MCMC convergence on manifolds through coupling techniques

Elena Bortolato & Pierre E. Jacob. & Robin J. Ryder

Statistical methods and models for complex data — Poster Session: 21 September 2022

Abstract

Some problems in statistics and machine learning require sampling distributions on submanifolds embedded in $\mathbb{R}^D$. To target such distributions, in the last twenty years constrained Markov Chain Monte Carlo methods have been developed. Assessing the convergence of such algorithms still remains an open problem. We propose to apply coupling techniques that help monitoring the practical convergence of the chains. In particular, we derive couplings of Metropolis-Rosenbluth-Teller-Hastings-type (MRTH) and Hamiltonian Monte Carlo-type (HMC) algorithms on smooth manifolds and present some applications in the domain of likelihood-free inference.

Recipes for sampling on manifolds

Key objects

- $\mathcal{M} = \{x \in \mathbb{R}^D | \, q_j(x) = 0, \; j = 1, \dots, m\}$: submanifold of $\mathbb{R}^D$,
- D: dimension of the ambient space,
- $d = D - m$: dimension of the manifold,
- $q_j(x) = 0, \; j = 1, \dots, m$: smooth differentiable constraints,
- $T_x = \{x^* \in \mathbb{R}^D | \, \nabla q_j(x)^\top (x - x^*) = 0, \; j = 1, \dots, m\} \in \mathbb{R}^d$: tangent space to $\mathcal{M}$ at x,
- $T_x^\perp = \sum_{j=1}^m a_j \nabla q_j(x) \quad a_j \in \mathbb{R}, \quad j = 1, \dots, m$: normal (or cotangent) space to $\mathcal{M}$ at x.

Overall goal

Sampling from the distribution:

$$\pi(x) = \frac{f(x)\mathcal{H}^d(\mathcal{M})}{Z}, \; f(x) > 0, \; Z = \int f(x)\mathcal{H}^d(\mathcal{M})dx,$$

with $\mathcal{H}^d(\mathcal{M})$ d-dimensional Hausdorff measure of $\mathcal{M}$:

$$\mathcal{H}^d(\mathcal{M}) = \lim_{\epsilon \to 0} \inf \left\{ \frac{\alpha_d}{2} \sum_j \delta(U_j)^d, \; \mathcal{M} = \bigcup_j U_j, \; \delta(U_j) < \epsilon \right\},$$

$\alpha_d = \frac{\pi^{d/2}}{\Gamma(d/2+1)}$ is the volume of a unit d-ball,

$\delta(U_j) = sup\{\rho(s,t), \; s, t \in U_j\}$ is the diameter of U_j.

This can be done with MCMC algorithms, for instance

1. MRHT-type (Zappa *et al.*, 2018):
   - start at $x \in \mathcal{M}$,
   - draw $\nu \sim p(\nu)$ and drift away with a step of size ν following T_x from $\mathcal{M}$,
   - project back to the manifold following $T_x^\perp$ and find $y \in \mathcal{M}$,
   - if x can be reached from y with some ν' (**ensures the reversibility** of the Markov kernel) accept the proposal with probability $\min\left(1, \frac{f(y)p(\nu')}{f(x)p(\nu)}\right)$.
2. HMC-type RATTLE (Lelievre *et al.*, 2019):
   - define the Hamiltonian for the position and the momentum
   $$H(x, v) = U(x) + \frac{1}{2}v^\top v,$$
   $U(x) = -\log \pi(x)$,
   - add a penalization term to the dynamic
   $$H(x, v) + \lambda^\top q(x).$$

Specific goal

Design couplings of MCMC on manifolds, for guiding the tuning of algorithms and refining chains summaries.

Benefits of coupling

Coupling L-lagged Markov Chains evolving marginally with same target distribution π that eventually meet at a random but finite time τ enables:

- **Parallelization of MCMC computation** through unbiasedness of the posterior summaries (Jacob *et al.*, 2020, Heng and Jacob, 2019).
- **Assessing the convergence of the chains** by deriving bounds in Total Variation (TV) (Biswas *et al.*, 2019):
$$||\pi_t - \pi||_{TV} \leq \max\left\{0, E\left[\left\lceil \frac{\tau - L - t}{L} \right\rceil\right]\right\}.$$
- **Estimating the asymptotic variance** (Douc *et al.*, 2022).

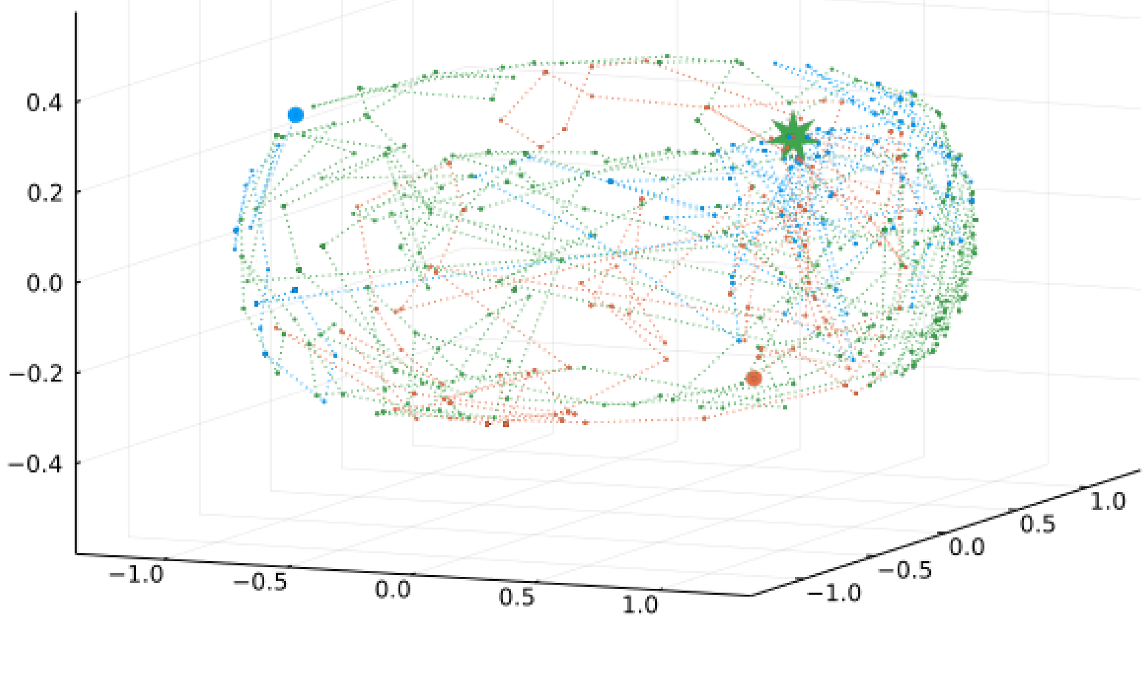

Figure 1. Trajectories of two chains (blue, red lines) on a torus $\mathbf{T}^2$, meeting at the point marked with the star and proceeding together thereafter (green line).

Designing coupled chains on manifolds

Coupling the proposals in MRHT algorithm

Key difficulty: construct a joint proposal kernel under which the meeting has positive probability.

Key idea: the probability of reaching the same point starting from $x, \tilde{x}$ depends on

- the probability of the proposals on the tangent spaces $p(\nu), p(\tilde{\nu})$,
- the probability of hitting $\mathcal{M}$ with a given direction from the tangent space (angles formed by projections) $\rightarrow J(y|x), J(y|\tilde{x})$,
- **the ratio $j = J(y|x)/J(y|\tilde{x})$ represents the infinitesimal expansion or contraction of the space under the change of variables.**

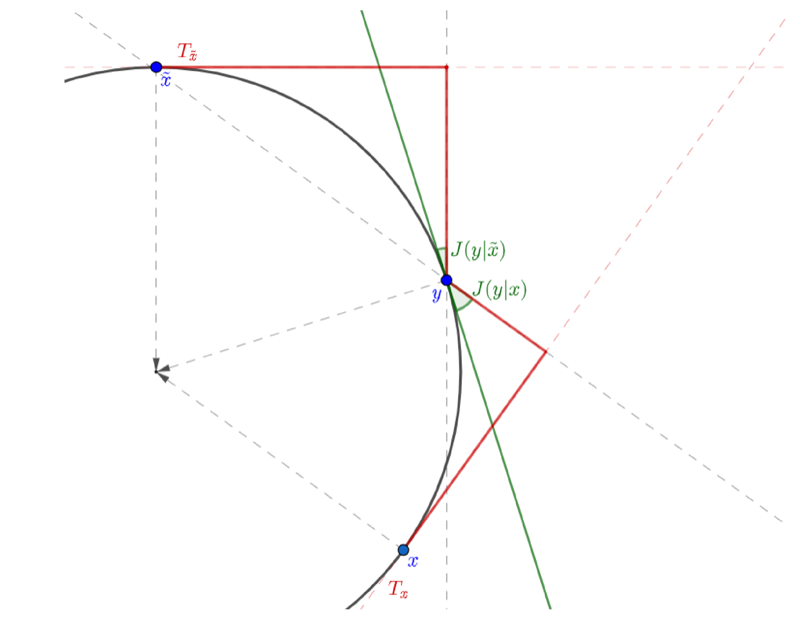

Figure 2. Example of projections from two different initial positions x and $\tilde{x}$.

Algorithm 1 Coupling of proposals

```
x, x̃, Ψ=1, Ψ'=1, no fail of projecting method
Draw ν ~ p(ν)
move along the tangent plane x + νT_x
find y = x + νT_x + αT_x^⊥ s.t. q(y) = 0
compute ν̃ ← T_x̃(y − x̃) for obtaining y from x̃
find ỹ = x̃ + ν̃T_x̃ + α̃T_x̃^⊥ s.t. q(ỹ) = 0
if y ∉ M or ỹ! = y then
    set Ψ = 0, j = 0
else j ← J(y|x)/J(y|x̃)
end if
Draw W ~ Uniform(0, 1)
if W < p(ν̃)·j·Ψ / p(ν) then
    return ν, ν̃
else Reject
    while Reject do
        Draw ν' ~ p(ν')
        move along the tangent plane x̃ + ν'T_x̃
        find y' = x̃ + ν'T_x̃ + α'T_x̃^⊥ s.t. q(y') = 0
        compute ν ← T_x(y' − x) for obtaining y' from x
        find y = x + νT_x + αT_x^⊥ s.t. q(y) = 0
        if y' ∉ M or y! = y' then
            set Ψ' = 0, j' = 0
        else j' ← J(y'|x̃)/J(y'|x)
        end if
        Draw W' ~ Uniform(0, 1),
        Reject= W' < p(ν)·j'·Ψ' / p(ν')
    end while
    return ν, ν'
end if
```

Constrained Hamiltonian Monte Carlo (CHMC)

- Use common random numbers for the chains to update the momenta.
- Adjust the posterior distribution in the dynamic evolution if $f(x)$ is flat on the manifold, to leverage common directions for exploring the space.

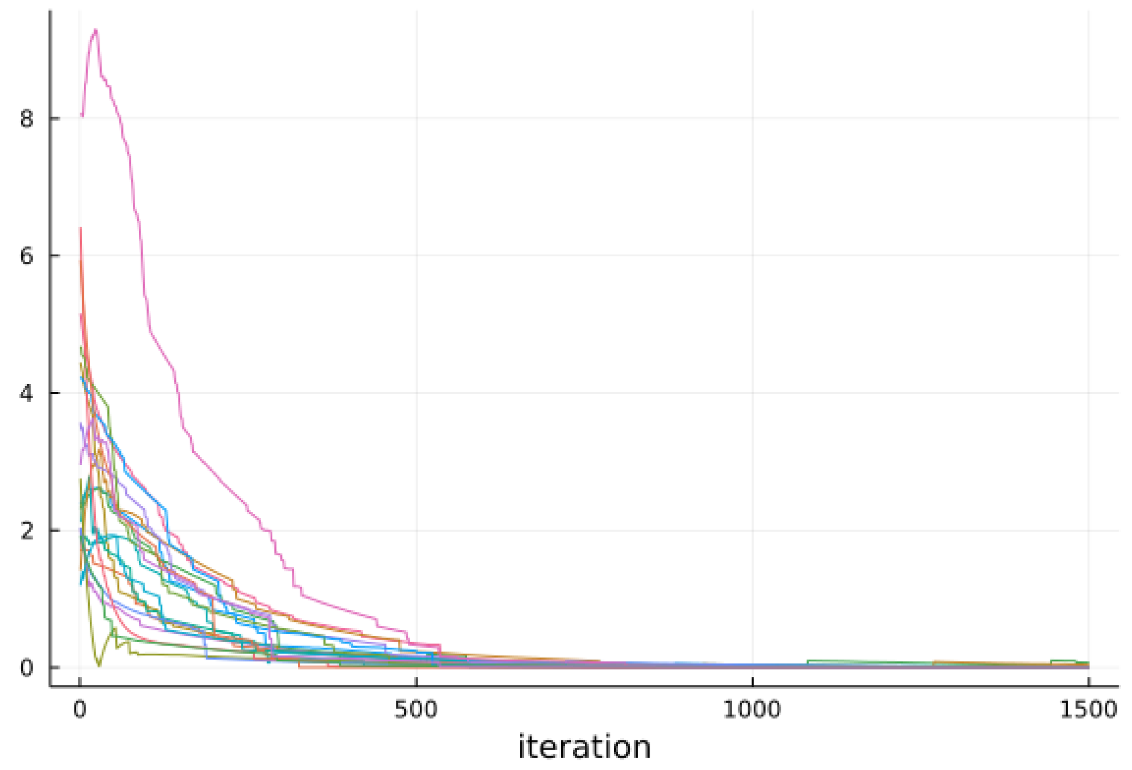

Figure 3. Distance among chains run with HMC and common random numbers.

- Combine the Metropolis with the Hamiltonian kernel in a mixture (Heng and Jacob, 2019):

$$K_{\text{MIX}}(y, \tilde{y}|x, \tilde{x}) = \epsilon K_{\text{MRTH}}(y, \tilde{y}|x, \tilde{x}) + (1 - \epsilon)K_{\text{CHMC}}(y, \tilde{y}|x, \tilde{x}).$$

Example: posterior distribution as model under manifold constraints

Objective: Sampling posterior distributions in double intractable models, where simulating from the data generating process (DGP) is feasible, evaluating the likelihood is not:

$$\pi(\theta|y^{\text{obs}}) \propto \pi(\theta) \underbrace{p(y^{\text{obs}}|\theta)}_{\text{intractable}},$$

formulated as in Graham and Storkey (2017):

1. sample jointly random inputs, u and parameter values, θ and apply a deterministic transformation:$y = g(\theta, u)$,
2. add the constraint $\mathcal{M} = \{(\theta, u) : g(\theta, u) - y^{\text{obs}} = 0\}$
3. take the marginal distributions of $\theta \in \mathcal{M}$: $\pi(\theta|y^{obs})$.

Sum of lognormals model ($D = 11$, $d = 6$, $m = 5$)

- $u_\ell \sim$ Normal$(\theta, 1)$: inputs,
- $\theta \sim$ Normal$(0, 1)$: parameter,
- $z_\ell = \exp(u_\ell), \quad \ell = 1, \dots, L,$
- $y = \sum_{\ell=1}^L z_\ell$: observable data point,
- $g : (\theta, u_{i1}, \dots, u_{iL}) \mapsto \sum_{\ell=1}^L \exp(\theta + u_{i\ell})$, $\forall i = 1, \dots, n,$

(the last four items: } deterministic transfomation)

- $g(\theta, u_{i1}, \dots, u_{iL}) - y_i^{\text{obs}} = 0$: manifold constraints, $\forall i = 1, \dots, n.$

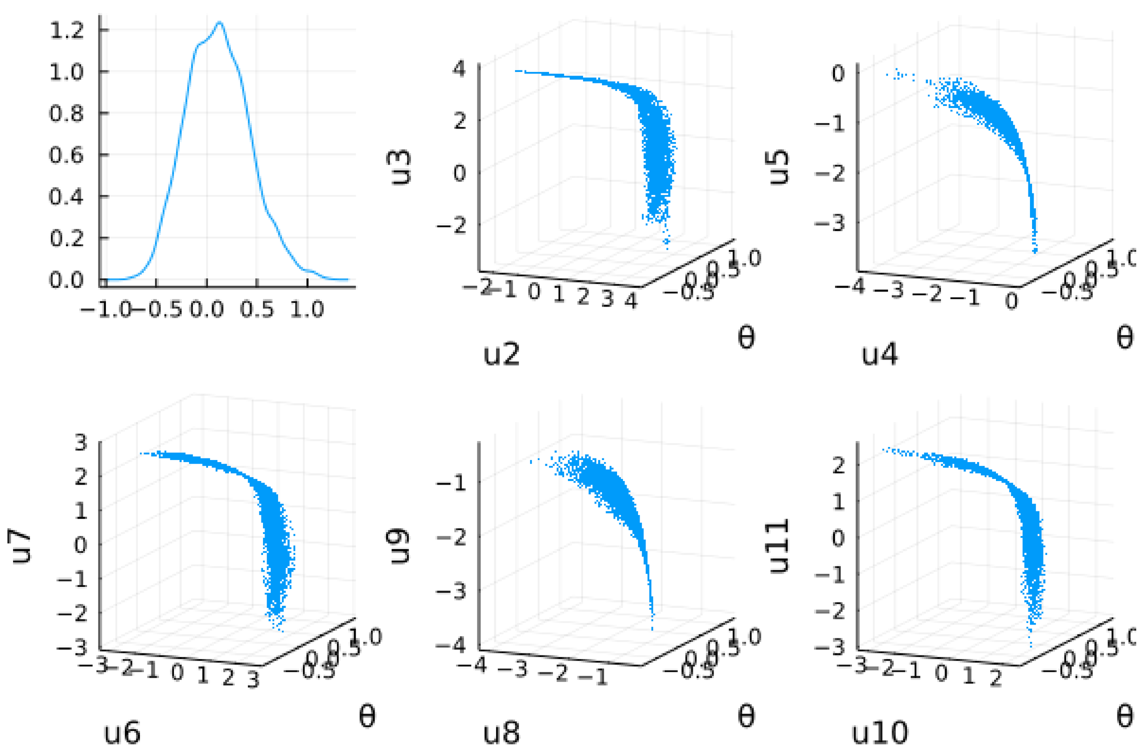

Figure 4. Posterior distribution for θ in the sum of lognormals model.

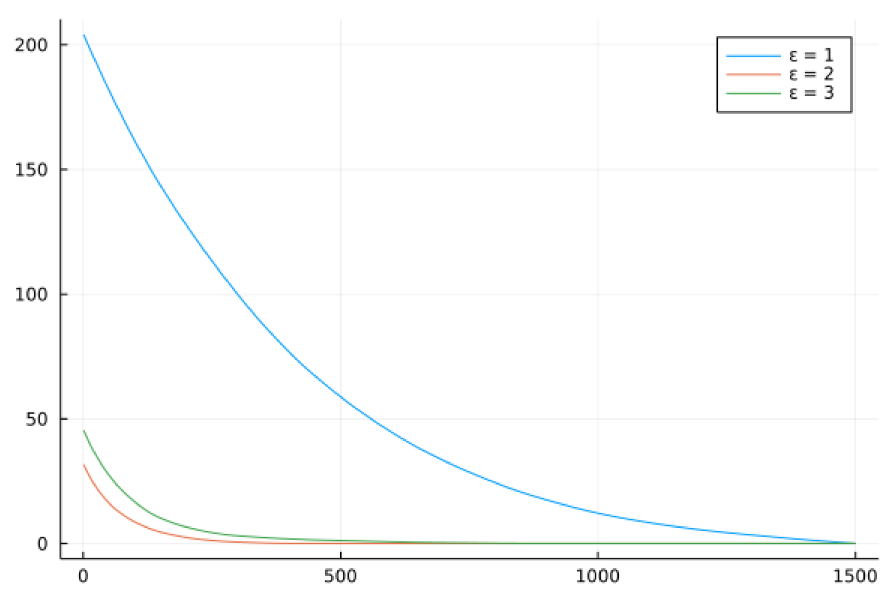

Figure 5. TV upper bounds for Random Walk stepsizes (ϵ=1,2,3).

References

Biswas, N., Jacob, P. E., Vanetti, P. (2019). Estimating convergence of Markov chains with L-lag couplings. Advances in Neural Information Processing Systems, 32.

Douc, R., Jacob, P. E., Lee, A., Vats, D. (2022). Solving the Poisson equation using coupled Markov chains. arXiv preprint arXiv:2206.05691.

Graham, M., Storkey, A. (2017). Asymptotically exact inference in differentiable generative models. In: Artificial Intelligence and Statistics (pp. 499-508). PMLR.

Heng, J., Jacob, P. E. (2019). Unbiased Hamiltonian Monte Carlo with couplings. Biometrika, 106(2), 287-302.

Jacob, P. E., O'Leary, J., Atchadé, Y. F. (2020) Unbiased Markov chain Monte Carlo methods with couplings. Journal of the Royal Statistical Society: Series B (Statistical Methodology) 82.3, 543-600.

Lelièvre, T., Rousset, M., Stoltz, G. (2019). Hybrid Monte Carlo methods for sampling probability measures on submanifolds. Numerische Mathematik, 143(2), 379-421.

Zappa, E., Holmes-Cerfon, M., Goodman, J. (2018). Monte Carlo on manifolds: sampling densities and integrating functions. Communications on Pure and Applied Mathematics, 71(12), 2609-2647.

Contact information

Elena Bortolato, Ph.D. student
Department of Statistical Sciences, University of Padova
elena.bortolato.1@phd.unipd.it
@ https://elenabortolato.github.io/