Approximate Manifold Sampling via the Hug Sampler

Mauro Camara Escudero, Christophe Andrieu, Mark Beaumont

University of Bristol

Table of contents

A General Set-Up

A General Set-Up

The following setup appears in many areas of statistics.

- $p(x)$ prior on $\mathcal{X}$.
- $f : \mathcal{X} \to \mathbb{R}$ smooth so $f^{-1}(y)$ is a submanifold of $\mathcal{X}$ for each $y \in \mathbb{R}$.

Interest often lies in sampling from the following distributions.

- Manifold Densities: Restricted prior onto $f^{-1}(y)$

$$\overline{p}(x \mid f(x) = y) \propto p(x)|J_f(x)J_f(x)^\top|^{-1/2}$$

- Filamentary Densities: Concentrated prior around $f^{-1}(y)$

$$p_\epsilon(x \mid y) \propto p(x)k_\epsilon(\|y - f(x)\|)$$

Typically filamentary densities are a relaxation of manifold densities

$$p_\epsilon(x \mid y) \longrightarrow \overline{p}(x \mid f(x) = y) \qquad \text{as} \qquad \epsilon \longrightarrow 0.$$

Bayesian Inverse Problems (BIP)

BIP - Set up I

Observed data $y \in \mathbb{R}$ is the output of forward function $h : \Theta \to \mathbb{R}$ of parameters $\theta \in \Theta$ perturbed by Gaussian observational noise $\eta \in \mathbb{R}$

$$y = f_\sigma(\theta, \eta) = h(\theta) + \sigma\eta \qquad \eta \sim \mathcal{N}(0, \mathrm{I}) \qquad \sigma > 0.$$

The forward function h encapsulates all the complexity of the model.

- forward problem (easy): Simulate y given θ.
- inverse problem (hard): Infer θ given y via BIP posterior $p_\sigma(\theta \mid y)$.

Inverse problems are under-determined due to

- $h^{-1}(y) \subset \Theta$ being a set when h is not injective.
- η introducing additional non-identifiability (small σ reduces it).

BIP - Set up II

Manifolds in BIP

When h is smooth then

- $h^{-1}(y)$ is a sub-manifold of Θ.
- $f_\sigma^{-1}(y)$ is a sub-manifold of $\Theta \times \mathbb{R}$ for any $\sigma > 0$.

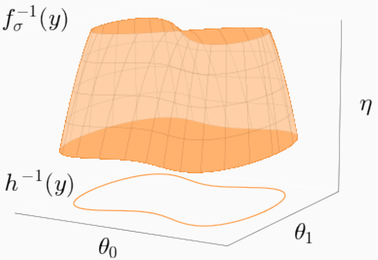
BIP Posteriors

- $p_\sigma(\theta \mid y) \propto p(\theta) p_\sigma(y \mid \theta)$ around $h^{-1}(y)$.
- $p_{\sigma,\epsilon}(\theta, \eta \mid y) \propto p_\sigma(\theta, \eta) k_\epsilon(\|y - f_\sigma(\theta, \eta)\|)$ around $f_\sigma^{-1}(y)$

Here k_ϵ is a smoothing kernel.

BIP - Toy Example I

- Observed data: $y = 1.0$
- Parameter: $\theta = (\theta_0, \theta_1)^\top \in \mathbb{R}^2$
- Forward Function[1]: $h(\theta) = \theta_1^2 + 3\theta_0^2(\theta_0^2 - 1)$

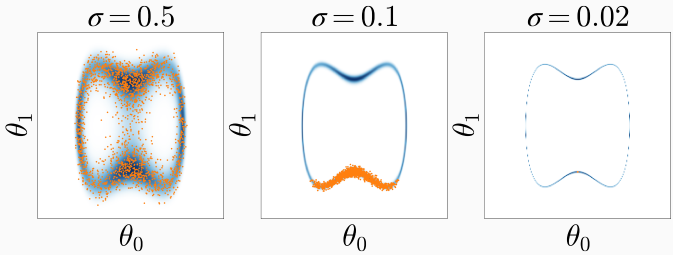


- Manifold $h^{-1}(y)$ independent of σ.
- Manifold $f_\sigma^{-1}(y)$ changes shape for different σ.

[1]Example taken from Au et al. (2021) and corresponding Jupyter Notebooks.

BIP - Toy Example II

Sampling from the posterior $p_\sigma(\theta \mid y)$ using HMC becomes harder as σ decreases.



Exact Manifold Sampling

Constrained HMC I (Lelièvre et al., 2019)

To target $\pi(dx) = \overline{p}(x)\mathcal{H}(dx)$ construct a constrained Hamiltonian system

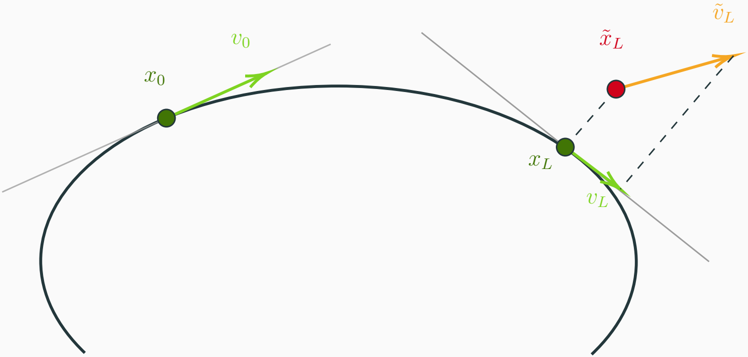
$$\begin{aligned} \dot{x} &= \nabla_v H(x, v) \\ \dot{v} &= -\nabla_x H(x, v) \\ f(x) &= 0. \end{aligned}$$

where $H(x, v) = -\log \overline{p}(x) + \|v\|^2/2$ and $M = \mathrm{I}$. This can be integrated with RATTLE/SHAKE which are constrained versions of the Leapfrog. Distinguishing features from Leapfrog are:

- Projections to enforce $x \in \mathcal{M}$.
- Reversibility checks (and reprojection).

Constrained HMC II (Lelièvre et al., 2019)

- Position Projections are non-linear and require a solver. Au et al. (2021) shows Newton/symmetric-Newton solvers works well in practice.



Constrained HMC III (Lelièvre et al., 2019)

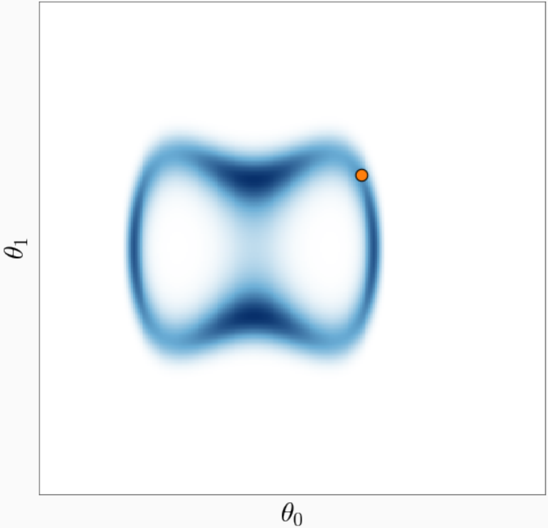
At each C-HMC integration step the following operations are expensive:

- Constraint Jacobian: $J_f(x)$
- Correction term: $\log\det(J_f(x)J_f(x)^\top)$
- Position projections and re-projections.

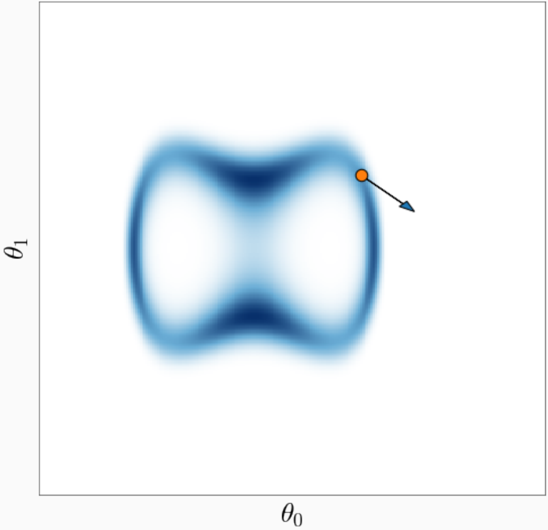
Position projections can potentially lead to many Jacobian evaluations. Next we see an approximate manifold sampling method to sample from $p_{\sigma,\epsilon}(\theta, \eta \mid y)$ and avoid these expensive operations.

Approximate Manifold Sampling

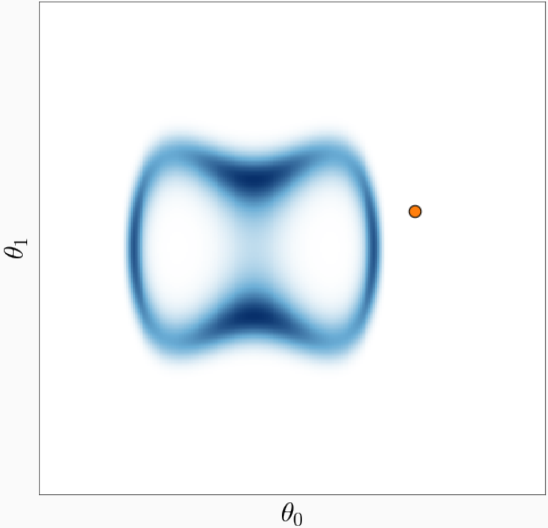
Visualization of Hug



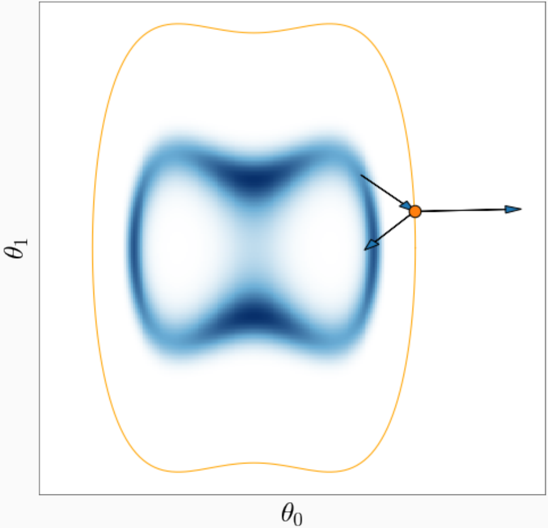
Visualization of Hug



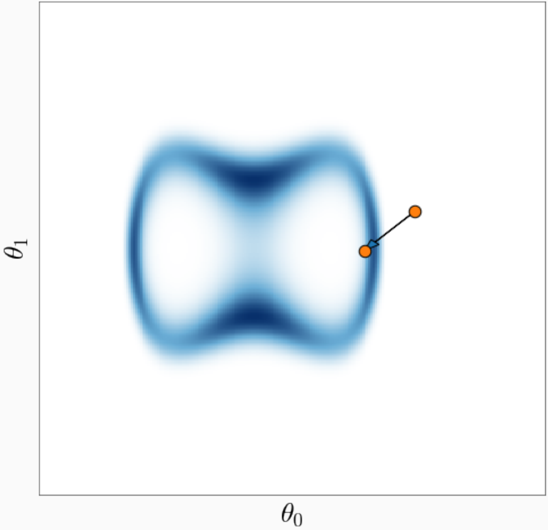
Visualization of Hug



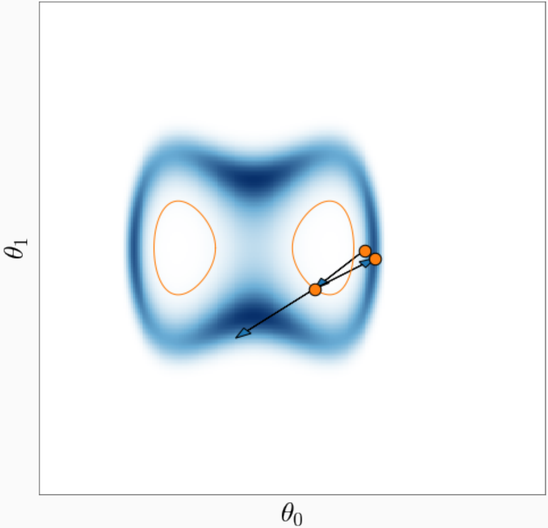
Visualization of Hug



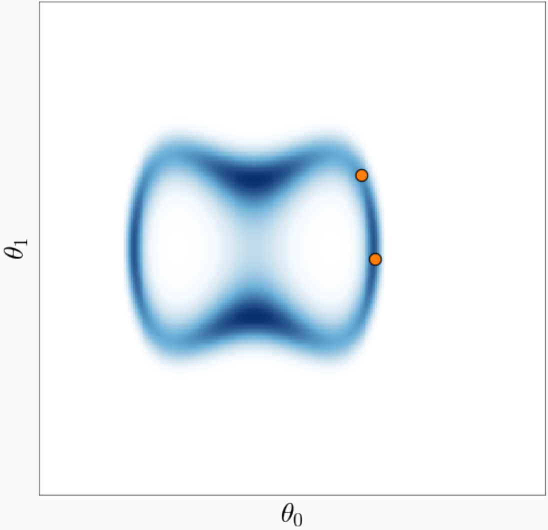
Visualization of Hug



Visualization of Hug



Visualization of Hug



Generalized Hug

- Generalization of the Hug algorithm by Ludkin & Sherlock (2019) originally used as an alternative to HMC to propose samples almost on the same contour of a density.
- Let $\ell(x)$ log-density of filamentary distribution around $f^{-1}(y)$ and $\hat{g}$ denote normalized gradient at a contour of f.

Algorithm 1: Generalized Hug Kernel (one iteration)

Sample auxiliary velocity variable $v_0 \sim \mathcal{N}(0, \mathrm{I})$.
for $b = 0, \ldots, B-1$ **do**
    Move: $x_{b+\delta/2} = x_b + (\delta/2)v_b$
    Reflect: $v_{b+1} = v_b - 2\hat{g}_{b+\delta/2}\hat{g}_{b+\delta/2}^\top v_b$
    Move: $x_{b+1} = x_{b+\delta/2} + (\delta/2)v_{b+1}$
end
With probability $a = \exp(\ell(x_B) - \ell(x_0))$ accept x_B, else stay at x_0.

Intuition Behind Generalized Hug I

Dynamic of a particle in $\mathbb{R}^n$ moving with constant speed on c-levelset of $f : \mathbb{R}^n \to \mathbb{R}$ with gradient g and Hessian H

$$\dot{x}_t = v_t$$
$$\dot{v}_t = -\frac{v_t^\top H_t v_t}{\|g_t\|}\hat{g}_t.$$

A position-Verlet-like discretization starting from (x_0, v_0) with $v_0 \perp g_0$

$$x_{t+\frac{\delta}{2}} = x_t + \frac{\delta}{2}v_t$$
$$v_{t+\delta} = v_t - \delta\frac{v_t^\top H_{t+\frac{\delta}{2}} v_t}{\|g_{t+\frac{\delta}{2}}\|}\hat{g}_{t+\frac{\delta}{2}}$$
$$x_{t+\delta} = x_{t+\frac{\delta}{2}} + \frac{\delta}{2}v_{t+\delta}$$

Intuition Behind Generalized Hug II

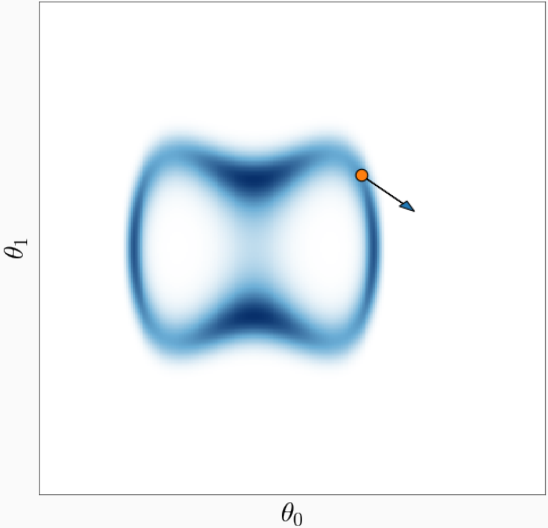
Bounce mechanism is an approximation of curvature information

$$\delta \frac{v_t^\top H_{t+\delta/2} v_t}{\|g_{t+\delta/2}\|} = 2 v_t^\top \hat{g}_{t+\delta/2} + \mathcal{O}(\delta^2)$$

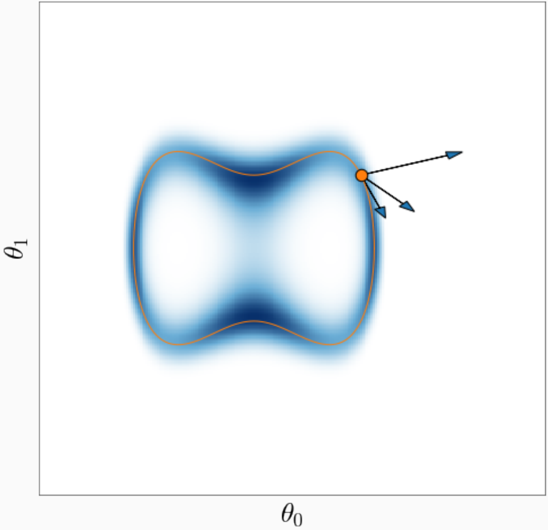
- Proposal mechanism can be thought of as approximate discretization.
- Continuous-time dynamic is a Constrained Hamiltonian system that has been solved for its Lagrange multipliers[2].
- Performance can deteriorate for highly concentrated filamentary distributions.

[2] Shout-out to Sam Livingstone for noticing this. See also Leimkuhler & Reich (2005).

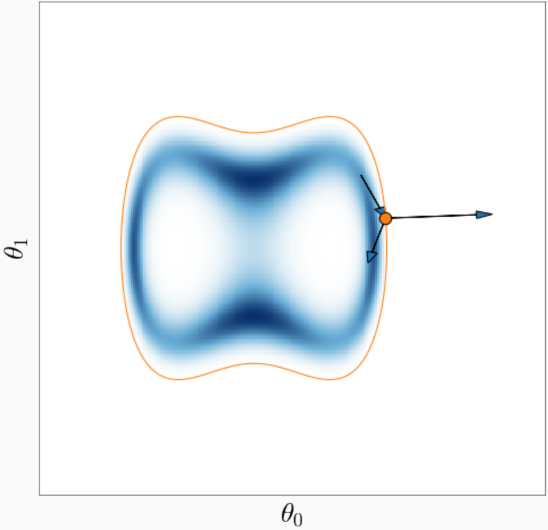
Visualization of Thug



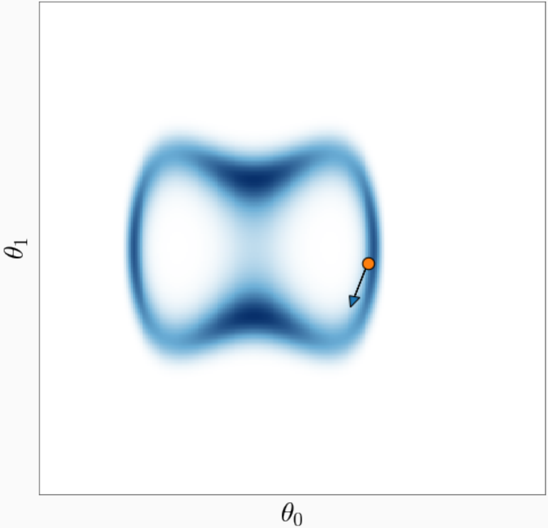
Visualization of Thug



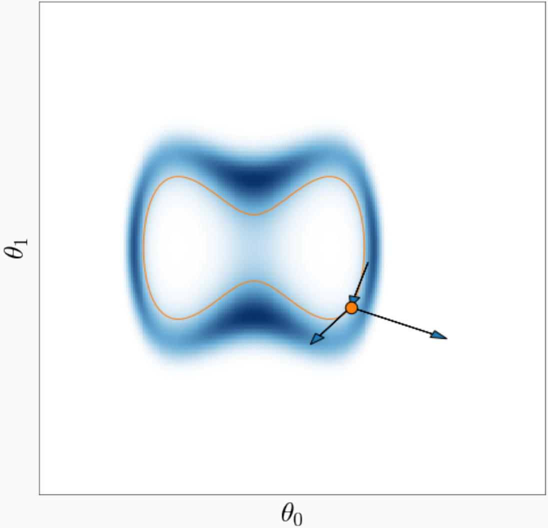
Visualization of Thug



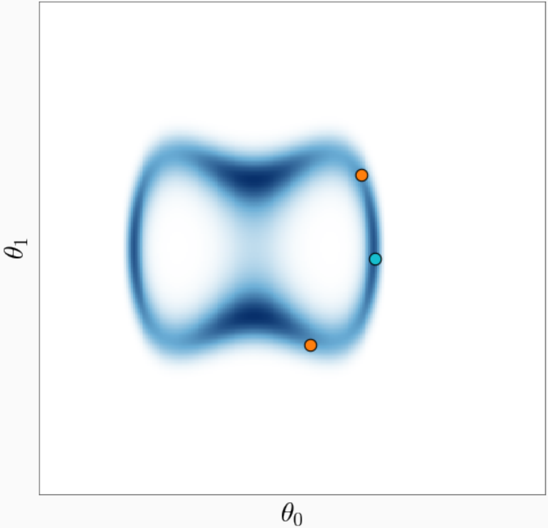
Visualization of Thug



Visualization of Thug



Visualization of Thug



Comments about Squeezing the Velocity

- Velocity squeezing for $\alpha \in [0, 1)$ is a dumpened velocity projection

$$w_t = \left(\mathrm{I} - \alpha \hat{g}_t \hat{g}_t^\top\right) v_t =: S_{\alpha,t} v_t$$

- Velocity distribution now concentrated around tangent plane

$$w_t \sim \mathcal{N}(0, S_{\alpha,t} S_{\alpha,t}^\top)$$

has variance $(1-\alpha)^2$ along $\hat{g}_t$ and variance 1 along any $\hat{t}_t \perp \hat{g}_t$.

- Final velocity needs to be unsqueezed for reversibility

$$v_{t+\delta} = \left(\mathrm{I} + \frac{\alpha}{1-\alpha} \hat{g}_{t+\delta} \hat{g}_{t+\delta}^\top\right) w_{t+\delta}.$$

Now $\|v_t\| \neq \|v_{t+\delta}\|$ inducing a reduction in acceptance probability.

Thug Algorithm

Algorithm 2: Thug Kernel (one iteration)

Sample auxiliary velocity variable $v_0 \sim \mathcal{N}(0, \mathrm{I})$.
Squeeze: $w_0 = v_0 - \alpha \hat{g}_0 \hat{g}_0^\top v_0$.
for $b = 0, \ldots, B-1$ **do**
    Move: $x_{b+\delta/2} = x_b + (\delta/2) w_b$
    Reflect: $w_{b+1} = w_b - 2\hat{g}_{b+\delta/2} \hat{g}_{b+\delta/2}^\top w_b$
    Move: $x_{b+1} = x_{b+\delta/2} + (\delta/2) w_{b+1}$
end
Unsqueeze: $v_B = w_B + (\alpha/(1-\alpha)) \hat{g}_B \hat{g}_B^\top w_B$.
With probability $a = \exp(\ell(x_B) - \ell(x_0) - \|v_B\|^2/2 + \|v_0\|^2/2)$ accept x_B, otherwise stay at x_0.

The Trade-off of Squeezing the Velocity

- Optimal value of α unknown. The larger α, the closer we stay to the manifold, however the larger the reduction in acceptance probability due to the mismatch between $\|v_0\|$ and $\|v_B\|$.
- When target is highly concentrated experiments show AP increase due to higher precision outweights AP decrease due to mismatch.

Strategy

When targeting the BIP posterior

$$p_{\sigma,\epsilon}(\theta, \eta \mid y) \propto p_\sigma(\theta, \eta) k_\epsilon(\|y - f_\sigma(\theta, \eta)\|),$$

embed Thug in SMC sampler, start with $\alpha_0 = 0$ and increase it adaptively (Andrieu & Thoms, 2008)

$$\tau_{i+1} = \tau_i - \gamma_{i+1}(\hat{a}_i - a^*),$$

where $\tau_i = \text{logit}(\alpha_i)$, $\hat{a}_i$ is the estimate of the acceptance probability from previous round, and a^* is target acceptance probability.

Experiments

Computational Time

Aim of Experiment

Determine if HUG/THUG bring noticeable computational savings with respect to C-HMC.

Algorithm	$f + \nabla f$	ESS	Cost per ESS
THUG	7002	74.68	**93.76**
HUG	6001	54.87	109.37
CHMC	60442	39.24 (**332.29**)	1540.37 (181.90)
HMC	6462	3.03	2135.66

True Posterior Distribution

$$\overline{p}(\theta, \eta \mid y) \propto p_\sigma(\theta, \eta) |\det J_f(\theta, \eta) J_f(\theta, \eta)^\top|^{-1/2}$$

Approximate Posterior Distribution

$$p_{\epsilon,\sigma}(\theta, \eta \mid y) \propto p_\sigma(\theta, \eta) k_\epsilon(\|f_\sigma(\theta, \eta) - y\|).$$

Here k_ϵ is Gaussian, $y = 1$, $\sigma = 0.02$, $\delta = 0.05$, $L = B = 5$, and $\epsilon = 0.001$.

Thug MCMC I

Aim of Experiment

Determine how acceptance probability deteriorates as step size δ decreases.

- Run Thug, Hug and HMC on lifted filamentary posterior

$$p_{\sigma,\epsilon}(\theta, \eta \mid y) \propto p_\sigma(\theta, \eta) k_\epsilon(\|y - f_\sigma(\theta, \eta)\|)$$

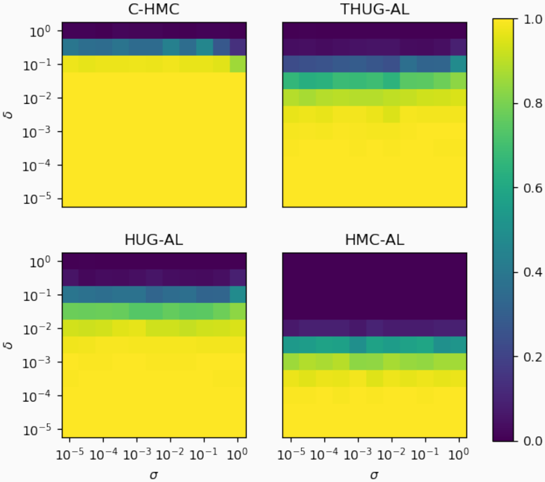
  for $\epsilon = 0.02$ and k_ϵ Epanechnikov, and C-HMC on lifted manifold posterior

$$\overline{p}(\theta, \eta \mid y) \propto p_\sigma(\theta, \eta) |J_{f_\sigma}(\theta, \eta) J_{f_\sigma}(\theta, \eta)^\top|^{-1/2}.$$

- Run across a grid of noise scale $\sigma \in (1 \times 10^{-5}, 1.0)$ and step-sizes $\delta \in (1 \times 10^{-5}, 1.0)$, keeping number of steps/bounces per iteration $B = L = 20$ fixed.
- Average acceptance probability across 10 runs of 50 samples.

Thug MCMC II

Acceptance probability decoupled from noise level. Thug/Hug can use up to 2 orders of magnitude larger step-sizes than HMC.



SMC-Thug I

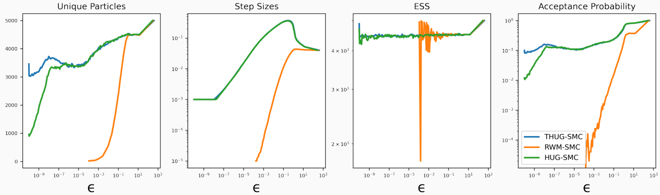
Aim of Experiment

Does SMC provide a good framework to adaptively tune Thug?

- Run RWM-SMC, HUG-SMC and adaptive THUG-SMC targeting $p_{\sigma,\epsilon}(\theta, \eta \mid y)$ for $\sigma = 1 \times 10^{-8}$.
- Use $N = 5000$ particles initialized from the prior and multinomial resampling at each step.
- Tune step-size based on estimated acceptance probability with minimum allowed stepsize of $\delta = 1 \times 10^{-3}$.
- Adaptively choose next ϵ_n using number of unique particles.
- Stop SMC samplers either after $\epsilon \leq 1 \times 10^{-10}$, after 200 iterations or when acceptance probability drops to zero.

SMC-Thug II

HUG-SMC and THUG-SMC outperform RWM-SMC in terms of ESS and acceptance probability. Surprisingly, they manage to reach $\epsilon = 1 \times 10^{-10}$ keeping $\delta = 1 \times 10^{-3}$. Importantly when $p_{\sigma,\epsilon}$ is concentrated enough, THUG-SMC outperforms HUG-SMC.



Limitations and Future Directions

- How to compare samples from manifold and filamentary distributions?
- Does the ESS make sense on manifolds? Could derive a better metric.
- Develop approximate manifold sampling algorithms for $\dim(\mathcal{Y}) \gg 1$.
- Experiments with models where $\dim(\Theta)$ large.
- Apply these methods to other areas (ABC, SSM, motion control, etc).

Thank you

References

Andrieu, C. and Thoms, J. A tutorial on adaptive mcmc. *Statistics and Computing*, 18(4):343–373, Dec 2008. ISSN 1573-1375. doi: 10.1007/s11222-008-9110-y. URL `https://doi.org/10.1007/s11222-008-9110-y`.

Au, K. X., Graham, M. M., and Thiery, A. H. Manifold lifting: scaling mcmc to the vanishing noise regime, 2021.

Federer, H. *Geometric Measure Theory*. Classics in Mathematics. Springer Berlin Heidelberg, 2014. ISBN 9783642620102. URL `https://books.google.co.uk/books?id=jld-BgAAQBAJ`.

Leimkuhler, B. and Reich, S. *Simulating Hamiltonian Dynamics*. Cambridge Monographs on Applied and Computational Mathematics. Cambridge University Press, 2005. doi: 10.1017/CBO9780511614118.

Lelièvre, T., Rousset, M., and Stoltz, G. Hybrid monte carlo methods for sampling probability measures on submanifolds, 2019.

Ludkin, M. and Sherlock, C. Hug and hop: a discrete-time, non-reversible markov chain monte-carlo algorithm, 2019. URL `https://arxiv.org/abs/1907.13570`.

THUG MCMC III

Aim of Experiment

Does the acceptance probability of Hug/Thug deteriorate at slower rate than HMC/RM-HMC with respect to step size?

- Run Thug, Hug, RM-HMC and HMC to target filamentary posterior

$$p_\sigma(\theta \mid y) \propto p(\theta)\mathcal{N}(h(\theta), \sigma^2 \mathrm{I}).$$

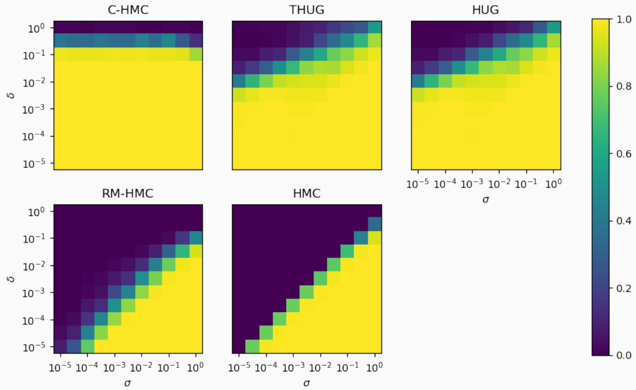
  and C-HMC to target lifted manifold posterior

$$\overline{p}(\theta, \eta \mid y) \propto p(\theta)p(\eta)|J_{f_\sigma}(\theta, \eta)J_{f_\sigma}(\theta, \eta)^\top|^{-1/2}$$

- Run across a grid of noise scale $\sigma \in (1 \times 10^{-5}, 1.0)$ and step-sizes $\delta \in (1 \times 10^{-5}, 1.0)$, keeping number of steps/bounces per iteration $B = L = 20$ fixed.
- Average acceptance probability across 10 runs of 50 samples.

Thug MCMC IV

HMC and RM-HMC need $\mathcal{O}(\delta) = \mathcal{O}(\sigma)$ for a good acceptance probability. Hug and Thug can achieve the same acceptance probability with 3 order of magnitude larger step-size.



Acceptance Probability vs Discretization Order

- Ludkin & Sherlock (2019) showed that when $f = \ell$, H is γ-Lipschitz and bounded above by $\beta > 0$ Hug satisfies

$$|\ell_B - \ell_0| \leq \frac{\delta^2}{8}\|v_0\|^2(2\beta + \gamma T\|v_0\|) =: \mathcal{B}_{\text{HUG}}$$

  Thug satisfies a tighter bound when $\alpha > 0$ and $\hat{g}_0^\top v_0 \neq 0$

$$|\ell_B - \ell_0| \leq \mathcal{B}_{\text{HUG}} - \frac{\alpha(2-\alpha)\delta^2(\hat{g}_0^\top v_0)^2}{8}(2\beta + \gamma T\|v_0\|) =: \mathcal{B}_{\text{THUG}}.$$

- When $f \neq \ell$ will require assumptions on relationship between f and ℓ

$$p_\epsilon(x \mid y) \propto p(x)k_\epsilon(\|y - f(x)\|)$$

GPV for Partitioned Systems

For a Partitioned ODE

$$\dot{x} = F_1(x, v)$$
$$\dot{v} = F_2(x, v)$$

the Generalized Position Verlet (GPV) integrator

$$
\begin{aligned}
x_{n+1/2} &= x_n + \frac{\delta}{2} F_1(x_{n+1/2}, v_n) \\
v_{n+1} &= v_n + \frac{\delta}{2} \left[F_2(x_{n+1/2}, v_n) + F_2(x_{n+1/2}, v_{n+1}) \right] \\
x_{n+1} &= x_{n+1/2} + \frac{\delta}{2} F_1(x_{n+1/2}, v_{n+1})
\end{aligned}
$$

is **implicit**, second-order, symmetric and symplectic.

GPV for Separable Systems

For a Separable ODE

$$\dot{x} = F_1(v)$$
$$\dot{v} = F_2(x)$$

the GPV integrator

$$x_{n+1/2} = x_n + \frac{\delta}{2} F_1(v_n)$$
$$v_{n+1} = v_n + \delta F_2(x_{n+1/2})$$
$$x_{n+1} = x_{n+1/2} + \frac{\delta}{2} F_1(v_{n+1})$$

is **explicit**, second-order, symmetric and symplectic.

Alternative Integrator I

Although in general the Generalized Position Verlet for a non-separable system is implicit, it turns out that one can actually solve explicitly for v_{n+1} in the velocity update.

$$v_{n+1} = \underbrace{v_n - \frac{\delta}{2}\frac{v_n^\top H_F(x_{n+1/2}v_n}{\|\nabla_x F(x_{n+1/2})\|}\widehat{\nabla_x F}(x_{n+1/2})}_{:=a}$$

$$\underbrace{-\frac{\delta}{2}\frac{\widehat{\nabla_x F}(x_{n+1/2})}{\|\nabla_x F(x_{n+1/2})\|}}_{=:b} v_{n+1}^\top H_F(x_{n+1/2})v_{n+1},$$

then the expression has the form (we write $H_{n+1/2} = H(x_{n+1/2})$)

$$v_{n+1} = a + b v_{n+1}^\top H_{n+1/2} v_{n+1}.$$

This can be solved by solving a simple quadratic equation for ϑ

$$c_1\vartheta^2 + (2c_2 - 1)\vartheta + c_3 = 0$$

Alternative Integrator II

where

$$c_1 = b^\top H_{n+1/2} b$$
$$c_2 = a^\top H_{n+1/2} b$$
$$c_3 = a^\top H_{n+1/2} a.$$

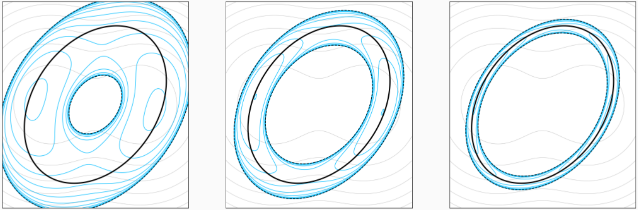
Interestingly, this discretization works well for sampling from filamentary distributions only when the initial velocity is perpendicular to the gradient at the initial position $v_0 \perp \hat{g}_0$, otherwise it quickly blows up. This is in contrast with the generalised Hug algorithm which remains stable thanks to the BPS reflection mechanism.

A General Approximate Manifold Sampling Framework

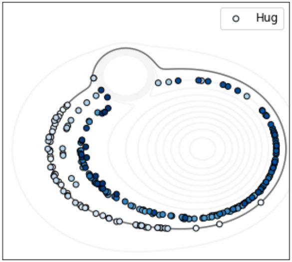
Let π be a filamentary distribution whose limiting manifold distribution is $\overline{\pi}$. A general approximate manifold sampling algorithm consists of a triplet (H_p, Φ, H_a) where

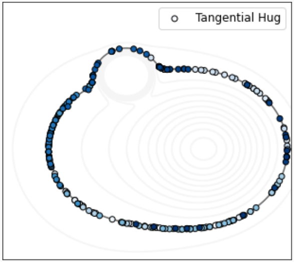
- H_p is a Hamiltonian system that forms the base of our proposal mechanism. A good H_p would follow/stay close to $\mathcal{M}$ and perhaps be a good Hamiltonian system for $\overline{\pi}$.
- Φ is a reversible (or skew-reversible) integrator for H_p of suitably high order and preferably with $|\det J_\Phi| = 1$, symplecticity is desired but not needed.
- H_a is a Hamiltonian that determines which samples get accepted or rejected. This should include π for the algorithm to be correct.

Contours of Filamentary Distribution



Tangential Hug Stays closer





Manifold Distributions I

Transformation of Random Variable by Diffeomorphism

Let X be an $\mathbb{R}^n$-valued random vector with density p_X. Let $f : \mathbb{R}^n \to \mathbb{R}^n$ be a diffeomorphism and $Y = f(X)$. Then

$$p_Y(y)dy = p_X(f^{-1}(y))|\det J_{f^{-1}}(y)|dy$$

The Co-Area formula for Lipschitz functions generalizes the above results to non-injective functions (see Theorem 5.3.9 in Federer (2014)).

Conditional Density of Random Variable on Submanifold

Let X be an $\mathbb{R}^n$-valued random vector with density p_X. Let $f : \mathbb{R}^n \to \mathbb{R}^m$ be a smooth function with $n > m$, and let $y \in \mathbb{R}^m$. Then on the sub-manifold $f^{-1}(y)$

$$p(x \mid f(x) = y)\mathcal{H}^{n-m}(dx) \propto p_X(x)|\det(J_f(x)J_f(x)^\top)|^{-1/2}\mathcal{H}^{n-m}(dx)$$

Manifold Distributions II

Assumption 2

π admits a density with respect to the Hausdorff measure on $\mathcal{M}$.

Manifold Distribution

Let $X : \Omega \to \mathbb{R}^n$ be a vector-valued random variable with distribution π and finite covariance matrix $\mathbb{V}_\pi[X]$, and let $f : \mathbb{R}^n \to \mathbb{R}$ be a smooth function. Consider $y \in \mathbb{R}$ fixed, then at any point $\xi \in f^{-1}(y)$ we denote by $\hat{g}(\xi)$ the normalized gradient of f and by $\mathsf{T}(\xi) = \{\hat{t}_1(\xi), \dots, \hat{t}_{n-1}(\xi)\}$ a basis for the tangent space at ξ. Then π is a manifold distribution if $\forall \xi \in \mathbb{R}^n$ and $\forall \hat{t}_i(\xi) \in \mathsf{T}(\xi)$

$$\hat{g}(\xi)^\top \mathbb{V}_\pi[X]\hat{g}(\xi) = 0 \qquad \text{and} \qquad \hat{t}_i(\xi)^\top \mathbb{V}_\pi[X]\hat{t}_i(\xi) > 0.$$

Typically obtained as limiting posterior density as some scale parameter goes to zero.

Manifold Distribution III

Manifold Distribution

Let U be an $\mathbb{R}^n$-valued random variable, and $f : \mathbb{R}^n \to \mathbb{R}^m$ be smooth. Let $\mathrm{K}(y, du)$ be a regular conditional distribution of U given $\sigma(f(U))$ and let $\mathcal{H}_y^{n-m}$ be the Hausdorff measure on $f^{-1}(y)$. If $\mathrm{K}(y, \cdot) \ll \mathcal{H}_y^{n-m}$ then $\pi = \mathrm{K}(y, \cdot)$ is a manifold distribution.

Manifold Distribution IV

Graham's Theorem Revisited

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space, $U : \Omega \to \mathbb{R}^n$ be a random vector with distribution P_U and density p_U with respect to λ^n, the Lebesgue measure on $(\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n))$. Let n >m and $f : \mathbb{R}^n \to \mathbb{R}^m$ be a smooth function with Jacobian matrix $J_f(u)$ having full row-rank λ^n-almost everywhere, and let $f \circ U$ have distribution $P_f = P_U \circ f^{-1}$. Let $\sigma(f)$ be the sigma-algebra generated by $f \circ U$, and let $\phi : \mathbb{R}^n \to \mathbb{R}$ be a $\mathcal{B}(\mathbb{R}^n)$-measurable test function. Let $\mathrm{K}(y, du)$ be a RCD of U given $\sigma(f)$ from $(\mathbb{R}^m, \sigma(f))$ to $(\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n))$, such that $\mathrm{K}(y, \cdot) \ll \mathcal{H}^{n-m}$. Then expectations with respect to K can be written as

$$\mathbb{E}[\phi(U) \mid f(U) = y] = \int_{f^{-1}(\{y\})} \phi(u) k_y(u) \mathcal{H}^{n-m}(du)$$

where $k_y(u)$ is the density of $\mathrm{K}(y, \cdot)$ on $f^{-1}(\{y\})$ with respect to $\mathcal{H}^{n-m}$, given by

$$k_y(u) \propto p_U(u) \left| \det J_f(u) J_f(u)^\top \right|^{-1/2}.$$

When is $f^{-1}(y)$ a submanifold?

Let $\mathcal{X}$ and $\mathcal{Y}$ be manifolds and $f : \mathcal{X} \to \mathcal{Y}$ be smooth.

Regular Value

Then $y \in \mathcal{Y}$ is a **regular value** for f if for all $x \in f^{-1}(y)$ the differential $df_x : \mathcal{T}_x\mathcal{X} \to \mathcal{T}_y\mathcal{Y}$ is surjective. (alternatively, f is a submersion at every $x \in f^{-1}(y)$).

Preimage Theorem

If $y \in \mathcal{Y}$ is a regular value of f then $f^{-1}(y)$ is a submanifold of $\mathcal{X}$.

Filamentary Distributions

Assumption 1

Manifold of interest has co-dimension 1.

Filamentary Distribution

Let $X : \Omega \to \mathbb{R}^n$ be a vector-valued random variable with distribution π and finite covariance matrix $\mathbb{V}_\pi[X]$, and let $f : \mathbb{R}^n \to \mathbb{R}$ be a smooth function. Consider $y \in \mathbb{R}$ fixed, then at any point $\xi \in f^{-1}(y)$ we denote by $\hat{g}(\xi)$ the normalized gradient of f and by $\mathsf{T}(\xi) = \{\hat{t}_1(\xi), \ldots, \hat{t}_{n-1}(\xi)\}$ a basis for the tangent space at ξ. We say that π is a filamentary distribution if

$$\forall \xi \in \mathbb{R}^n, \quad \forall \hat{t}_i(\xi) \in \mathsf{T}(\xi) \qquad 0 < \hat{g}(\xi)^\top \mathbb{V}_\pi[X] \hat{g}(\xi) \ll \hat{t}_i(\xi)^\top \mathbb{V}_\pi[X] \hat{t}_i(\xi).$$

In practice one doesn't need to check the definition, it will be clear if the posterior has a filamentary structure.

- Filamentary distributions are highly concentrated around a submanifold.
- Orthogonal scaling $\ll$ tangential scaling.