Mixture of noises and sampling of non-log-concave posterior distributions

Pierre Palud, Pierre Chainais, Franck Le Petit, Emeric Bron, Maxime Vono, Lucas Einig, Miriam Garcia Santa-Maria, Mathilde Gaudel, Jan Orkisz, Victor de Souza Magalhaes, et al.

Mixture of noises and sampling of non-log-concave posterior distributions

Pierre Palud[1,2], Pierre Chainais[1], Franck Le Petit[2], Emeric Bron[2], Pierre-Antoine Thouvenin[1], Maxime Vono[3], L. Einig[4], M. Garcia Santa-Maria[4], M. Gaudel[4], J. H. Orkisz[4], V. de Souza Magalhaes[4], S. Bardeau[4], M. Gerin[4], J. R. Goicoechea[4], P. Gratier[4], V. V. Guzman[4], J. Kainulainen[4], F. Levrier[4], N. Peretto[4], J. Pety[4], A. Roueff[4], A. Sievers[4]

[1]Univ. Lille, CNRS, Centrale Lille, UMR 9189 CRIStAL, F-59000 Lille, France
[2]LERMA, Observatoire de Paris, PSL Research University, CNRS, Sorbonne Universités, 92190 Meudon, France
[3]Criteo AI Lab, Paris, France
[4]Consortium ORION-B (https://www.iram.fr/~pety/ORION-B/team.html)

***Abstract*—This work considers a challenging radio-astronomy inverse problem of physical parameter inference from multispectral observations. The forward model underlying this problem is a computationally expensive numerical simulation. In addition, the observation model mixes different sources of noise yielding a non-concave log-likelihood function. To overcome these issues, we introduce a likelihood approximation with controlled error. Given the absence of ground truth, parameter inference is conducted with a Markov chain Monte Carlo (MCMC) algorithm to provide credibility intervals along with point estimates. To this aim, we propose a new sampler that addresses the numerical challenges induced by the observation model, in particular the non-log-concavity of the posterior distribution. The efficiency of the proposed method is demonstrated on synthetic yet realistic astrophysical data. We believe that the proposed approach is very general and can be adapted to many similar difficult inverse problems.**

***Index Terms*—Inverse problem, Bayesian inference, Markov chain Monte Carlo algorithm, multiplicative noise**

I. Introduction

Molecular clouds are of particular interest for astrophysicists, since the gravitational collapse of a dense core is the first step towards star formation. Once formed, a massive star emits UV photons that heat its parent cloud and greatly affects its physical conditions. Such regions are called Photo-Dissociation Regions (PDR), and can be simulated with the Meudon PDR code [1]. The ORION-B consortium provides large maps of molecular line emission intensities of such regions [2], and future observations will be provided by the James Webb Space Telescope. The problem at the roots of this work is to infer a PDR physical parameters from those maps.

In most classic inverse problems, the likelihood function is explicit with either additive or multiplicative noise, and the observables span a limited number of decades in magnitude. This radio-astronomy problem is challenging because none of these three criteria is met. The Meudon PDR code is a highly non-linear numerical simulation that takes into account many complex physical phenomena: it leads to a non-explicit, non-concave and potentially multimodal log-likelihood function. Emission intensities can typically cover more than 10 decades in magnitude, which leads to regularity issues with no global Lipschitz constant. Due to the complexity of the measurement process and the limited sensitivity of the telescope, observations are degraded by noises of different natures (additive and multiplicative) and can be censored. Multiplicative noise dominates at very high intensities and additive noise at low intensities. However, since the measured intensities can be very small or very large over several orders of magnitude, the noise mixture must be addressed at once. In addition, there is no ground truth in radio-astronomy. In this respect, we use a Bayesian framework to provide uncertainty quantifications along with point estimates. A Monte Carlo Markov Chain (MCMC) method generates samples from the posterior distribution to form estimators along with credibility intervals.

A reduced version of the forward model tackles the expensive evaluation of the likelihood. Then we design a smooth approximate likelihood function that mixes the additive and multiplicative noises. The prior is guided by astrophysicists expertise (spatial smoothness, validity intervals). Due to the extremely wide range of amplitudes at stake in astrophysics, the Lipschitz constant of the resulting log-posterior gradient is potentially huge. To tackle this regularity issue, we use a preconditioned Metropolis Adjusted Langevin Algorithm (P-MALA) [3], [4], which requires the log-posterior to be $\mathscr{C}^2$, i.e., twice continuously differentiable. The high non-linearity of the forward model results in a non-concave log-posterior so that we combine P-MALA with an independent Multiple-Try Metropolis (MTM) algorithm [5]. Illustrations on a synthetic yet realistic astrophysical case show that this approach yields reliable estimators. We believe that the proposed approach is sufficiently general to remain relevant for a variety of difficult inverse problems that feature non-Lipschitz regularity, non-log-concavity and a mixture of noises.

The paper is organized as follows. The proposed Bayesian model is presented in Section II. The sampler used to solve the inverse problem is introduced in Section III. The method is then assessed on synthetic astrophysical data in Section IV. Conclusion and research perspectives are given in Section V.

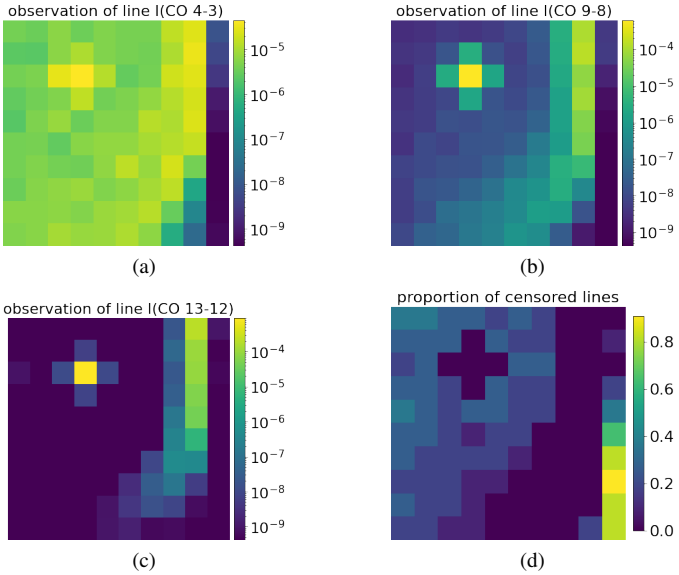

Fig. 1. Illustration of typical observation maps. (a), (b), (c): Synthetic maps of observation of some emission lines of ^{12}CO. (d) Proportion of lines with intensity below the telescope sensitivity limit ω (out of $L = 10$ emission lines of ^{12}CO). Lines intensities are given in erg cm^{-2} s^{-1} sr^{-1}.

II. Bayesian Model

A. Problem Statement

We consider the observation of feature maps $Y \in \mathbb{R}^{N\times L}$, where N is the number of components and L the number of features per component. The map of parameters to be inferred is denoted by $\mathbf{\Theta} = (\boldsymbol{\theta}_n)_{n\in[\![1,N]\!]} \in \mathbb{R}^{N\times D}$. Let $f : \mathbb{R}^D \to \mathbb{R}^L$ be a black-box forward model. Each individual observation $y_{n,\ell}$ is assumed to be corrupted by an additive noise $\epsilon^{(a)}_{n,\ell} \sim \mathcal{N}(0, \sigma_a^2)$ and a multiplicative noise $\epsilon^{(m)}_{n,\ell} \sim \log\mathcal{N}(0, \sigma_m^2)$. On top on that, the sensors have a sensitivity limit ω below which observations are censored, that is, for all $n \in [\![1, N]\!]$, $\ell \in [\![1, L]\!]$:

$$y_{n,\ell} = \max\left\{\omega, \epsilon^{(m)}_{n,\ell} f_\ell(\boldsymbol{\theta}_n) + \epsilon^{(a)}_{n,\ell}\right\}. \tag{1}$$

The noise terms $\epsilon^{(a)}_{n,\ell}$ and $\epsilon^{(m)}_{n,\ell}$ are supposed independent, with known variances σ_a^2 and σ_m^2, respectively. Note that we do not assume $\mathbb{E}[\epsilon^{(m)}_{n,\ell}] = 1$, in contrast with a usual assumption from the literature [6], [7]. This will allow the surrogate likelihood model introduced in II-C to be derived by matching the first two moments of the likelihood in (1).

Back to our motivating example in astrophysics, Y will be a multispectral image where each pixel features L molecular spectral lines of ^{12}CO, see Fig. 1. The forward model f will be encoded by the Meudon PDR code [1]. In particular, for $L = 10$ emission lines of ^{12}CO, Fig. 1d shows the proportion of censored lines below the sensitivity level $\omega = 3\sigma_a$. Further details are given in Section IV.

B. Model Reduction

When the evaluation of the likelihood is expensive, one usually uses a reduced model that is computationally cheaper. In the example, given a set of parameters $\boldsymbol{\theta} \in \mathbb{R}^D$ that define the physical conditions of a PDR, the Meudon PDR code [1] iteratively solves large systems of equations of thermal balance, radiative transfer and chemistry evolution to compute molecular emission line intensities. It is therefore a positive function, i.e. for all $\ell \in [\![1, L]\!]$ and $\boldsymbol{\theta} \in \mathbb{R}^D$, $f_\ell(\boldsymbol{\theta}) > 0$. However, it is very expensive to evaluate (from 2 to 6 hours per simulation). To avoid the use of costly likelihood-free methods such as Approximate Bayesian Computation (ABC) algorithms [8], which would require several evaluations of the Meudon PDR code, we approximate it with a closed-form function. To reflect physical considerations, all f_ℓ are assumed to be smooth: any approximation $\tilde{f}_\ell$ of f_ℓ such that the gradients are easy to compute would thus be suitable. For the sake of simplicity in this paper, we propose a polynomial approximation of the log-intensities:

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

The coefficients of the polynomials $\tilde{P}_\ell$ are estimated using ridge regression based on a grid of 1362 simulations. Using a leave-one-out cross-validation, we estimate an average error of a factor 1.09 (in linear scale) for a polynomial $\tilde{P}_\ell$ of degree 8. Guided by the application, we define a validity set for $\boldsymbol{\theta}$: $\mathcal{C} = [l_0, u_0] \times \ldots \times [l_{D-1}, u_{D-1}] \subset \mathbb{R}^D$, where $l_d, u_d \in \mathbb{R}$ are lower and upper bounds of θ_d, respectively.

C. Likelihood with Mixture of Noises

For the sake of simplicity, we introduce here the uncensored case only, i.e., the right side of (1). To the best of the authors' knowledge, no model from the literature can exactly handle this mixture of noises with MCMC methods. Exact marginalization is intractable, and the separable model [9] is dedicated to linear inverse problems (e.g., deblurring). In general, a single global approximation using either purely additive or multiplicative model is considered in the likelihood [6]. The additive noise (resp. the multiplicative noise) involved in (1) can be neglected when $f_\ell(\boldsymbol{\theta}) \to +\infty$ (resp. when $f_\ell(\boldsymbol{\theta}) \to 0$). We thus consider a mixture of a Gaussian and a lognormal approximation of the overall distribution of $y_{n,\ell}$, controlled by $\tilde{f}_\ell(\boldsymbol{\theta}_n)$. The former gives a better description of low intensity observations, whereas the latter better approximates high intensity observations. In the spirit of [6], the Gaussian approximation is given by

$$y_{n,\ell} \simeq \tilde{f}_\ell(\boldsymbol{\theta}_n) + e^{(a)}_{n,\ell}, \quad e^{(a)}_{n,\ell} \sim \mathcal{N}(m_{a,n,\ell}, s^2_{a,n,\ell}), \tag{3}$$

where $m_{a,n,\ell}$ and $s^2_{a,n,\ell}$ are obtained by matching the first two moments with (1). Proceeding similarly for the lognormal approximation

$$y_{n,\ell} \simeq e^{(m)}_{n,\ell} \tilde{f}_\ell(\boldsymbol{\theta}_n), \quad e^{(m)}_{n,\ell} \sim \log\mathcal{N}(m_{m,n,\ell}, s^2_{m,n,\ell}). \tag{4}$$

The resulting additive and multiplicative approximations define two likelihood functions denoted by $\pi^{(a)}(y_{n,\ell}|\cdot)$ and $\pi^{(m)}(y_{n,\ell}|\cdot)$, respectively. Then, we approximate the true likelihood function $\pi(y_{n,\ell}|\cdot)$ by combining (3) and (4)

$$\tilde{\pi}(y_{n,\ell}|\boldsymbol{\theta}_n) \propto \pi^{(a)}(y_{n,\ell}|\boldsymbol{\theta}_n)^{\lambda(\boldsymbol{\theta}_n)} \pi^{(m)}(y_{n,\ell}|\boldsymbol{\theta}_n)^{1-\lambda(\boldsymbol{\theta}_n)}, \tag{5}$$

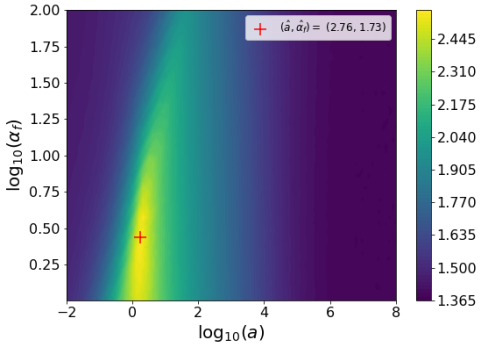

Fig. 2. Evolution of $-\log_{10} g$ defined in (7) for the target astrophysical example. We consider $L = 10$ emission lines with $\sigma_a = 1.4 \times 10^{-10}$ erg cm^{-2} s^{-1} sr^{-1} and $\sigma_m = \log(1.3)$. Note that $a \to +\infty$ corresponds to a purely additive approximation, $a \to 0$ to a purely multiplicative approximation; the red cross indicates the minimum of g.

where $\lambda \in \mathscr{C}^2(\mathbb{R}^D, [0,1])$ is an S-shaped function that controls the model mixing and verifies $\lambda(\boldsymbol{\theta}) \to 1$ when $\tilde{f}_\ell(\boldsymbol{\theta}) \to 0$, and $\lambda(\boldsymbol{\theta}) \to 0$ when $\tilde{f}_\ell(\boldsymbol{\theta}) \to +\infty$. For flexibility, λ is parametrized by two scalars $a > 0$ and $\alpha_f > 1$ that respectively control the location and speed of the transition between additive and multiplicative approximations. When $a \to 0$ (resp. when $a \to +\infty$), one only considers a multiplicative model (resp. an additive model). Similarly, the higher α_f, the steeper the transition.

D. Minimization of the likelihood misspecification

To evaluate the relevance of the proposed likelihood approximation for (σ_a, σ_m) and (a, α_f) fixed and compare it to purely additive or multiplicative approximations, we define the following model-dependent quantitative criterion. First, we consider for each line the Kolmogorov-Smirnov distance $D^{(\ell)}_{\text{K-S}}$ between the true marginal likelihood and the proposed approximation:

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

where $F^{(\ell)}$ (resp. $\tilde{F}_{(a,\alpha_f)}$) is the cumulative density function (cdf) of the marginal of feature ℓ for the true likelihood (resp. the proposed approximation). Assuming $\boldsymbol{\theta}$ follows a uniform distribution on $\mathcal{C}$, we consider:

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

The values of (a, α_f) are adjusted by minimizing g. Fig. 2 illustrates the evolution of this criterion with a and α_f for the target astrophysical application. In particular, one can see that the approximation error of the best (a, α_f) is more than 10 times lower than those of pure additive ($a \to +\infty$) or multiplicative ($a \to 0$) models, which demonstrates the value of our approach.

E. Regularization and posterior distribution

To perform Bayesian inference, the following prior information is combined with the approximate likelihood function. First, a spatial smoothness prior on the parameters is considered to reflect physical considerations. Second, since $\tilde{f}$ is defined from a grid on $\mathcal{C}$, the true parameters associated with each pixel are assumed to belong to this set, i.e., $\boldsymbol{\Theta} \in \mathcal{C}^N$. The considered prior is thus of the form

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

where $h(\boldsymbol{\Theta}, \boldsymbol{\beta})$ is a squared ℓ_2 regularization on the gradients of the maps of each physical parameter; $\boldsymbol{\beta} \in \mathbb{R}^D$ contains the corresponding weights; $\iota_{\mathcal{C}^N}$ is the indicator function of the full parameter validity set $\mathcal{C}^N$, i.e., $\iota_{\mathcal{C}^N}(\boldsymbol{\Theta}) = 0$ when $\boldsymbol{\Theta} \in \mathcal{C}^N$ and $+\infty$ otherwise. To ensure the log of the prior is $\mathscr{C}^2$, we approximate $\iota_{\mathcal{C}^N}$ with a smooth penalty function $\tilde{\iota}_{\mathcal{C}^N}$, defined with a polynomial of degree 4 and a scale parameter that controls the approximation error. We denote $\tilde{\pi}(\boldsymbol{\Theta}, \boldsymbol{\beta})$ the corresponding approximation of $\pi(\boldsymbol{\Theta}, \boldsymbol{\beta})$. The posterior distribution combining (5) and (8) is of the form

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

III. MCMC Sampler

The forward model spans several decades and the log-posterior is non-concave: the resulting posterior distribution is thus challenging to sample, especially for large numbers of pixels. To address this issue, we introduce a new sampler relying on a random mixture of two transition kernels, with mixing parameter $p \in]0,1[$: the MTM kernel is active with probability p while the P-MALA kernel is active with probability $1-p$. The former detects modes of the posterior distribution on-the-fly and allows jumps between them, and the latter efficiently explores the local geometry. Note that, since both kernels satisfy the detailed balance criterion, so does their mixture. This property ensures that the posterior distribution is the stationary distribution of any Markov chain drawn using the proposed sampler.

A. P-MALA Sampling Kernel

In the considered problem, the forward model $\tilde{f}$ and its derivatives cover many decades, which causes the gradient of the likelihood function to have potentially huge Lipschitz constant. This is problematic for classic sampling algorithms like MALA [10], in that it requires a step size valid globally, which is inversely proportional to this Lipschitz constant. To address this issue, a solution consists in using a step-size which better reflects the local geometry of the log-posterior distribution around the current iterate. To do so, we consider the P-MALA kernel [3], [4], which hinges on a position-dependent preconditioning. We used the RMSProp preconditioner, which has already been applied in the MCMC literature [11]. It relies on three hyperparameters that directly or indirectly define the considered neighborhood: $\alpha_0 \in]0,1[$ (the exponential decay of the pre-conditioner memory), $\epsilon_0 > 0$ (a raw step size) and $\lambda_0 > 0$ (a damping parameter that avoids zero-divisions).

B. MTM Sampling Kernel

A second major difficulty coming from the non-linearity of f is the non-convexity of the negative log pdf of the posterior, which can even be multimodal. Sampling algorithms such as the Metropolis-Hastings (MH) [12], MALA and P-MALA algorithms fail to efficiently explore such distributions, as they get stuck in one local minimum. Alternative samplers dedicated to multimodal distributions exist (Equi-Energy Sampler [13], evolutionary MC [14], [15], darting MC [16], wormhole MC [17], etc.), but most of them are either based on interacting Markov chains or modes fitting (prior to sampling or on the fly), which is computationally intensive.

In the considered case, an independent MH kernel with a uniform proposal distribution on $\mathcal{C}^N$ would be a simple yet inefficient sampler. We improve this basic sampler in three ways. First, instead of considering $\mathcal{C}^N \subset \mathbb{R}^{N\times D}$, at each step we sample one $n \in [\![1, N]\!]$ with uniform weights and then sample in $\mathcal{C} \subset \mathbb{R}^D$ in a component-wise MH fashion [18]. Second, instead of drawing only one candidate that would have a very low acceptance probability, we use a MTM approach [5] that generates K candidates in $\mathcal{C}$, selects one based on the posterior pdf and then performs an accept-reject step. Though using $K > 1$ candidates increases the computational time needed for one step, it also dramatically improves the acceptance probabilities. Third, instead of using independent samples of the uniform distribution in our MTM kernel, we use a stratified Monte Carlo (SMC) algorithm [19] that better covers $\mathcal{C}$.

IV. Experiments

A. Experimental setup

The potential of the proposed approach is illustrated on a synthetic astrophysical problem. The purpose is to infer physical parameters of a molecular cloud from its observation simulated from the above-mentioned reduced polynomial model $\tilde{f}$ of the Meudon PDR code. For each pixel, the forward model $\tilde{f}$ generates $L = 10$ molecular emission lines of ^{12}CO (from $J = 4-3$ to $J = 13-12$), see Fig. 1. Note that excited ^{12}CO lines were at the core of the observations of the Herschel satellite in the Milky Way and other galaxies; astrophysicists understand well ^{12}CO dominant physical processes and inference capabilities. The parameters map $\boldsymbol{\Theta}$ consists of $N = 10 \times 10$ parameter vectors $\boldsymbol{\theta}_n = (\kappa, P_{th}, G_0, A_V)$. The geometrical scaling factor κ is related to the conditions of observation (beam dilution, angle of view); here $\kappa = 1$ over the whole map. The main physical parameters of interest for each pixel are the thermal pressure P_{th}, the intensity of the UV radiative field G_0, the visual extinction along the line of sight A_V (related to the cloud depth). From Fig. 3, 1st column, one can see that $\boldsymbol{\Theta}$ contains a wide range of physical environments.

Both additive and multiplicative noises, which correspond respectively to thermal noise and calibration errors, affect the simulated observations. The intensity of the additive noise is fixed to the median of the $J = 13-12$ line, i.e., $\sigma_a = 1.4 \times 10^{-10}$ erg cm^{-2} s^{-1} sr^{-1}, and the level of the multiplicative noise is $\sigma_m = \log(1.3)$. These values are consistent with real datasets [2]. Considering the convex envelope of the grid used to define $\tilde{f}$, we define $\mathcal{C}$ such that $\kappa \in [10^{-1}, 10^{1}]$, $P_{th} \in [10^{5}, 10^{9}]$ K.cm^{-3}, $G_0 \in [10^{0}, 10^{5}]$, $A_V \in [1, 40]$ mag. Due to the wide range of the considered astrophysical parameters, the sampling is performed on the logarithmic quantities. Estimators are defined in the log space as well. The selection probability of the MTM kernel is set to $p = 0.2$, see section III, and it uses $K = 6^D = 1296$ candidates. The parameters of the P-MALA kernel are $\epsilon_0 = 2 \times 10^{-3}$, $\lambda_0 = 10^{-5}$ and $\alpha_0 = 0.995$. A set of $N_{\text{sim}} = 5$ Markov chains is run with 10^5 iterations including 2×10^4 burn-in samples. To make our approach auto-tuned, the regularization weights $\boldsymbol{\beta}$ in (8) are also sampled together with $\boldsymbol{\Theta}$ following [20].

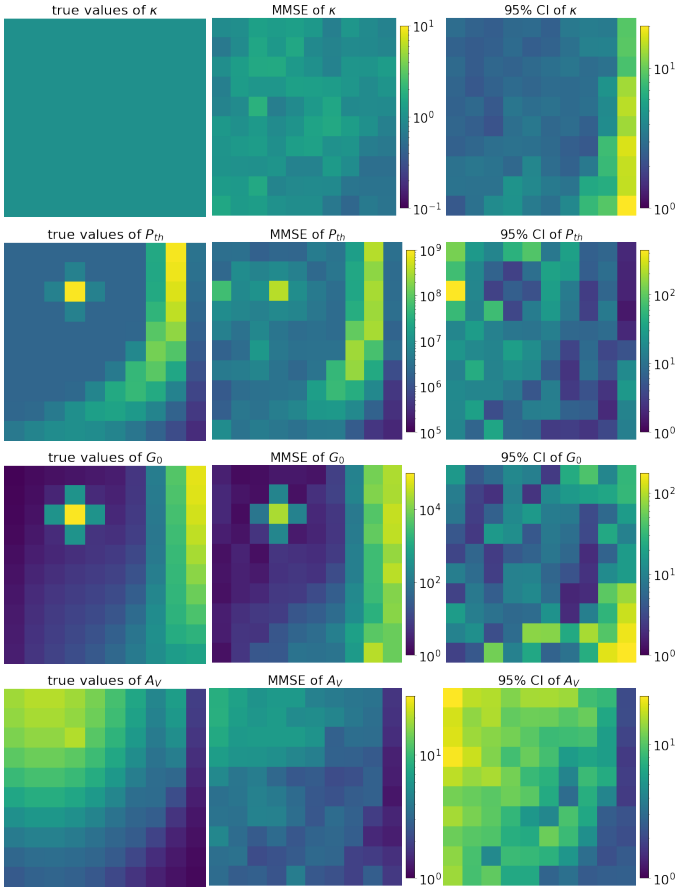

Fig. 3. Inference Results. The first column represents the true $N = 10 \times 10$ maps of $\boldsymbol{\Theta}$, the second column our MMSE estimation and the third shows the size of the 95% credibility intervals on estimations $\log_{10} q_{97.5} - \log_{10} q_{2.5}$, with q_u the percentile u in linear scale.

B. Results

Since the proposed approach produces samples from the posterior distributions, it simultaneously yields an estimate of the physical parameters and the corresponding credibility intervals (C.I.). The 2nd and 3rd columns of Fig. 3 show the point estimates of each $\theta_{n,d}$ as well as the sizes of their 95% C.I.. Qualitatively, they are both remarkably consistent with the underlying physics. First, it is a well known fact

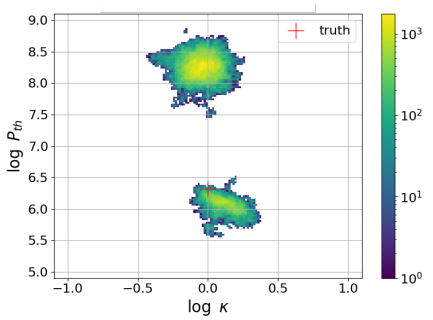

Fig. 4. Marginal 2D histogram of (κ, P_{th}) for the pixel $n = 20$. The sampler successfully explored two distinct modes thanks to the proposed MTM/P-MALA mixture.

for astrophysicists that this inverse problem, with ^{12}CO lines observations, is very ill-posed for A_V above some threshold. In Fig. 3, the size of the A_V C.I. almost covers the whole valid interval for high true values. For κ, P_{th} and G_0, the estimation problem is very ill-posed in the low SNR limit only. Our synthetic maps were designed to include a region at the bottom right with most spectral lines censored, see Fig. 1d. Even there, the MMSE estimations are quite close to the true values, though the credibility intervals are larger than on the rest of the map, as expected.

Let us now focus on the pixel corresponding to $n = 20$ that has a C.I. for the pressure P_{th} that is 10 times larger than for the rest of the map (see Fig. 3, 3rd column, 2nd row). Indeed it appears that the marginal distribution on this pixel is bimodal, see Fig. 4: it admits two relatively distant modes so that the C.I. includes very low probability regions. Such situations lead to an artificially overestimated uncertainty. This is not really a problem since the physical expertise (not used here) would permit to eliminate the physically inconsistent mode in a simple manner. In principle, one could expect the prior distribution to eliminate this non-physical mode by smoothing it out. However, the true pressure map is not as smooth as for the other parameters. Therefore the estimated regularization weight is lower for the pressure, and the prior is not strong enough to exclude the higher pressure mode.

V. Conclusion

Motivated by a real inverse problem from astrophysics, a new likelihood approximation has been derived to handle mixtures of additive and multiplicative noises. It enables physical parameters to be better inferred from observation maps covering a wide range of intensities as well as censored regions. To solve this inverse problem that involves a non-concave log-posterior with non-Lipschitz regularity, an original MCMC sampler has been introduced. The proposed approach deals with both issues by mixing a preconditioned MALA and an MTM algorithm that even handles multimodal local landscapes. The method provides good estimates as well as relevant credibility intervals. An application to synthetic yet realistic radio-astronomy data shows the efficiency of the proposed approach. We believe that it is sufficiently general to be adapted to many similar difficult inverse problems. Perspectives include the application of this approach to real star forming region observations, including real data such as the IRAM Orion B Large Program maps [2].

Acknowledgment

This work was partly supported by the ANR project "Chaire IA Sherlock" ANR-20-CHIA-0031-01 hold by P. Chainais, and by the national support within the *programme d'investissements d'avenir* ANR-16-IDEX-0004 ULNE and Région HDF.

References

[1] F. Le Petit, C. Nehme, J. Le Bourlot, and E. Roueff, "A model for atomic and molecular interstellar gas: the Meudon PDR Code," *The Astrophys. J. Supp. Series*, vol. 164, no. 2, pp. 506–529, 2006.

[2] Pety, J. *et al.*, "The anatomy of the Orion B giant molecular cloud: A local template for studies of nearby galaxies," *Astro. & Astrophys.*, vol. 599, p. A98, 2017.

[3] M. Girolami and B. Calderhead, "Riemann manifold Langevin and Hamiltonian Monte Carlo methods," *J. Royal Stat. Soc.: Series B*, vol. 73, no. 2, pp. 123–214, 2011.

[4] T. Xifara, C. Sherlock, S. Livingstone, S. Byrne, and M. Girolami, "Langevin diffusions and the Metropolis-adjusted Langevin algorithm," *Stat. & Proba. Lett.*, vol. 91, pp. 14–19, 2014.

[5] J. S. Liu, F. Liang, and W. H. Wong, "The Multiple-Try method and local optimization in Metropolis sampling," *J. of the Am. Stat. Assoc.*, vol. 95, no. 449, pp. 121–134, 2000.

[6] R. Nicholson and J. P. Kaipio, "An additive approximation to multiplicative noise," *J. of Math. Im. and Vis.*, vol. 62, no. 9, pp. 1227–1237, 2020.

[7] S. Durand, J. Fadili, and M. Nikolova, "Multiplicative noise removal using l1 fidelity on frame coefficients," *J. of Math. Im. and Vis.*, vol. 36, no. 3, pp. 201–226, 2010.

[8] M. A. Beaumont, W. Zhang, and D. J. Balding, "Approximate Bayesian computation in population genetics," *Genetics*, vol. 162, no. 4, pp. 2025–2035, 2002.

[9] Y. Huang, M. Ng, and T. Zeng, "The convex relaxation method on deconvolution model with multiplicative noise," *Comm. in Comput. Phys.*, vol. 13, no. 4, pp. 1066–1092, 2013.

[10] G. O. Roberts and O. Stramer, "Langevin diffusions and Metropolis-Hastings algorithms," *Method. & Comp. in App. Proba.*, vol. 4, no. 4, pp. 337–357, 2002.

[11] C. Li, C. Chen, D. Carlson, and L. Carin, "Preconditioned stochastic gradient Langevin dynamics for deep neural networks," *Proc. of the AAAI Conf. on Artif. Intell.*, vol. 30, no. 1, 2016.

[12] W. K. Hastings, "Monte Carlo sampling methods using Markov chains and their applications," *Biometrika*, vol. 57, no. 1, pp. 97–109, 1970.

[13] S. C. Kou, Q. Zhou, and W. H. Wong, "Equi-energy sampler with applications in statistical inference and statistical mechanics," *The Annals of Stat.*, vol. 34, no. 4, pp. 1581–1619, 2006.

[14] F. Liang and W. H. Wong, "Evolutionary Monte Carlo: applications to Cp model sampling and change point problem," *Statistica Sinica*, vol. 10, no. 2, pp. 317–342, 2000.

[15] B. Hu and K.-W. Tsui, "Distributed evolutionary Monte Carlo for Bayesian computing," *Comput. Stat. & Data Anal.*, vol. 54, no. 3, pp. 688–697, 2010.

[16] S. Ahn, Y. Chen, and M. Welling, "Distributed and adaptive darting monte carlo through regenerations," in *Proc. of 16th Int. Conf. on Artif. Intell. and Stat.*, ser. PMLR, vol. 31, 2013, pp. 108–116.

[17] S. Lan, J. Streets, and B. Shahbaba, "Wormhole hamiltonian monte carlo," in *Proc. of the Nat. Conf. on Art. Intell.*, ser. Proc. of the Nat. Conf. on Art. Intell., 2014, pp. 1953–1959.

[18] A. A. Johnson, G. L. Jones, and R. C. Neath, "Component-wise Markov chain Monte Carlo: uniform and geometric ergodicity under mixing and composition," *Stat. Science*, vol. 28, no. 3, pp. 360–375, 2013.

[19] S. Haber, "A modified Monte-Carlo quadrature," *Mathematics of Computation*, vol. 20, no. 95, pp. 361–368, 1966.

[20] M. Pereyra, J. M. Bioucas-Dias, and M. A. T. Figueiredo, "Maximum-a-posteriori estimation with unknown regularisation parameters," in *2015 23rd Euro. Sig. Proc. Conf. (EUSIPCO)*, 2015, pp. 230–234.