# A Metropolis-Adjusted Langevin Algorithm for Sampling Jeffreys Prior

Yibo Shi, Braghadeesh Lakshminarayanan and Cristian R. Rojas

## I. Introduction

### A. Background

Inference and estimation are fundamental aspects of statistics, system identification, and machine learning. For most inference problems, prior knowledge is available on the system to be modeled, and Bayesian analysis is a natural framework to impose such prior information in the form of a prior distribution. However, in many situations, coming out with a fully specified prior distribution is not easy, as prior knowledge might be too vague, so practitioners prefer to use a prior distribution that is as "ignorant" or "uninformative" as possible, in the sense of not imposing subjective beliefs, while still supporting reliable statistical analysis.

*Jeffreys prior* [1] is an appealing uninformative prior because it offers two important benefits: (i) it is invariant under any re-parameterization of the model, (ii) it encodes the intrinsic geometric structure of the parameter space through the *Fisher information matrix* (FIM), which enhances the diversity of parameter samples. Despite these benefits, drawing samples from Jeffreys prior is challenging. In this paper, we propose a sampling scheme using the *Metropolis-Adjusted Langevin Algorithm* (MALA) [2] that enables sampling from Jeffreys prior, and provide numerical illustrations of our approach through an example.

### B. Problem Statement

Consider a family of probability distributions $\{p(\cdot;\boldsymbol{\theta})\colon \boldsymbol{\theta} \in \Theta\}$ defined on a sample space $\mathcal{Y}$, where $\Theta \subseteq \mathbb{R}^d$ is the parameter space.[1] The constrained parameter space for sampling is defined as $\Theta_c \subseteq \Theta$. The FIM at a given parameter value $\boldsymbol{\theta}$, denoted by $\boldsymbol{J}_{\boldsymbol{\theta}}$, is defined as

$$\boldsymbol{J}_{\boldsymbol{\theta}} \;=\; \mathbb{E}_{y\sim p(\cdot;\boldsymbol{\theta})}\Big[\nabla_{\boldsymbol{\theta}} \ln p(y;\boldsymbol{\theta})\, \nabla_{\boldsymbol{\theta}^\top} \ln p(y;\boldsymbol{\theta})\Big], \quad (1)$$

where $y \in \mathcal{Y}$ denotes the observations, $\nabla_{\boldsymbol{\theta}}$ denotes the gradient with respect to $\boldsymbol{\theta}$, and $\mathbb{E}_{y\sim p(\cdot;\boldsymbol{\theta})}$ the expectation operator w.r.t $y$. Jeffreys prior, denoted $\pi(\boldsymbol{\theta})$, is then given (up to a constant factor) by

$$\pi(\boldsymbol{\theta}) \;\propto\; \sqrt{\det\big(\boldsymbol{J}_{\boldsymbol{\theta}}\big)}. \quad (2)$$

The difficulty of sampling from Jeffreys prior includes: (i) the cost of computing or approximating $\boldsymbol{J}_{\boldsymbol{\theta}}$, and that (ii) $\det\big(\boldsymbol{J}_{\boldsymbol{\theta}}\big)$ may not integrate to a known normalizing constant.

This paper is based on the result in arXiv:2504.06372, which was submitted to the 2025 IEEE Conference on Decision and Control (CDC). The authors are with the Division of Decision and Control Systems, KTH Royal Institute of Technology, SE-100 44 Stockholm, Sweden. Emails: {yibos, blak, crro}@kth.se

[1]Throughout this paper, we use boldface fonts (e.g., $\boldsymbol{\theta}$) to refer to vector or matrix variables and normal fonts (e.g., $\theta$) for scalar variables.

## II. Methodology

The MALA is an advanced *Langevin-based Monte Carlo* (LMC) variant. In this paper, we generate samples distributed according to

$$\boldsymbol{\theta} \sim \pi(\boldsymbol{\theta}) \propto \exp(-V(\boldsymbol{\theta})) \quad (3)$$

by simulating the following *Langevin stochastic differential equation* (SDE):

$$d\boldsymbol{\theta}_t = -\nabla_{\boldsymbol{\theta}} V(\boldsymbol{\theta}_t)dt + \sqrt{2}d\boldsymbol{w}_t, \quad (4)$$

where $\boldsymbol{w}_t$ denotes standard Brownian motion in $\mathbb{R}^d$, and $V\colon \mathbb{R}^d \to \mathbb{R}$ is a differentiable potential function. A standard numerical method to simulate (4) is the *Euler–Maruyama* method. Discretizing time in steps of size $\tau$, we obtain

$$\boldsymbol{\theta}_{i+1} = \boldsymbol{\theta}_i - \tau\nabla_{\boldsymbol{\theta}} V(\boldsymbol{\theta}_i) + \sqrt{2\tau}\boldsymbol{\xi}_i, \quad (5)$$

where $\boldsymbol{\xi}_i \sim \mathcal{N}(\mathbf{0}_d, \boldsymbol{I}_d)$. The proposed variable $\boldsymbol{\theta}_{i+1}$ in (5) follows Gaussian distribution. Therefore, the proposal density can be explicitly written as

$$q\big(\boldsymbol{\theta}_{i+1} \mid \boldsymbol{\theta}_i\big) \propto \exp\Big(-\frac{\|\,\boldsymbol{\theta}_{i+1} - \boldsymbol{\theta}_i + \tau\nabla_{\boldsymbol{\theta}} V(\boldsymbol{\theta}_i)\|_2^2}{4\,\tau}\Big). \quad (6)$$

Given a sample $\boldsymbol{\theta}'$, MALA accepts it with probability

$$\rho^{\text{MALA}}(\boldsymbol{\theta}', \boldsymbol{\theta}_i) \;=\; \min\Big\{1,\; \frac{\exp(-V(\boldsymbol{\theta}'))\, q\big(\boldsymbol{\theta}_i \mid \boldsymbol{\theta}'\big)}{\exp(-V(\boldsymbol{\theta}_i))\, q\big(\boldsymbol{\theta}' \mid \boldsymbol{\theta}_i\big)}\Big\}. \quad (7)$$

If the proposal is accepted, the chain advances as $\boldsymbol{\theta}_{i+1} = \boldsymbol{\theta}'$; otherwise, it remains at the current position, $\boldsymbol{\theta}_{i+1} = \boldsymbol{\theta}_t$. Furthermore, MALA inherently accommodates constrained parameter spaces $\boldsymbol{\theta} \in \Theta_{\text{c}}$ by modifying the acceptance probability as follows:

$$\rho_c^{\text{MALA}}\left(\boldsymbol{\theta}', \boldsymbol{\theta}_i\right) = \begin{cases} \rho^{\text{MALA}}\left(\boldsymbol{\theta}', \boldsymbol{\theta}_i\right), & \text{if } \boldsymbol{\theta}' \in \Theta_c, \\ 0, & \text{otherwise .} \end{cases} \quad (8)$$

For certain systems with an analytical form of the FIM, we can directly define our potential function as

$$V(\boldsymbol{\theta}) = -\frac{1}{2}\ln\det(\boldsymbol{J}_{\boldsymbol{\theta}}), \quad (9)$$

with its gradient as

$$\nabla_{\boldsymbol{\theta}} V(\boldsymbol{\theta}) = -\frac{1}{2}\text{tr}\left[\boldsymbol{J}_{\boldsymbol{\theta}}^{-1}\frac{\partial \boldsymbol{J}_{\boldsymbol{\theta}}}{\partial \boldsymbol{\theta}}\right], \quad (10)$$

to sample from Jeffreys prior via MALA.

For nonlinear state-space (NLSS) models, closed-form expressions for FIMs are not available. In this paper, we adopt a particle-filter–based approach to estimate the FIM, following [3], whereby Forward Filtering–Backward Smoothing (FFBSm) provides an unbiased Monte Carlo approximation of

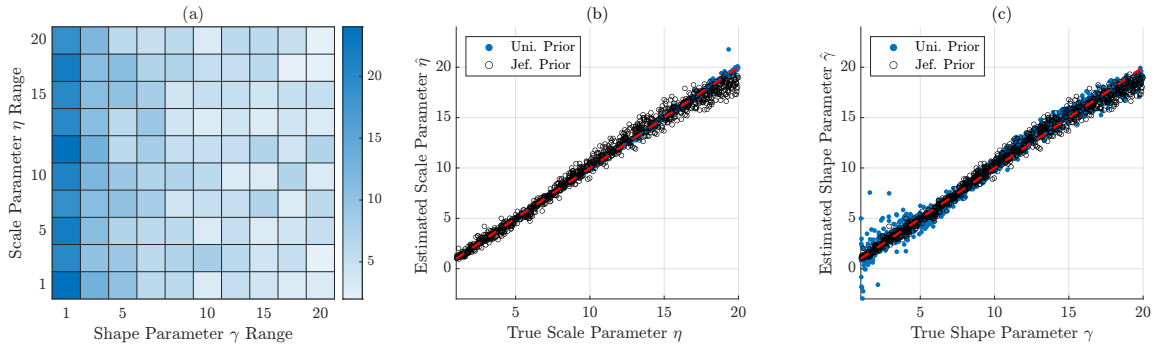


Fig. 1. (a) Heatmap of 1000 samples from Jeffreys prior of ($\eta$) and ($\gamma$) parameters from the Weibull distribution; (b) Scatter plot of the estimation $\hat{\eta}$ based on the uniform prior and Jeffreys prior vs. its true value; (c) Scatter plot of the estimation $\hat{\gamma}$ based on the uniform prior and Jeffreys prior vs. its true value. The red dashed line corresponds to an oracle estimate, which knows the true value of the parameter.

the score function, and consequently, the FIM. Given the estimated FIM $\hat{\boldsymbol{J}}_{\boldsymbol{\theta}}$, the gradient $\partial \boldsymbol{J}_{\boldsymbol{\theta}}/\partial \boldsymbol{\theta}$ can be approximated by *one-point unbiased estimate*. Specifically, we introduce a random perturbation $\boldsymbol{\mu} \sim \mathcal{N}(\mathbf{0}_d, \boldsymbol{I}_d)$ and approximate the derivative as

$$\frac{\partial \boldsymbol{J}_{\boldsymbol{\theta}}}{\partial \theta_j} \approx \frac{\mu_j}{\delta}\left(\hat{\boldsymbol{J}}_{\boldsymbol{\theta}+\delta\boldsymbol{\mu}} - \hat{\boldsymbol{J}}_{\boldsymbol{\theta}}\right), \quad j = 1, \ldots, d, \tag{11}$$

where $\delta > 0$ is a small step size, and $\hat{\boldsymbol{J}}_{\boldsymbol{\theta}+\delta\boldsymbol{\mu}}$ is computed using the same estimation procedure as for $\hat{\boldsymbol{J}}_{\boldsymbol{\theta}}$. Algorithm 1 integrates the FIM estimation with the gradient approximation within MALA. More details can be found in [4].

**Algorithm 1** Sample from Jeffreys Prior Distribution

**Require:** Initial guess $\boldsymbol{\theta}_0$, parameter space $\Theta_c$, step size $\tau$, iteration number $N$, finite-difference parameter $\delta > 0$

**for** $n = 0, 1, \ldots, N-1$ **do**
  Compute or estimate $\boldsymbol{J}_{\boldsymbol{\theta}_n}$
  Compute $\nabla_{\boldsymbol{\theta}} V(\boldsymbol{\theta}_n)$ using (10) **or** run
    Draw a random direction $\boldsymbol{\mu} \sim \mathcal{N}(\mathbf{0}_d, \boldsymbol{I}_d)$
    Estimate $\nabla_{\boldsymbol{\theta}} V(\boldsymbol{\theta})$ using (11)
    Estimate $\boldsymbol{J}_{\boldsymbol{\theta}+\delta\boldsymbol{\mu}}$
    Compute $\nabla_{\boldsymbol{\theta}} V(\boldsymbol{\theta}_n)$ using (10)
  Sample $\boldsymbol{\xi}_n \sim \mathcal{N}(\mathbf{0}_d, \boldsymbol{I}_d)$ and $U \sim \mathcal{U}(0,1)$
  $\boldsymbol{\theta}' \leftarrow \boldsymbol{\theta}_n - \tau \nabla_\theta V(\boldsymbol{\theta}_n) + \sqrt{2\tau}\boldsymbol{\xi}_n$,
  $\rho_n \leftarrow \rho^{\text{MALA}}(\boldsymbol{\theta}', \boldsymbol{\theta}_n)$ from (8)
  $\boldsymbol{\theta}_{n+1} \leftarrow \begin{cases} \boldsymbol{\theta}', & \text{If } U < \rho_n, \\ \boldsymbol{\theta}_n, & \text{Otherwise.} \end{cases}$
**end for**
**return** $\{\boldsymbol{\theta}_n\}_{n=1}^N$

## III. Numerical Illustration and Conclusions

We sample Jeffreys prior using Algorithm 1 and then illustrate its advantages within the *Two-Stage* (TS) estimation framework[2], using parameter estimation for a Weibull distribution as a test case, whose probability density function is given by

$$f(A; \eta, \gamma) = \frac{\gamma}{\eta}\left(\frac{A}{\eta}\right)^{\gamma-1} \exp\left[-\left(\frac{A}{\eta}\right)^{\gamma}\right], \quad A \geq 0,$$

where $\eta > 0$ is the scale parameter and $\gamma > 0$ is the shape parameter. Thus, the parameter vector is $\boldsymbol{\theta} = [\eta, \gamma]^T$.

We validate the TS estimators trained under uniform and Jeffreys priors using a validation set consisting of 1000 parameter points $\boldsymbol{\theta}_\ell = [\eta_\ell, \gamma_\ell]^T$ uniformly sampled from $[1, 20] \times [1, 20]$. For each $\boldsymbol{\theta}_\ell$ synthetic data $\{y_\ell^i\}_{i=1}^M$ from the Weibull model. We evaluate two classes of TS estimators based on the samples from the uniform and Jeffreys priors, respectively.

Fig. 1(a) shows the sampled Jeffreys prior distribution in the $(\gamma, \eta)$ parameter space. Estimation performances for $\eta$ and $\gamma$ are compared in Figs. 1(b) and 1(c), respectively. In Fig. 1(b), both uniform and Jeffreys-based estimators produce accurate and similar results of the scale parameter across its entire range. However, notable differences arise for the shape parameter $\gamma$, as seen in Fig. 1(c). Near low values of $\gamma < 5$, the uniform-based estimator exhibits significant variance and bias. In contrast, the Jeffreys-based estimator demonstrates considerably improved accuracy in this region.

Overall, these results demonstrate the validity of our sampling framework and highlight the advantage of leveraging the geometry encoded by the Jeffreys prior for parameter inference and experimental design.

[2]For additional details on the implementation setup of TS, we refer the reader to [5].

## References


[1] H. Jeffreys, "An invariant form for the prior probability in estimation problems," *Proceedings of the Royal Society of London. Series A. Mathematical and Physical Sciences*, vol. 186, no. 1007, pp. 453–461, 1946.

[2] G. O. Roberts and O. Stramer, "Langevin diffusions and metropolis-hastings algorithms," *Methodology and Computing in Applied Probability*, vol. 4, pp. 337–357, 2002.

[3] P. E. Valenzuela, J. Dahlin, C. R. Rojas, and T. B. Schön, "On robust input design for nonlinear dynamical models," *Automatica*, vol. 77, pp. 268–278, 2017.

[4] Y. Shi, B. Lakshminarayanan, and C. R. Rojas, "A metropolis-adjusted langevin algorithm for sampling jeffreys prior," *arXiv preprint arXiv:2504.06372*, 2025.

[5] B. Lakshminarayanan and C. R. Rojas, "A statistical decision-theoretical perspective on the two-stage approach to parameter estimation," in *Proceedings of the 61st IEEE Conference on Decision and Control (CDC)*, pp. 5369–5374, 2022.