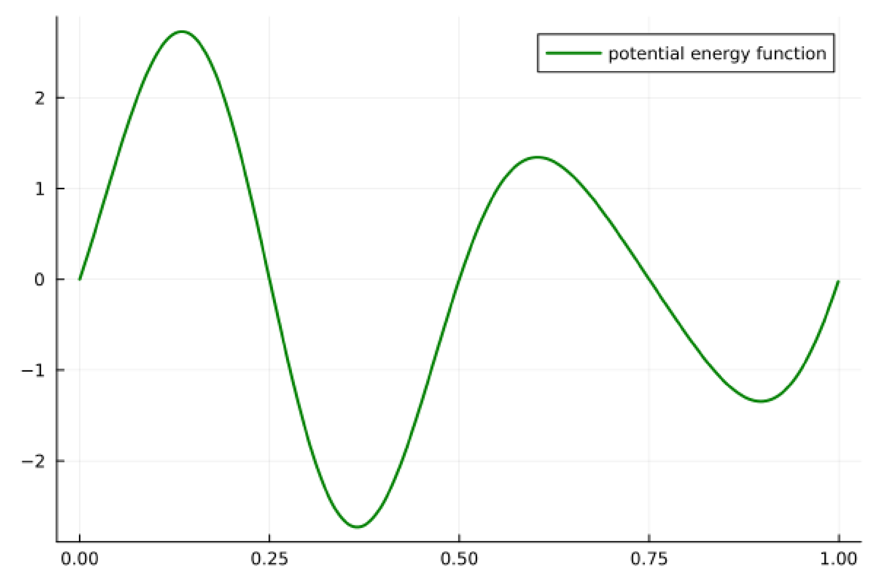
Optimizing the diffusion of overdamped Langevin dynamics

T. LELIÈVRE[1], G. PAVLIOTIS[2], G. ROBIN[3], RÉGIS SANTET[1], G. STOLTZ[1]

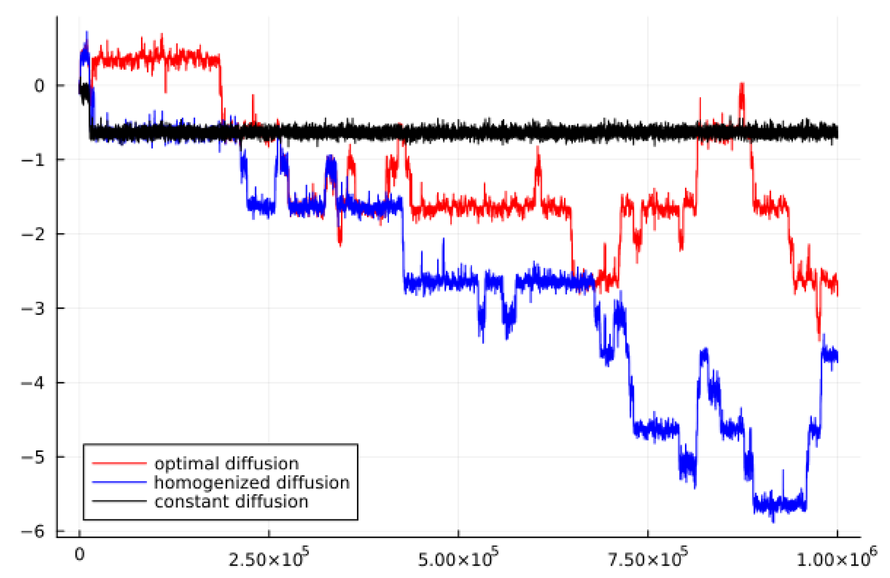
[1]CERMICS, Ecole des Ponts, Marne-la-Vallée, France & MATHERIALS project-team, Inria Paris, France [2]Department of Mathematics, Imperial College, London, United-Kingdom

[3]CNRS & Laboratoire de Mathématiques et Modélisation d'Évry, Évry, France

Improve sampling for metastable dynamics

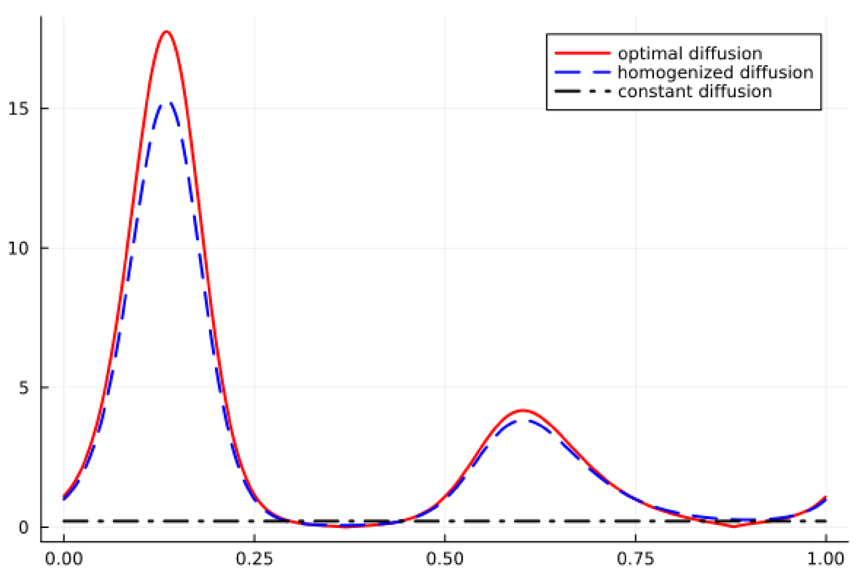


$V(q) = \sin(4\pi q)(2 + \sin(2\pi q))$

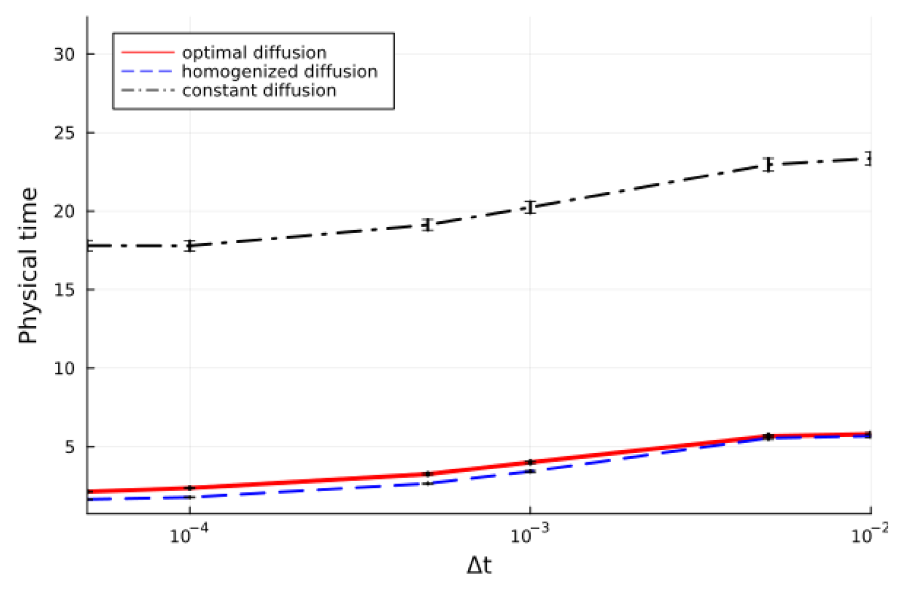


Markov chains

typical trajectories (RWMH, unperiodized)



Diffusion coefficients



Transition times

from the deepest well to its nearby copy, average of 10^5 transitions

Diffusion dependent Overdamped Langevin dynamics

- **Aim**: Improve sampling efficiency when estimating

$$\mathbb{E}_\pi[f] = \int_{\mathbb{T}^d} f(q)\pi(q)\,\mathrm{d}q \quad \text{with} \quad \widehat{I}_N = \frac{1}{N}\sum_{i=1}^N f(q^i), \qquad q^i \sim \pi \propto \mathrm{e}^{-\beta V}$$

- **Method**: Euler–Maruyama scheme + Metropolis–Hastings correction:

$$\mathrm{d}q_t = \left(-\mathcal{D}(q_t)\nabla V(q_t) + \beta^{-1}\operatorname{div}\mathcal{D}(q_t)\right)\mathrm{d}t + \sqrt{2\beta^{-1}\mathcal{D}(q_t)}\,\mathrm{d}W_t \qquad (1)$$

→ Map $\mathcal{D}\colon \mathbb{T}^d \to \mathcal{S}_d^+(\mathbb{R})$ introduced to favor exploration in **anisotropic** or **metastable** potential landscapes[2]. If $\mathcal{D} \equiv \mathrm{I}_d$, this is MALA[4]

- **Main property**: convergence quantified as[3]

$$\left\| \frac{\pi_t}{\pi} - 1 \right\|_{L^2(\pi)} \leqslant \mathrm{e}^{-\Lambda(\mathcal{D})\beta^{-1}t} \left\| \frac{\pi_0}{\pi} - 1 \right\|_{L^2(\pi)}, \qquad q_t \sim \pi_t$$

→ $\mathcal{L}_\mathcal{D} = -\beta^{-1}\nabla^*\mathcal{D}\nabla$ on $L^2(\pi)$ is the generator of (1)

→ $\Lambda(\mathcal{D})$ is the spectral gap of $-\beta\mathcal{L}_\mathcal{D} \geqslant 0$

- **Idea**: Compute $\mathcal{D}^\star$ leading to a **large spectral gap**

→ Numerically via an optimization procedure

→ Explicitly via a homogenization procedure

Optimization problem

Objective function

From min-max principle:

$$\Lambda(\mathcal{D}) = \min_{\substack{u\in H^1(\mathbb{T}^d)\\ u\neq 0}} \left\{ \frac{\int_{\mathbb{T}^d} \nabla u^\top \mathcal{D}\nabla u \,\mathrm{d}\pi}{\int_{\mathbb{T}^d} u^2\,\mathrm{d}\pi} \;\middle|\; \int_{\mathbb{T}^d} u\,\mathrm{d}\pi = 0 \right\}$$

→ Need to normalize $\mathcal{D}$. If $\|\mathcal{D}\| \uparrow$, then $\Lambda(\mathcal{D}) \uparrow$, but $\Delta t\mathcal{D}$ is what appears in (1): timestep has to compensate $\Delta t \downarrow$

L^p Constraint

$$\mathcal{D} \in \mathfrak{D}_p^{a,b} = \left\{\mathcal{D} \in L_\pi^\infty(\mathbb{T}^d, \mathcal{M}_{a,b}) \;\middle|\; \|\mathcal{D}\|_{L_\pi^p} \leqslant 1\right\}$$

$$\mathcal{M}_{a,b} = \left\{ M \in \mathcal{S}_d^+ \;\middle|\; \forall \xi, a\,|\xi|^2 \leqslant \xi^\top M \xi \leqslant b^{-1}|\xi|^2 \right\}$$

$$L_\pi^p(\mathbb{T}^d, \mathcal{M}_{a,b}) = \left\{ \mathcal{D},\ \mathrm{e}^{-\beta V(q)}\mathcal{D}(q) \in \mathcal{M}_{a,b} \text{ a.e.} \right\}$$

$$\|\mathcal{D}\|_{L_\pi^p} = \left(\int_{\mathbb{T}^d} |\mathcal{D}(q)|_{\mathrm{F}}^p\, \mathrm{e}^{-\beta p V(q)}\mathrm{d}q \right)^{1/p}$$

Theoretical analysis

- $V \in \mathcal{C}^\infty(\mathbb{T}^d)$: V and π bounded
- π satisfies a Poincaré inequality
- $\mathfrak{D}_p^{a,b}$ weakly closed for L_π^p
- $\mathcal{D} \mapsto \Lambda(\mathcal{D})$ is concave, semi-lower continuous

→ Similar result for discretized optimization problem

THEOREM. For any $p \in [1, +\infty)$ there exists

$$\mathcal{D}_p^\star = \underset{\mathcal{D}\in\mathfrak{D}_p^{a,b}}{\arg\max}\, \Lambda(\mathcal{D})$$

such that $\|\mathcal{D}_p^\star\|_{L_\pi^p} = 1$ and $\mathcal{D}_p^\star \neq 0$ a.e.

Characterization of a maximizer

Euler–Lagrange equation leads to $\left[\ker\left(\mathcal{L}_{\mathcal{D}_p^\star} - \Lambda(\mathcal{D}_p^\star)\mathrm{I}\right) = \operatorname{Span}\left(u_{\mathcal{D}_p^\star}^i\right)_{1\leqslant i\leqslant N}\right]$

$$\mathcal{D}_p^\star \propto \left|\mathcal{D}_p^\star\right|_{\mathrm{F}}^{2-p} \mathrm{e}^{\beta(p-1)V} \sum_{i=1}^N \nabla u_{\mathcal{D}_p^\star}^i \otimes \nabla u_{\mathcal{D}_p^\star}^i$$

If $\Lambda(\mathcal{D}_p^\star)$ is isolated: $\mathcal{D}_p^\star$ is of rank 1 a.e, and has to vanish on $\mathbb{T}^d$

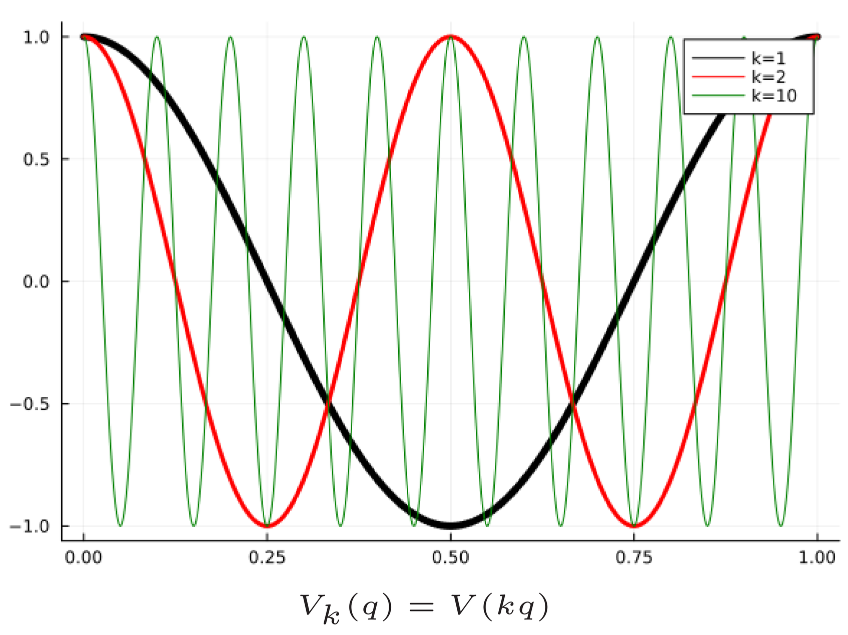
- $\mathbb{P}_1$ FEM to compute $(\Lambda(\mathcal{D}), u_\mathcal{D})$: $A(\mathcal{D})u_\mathcal{D} = \Lambda(\mathcal{D})Bu_\mathcal{D}$

Homogenization

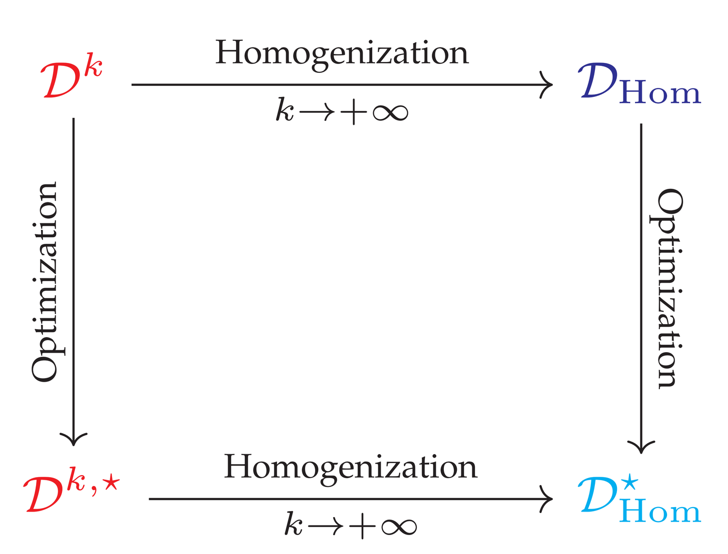
- **Issue:** Optimization numerical procedure only helpful in low dimensions
- **Goal:** Obtain a good approximation/proxy

→ Asymptotic behaviour of the optimal diffusion in the homogenized limit[1]

→ Optimize the periodic homogenization limit



$V_k(q) = V(kq)$



Homogenization vs. Optimization

$V(q) = \sin(4\pi q)(2 + \sin(2\pi q)),\ a = b = 0,\ p = 2$

Diffusion coefficient	Constant	Homogenized	Optimal
Spectral gap	2.16	10.57	11.23

THEOREM. If $d = 1$, then $\mathcal{D}_{\mathrm{Hom}}^\star = \mathrm{e}^{\beta V}$ is a maximizer.

→ Suggests using $\mathcal{D} = \mathrm{e}^{\beta V}\mathrm{I}_d$ or $\mathcal{D} = \mathrm{e}^{\beta F}\mathrm{I}_d$, F being the free energy

References

[1] G. Allaire. *Shape Optimization by the Homogenization Method*. Applied Mathematical Sciences. Springer-Verlag New York, 2002.

[2] N. Bou-Rabee, A. Donev, and E. Vanden-Eijnden. "Metropolis integration schemes for self-adjoint diffusions". In: *Multiscale Modeling & Simulation* (2014).

[3] T. Lelièvre, F. Nier, and G. A. Pavliotis. "Optimal Non-reversible Linear Drift for the Convergence to Equilibrium of a Diffusion". In: *Journal of Statistical Physics* (2013).

[4] G. O. Roberts and R. L. Tweedie. "Exponential convergence of Langevin distributions and their discrete approximations". In: *Bernoulli* (1996).