theta) - \tilde{F}_{(a,\alpha_f)}^{(\ell)}(y|\theta) \right| \quad (6)$$

où $F^{(\ell)}$ (resp. $\tilde{F}_{(a,\alpha_f)}^{(\ell)}$) est la fonction de répartition de la marginale de la raie ℓ pour la vraisemblance cible (resp. l'approximation proposée). En supposant que θ suit une loi uniforme sur \mathcal{C} , nous considérons :

$$g(a, \alpha_f) = \frac{1}{L} \sum_{\ell=1}^L \mathbb{E} \left[D_{\text{K-S}}^{(\ell)}(\pi, \tilde{\pi}_{(a,\alpha_f)}) (\theta) \right]. \quad (7)$$

Les valeurs de (a, α_f) sont ajustées en minimisant g . La Fig. 2 illustre l'évolution de g avec a et α_f dans l'exemple considéré. En particulier, l'erreur d'approximation associée au meilleur couple (a, α_f) est environ 8 fois (resp. 20 fois) inférieure à celle d'une approximation purement gaussienne (resp. lognormale), ce qui prouve l'intérêt de notre approche.

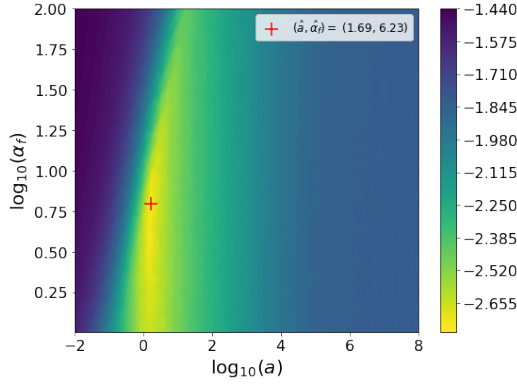


FIGURE 2 – Évolution de $\log_{10} g$ définie dans (7) pour l'exemple « fil rouge ». $a \rightarrow +\infty$ correspond à un modèle purement additif, et $a \rightarrow 0$ à un modèle purement multiplicatif. La croix rouge indique le minimum de $\log_{10} g$.

2.5 Régularisation et distribution a posteriori

La loi a priori suivante est combinée avec la vraisemblance approchée proposée. Un lissage spatial traduit des considérations physiques. Puisque \tilde{f} est définie à partir d'une grille sur \mathcal{C} , les paramètres de chaque pixel sont supposés appartenir à cet ensemble : $\Theta \in \mathcal{C}^N$. La prior considérée est ainsi de la forme

$$\pi(\Theta, \beta) \propto \exp(-h(\Theta, \beta) - \iota_{\mathcal{C}^N}(\Theta)), \quad (8)$$

où $h(\Theta, \beta)$ est pour l'exemple considéré une régularisation ℓ_2 au carré sur les cartes de gradients de chaque paramètre physique ; $\beta \in \mathbb{R}^D$ correspond aux paramètres de cette régularisation ; $\iota_{\mathcal{C}^N}$ est la fonction indicatrice sur l'ensemble de validité complet \mathcal{C}^N : $\iota_{\mathcal{C}^N}(\Theta) = 0$ quand $\Theta \in \mathcal{C}^N$ et $+\infty$ sinon. Pour assurer la régularité \mathcal{C}^2 nécessaire au noyau P-MALA, nous utilisons une approximation lisse de $\iota_{\mathcal{C}^N}$ dont on contrôle l'erreur. L'approximation de $\pi(\Theta, \beta)$ correspondante est notée $\tilde{\pi}(\Theta, \beta)$. La posterior obtenue en combinant (5) et (8), après marginalisation des hyperparamètres, est de la forme :

$$\pi(\Theta|Y) \propto \left[\prod_{n=1}^N \prod_{\ell=1}^L \tilde{\pi}(y_{n,\ell}|\theta_n) \right] \int \tilde{\pi}(\Theta, \beta) d\beta. \quad (9)$$

3 Échantillonneur MCMC

Dans la mesure où le modèle direct approché \tilde{f} et ses dérivées couvrent de nombreuses décades, les constantes de Lipschitz des gradients partiels de la log-posterior sont très grandes. C'est une difficulté majeure pour des algorithmes d'échantillonnage classiques comme MALA [6], dont le pas est inversement proportionnel à cette constante de Lipschitz. Une solution consiste à utiliser un pas exploitant la géométrie locale de la log-posterior au voisinage de l'itéré considéré. Pour ce faire, nous utilisons une version préconditionnée de MALA, P-MALA [1] avec un préconditionneur RMSProp. Ce noyau a déjà été utilisé dans la littérature [7].

Une seconde difficulté majeure est la non log-concavité de la posterior, potentiellement multimodale. Des échantillonneurs

comme Metropolis-Hastings (MH) [8], MALA ou P-MALA ne parviennent pas à explorer efficacement de telles distributions et restent bloqués dans un minimum local. Des échantillonneurs alternatifs ont été proposés (darting MC [9], wormhole MC [10], etc.), mais la plupart sont basés sur l'interaction entre plusieurs chaînes de Markov ou une identification des modes avant ou pendant l'échantillonnage, ce qui est coûteux en temps de calcul. Dans le cas étudié, un noyau MH avec une distribution uniforme sur \mathcal{C}^N serait simple mais inefficace : nous améliorons ce noyau MH de trois manières. Sur une itération, au lieu de considérer l'ensemble $\mathcal{C}^N \subset \mathbb{R}^{N \times D}$, nous choisissons aléatoirement une composante $n \in \llbracket 1, N \rrbracket$ et échantillonnons sur $\mathcal{C} \subset \mathbb{R}^D$. Ensuite, nous utilisons une approche MTM (Multiple-Try Metropolis) [2] qui génère $K > 1$ candidats, en sélectionne un et effectue une étape d'acceptation-rejet. Bien qu'utiliser K candidats augmente la durée d'une itération, cela permet d'augmenter considérablement la probabilité d'acceptation. Enfin, au lieu de générer des candidats indépendants avec une loi uniforme sur \mathcal{C} , nous utilisons un algorithme de Monte Carlo Stratifié [11] qui couvre mieux cet ensemble.

À chaque itération, un de ces deux noyaux est utilisé, avec une probabilité $p \in]0, 1[$ pour le noyau MTM et $1 - p$ pour le noyau P-MALA, ce qui permet de combiner leurs avantages respectifs. Cet échantillonneur peut explorer efficacement une loi a posteriori très irrégulière et potentiellement multimodale.

4 Illustration

Le potentiel de notre approche est illustré sur notre problème astrophysique synthétique « fil rouge ». La carte des paramètres à estimer Θ , représentée en colonne 1 de la Fig. 3, contient $N = 10 \times 10$ vecteurs de paramètres $\theta_n = (\kappa, P_{th}, G_0, A_V)$. Le facteur κ , fixé à 1 sur toute la carte, est relié aux conditions d'observation (facteur de dilution, angle de vue, etc.). Les principaux paramètres d'intérêt sont la pression thermique P_{th} , l'intensité du flux radiatif UV G_0 et l'extinction visuelle A_V . Pour chaque pixel n , le modèle direct réduit \tilde{f} génère $L = 10$ intensités de raies d'émission de ^{12}CO , dont certaines sont représentées Fig. 1. Pour se placer dans des conditions réalistes inspirées de [12], les observations sont dégradées avec des bruits additif et multiplicatif correspondant à du bruit thermique et à des erreurs de calibration. L'ensemble \mathcal{C} est défini par $\kappa \in [10^{-1}, 10^1]$, $P_{th} \in [10^5, 10^9] \text{ K.cm}^{-3}$, $G_0 \in [10^0, 10^5]$, $A_V \in [1, 40] \text{ mag}$. Puisque les paramètres d'intérêts parcourent plusieurs décades, l'échantillonnage et les estimations sont faites dans l'espace logarithmique. Enfin, pour effectuer la marginalisation décrite dans (9), les paramètres de régularisation β et les paramètres Θ sont échantillonnés conjointement en suivant [13]. Puisque l'approche présentée produit des échantillons de la posterior, elle donne simultanément accès à des estimations ponctuelles et aux intervalles de crédibilité correspondant, présentés dans les colonnes 2 et 3 de la Fig. 3. Les résultats sont remarquablement cohérents avec la physique sous-jacente.

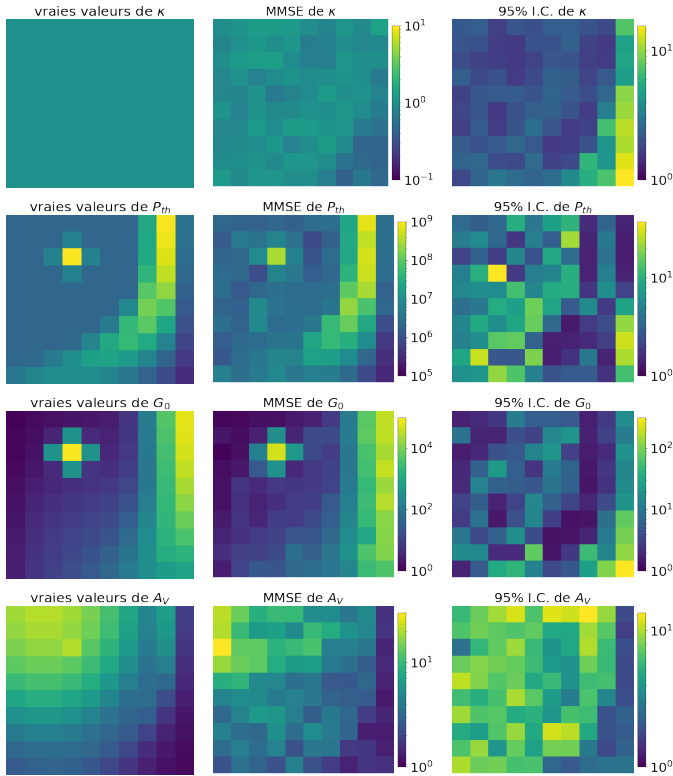


FIGURE 3 – À gauche : cartes 10×10 de Θ ; au milieu : MMSE estimé; à droite : taille des intervalles de crédibilité (I.C.) à 95%, définie comme $\log_{10} q_{97.5} - \log_{10} q_{2.5}$.

5 Conclusion

En partant d'un problème inverse de radio-astronomie impliquant un mélange potentiellement censuré de bruits additifs et multiplicatifs et des cartes d'intensités couvrant plusieurs ordres de grandeur, nous avons construit une distribution a posteriori approchée qui permet de bien estimer des paramètres physiques. Afin de résoudre ce problème inverse, décrit par une log-posterior non concave et sans régularité exploitable, un échantillonneur MCMC combinant des noyaux MTM et P-MALA a été introduit. Cette méthode fournit de bonnes estimations et des intervalles de crédibilité pertinents, comme illustré sur un exemple synthétique. La généralité de l'approche la rend adaptable à une grande variété de problèmes inverses difficiles.

Remerciements

Ce travail est soutenu par un projet 80Prime du CNRS, et par le Programme National « Physique et Chimie du Milieu Interstellaire » (PCMI) du CNRS/INSU, INC/INP, co-financé par le CEA et le CNES. Nous remercions également la Chaire IA Sherlock ANR-20-CHIA-0031-01 portée par P. Chainais, ainsi que par le programme national d'investissement d'avenir ANR-16-IDEX-0004 ULNE et la Région Hauts-de-France.

Références

- [1] T. XIFARA, C. SHERLOCK, S. LIVINGSTONE et al., « Langevin diffusions and the Metropolis-adjusted Langevin algorithm, » *Stat. & Proba. Lett.*, t. 91, p. 14-19, 2014.
- [2] J. S. LIU, F. LIANG et W. H. WONG, « The Multiple-Try method and local optimization in Metropolis sampling, » *J. of the Am. Stat. Assoc.*, t. 95, n° 449, p. 121-134, 2000.
- [3] R. NICHOLSON et J. P. KAIPIO, « An additive approximation to multiplicative noise, » *J. of Math. Im. and Vis.*, t. 62, n° 9, p. 1227-1237, 2020.
- [4] F. LE PETIT, C. NEHME, J. LE BOURLOT et al., « A model for atomic and molecular interstellar gas : the Meudon PDR Code, » *The Astrophys. J. Supp. Series*, t. 164, n° 2, p. 506-529, 2006.
- [5] M. A. BEAUMONT, W. ZHANG et D. J. BALDING, « Approximate Bayesian computation in population genetics, » *Genetics*, t. 162, n° 4, p. 2025-2035, 2002.
- [6] G. O. ROBERTS et O. STRAMER, « Langevin diffusions and Metropolis-Hastings algorithms, » *Method. & Comp. in App. Proba.*, t. 4, n° 4, p. 337-357, 2002.
- [7] C. LI, C. CHEN, D. CARLSON et al., « Preconditioned Stochastic Gradient Langevin Dynamics for Deep Neural Networks, » *Proc. of the AAAI Conf. on Artif. Intell.*, t. 30, n° 1, 2016.
- [8] W. K. HASTINGS, « Monte Carlo sampling methods using Markov chains and their applications, » *Biometrika*, t. 57, n° 1, p. 97-109, 1970.
- [9] S. AHN, Y. CHEN et M. WELLING, « Distributed and adaptive darting Monte Carlo through regenerations, » in *Proc. of 16th Int. Conf. on Artif. Intell. and Stat.*, sér. PMLR, t. 31, 2013, p. 108-116.
- [10] S. LAN, J. STREETS et B. SHAHBABA, « Wormhole Hamiltonian Monte Carlo, » in *Proc. of the Nat. Conf. on Art. Intell.*, sér. Proc. of the Nat. Conf. on Art. Intell. 2014, p. 1953-1959.
- [11] S. HABER, « A modified Monte-Carlo quadrature, » *Mathematics of Computation*, t. 20, n° 95, p. 361-368, 1966.
- [12] PETY, J., GUZMÁN, VIVIANA V., ORKISZ, JAN H. et al., « The anatomy of the Orion B giant molecular cloud : A local template for studies of nearby galaxies, » *Astro. & Astrophys.*, t. 599, A98, 2017.
- [13] M. PEREYRA, J. M. BIOCAS-DIAS et M. A. T. FIGUEIREDO, « Maximum-a-posteriori estimation with unknown regularisation parameters, » in *2015 23rd Euro. Sig. Proc. Conf. (EUSIPCO)*, 2015, p. 230-234.