Hybrid Monte Carlo methods for sampling on submanifolds

Gabriel STOLTZ

(CERMICS, Ecole des Ponts & MATHERIALS team, INRIA Paris)

In collaboration with T. Lelièvre and M. Rousset

Work also supported by ANR Funding ANR-14-CE23-0012 ("COSMOS")

Aim: sampling probability measures with constraints

• **Typical probability measures in stat. physics/Bayesian statistics**: unknowns = parameters in statistics, atomic coordinates for stat phys

$$\nu(dp) = Z^{-1} \mathrm{e}^{-\beta V(q)} \, dq$$

• Equality constraints: molecular constraints (fixed bond lengths, angles, etc), fixed values of reaction coordinates $\xi(q)$ (free energy)

$$\mathcal{M} = \left\{ q \in \mathbb{R}^d, \, \xi(q) = 0 \in \mathbb{R}^m \right\}$$

Target measure to sample ν

$$\nu(dq) = Z_\nu^{-1} \, \mathrm{e}^{-V(q)} \, \sigma_{\mathcal{M}}(dq)$$

where $\sigma_{\mathcal{M}}(dq)$ Lebesgue measure on $\mathcal{M}$ induced by canonical scalar product on $\mathbb{R}^d$

General perspective

- **Constrained overdamped Langevin dynamics**

$$dq_t = -\nabla V(q_t)\,dt + \sqrt{2}\,dW_t + \nabla\xi(q_t)d\lambda_t, \qquad \xi(q_t) = 0$$

Discretization leads to a bias with respect to the timestep...

- **Phase-space formulation + Hybrid Monte Carlo** to remove bias:

$$T_q^*\mathcal{M} = \left\{ p \in \mathbb{R}^d,\, [\nabla\xi(q)]^T\, M^{-1} p = 0 \right\} \subset \mathbb{R}^d,$$

Hamiltonian $H(q,p) = V(q) + |p|^2/2$, sample

$$\mu(dq\,dp) = Z_\mu^{-1} \mathrm{e}^{-H(q,p)}\,\sigma_{T^*\mathcal{M}}(dq\,dp) = \nu(dq)\,\kappa_q(dp),$$

with phase space Liouville measure $\sigma_{T^*\mathcal{M}}(dq\,dp) = \sigma_{\mathcal{M}}(dq)\,\sigma_{T_q^*\mathcal{M}}(dp)$

Constrained Langevin dynamics

$$\begin{cases} dq_t = p_t\,dt\,, \\ dp_t = -\nabla V(q_t)\,dt - \gamma p_t\,dt + \sqrt{2\gamma}\,dW_t + \nabla\xi(q_t)\,d\lambda_t\,, \end{cases} \qquad \text{with } \xi(q_t) = 0$$

Generalized HMC

Generalized Hybrid Monte Carlo

- **Aim:** sample the phase-space measure through momentum resampling + Hamiltonian dynamics

$$\begin{cases} \dot{q}(t) = M^{-1} p(t), \\ \dot{p}(t) = -\nabla V(q(t)) \end{cases}$$

Reversibility: $\phi_t \circ S = S \circ \phi_{-t}$ where $S(q,p) = (q,-p)$ and ϕ_t flow

- In practice, discretization using a reversible scheme, e.g. Verlet

$$\begin{cases} p^{n+1/2} = p^n - \dfrac{\Delta t}{2} \nabla V(q^n) \\ q^{n+1} = q^n + \Delta t \, M^{-1} p^{n+1/2} \\ p^{n+1} = p^{n+1/2} - \dfrac{\Delta t}{2} \nabla V(q^{n+1}) \end{cases}$$

- Two importants properties of the scheme: reversible and preserves the Lebesgue measure

Generalized Hybrid Monte Carlo (2)

- Transition kernel $T(x, x')$ with $x = (q, p)$
- Assume that $r(x, x') = \dfrac{T(S(x'), S(dx))\, \pi(dx')}{T(x, dx')\, \pi(dx)}$ is defined and positive[1]

Generalized Hybrid Monte Carlo (Horowitz, 1991)

- given x^n, propose a new state $\tilde{x}^{n+1}$ from x^n according to $T(x^n, \cdot)$;
- accept the move with probability $\min\left(1, r(x^n, \tilde{x}^{n+1})\right)$, and set in this case $x^{n+1} = \tilde{x}^{n+1}$; otherwise, set $x^{n+1} = S(x^n)$.

- Reversibility up to S, *i.e.* $P(x, dx')\, \mu(dx) = P(S(x'), S(dx))\, \mu(dx')$

- Standard HMC: $T(q, dq') = \delta_{\Phi_\tau(q)}(dq')$, momentum reversal upon rejection (not important since momenta are resampled, but is important when momenta are partially resampled)

[1] T. Lelièvre, M. Rousset, and G. Stoltz, *Free Energy Computations: A Mathematical Perspective*

Generalized Hybrid Monte Carlo (3)

Complete algorithm: starting from (q^0, p^0),

1. update the momentum as $\widetilde{p}^{n+1} = \alpha p^n + \sqrt{\frac{m(1-\alpha^2)}{\beta}}\, G^n$
2. propose $(\widetilde{q}^{n+1}, p^{n+1}) = \Phi_{\Delta t}(q^n, \widetilde{p}^{n+1})$
3. accept with probability $\min\left(1, \mathrm{e}^{-\beta[H(\widetilde{q}^{n+1},p^{n+1})-H(q^n,p^n)]}\right)$ and set $(q^{n+1}, p^{n+1}) = (\widetilde{q}^{n+1}, p^{n+1})$ in this case; otherwise set $(q^{n+1}, p^{n+1}) = (q^n, -\widetilde{p}^{n+1})$

• **Limiting case $\alpha = 0$:** one-step HMC = MALA = Euler-Maruyama discretization of the overdamped Langevin dynamics + Metropolis

$$\widetilde{q}^{n+1} = q^n - h\nabla V(q^n) + \sqrt{\frac{2h}{\beta}}\, G^n, \qquad h = \frac{\Delta t^2}{2}$$

• Possible application: sampling eigenvalues of random matrices[2]

[2] D. Chafaï and G. Ferré, *J. Stat. Phys.* (2018)

(Truly) Reversible RATTLE dynamics

The RATTLE integrator

- Second order discretization of the constrained Hamiltonian dynamics

$$\begin{cases} dq_t = M^{-1} p_t \, dt \,, \\ dp_t = -\nabla V(q_t) \, dt + \nabla \xi(q_t) \, d\lambda_t \,, \\ \xi(q_t) = 0 \end{cases}$$

RATTLE scheme (Andersen, 1983)

$$\begin{cases} p^{n+1/2} = p^n - \dfrac{\Delta t}{2} \nabla V(q^n) + \nabla \xi(q^n) \, \lambda^{n+1/2}, \\ q^{n+1} = q^n + \Delta t \, M^{-1} \, p^{n+1/2}, \\ \xi(q^{n+1}) = 0, & (C_q) \\ p^{n+1} = p^{n+1/2} - \dfrac{\Delta t}{2} \nabla V(q^{n+1}) + \nabla \xi(q^{n+1}) \, \lambda^{n+1}, \\ \left[\nabla \xi(q^{n+1}) \right]^T M^{-1} p^{n+1} = 0, & (C_p) \end{cases}$$

- Momentum constraint always satisfied, but not the position constraint

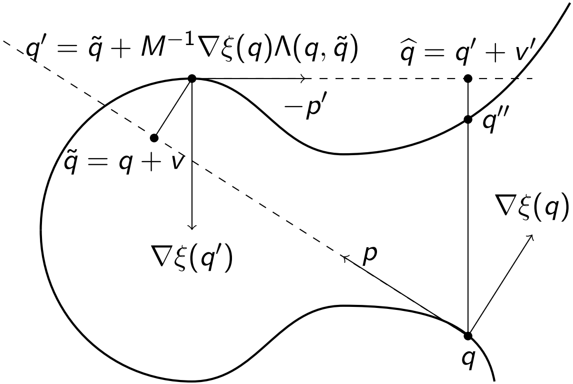
Formal reversibility of RATTLE

- Start from $(q^{n+1}, -p^{n+1})$ and go to $(q^n, -p^n)$
- Initially $\left[\nabla\xi(q^{n+1})\right]^T M^{-1} p^{n+1} = 0$ and $\xi(q^{n+1}) = 0$
- Call $\widetilde{\lambda}^{n+1}$ and $\widetilde{\lambda}^{n+1/2}$ the Lagrange multipliers

$$
\begin{cases}
\left[\nabla\xi(q^{n+1})\right]^T M^{-1} p^{n+1} = 0, \\
-p^{n+1/2} = -p^{n+1} - \dfrac{\Delta t}{2}\nabla V(q^{n+1}) + \nabla\xi(q^{n+1})\,\widetilde{\lambda}^{n+1/2}, \\
\xi(q^{n+1}) = 0, \\
q^n = q^{n+1} - \Delta t\, M^{-1}\, p^{n+1/2}, \\
\xi(q^n) = 0, & (C_q) \\
-p^n = -p^{n+1/2} - \dfrac{\Delta t}{2}\nabla V(q^n) + \nabla\xi(q^n)\,\widetilde{\lambda}^{n+1}, \\
\left[\nabla\xi(q^n)\right]^T M^{-1} p^n = 0, & (C_p)
\end{cases}
$$

- Suggests $\widetilde{\lambda}^{n+1} = \lambda^{n+1/2}$ and $\widetilde{\lambda}^{n+1/2} = \lambda^{n+1}$

What can go wrong with the projection?



The projection may not exist, or may not be unique

RATTLE may not be reversible for large timesteps due to the choice of projection $\rightarrow$ **This prevents using the Metropolis algorithm!!**

Admissible Lagrange multiplier functions

- **Assumption:** $G_M(q) = [\nabla\xi(q)]^T M^{-1}\nabla\xi(q) \in \mathbb{R}^{m\times m}$ is invertible in a neighborhood of $\mathcal{M}$ in $\mathbb{R}^d$
- Note that $q^{n+1} = \tilde{q}^n + \Delta t\, M^{-1}\nabla\xi(q^n)\lambda^{n+1/2}$ (unconstrained move $\tilde{q}^n$)
- Lagrange multipliers $\Delta t\lambda^{n+1/2} = \Lambda(q^n,\tilde{q}^n)$ function of current position (direction of projection) and unconstrained move $\tilde{q}^n$ (can be far off q^n)

Admissible Lagrange multiplier function Λ

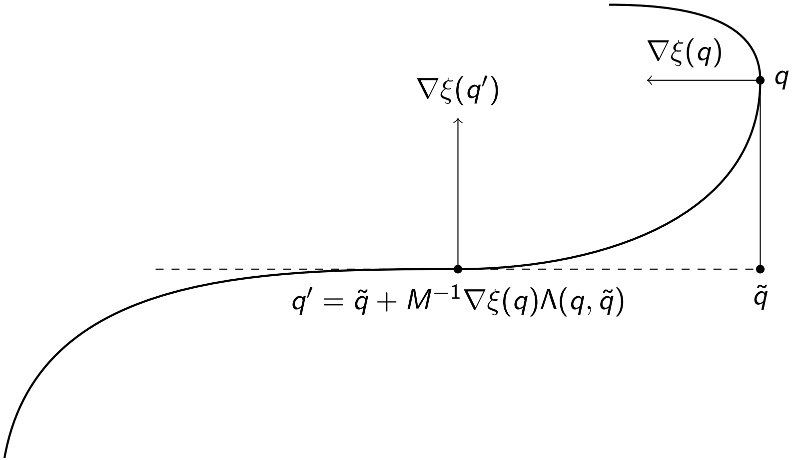
C^1 function defined on an open set $\mathcal{D}$ of $\mathcal{M}\times\mathbb{R}^d$ with values in $\mathbb{R}^m$ with

- projection property: $\forall(q,\tilde{q})\in\mathcal{D}, \quad \tilde{q} + M^{-1}\nabla\xi(q)\Lambda(q,\tilde{q}) \in \mathcal{M}$
- non-tangential projection property: for all $(q,\tilde{q})\in\mathcal{D}$,

$$\left[\nabla\xi\left(\tilde{q} + M^{-1}\nabla\xi(q)\Lambda(q,\tilde{q})\right)\right]^T M^{-1}\nabla\xi(q) \in \mathbb{R}^{m\times m} \text{ is invertible.}$$

- $\mathcal{D}$ contain elements $(q,\tilde{q})\in\mathcal{M}\times\mathcal{M}$ for which $[\nabla\xi(\tilde{q})]^T M^{-1}\nabla\xi(q)$ is invertible (in this case, $\Lambda(q,\tilde{q}) = 0$)

About the non-tangential property



There may be infinitely many possible projections (not isolated points)

Towards a reversible RATTLE scheme

• Composing RATTLE with momentum reversal (involution = good for Metropolis!)

• Admissible set (open) of moves which can be projected back onto $\mathcal{M}$

$$A = \left\{ (q,p) \in T^*\mathcal{M}, \ \left(q, q + \Delta t \, M^{-1} \left[p - \frac{\Delta t}{2} \nabla V(q) \right] \right) \in \mathcal{D} \right\}$$

Can be proved to be non-empty!

• Define $\Psi_{\Delta t}(q,p) = (q^1, -p^1)$ for $(q,p) \in A$ where (q^1, p^1) is obtained from (q,p) by one step of the RATTLE scheme

Properties of $\Psi_{\Delta t}$

The application $\Psi_{\Delta t} : A \to T^*\mathcal{M}$ is a C^1 local diffeomorphism, locally preserving the phase-space measure $\sigma_{T^*\mathcal{M}}(dq\,dp)$

The reversible RATTLE scheme

- Difficulty: analysis at fixed Δt, for all configurations (q, p)

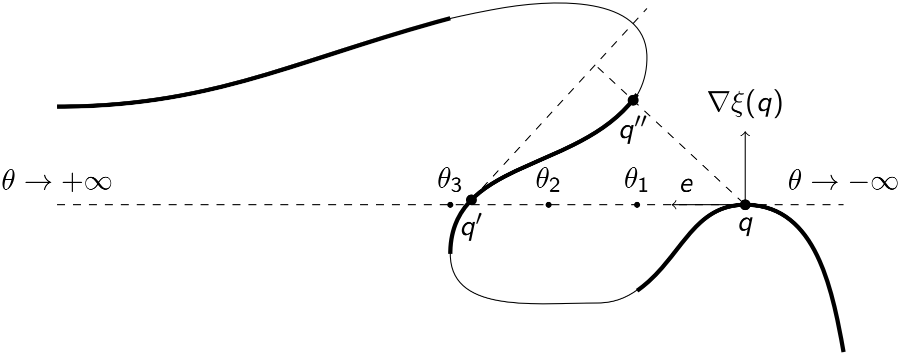
Reversible RATTLE scheme

Define $\Psi_{\Delta t}^{\mathrm{rev}}(q, p) = \Psi_{\Delta t}(q, p) 1_{\{(q,p) \in B\}} + (q, p) 1_{\{(q,p) \notin B\}}$ where

$$B = \Big\{ (q, p) \in A, \ \Psi_{\Delta t}(q, p) \in A \text{ and } (\Psi_{\Delta t} \circ \Psi_{\Delta t})(q, p) = (q, p) \Big\}$$

- Explicitly, for any $(q, p) \in T^* \mathcal{M}$,
  - check if (q, p) is in A; if not return (q, p);
  - when $(q, p) \in A$, compute the configuration (q^1, p^1) obtained by one step of the RATTLE scheme;
  - check if $(q^1, -p^1)$ is in A; if not, return (q, p);
  - compute the configuration $(q^2, -p^2)$ obtained by one step of the RATTLE scheme starting from $(q^1, -p^1)$;
  - if $(q^2, p^2) = (q, p)$, return $(q^1, -p^1)$; otherwise return (q, p).

Illustrating the reverse projection check



Reverse projection check...

- not successful for increments corresponding to $\theta \in (\theta_2, \theta_3)$
- successful for small increments (corresponding to $\theta < \theta_2$) or for sufficiently large ones (corresponding to $\theta > \theta_3$)

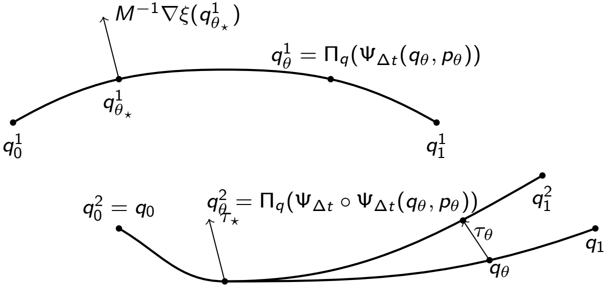
Properties of the reversible RATTLE integrator (1)

On the structure of the set B

Let C be a path connected component of $A \cap \Psi_{\Delta t}^{-1}(A)$. If there is $(q,p) \in C$ such that $(\Psi_{\Delta t} \circ \Psi_{\Delta t})(q,p) = (q,p)$, then

$$\forall (q,p) \in C, \qquad (\Psi_{\Delta t} \circ \Psi_{\Delta t})(q,p) = (q,p).$$

As a corollary, the set B is the union of path connected components of the open set $A \cap \Psi_{\Delta t}^{-1}(A)$. In particular, it is an open set of $T^*\mathcal{M}$.



Properties of the reversible RATTLE integrator (2)

Reversibility and measure preservation

The map $\Psi_{\Delta t}^{\rm rev} : T^*\mathcal{M} \to T^*\mathcal{M}$ is globally well defined, and satisfies

$$\Psi_{\Delta t}^{\rm rev} \circ \Psi_{\Delta t}^{\rm rev} = \mathrm{Id}.$$

Moreover, $\Psi_{\Delta t}^{\rm rev} : B \to B$ and $\Psi_{\Delta t}^{\rm rev} : B^c \to B^c$ are C^1-diffeomorphisms preserving $\sigma_{T^*\mathcal{M}}(dq\,dp)$. As a consequence, $\Psi_{\Delta t}^{\rm rev} : T^*\mathcal{M} \to T^*\mathcal{M}$ globally preserves the measure $\sigma_{T^*\mathcal{M}}(dq\,dp)$.

- **In practice...**
  - Theoretical construction of sets $\mathcal{D}$, A, B from implicit function theorem (however, points q and $\tilde{q}$ not required to be close; but $\tilde{q}$ should still be close to $\mathcal{M}$)
  - Newton algorithm to bring $\tilde{q}$ in the "implicit function theorem" regime
  - Metropolis acceptance/rejection + momentum resampling

The method in practice

Sampling the constrained Gibbs measure

Algorithm: Starting from $(q^n, p^n) \in T^*\mathcal{M}$,

1. Update momenta (with $\alpha = \mathrm{e}^{-\gamma \Delta t}$)

$$p^{n+1/4} = P(q^n)\left[\alpha p^n + \sqrt{1-\alpha^2}\, G^n\right], \quad P(q) = \mathrm{Id} - \frac{\nabla\xi(q) \otimes \nabla\xi(q)}{|\nabla\xi(q)|^2}$$

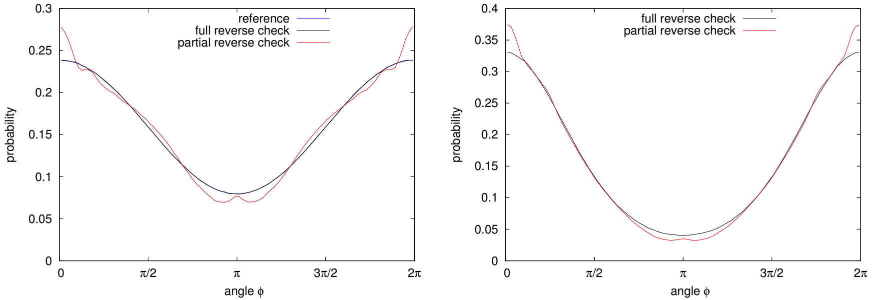
2. Evolve with reversible RATTLE: $(\tilde{q}^{n+1}, \tilde{p}^{n+3/4}) = \Psi^{\mathrm{rev}}_{\Delta t}(q^n, p^{n+1/4})$
3. Draw a random variable U^n with uniform law on $(0,1)$:
   - if $U^n \leqslant \exp(-H(\tilde{q}^{n+1}, \tilde{p}^{n+3/4}) + H(q^n, p^{n+1/4}))$, accept the proposal
   - else reject the proposal: $(q^{n+1}, p^{n+3/4}) = (q^n, p^{n+1/4})$.
4. Reverse momenta $p^{n+1} = -p^{n+3/4}$.

Test partial reverse check ($\Psi_{\Delta t} \circ \Psi_{\Delta \mathrm{t}}$ is well defined but do not check whether $\Psi_{\Delta t} \circ \Psi_{\Delta \mathrm{t}}(q, p) = (q, p)$)

The need for reversibility checks

Simple example: Three-dimensional system, one-dimensional constraint

- $q = (x, y, z) \in \mathbb{R}^3$ and $\xi(q) = \left(R - \sqrt{x^2 + y^2}\right)^2 + z^2 - r^2$
- Potential $V(q) = k|q|^2/2$

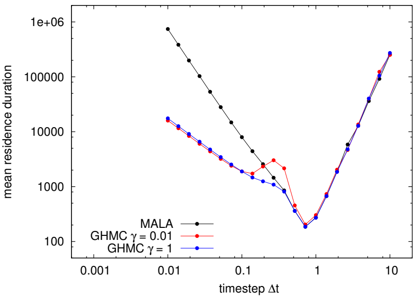


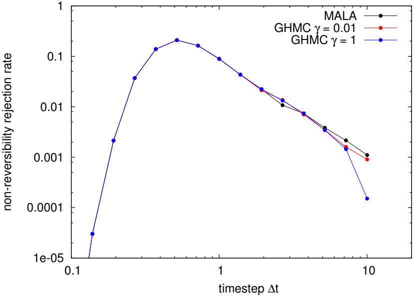
Histograms of the sampled angles ϕ with the GHMC scheme, with full or partial reverse projection check, for $\Delta t = 1$. Left: $k = 0$. Right: $k = 1$.

Analysis of the rejection rate

Method	Total	Newton	Newton rev.	non-rev.	Metropolis
MRW $\Delta t = 1$	0.675	0.562	$3.02 \cdot 10^{-4}$	0.0742	0.0385
MALA $\Delta t = 1$	0.675	0.509	$5.83 \cdot 10^{-4}$	0.149	0.0167
GHMC $\Delta t = 1$, $\alpha = 0.1$	0.675	0.509	$5.83 \cdot 10^{-4}$	0.149	0.0167
GHMC $\Delta t = 1$, $\alpha = 0.5$	0.675	0.509	$5.83 \cdot 10^{-4}$	0.149	0.0167
GHMC $\Delta t = 1$, $\alpha = 0.9$	0.675	0.509	$5.83 \cdot 10^{-4}$	0.149	0.0167
MRW $\Delta t = 0.3$	0.158	0.0803	$1.06 \cdot 10^{-4}$	0.0127	0.0652
MALA $\Delta t = 0.3$	0.107	0.0763	$1.22 \cdot 10^{-4}$	0.0138	0.0168
GHMC $\Delta t = 0.3$, $\alpha = 0.1$	0.107	0.0763	$1.22 \cdot 10^{-4}$	0.0138	0.0168
GHMC $\Delta t = 0.3$, $\alpha = 0.5$	0.107	0.0763	$1.22 \cdot 10^{-4}$	0.0138	0.0168
GHMC $\Delta t = 0.3$, $\alpha = 0.9$	0.107	0.0763	$1.22 \cdot 10^{-4}$	0.0138	0.0168
MRW $\Delta t = 0.1$	0.0259	$5 \cdot 10^{-7}$	0	$7 \cdot 10^{-8}$	0.0259
MALA $\Delta t = 0.1$	$6.73 \cdot 10^{-4}$	$5 \cdot 10^{-7}$	10^{-9}	$5 \cdot 10^{-8}$	$6.73 \cdot 10^{-4}$
GHMC $\Delta t = 0.1$, $\alpha = 0.1$	$6.72 \cdot 10^{-4}$	$5 \cdot 10^{-7}$	10^{-9}	$6 \cdot 10^{-8}$	$6.72 \cdot 10^{-4}$
GHMC $\Delta t = 0.1$, $\alpha = 0.5$	$6.73 \cdot 10^{-4}$	$5 \cdot 10^{-7}$	$2 \cdot 10^{-9}$	$8 \cdot 10^{-8}$	$6.72 \cdot 10^{-4}$
GHMC $\Delta t = 0.1$, $\alpha = 0.9$	$6.74 \cdot 10^{-4}$	$5 \cdot 10^{-7}$	0	$7 \cdot 10^{-8}$	$6.73 \cdot 10^{-4}$

Metastability analysis for a double-well potential





Left: mean residence duration as a function of the timestep. Right: non-reversibility rejection rate

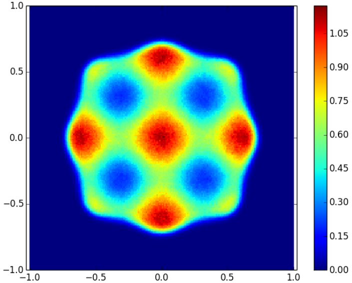
Maximal (and non negligible) non reversibility rejection rate at the optimal timestep!

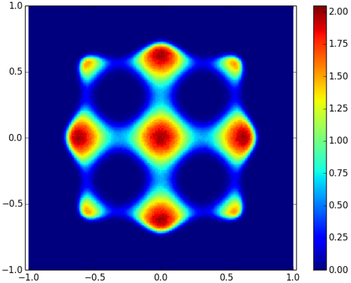
Sampling eigenvalues of random matrices with constraints

Ginibre ensemble (hermitian matrices with Gaussian entries): $d = 2$, $V(x) = |x|^2$ and $w(x) = -\log|x|$, at inverse temperature $\beta_n = n^2$

$$H_n(x_1, \dots, x_n) = \frac{1}{n}\sum_{i=1}^{n} V(x_i) + \frac{1}{n^2}\sum_{i \neq j} w(x_i - x_j)$$

with $\xi_n(x_1, \dots, x_n) = \dfrac{1}{n}\displaystyle\sum^{n} \varphi(x_i) - c$ and $\varphi(x) = \dfrac{\cos(5x_1) + \cos(5x_2)}{2}$





References

This work:

T. Lelièvre, M. Rousset, and G. Stoltz, Hybrid Monte Carlo methods for sampling probability measures on submanifolds, *Numer. Math.* (2019)

Motivated by:

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

Application to sampling random matrix models:

D. Chafaï, G. Ferré and G. Stoltz, Coulomb gases under constraint: some theoretical and numerical results, *arXiv preprint* **1907.05803** (2019)

Building on the previous works:

T. Lelièvre, M. Rousset, and G. Stoltz, *Free-energy Computations: A Mathematical Perspective*, Imperial College Press (2010)

T. Lelièvre, M. Rousset, and G. Stoltz, Langevin dynamics with constraints and computation of free energy differences, *Math. Comput.* (2012)