# Optimizing the diffusion for sampling with overdamped Langevin dynamics

**Gabriel STOLTZ**

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

In collaboration with Tony Lelièvre, Greg Pavliotis, Geneviève Robin and Régis Santet

Project funded by ANR SINEQ and ERC Synergy EMC2

*Program "Probabilistic sampling for physics" at Institut Pascal*

# Outline

- **Overdamped Langevin dynamics**
  - General diffusion coefficients
  - Convergence rates

**Characterization of the optimal diffusion**

- Normalization of the diffusion
- Necessary conditions
- Approximation in the homogenized limit

**Numerical results**

- Numerical approximation of the optimal diffusion
- Numerical integration of overdamped Langevin dynamics
- Efficiency gains from optimized diffusions

# Overdamped Langevin dynamics

# Computing average properties

**Aim:** Sample target measure $\mu(dq) = Z_\mu^{-1} \mathrm{e}^{-\beta V(q)} \, dq$ on $\mathcal{Q}$
(assume $Z_\mu = 1$ in the remainder)

## Main issue

Computation of high-dimensional integrals... Ergodic averages

$$\mathbb{E}_\mu(\varphi) = \lim_{t \to +\infty} \widehat{\varphi}_t, \qquad \widehat{\varphi}_t = \frac{1}{t} \int_0^t \varphi(q_s) \, ds$$

- One possible choice: overdamped Langevin dynamics

= Stochastic perturbation of gradient dynamics

$$dq_t = -\nabla V(q_t) \, dt + \sqrt{\frac{2}{\beta}} \, dW_t$$

- Other choices include Metropolis-like schemes

# Properties of the standard overdamped Langevin dynamics

**Generator** $\mathcal{L} = -\nabla V(q) \cdot \nabla_q + \frac{1}{\beta}\Delta_q$

- elliptic generator hence irreducibility and ergodicity
- adjoints on $L^2(\mathcal{Q})$ versus $L^2(\mu)$

$$\int_{\mathcal{Q}} (\mathcal{L}f)\, g = \int_{\mathcal{Q}} f\left(\mathcal{L}^\dagger g\right), \qquad \int_{\mathcal{Q}} (\mathcal{L}f)\, g\, d\mu = \int_{\mathcal{Q}} f\,(\mathcal{L}^\star g)\, d\mu$$

- flat adjoint $\mathcal{L}^\dagger \varphi = \mathrm{div}_q\left((\nabla V)\varphi + \frac{1}{\beta}\nabla_q \varphi\right)$
- self-adjoint operator on $L^2(\mu)$, hence reversibility

$$\mathcal{L} = -\frac{1}{\beta}\nabla_q^\star \nabla_q = \mathcal{L}^\star, \qquad \partial_{q_i}^\star = -\partial_{q_i} + \beta \partial_{q_i} V$$

Invariance of canonical measure encoded as $\mathcal{L}^\dagger \mu$ or $\mathcal{L}^\star \mathbf{1} = 0$

$$\frac{d}{dt}\left[\mathbf{E}_\mu\left(\varphi(X_t)\right)\right] = \frac{d}{dt}\left(\int_{\mathcal{Q}} \mathrm{e}^{t\mathcal{L}}\varphi\, d\mu\right) = \int_{\mathcal{Q}} \mathcal{L}\left(\mathrm{e}^{t\mathcal{L}}\varphi\right) d\mu = 0$$

# Overdamped Langevin dynamics with multiplicative noise

**Diffusion matrix** $\mathcal{D}(q) \in \mathbb{R}^d$ (symmetric positive, not necessarily definite)

$$dq_t = \left(-\mathcal{D}(q_t)\nabla V(q_t) + \frac{1}{\beta}\mathrm{div}\mathcal{D}(q_t)\right) dt + \sqrt{\frac{2}{\beta}}\mathcal{D}^{1/2}(q_t)\, dW_t$$

with $\mathrm{div}\,\mathcal{D}$ the vector whose $i$-th component is the divergence of the $i$-th column of the matrix $\mathcal{D} = [\mathcal{D}_1, \dots, \mathcal{D}_d]$

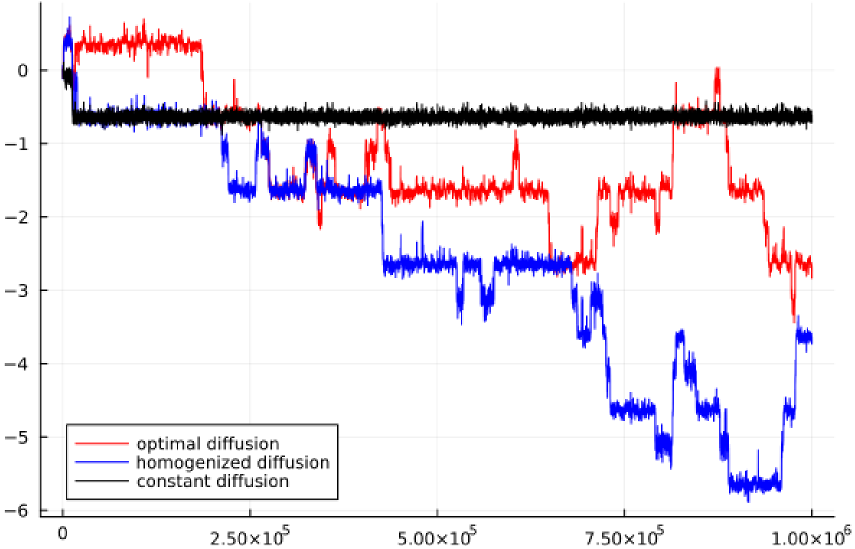
**Two possible motivations:**

- Compensate for anisotropic potential energy landscapes
- Reduce metastability

Generator still self-adjoint, invariant probability measure $\mu$

$$\mathcal{L}_{\mathcal{D}} = -\frac{1}{\beta}\nabla^\star \mathcal{D}\nabla = -\frac{1}{\beta}\sum_{i,j=1}^d \partial_{q_j}^\star \mathcal{D}_{i,j}\partial_{q_i}$$

# Behavior of overdamped Langevin dynamics for various $\mathcal{D}$



# Convergence of overdamped Langevin dynamics

**Various measures of convergence**, for instance

- asymptotic variance in central limit theorem
- convergence of the law at time $t$ to the stationary distribution
- average exit time of a metastable well

Here: second option, in a $L^2(\mu)$ framework

Law at time $t$ written as $\psi(t) = f(t)\mu$, so that $f(t) = \mathrm{e}^{t\mathcal{L}_{\mathcal{D}}} f(0)$

$$\mathbf{E}_{\psi(0)}\left(\varphi(X_t)\right) = \int_{\mathcal{Q}} \varphi\, f(t)\, d\mu = \int_{\mathcal{Q}} \mathrm{e}^{t\mathcal{L}_{\mathcal{D}}}\varphi\, f(0)\, d\mu$$

**Typical convergence result:** exponential convergence rate for $\mathrm{e}^{t\mathcal{L}_{\mathcal{D}}}$

$$\|f(t) - \mathbf{1}\|_{L^2(\mu)} = \left\|\mathrm{e}^{t\mathcal{L}_{\mathcal{D}}}\left(f(0) - \mathbf{1}\right)\right\|_{L^2(\mu)} \leqslant \mathrm{e}^{-\Lambda(\mathcal{D})t/\beta}\|f(0) - \mathbf{1}\|_{L^2(\mu)},$$

Implies bounds on the asymptotic variance

# Obtaining an exponential rate of convergence

**Spectral gap** on $H^1_0(\mu) = \left\{ u \in H^1(\mu) \,\middle|\, \int_{\mathbb{T}^d} u(q)\,\mu(q)\,dq = 0 \right\}$

$$\Lambda(\mathcal{D}) = \inf_{u \in H^1_0(\mu)\setminus\{0\}} \frac{\displaystyle\int_{\mathbb{T}^d} \nabla u(q)^\top \mathcal{D}(q) \nabla u(q)\,\mu(q)\,dq}{\displaystyle\int_{\mathbb{T}^d} u(q)^2 \mu(q)\,dq}$$

Desired inequality follows from a Gronwall estimate and

$$\frac{d}{dt}\left(\frac{1}{2}\left\|\mathrm{e}^{t\mathcal{L}_{\mathcal{D}}}\varphi\right\|^2_{L^2(\mu)}\right) = \left\langle \mathrm{e}^{t\mathcal{L}_{\mathcal{D}}}\varphi, \mathcal{L}_{\mathcal{D}}\,\mathrm{e}^{t\mathcal{L}_{\mathcal{D}}}\varphi \right\rangle_{L^2(\mu)} \leqslant -\frac{\Lambda(\mathcal{D})}{\beta}\left\|\mathrm{e}^{t\mathcal{L}_{\mathcal{D}}}\varphi\right\|^2_{L^2(\mu)}$$

**Criterion to choose $\mathcal{D}$**

Maximize the spectral gap $\Lambda(\mathcal{D})$

**Possible choices:**

- $\mathcal{D} = \left(\nabla^2 V\right)^{-1}$ for strictly convex potentials [Girolami/Calderhead 2011]
- $\mathcal{D} = \mathrm{e}^{\beta V}$ [Roberts/Stramer 2002, Ghimenti/van Wijland/... 2023]

# Characterization of the optimal diffusion

# Need for normalization

**Motivation:** $\Lambda(\alpha\mathcal{D}) = \alpha\Lambda(\mathcal{D})$ and large $\mathcal{D}$ requires smaller timesteps

$L^\infty$ bounds trivial (saturate the constraint)

**Chosen normalization:** $L^p_\mu(\mathbb{T}^d, \mathcal{M}_{a,b})$ (note $\mathcal{Q} = \mathbb{T}^d$), with associated norm

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

and requirement $\mathrm{e}^{\beta V}\mathcal{D} \in \mathcal{M}_{a,b}$ with (for $a, b \geqslant 0$)

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

Matrix norm compatible with order on symmetric positive matrices

**Maximization performed on**

$$\mathfrak{D}_p^{a,b} = \left\{ \mathcal{D} \in L^\infty_\mu(\mathbb{T}^d, \mathcal{M}_{a,b}) \,\middle|\, \int_{\mathbb{T}^d} |\mathcal{D}(q)|_{\mathrm{F}}^p \, \mathrm{e}^{-\beta p V(q)} \, dq \leqslant 1 \right\}$$

# Well posedness of the maximization problem

**Existence of maximizer for** $p \in [1, +\infty)$: For any $a \in [0, |\mathrm{Id}_d|_{\mathrm{F}}^{-1}]$ and $b > 0$ such that $ab \leqslant 1$, there exists $\mathcal{D}_p^\star \in \mathfrak{D}_p^{a,b}$ such that

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

In addition,

- For any open set $\Omega \subset \mathbb{T}^d$, there exists $q \in \Omega$ such that $\mathcal{D}_p^\star(q) \neq 0$
- $\displaystyle \int_{\mathbb{T}^d} \left|\mathcal{D}_p^\star(q)\right|_{\mathrm{F}}^p \, \mathrm{e}^{-\beta p V(q)} \, dq = 1$

**Main arguments/properties:**

- $\Lambda$ is bounded (Poincaré inequality)
- $\Lambda$ is concave (sup of linear functions in $\mathcal{D}$)
- $\Lambda$ is upper semicontinuous for the weak-* $L^\infty_\mu$ topology $(b > 0)$
- the set $\mathfrak{D}_p^{a,b}$ is compact for the weak-* $L^\infty_\mu$ topology

# Characterization for non-degenerate eigenvalue $\Lambda(\mathcal{D}_p^\star)$

**Main result:** The maximizer $\mathcal{D}_p^\star$ cannot be of full rank on $\mathbb{T}^d$

**Precise statement:**

- Frobenius norm $|\cdot|_{\mathrm{F}}$, Lebesgue exponent $p \in (1, +\infty)$ and $a = b = 0$
- assume that there exists a maximizer $\mathcal{D}_p^\star$ of $\Lambda$ on $\mathfrak{D}_p^{a,b}$ and an eigenvector $u_{\mathcal{D}_p^\star}$ satisfying $-\beta \mathcal{L}_{\mathcal{D}_p^\star} u_{\mathcal{D}_p^\star} = \Lambda(\mathcal{D}_p^\star) u_{\mathcal{D}_p^\star}$
- additionally $\mathcal{D}_p^\star \in C^0(\mathbb{T}, \mathbb{R}_+)$ when $d = 1$

If $\mathcal{D}_p^\star \geqslant c\mathrm{Id}_d$, then $\Lambda(\mathcal{D}_p^\star)$ is a degenerate eigenvalue of $-\beta \mathcal{L}_{\mathcal{D}_p^\star}$.

**Idea of proof:** From the Euler–Lagrange equation (regular perturbation theory)

$$\int_{\mathbb{T}^d} \delta\mathcal{D}(q) : \left( \nabla u_{\mathcal{D}_p^\star} \otimes \nabla u_{\mathcal{D}_p^\star} \right) \mu(q)\, dq = p\gamma \int_{\mathbb{T}^d} \left| \mathcal{D}_p^\star(q) \right|_{\mathrm{F}}^{p-2} \mathcal{D}_p^\star(q) : \delta\mathcal{D}(q)\, \mathrm{e}^{-\beta p V(q)}\, dq,$$

so that $\mathcal{D}_p^\star = \alpha_p \left| \mathcal{D}_p^\star \right|_{\mathrm{F}}^{2-p} \mathrm{e}^{\beta(p-1)V} \nabla u_{\mathcal{D}_p^\star} \otimes \nabla u_{\mathcal{D}_p^\star}$, contradicting $\mathcal{D}_p^\star(q) \geqslant c\mathrm{Id}_d$

# Characterization for degenerate eigenvalue $\Lambda(\mathcal{D}_p^\star)$

**Difficulty:** Cannot directly rely on Euler–Lagrange equation

**Strategy:** $\max_{\mathcal{D}\in\mathfrak{D}_p^{a,b}} f_\alpha$ = regularize using softmax and pass to the limit[1]

$$f_\alpha(\mathcal{D}) = \frac{\mathrm{Tr}_{L^2(\mu)}\left(\mathcal{L}_\mathcal{D}\mathrm{e}^{\alpha\mathcal{L}_\mathcal{D}}\right)}{\mathrm{Tr}_{L^2(\mu)}\left(\mathrm{e}^{\alpha\mathcal{L}_\mathcal{D}}\right)-1} = \frac{N_2\lambda_2 + \sum_{i\geqslant 3} N_i\lambda_i \mathrm{e}^{\alpha(\lambda_i-\lambda_2)}}{N_2 + \sum_{i\geqslant 3} N_i \mathrm{e}^{\alpha(\lambda_i-\lambda_2)}} \xrightarrow[\alpha\to+\infty]{} \lambda_2$$

Can write Euler–Lagrange condition for $f_\alpha$ using spectral calculus

$$\mathcal{D}_{p,\alpha}^\star = \gamma_{p,\alpha}\left|\mathcal{D}_{p,\alpha}^\star\right|_{\mathrm{F}}^{2-p}\mathrm{e}^{\beta(p-1)V}\sum_{k\geqslant 2}\left[\frac{G_\alpha(1+\alpha\lambda_{k,\alpha})-\alpha H_\alpha}{G_\alpha^2}\mathrm{e}^{\alpha\lambda_{k,\alpha}}\right]\nabla e_{k,\alpha}\otimes\nabla e_{k,\alpha}$$

with $G_\alpha = \sum_{j\geqslant 2} N_j \mathrm{e}^{\alpha\lambda_{j,\alpha}}$, $H_\alpha = \sum_{j\geqslant 2} N_j\lambda_j\mathrm{e}^{\alpha\lambda_{j,\alpha}}$; limit depends on $\lim_{\alpha\to+\infty}\alpha(\lambda_{j,\alpha}-\lambda_{2,\alpha})$

Typical example: $d=1$, degeneracy of order 2 of first non zero eigenvalue

$$\mathcal{D}_{p,\infty}^\star(q) = \tilde{\gamma}_{p,\infty}\mathrm{e}^{\beta V(q)}\left(\left|e_{2,\infty}'(q)\right|^2 + \frac{\mathrm{e}^\eta(1+\mathrm{e}^\eta+\eta)}{1+\mathrm{e}^\eta-\eta\mathrm{e}^\eta}\left|e_{3,\infty}'(q)\right|^2\right)^{1/(p-1)}$$

---

[1] Thank you Danny Perez for suggesting this!!

# Approximation by homogenization theory

**Homogenized limit:** for fixed $\mathcal{D}$,

- decrease the period: $\mathcal{D}_{\#,k}(q) = \mathcal{D}(kq)$ and $V_{\#,k}(q) = V(kq)$
- associated spectral gap

$$\Lambda_{\#,k}(\mathcal{D}) = \min_{u \in H^1(\mathbb{T}^d)\setminus\{0\}} \left\{ \frac{\displaystyle\int_{\mathbb{T}^d} \nabla u^\top \mathcal{D}_{\#,k} \nabla u \, \mathrm{e}^{-\beta V_{\#,k}}}{\displaystyle\int_{\mathbb{T}^d} u^2 \, \mathrm{e}^{-\beta V_{\#,k}}} \;\middle|\; \int_{\mathbb{T}^d} u \, \mathrm{e}^{-\beta V_{\#,k}} = 0 \right\}$$

- converges to $\Lambda_{\mathrm{hom}}(\mathcal{D})$, spectral gap of $-\mathcal{L}_{\overline{\mathcal{D}}}$ on $L^2(\mathbb{T}^d)$ with (1D case)

$$\overline{\mathcal{D}} = \int_{\mathbb{T}^d} \mathcal{D}(q) \left(1 - w_{\mathcal{D}}'(q)^2\right) \mu(q) \, dq, \qquad \left[\mathrm{e}^{-\beta V} \mathcal{D}(1 + w_{\mathcal{D}}')\right]' = 0$$

**Commutation optimization/homogenization**: maximize $\Lambda_{\mathrm{hom}}(\mathcal{D})$

$$\mathcal{D}^\star_{\mathrm{hom}}(q) = \mathrm{e}^{\beta V(q)}$$

# Numerical results

**Maximization of the spectral gap**

- $D$ piecewise constant, on uniform mesh
- finite element approximation of test functions/eigenfunctions
- Sequential Least Squares Quadratic Programming algorithm for nonlinear eigenvalue problem with constraints

$$A(D)U_D = \lambda(D)BU_D, \qquad U_D^\top BU_D = \mathrm{Id}$$

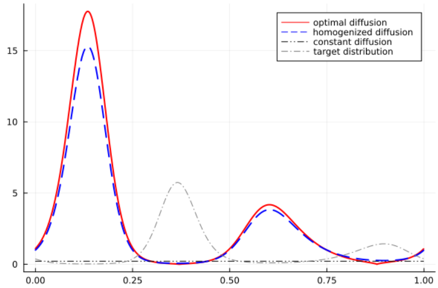
**Discretization of the SDE**

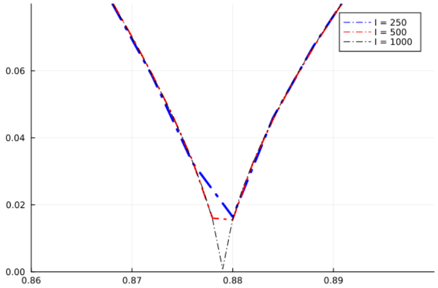
- use Metropolis acceptance/rejection to ensure unbiased sampling
- rejection probability $\mathrm{O}(\sqrt{\Delta t})$ for proposals based on naive Euler–Maruyama discretization
- lowered to $\mathrm{O}(\Delta t^{3/2})$ with dedicated (implicit) HMC algorithms[2]

[2] Noble/De Bortoli/Durmus (2022), Lelièvre/Santet/Stoltz (2023)

# Optimal diffusion for non-degenerate dominant eigenvalue

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

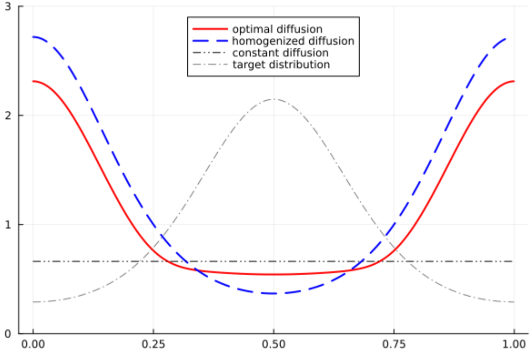



Spectral gaps: 0.81 (constant), 10.6 (homogenized), 11.2 (optimal)

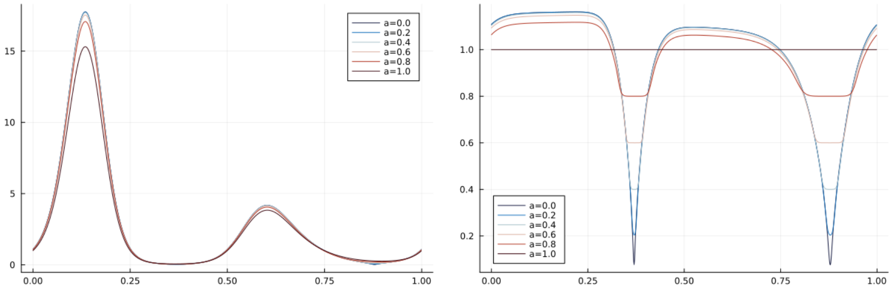
# Optimal diffusion for degenerate dominant eigenvalue

Potential $V(q) = \cos(2\pi q)$




Spectral gaps: 30.47 (constant), 32.43 (homogenized), 36.75 (optimal)

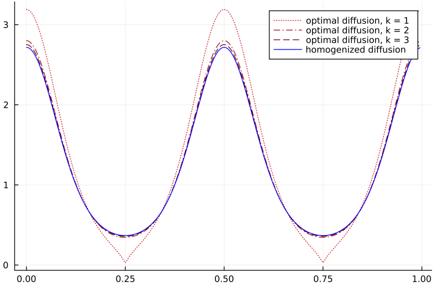
# Influence of the lower bound

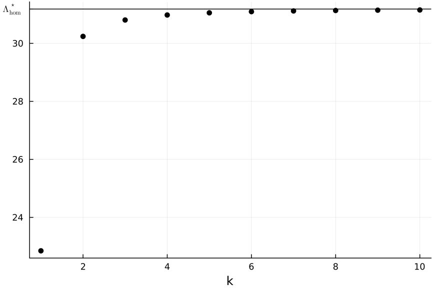



Spectral gap for various lower bounds $a$

| Lower bound $a$ | 0.0 | 0.2 | 0.4 | 0.6 | 0.8 | 1.0 |
|---|---|---|---|---|---|---|
| Spectral gap | 11.227 | 11.226 | 11.208 | 11.145 | 10.983 | 10.572 |

# Approximation by homogenized limit

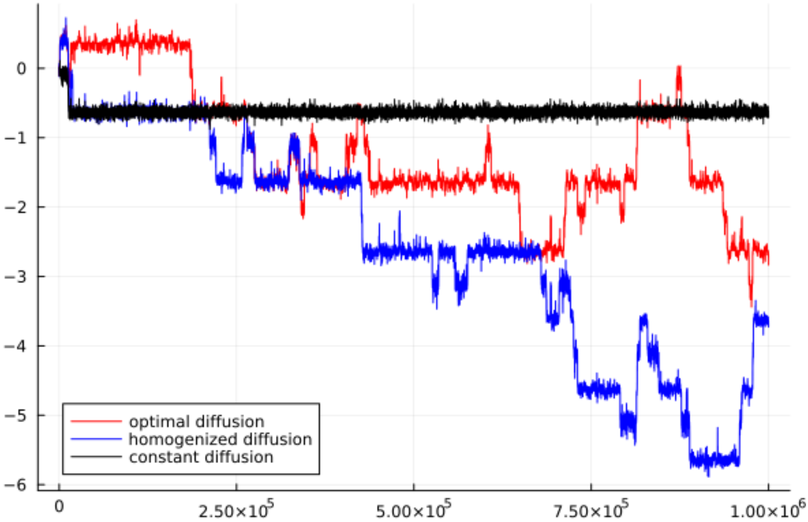



Positive diffusion when periodizing

Fast convergence to the homogenized limit

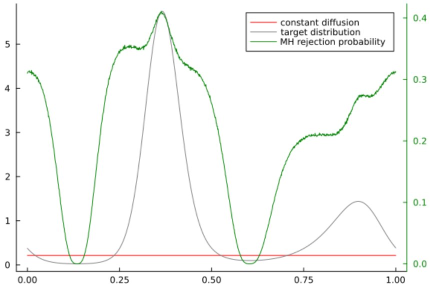
# Simulation of overdamped Langevin dynamics

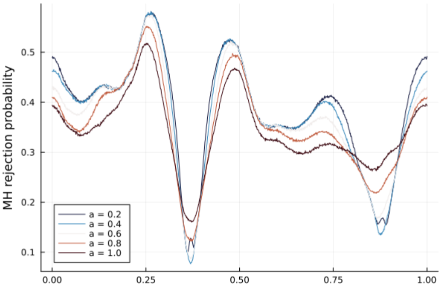



Spectral gaps: 0.81 (constant), 10.6 (homogenized), 11.2 (optimal)

# Metropolis rejection probabilities

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




Rejection probabilities for constant diffusion mostly where $V$ maximal

Rejection probabilities for optimized diffusion mostly where $V$ minimal

# Conclusion and perspectives

# Extensions

**Normalization:** numerical criterion (e.g. Metropolis rejection probability)

**Scaling with dimension:**

- diffusion depending only on some metastable degrees of freedom, e.g.

$$D(q) = D_0 \mathrm{e}^{\beta F(\xi(q))}$$

  or ad-hoc choices (cf. hackathon)

- beyond isotropic diffusions: genuine matrix shape?

**Underdamped Langevin dynamics:**

- no variational framework
- optimization of constant diffusion[3]

[3] Chak/Kantas/Pavliotis/Lelièvre (2021)