Optimizing the diffusion of overdamped Langevin dynamics

Régis SANTET

(CERMICS, École des Ponts & MATHERIALS Team, Inria Paris)

Joint work with: T. Lelièvre, G. Pavliotis, G. Robin, G. Stoltz

4th Italian Meeting on Probability and Mathematics Statistics, June 2024

I. Overdamped Langevin dynamics

II. Optimal problem

III. Numerical results

IV. Going beyond

Target distribution

$$\boxed{\mathrm{d}\pi = Z_\pi^{-1} \mathrm{e}^{-V(q)} \mathrm{d}q}$$

Target distribution

$$\boxed{\mathrm{d}\pi = Z_\pi^{-1} \mathrm{e}^{-V(q)} \mathrm{d}q}$$

Use **overdamped Langevin dynamics**: $q_t \in \mathbb{T}^d$

$$\boxed{\mathrm{d}q_t = -\nabla V(q_t)\, \mathrm{d}t + \sqrt{2}\, \mathrm{d}W_t} \tag{1}$$

- This is simply a stochastic perturbation of gradient dynamics

Target distribution

$$\boxed{\mathrm{d}\pi = Z_\pi^{-1}\mathrm{e}^{-V(q)}\mathrm{d}q}$$

Use **overdamped Langevin dynamics**: $q_t \in \mathbb{T}^d$

$$\boxed{\mathrm{d}q_t = -\nabla V(q_t)\,\mathrm{d}t + \sqrt{2}\,\mathrm{d}W_t} \tag{1}$$

- This is simply a stochastic perturbation of gradient dynamics

Ergodic average

$$\mathbb{E}_\pi[f] \approx \frac{1}{N}\sum_{i=1}^{N} f(q^i)$$

with (q^i) samples from a trajectory solving (1)

Diffusion matrix $\mathcal{D}(q) \in \mathcal{S}_d^+(\mathbb{R})$[1]

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

$\rightarrow$ **Generator $\mathcal{L}_{\mathcal{D}}$**

[1] Jardat/Bernard/Turq/Kneller (1999), Bou-Rabee/Donev/Vanden-Eijnden (2014)

Diffusion matrix $\mathcal{D}(q) \in \mathcal{S}_d^+(\mathbb{R})$[1]

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

$\rightarrow$ **Generator** $\mathcal{L}_{\mathcal{D}}$

Observations

- $\mathcal{D} = \mathrm{I}_d$: sensible generalization
- $\mathcal{L}_{\mathcal{D}} = -\nabla^*\mathcal{D}\nabla = -\sum_{i,j=1}^{d} \partial_{q_j}^* \mathcal{D}_{i,j}\partial_{q_i}$ self-adjoint
- $\mathcal{L}_{\mathcal{D}}\mathbf{1} = 0$: π is an invariant probability measure

[1] Jardat/Bernard/Turq/Kneller (1999), Bou-Rabee/Donev/Vanden-Eijnden (2014)

Diffusion matrix $\mathcal{D}(q) \in \mathcal{S}_d^+(\mathbb{R})$[1]

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

$\rightarrow$ **Generator** $\mathcal{L}_{\mathcal{D}}$

Observations

- $\mathcal{D} = \mathrm{I}_d$: sensible generalization
- $\mathcal{L}_{\mathcal{D}} = -\nabla^*\mathcal{D}\nabla = -\sum_{i,j=1}^{d} \partial_{q_j}^*\mathcal{D}_{i,j}\partial_{q_i}$ self-adjoint
- $\mathcal{L}_{\mathcal{D}}\mathbf{1} = 0$: π is an invariant probability measure

$\mathcal{D} \equiv$ inverse of position-dependent mass tensor/metric (RMHMC)[3]

[1] Jardat/Bernard/Turq/Kneller (1999), Bou-Rabee/Donev/Vanden-Eijnden (2014)

[2] Girolami/Calderhead (2011), Noble/De Bortoli/Durmus (2022), Lelièvre/RS/Stoltz (2023)

Diffusion matrix $\mathcal{D}(q) \in \mathcal{S}_d^+(\mathbb{R})$[1]

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

$\rightarrow$ **Generator** $\mathcal{L}_{\mathcal{D}}$

Observations

- $\mathcal{D} = \mathrm{I}_d$: sensible generalization
- $\mathcal{L}_{\mathcal{D}} = -\nabla^*\mathcal{D}\nabla = -\sum_{i,j=1}^{d} \partial_{q_j}^* \mathcal{D}_{i,j}\partial_{q_i}$ self-adjoint
- $\mathcal{L}_{\mathcal{D}}\mathbf{1} = 0$: π is an invariant probability measure

$\mathcal{D} \equiv$ inverse of position-dependent mass tensor/metric (RMHMC)[3]

Motivations

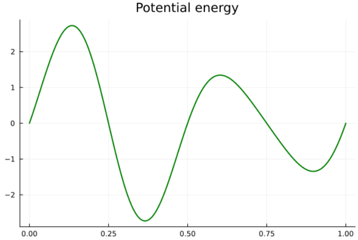
- Explore efficiently multimodal targets
- Compensate for anisotropic potential energy landscapes

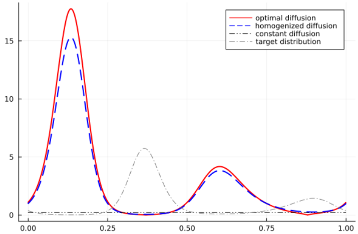
[1] Jardat/Bernard/Turq/Kneller (1999), Bou-Rabee/Donev/Vanden-Eijnden (2014)

[2] Girolami/Calderhead (2011), Noble/De Bortoli/Durmus (2022), Lelièvre/RS/Stoltz (2023)

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

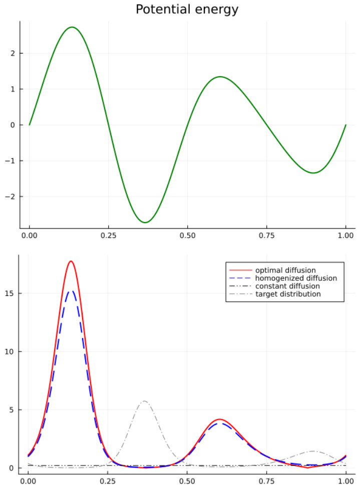
$\mathcal{D}_{\text{opt}}, \mathcal{D}_{\text{exp}} = \mathrm{e}^{V}, \mathcal{D}_{\text{cst}} = a \in \mathbb{R}$ (all three normalized in $L^2(\pi)$)

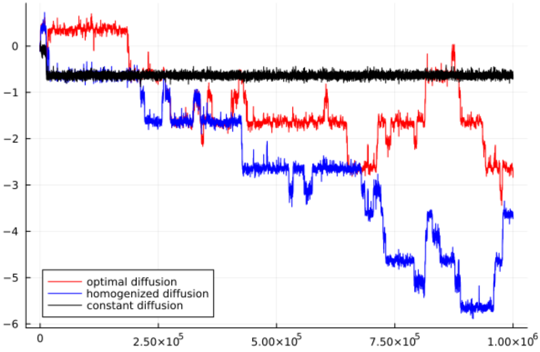




- Example with $V(q) = \sin(4\pi q)(2 + \sin(2\pi q))$

$\mathcal{D}_{\mathrm{opt}}, \mathcal{D}_{\mathrm{exp}} = \mathrm{e}^{V}, \mathcal{D}_{\mathrm{cst}} = a \in \mathbb{R}$ (all three normalized in $L^2(\pi)$)





RWMH example trajectories (same noise)

- 'Optimal' $\mathcal{D}$ helps to **cross energy barriers** (if $V \uparrow$, then $\mathcal{D} \uparrow$)

Measuring convergence

- asymptotic variance in CLT
- convergence of the law at time t towards the target distribution
- average exit time of a mode

This work $\rightarrow$ second option in a $L^2(\pi)$ framework

Measuring convergence

- asymptotic variance in CLT
- convergence of the law at time t towards the target distribution
- average exit time of a mode

This work $\rightarrow$ second option in a $L^2(\pi)$ framework

Main tool

Poincaré inequality satisfied by π

$$\forall \pi_0 \in L^2(\pi^{-1}), \qquad \int_{\mathbb{T}^d} \left(\frac{\pi_0}{\pi} - 1\right)^2 \mathrm{d}\pi \leqslant \frac{1}{\Lambda(\mathcal{D})} \int_{\mathbb{T}^d} \left|\nabla \left(\frac{\pi_0}{\pi}\right)\right|^2 \mathrm{d}\pi$$

$\Lambda(\mathcal{D})$: smallest nonzero eigenvalue of $-\mathcal{L}_{\mathcal{D}}$: the **spectral gap**

Claim: Exponential convergence towards the target measure

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

Claim: Exponential convergence towards the target measure

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

$$\Lambda(\mathcal{D}) = \inf_{\substack{u \in H^1(\pi) \setminus \{0\} \\ \int u \, \mathrm{d}\pi = 0}} \frac{\displaystyle\int_{\mathbb{T}^d} \nabla u^{\mathsf{T}} \mathcal{D} \nabla u \, \mathrm{d}\pi}{\displaystyle\int_{\mathbb{T}^d} u^2 \, \mathrm{d}\pi}$$

This work → **Maximize the spectral gap** $\Lambda(\mathcal{D})$ w.r.t. $\mathcal{D}$

Choices in the literature

- $\mathcal{D} = (\nabla^2 V)^{-1}$ for strictly convex potentials[3]

  (also Brascamb–Lieb inequality)
- $\mathcal{D} = \mathrm{e}^V$ 'Langevin tempered algorithms' [4]

[3]Girolami/Calderhead (2011)
[4]Roberts/Stramer (2002)

Choices in the literature

- $\mathcal{D} = (\nabla^2 V)^{-1}$ for strictly convex potentials[3]
  (also Brascamb–Lieb inequality)
- $\mathcal{D} = \mathrm{e}^V$ 'Langevin tempered algorithms' [4]

Previous works available

- Optimizing the diffusion in the case of the uniform distribution[5]
- Optimizing the diffusion for reversible MC on discrete spaces[6]

[3] Girolami/Calderhead (2011)
[4] Roberts/Stramer (2002)
[5] Henrot (2006)
[6] Boyd/Diaconis/Parrilo/Xiao (2009)

Choices in the literature

- $\mathcal{D} = (\nabla^2 V)^{-1}$ for strictly convex potentials[3]
  (also Brascamb–Lieb inequality)
- $\mathcal{D} = \mathrm{e}^V$ 'Langevin tempered algorithms' [4]

Previous works available

- Optimizing the diffusion in the case of the uniform distribution[5]
- Optimizing the diffusion for reversible MC on discrete spaces[6]

Very recent work

- Cui/Tong/Zahm, *Optimal Riemannian metric for Poincaré inequalities and how to ideally precondition Langevin dynamics* (2024)

---

[3] Girolami/Calderhead (2011)
[4] Roberts/Stramer (2002)
[5] Henrot (2006)
[6] Boyd/Diaconis/Parrilo/Xiao (2009)

Optimal problem

- $\Lambda(a\mathcal{D}) = a\Lambda(\mathcal{D}) \xrightarrow[a\to+\infty]{} +\infty$: large $\mathcal{D}$ requires smaller time steps
- L^∞ bounds trivial (simply saturates the constraint)

Optimal problem

- $\Lambda(a\mathcal{D}) = a\Lambda(\mathcal{D}) \xrightarrow[a\to+\infty]{} +\infty$: large $\mathcal{D}$ requires smaller time steps
- L^∞ bounds trivial (simply saturates the constraint)

Our choice: L^p constraint on $\mathcal{D}$: $1 \leqslant p \leqslant +\infty$, $a, b \geqslant 0$

$$\mathrm{e}^{-V(q)}\mathcal{D}(q) \in \mathcal{M}_{a,b} = \left\{ M \in \mathcal{S}_d^+ \,\middle|\, \forall \xi \in \mathbb{R}^d, a\,|\xi|^2 \leqslant \xi^\mathsf{T} M \xi \leqslant b^{-1}\,|\xi|^2 \right\} \text{ a.e.}$$

endowed with

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

Optimal problem

- $\Lambda(a\mathcal{D}) = a\Lambda(\mathcal{D}) \xrightarrow[a\to+\infty]{} +\infty$: large $\mathcal{D}$ requires smaller time steps
- L^∞ bounds trivial (simply saturates the constraint)

Our choice: L^p constraint on $\mathcal{D}$: $1 \leqslant p \leqslant +\infty$, $a, b \geqslant 0$

$$\mathrm{e}^{-V(q)}\mathcal{D}(q) \in \mathcal{M}_{a,b} = \left\{ M \in \mathcal{S}_d^+ \,\middle|\, \forall \xi \in \mathbb{R}^d, a\,|\xi|^2 \leqslant \xi^\mathsf{T} M \xi \leqslant b^{-1}\,|\xi|^2 \right\} \text{ a.e.}$$

endowed with

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

Maximization is performed on

$$\boxed{\mathfrak{D}_p^{a,b} = \left\{ \mathcal{D} \in L^\infty_\pi(\mathbb{T}^d, \mathcal{M}_{a,b}) \,\middle|\, \|\mathcal{D}\|_{L^p_\pi} \leqslant 1 \right\}}$$

Optimal problem

- $\Lambda(a\mathcal{D}) = a\Lambda(\mathcal{D}) \xrightarrow[a\to+\infty]{} +\infty$: large $\mathcal{D}$ requires smaller time steps
- L^∞ bounds trivial (simply saturates the constraint)

Our choice: L^p constraint on $\mathcal{D}$: $1 \leqslant p \leqslant +\infty$, $a, b \geqslant 0$

$$\mathrm{e}^{-V(q)}\mathcal{D}(q) \in \mathcal{M}_{a,b} = \left\{ M \in \mathcal{S}_d^+ \,\middle|\, \forall \xi \in \mathbb{R}^d, a\,|\xi|^2 \leqslant \xi^\mathsf{T} M \xi \leqslant b^{-1}\,|\xi|^2 \right\} \text{ a.e.}$$

endowed with

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

Maximization is performed on

$$\boxed{\mathfrak{D}_p^{a,b} = \left\{ \mathcal{D} \in L^\infty_\pi(\mathbb{T}^d, \mathcal{M}_{a,b}) \,\middle|\, \|\mathcal{D}\|_{L^p_\pi} \leqslant 1 \right\}}$$

- In Cui/Tong/Zahm, they choose (relation to Stein kernels)

$$\mathbb{E}_\pi[\mathrm{Tr}\, D] = \mathrm{Tr}[\mathrm{Cov}_\pi]$$

Existence of a maximizer for $p \in [1, +\infty)$, $a \in [0, |\mathrm{I}_d|_{\mathrm{F}}^{-1}]$, $b > 0$ such that $ab \leqslant 1$,

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

Existence of a maximizer for $p \in [1, +\infty)$, $a \in [0, |\mathrm{I}_d|_{\mathrm{F}}^{-1}]$, $b > 0$ such that $ab \leqslant 1$,

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

Moreover:

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

Existence of a maximizer for $p \in [1, +\infty)$, $a \in [0, |\mathrm{I}_d|_{\mathrm{F}}^{-1}]$, $b > 0$ such that $ab \leqslant 1$,

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

Moreover:

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

Claim: If $\mathcal{D}^\star$ is uniformly definite positive, then $\Lambda(\mathcal{D}^\star)$ is degenerate

Existence of a maximizer for $p \in [1, +\infty)$, $a \in [0, |\mathrm{I}_d|_{\mathrm{F}}^{-1}]$, $b > 0$ such that $ab \leqslant 1$,

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

Moreover:

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

Claim: If $\mathcal{D}^\star$ is uniformly definite positive, then $\Lambda(\mathcal{D}^\star)$ is degenerate

Difficulty: the map $\mathcal{D} \mapsto \Lambda(\mathcal{D})$ may therefore not be differentiable at $\mathcal{D}^\star$

Existence of a maximizer for $p \in [1, +\infty)$, $a \in [0, |\mathrm{I}_d|_{\mathrm{F}}^{-1}]$, $b > 0$ such that $ab \leqslant 1$,

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

Moreover:

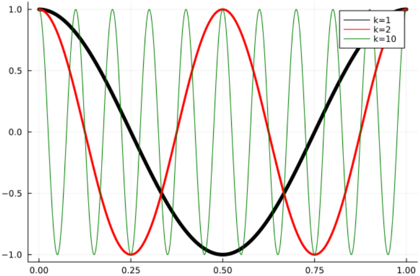
- $\displaystyle\int_{\mathbb{T}^d} |\mathcal{D}^\star(q)|_{\mathrm{F}}^p \, \mathrm{e}^{-pV(q)} \, \mathrm{d}q = 1$
- For any open set $\Omega \subset \mathbb{T}^d$, there exists $q \in \Omega$ such that $\mathcal{D}^\star(q) \neq 0$

Claim: If $\mathcal{D}^\star$ is uniformly definite positive, then $\Lambda(\mathcal{D}^\star)$ is degenerate

Difficulty: the map $\mathcal{D} \mapsto \Lambda(\mathcal{D})$ may therefore not be differentiable at $\mathcal{D}^\star$

We obtained a formal characterization using a smooth-maximum approach

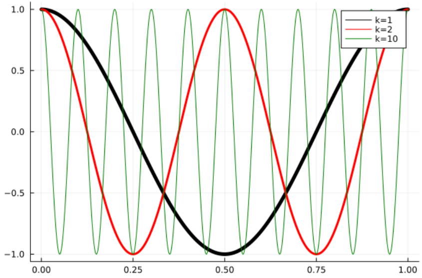
Homogenization: Obtaining a good approximation of $\mathcal{D}^\star$

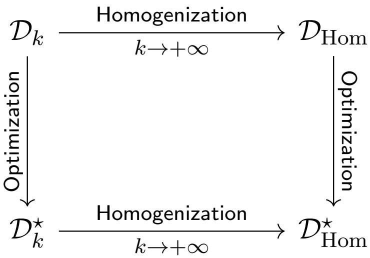


Homogenization: Obtaining a good approximation of $\mathcal{D}^\star$

• Idea 1: study the asymptotic behaviour of the optimal diffusion in the homogenized limit

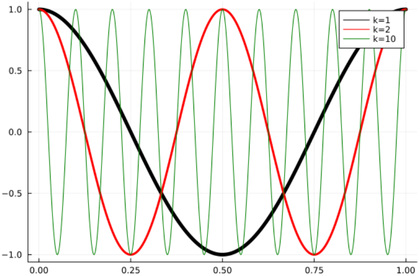
• Idea 2: optimize the periodic homogenization limit

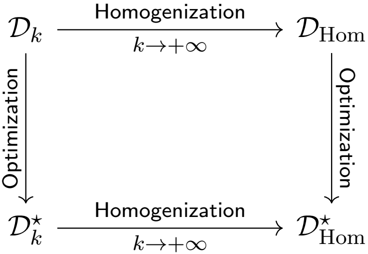




Homogenization: Obtaining a good approximation of $\mathcal{D}^\star$

- Idea 1: study the asymptotic behaviour of the optimal diffusion in the homogenized limit
- Idea 2: optimize the periodic homogenization limit





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

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

Maximization of the spectral gap

- D isotropic, piecewise constant, on uniform mesh
- Finite Element approximation of test functions/eigenfunctions

Maximization of the spectral gap

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

$$A(D)U_D = \lambda(D) B U_D, \qquad U_D^\mathsf{T} B U_D = \mathrm{I}_d$$

with

$$[A(D)]_{i,j} = \int_{\mathbb{T}^d} \nabla\varphi_j^\mathsf{T} D \nabla\varphi_i \, \mathrm{d}\pi, \qquad [B]_{i,j} = \int_{\mathbb{T}^d} \varphi_j \varphi_i \, \mathrm{d}\pi$$

$\rightarrow$ Generalized eigenvalue problem

Maximization of the spectral gap

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

$$A(D)U_D = \lambda(D)BU_D, \qquad U_D^{\mathsf{T}}BU_D = \mathrm{I}_d$$

with

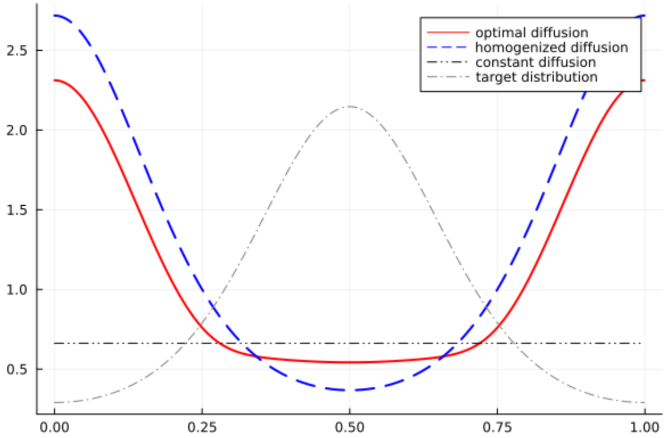
$$[A(D)]_{i,j} = \int_{\mathbb{T}^d} \nabla\varphi_j^{\mathsf{T}} D\nabla\varphi_i \, \mathrm{d}\pi, \qquad [B]_{i,j} = \int_{\mathbb{T}^d} \varphi_j\varphi_i \, \mathrm{d}\pi$$

$\to$ Generalized eigenvalue problem

Only practical if $d \leqslant 3$!

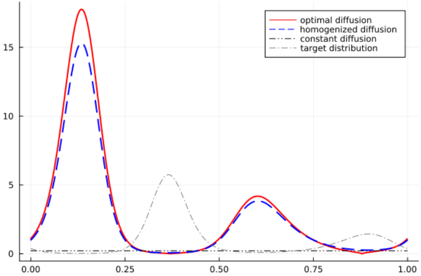
- Use approximation solution $\mathcal{D}^\star_{\mathrm{Hom}}$
- Use reaction coordinate/summary statistics $\xi : \mathbb{R}^d \to \mathbb{R}$

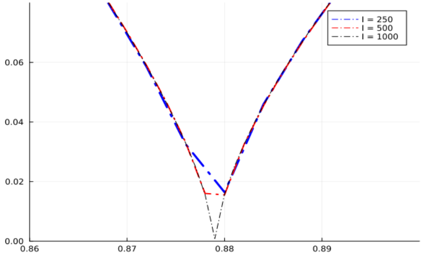
$$V(q) = \cos(2\pi q)$$



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

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





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

Discretization of the SDE

- **Random Walk** with 'guided variance', using Euler–Maruyama scheme

$$q^{i+1} = q^i + \sqrt{2\Delta t}\,\mathcal{D}^{1/2}(q^i)\,\mathrm{G}^{i+1}$$

- use Metropolis acceptance/rejection to ensure unbiased sampling

Discretization of the SDE

- **Random Walk** with 'guided variance', using Euler–Maruyama scheme

$$q^{i+1} = q^i + \sqrt{2\Delta t}\, \mathcal{D}^{1/2}(q^i)\, \mathrm{G}^{i+1}$$

- use Metropolis acceptance/rejection to ensure unbiased sampling

- rejection probability $\mathrm{O}(\Delta t^{1/2})$ for proposals based on naive Euler–Maruyama discretization of the continuous SDE[6]

[6] Fathi/Stoltz (2017)

Discretization of the SDE

- **Random Walk** with 'guided variance', using Euler–Maruyama scheme

$$q^{i+1} = q^i + \sqrt{2\Delta t}\,\mathcal{D}^{1/2}(q^i)\,\mathrm{G}^{i+1}$$

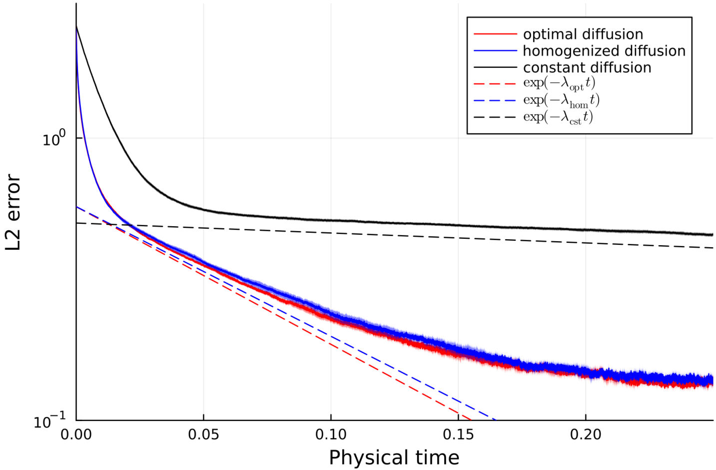
- use Metropolis acceptance/rejection to ensure unbiased sampling

- rejection probability $\mathrm{O}(\Delta t^{1/2})$ for proposals based on naive Euler–Maruyama discretization of the continuous SDE[6]
- lowered to $\mathrm{O}(\Delta t^{3/2})$ with dedicated (implicit) HMC algorithms[7] RMHMC algorithm:

$$H(q,p) = V(q) - \frac{1}{2}\ln\det\mathcal{D}(q) + \frac{1}{2}p^{\mathsf{T}}\mathcal{D}(q)p$$

[6] Fathi/Stoltz (2017)

[7] Noble/De Bortoli/Durmus (2022), Lelièvre/RS/Stoltz (2023)



L^2 error between empirical and target distributions

Settings: $\Delta t = 10^{-6}$, 10^4 samples generated uniformly on $\mathbb{T}$, 100 bins to approximate the L^2 norm, results averaged over 10 simulations

In high dimension ?

Reaction coordinate/summary statistics $\xi : \mathbb{T}^d \to \mathbb{R}$ 'slow variable'
Level set $\Sigma_z = \xi^{-1}(\{z\})$

In high dimension ?

Reaction coordinate/summary statistics $\xi : \mathbb{T}^d \to \mathbb{R}$ 'slow variable'
Level set $\Sigma_z = \xi^{-1}(\{z\})$

- Projection onto the tangent space $T_q\Sigma(z)$ to Σ_z at q

$$P^{\parallel} = \mathrm{I}_d - \frac{\nabla\xi \otimes \nabla\xi}{\|\nabla\xi\|^2}$$

In high dimension ?

Reaction coordinate/summary statistics $\xi : \mathbb{T}^d \to \mathbb{R}$ 'slow variable'
Level set $\Sigma_z = \xi^{-1}(\{z\})$

- Projection onto the tangent space $T_q\Sigma(z)$ to Σ_z at q

$$P^{\parallel} = \mathrm{I}_d - \frac{\nabla\xi \otimes \nabla\xi}{\|\nabla\xi\|^2}$$

$\to$ Change values perpendicular to the gradient of ξ:

$$\widetilde{\mathcal{D}}(q) = f(q)P^{\perp}(q) + P^{\parallel}(q), \qquad \mathcal{D} = \gamma^{-1}\widetilde{\mathcal{D}}$$

In high dimension ?

Reaction coordinate/summary statistics $\xi : \mathbb{T}^d \to \mathbb{R}$ 'slow variable'
Level set $\Sigma_z = \xi^{-1}(\{z\})$

- Projection onto the tangent space $T_q\Sigma(z)$ to Σ_z at q

$$P^{\parallel} = \mathrm{I}_d - \frac{\nabla\xi \otimes \nabla\xi}{\|\nabla\xi\|^2}$$

$\to$ Change values perpendicular to the gradient of ξ:

$$\widetilde{\mathcal{D}}(q) = f(q) P^{\perp}(q) + P^{\parallel}(q), \qquad \mathcal{D} = \gamma^{-1}\widetilde{\mathcal{D}}$$

Normalization constraint

Choose f as a function of $\xi(q) \equiv z$ and use the free energy F[8]

$$\int_{\mathbb{T}^d} |D(q)|_{\mathrm{F}}\, \mathrm{e}^{-V(q)}\, \mathrm{d}q = \int_{\Sigma} |D(z)|_{\mathrm{F}}\, \mathrm{e}^{-F(z)}\, \mathrm{d}z = 1$$

[8] Lelièvre/Rousset/Stoltz (2010)

In high dimension ?

Reaction coordinate/summary statistics $\xi : \mathbb{T}^d \to \mathbb{R}$ 'slow variable'
Level set $\Sigma_z = \xi^{-1}(\{z\})$

- Projection onto the tangent space $T_q\Sigma(z)$ to Σ_z at q

$$P^{\parallel} = \mathrm{I}_d - \frac{\nabla\xi \otimes \nabla\xi}{\|\nabla\xi\|^2}$$

$\to$ Change values perpendicular to the gradient of ξ:

$$\widetilde{\mathcal{D}}(q) = f(q) P^{\perp}(q) + P^{\parallel}(q), \qquad \mathcal{D} = \gamma^{-1}\widetilde{\mathcal{D}}$$

Normalization constraint

Choose f as a function of $\xi(q) \equiv z$ and use the free energy F[8]

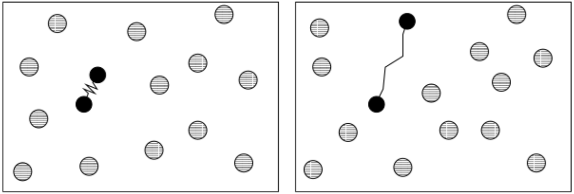
$$\int_{\mathbb{T}^d} |D(q)|_{\mathrm{F}}\, \mathrm{e}^{-V(q)}\, \mathrm{d}q = \int_{\Sigma} |D(z)|_{\mathrm{F}}\, \mathrm{e}^{-F(z)}\, \mathrm{d}z = 1$$

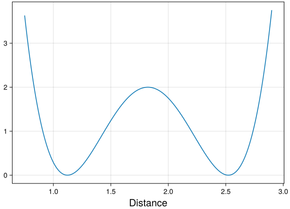
Divergence of D: need to have acess to $\nabla^2\xi(q)$, $F'(z)$

[8] Lelièvre/Rousset/Stoltz (2010)

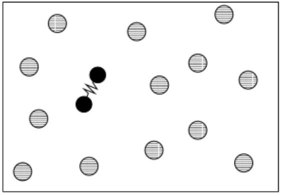
Numerical example: dimer in a solvent, $d = 32$

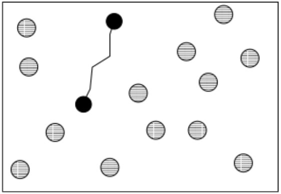
Numerical example: dimer in a solvent, $d = 32$

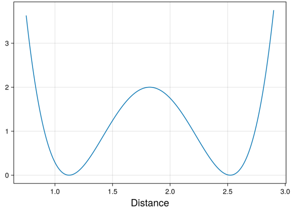




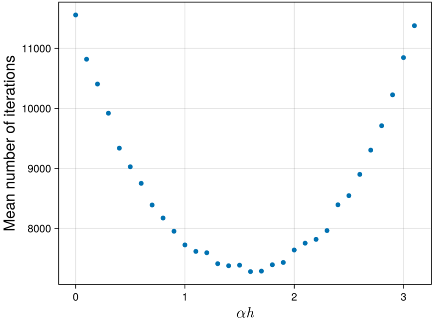
Numerical example: dimer in a solvent, $d = 32$







$$\xi(q) \equiv \xi(q_1, q_2) \propto \|q_2 - q_1\|\,, \qquad {\color{red} f_\alpha} = \mathrm{e}^{\alpha F \circ \xi}, \qquad F \sim h$$



Normalization: numerical criterion ? *e.g.* Metropolis rejection probability

Normalization: numerical criterion ? *e.g.* Metropolis rejection probability

Scaling with dimension

- Not relying on reaction coordinates ?
- diffusion matrix: isotropic diffusions to tackle anisotropic targets

Normalization: numerical criterion ? *e.g.* Metropolis rejection probability

Scaling with dimension

- Not relying on reaction coordinates ?
- diffusion matrix: isotropic diffusions to tackle anisotropic targets

Underdamped Langevin dynamics

- No variational framework
- optimization of constant diffusion (optimal friction matrix)[9]
  $\rightarrow$ Generalized HMC

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

Normalization: numerical criterion ? *e.g.* Metropolis rejection probability

Scaling with dimension

- Not relying on reaction coordinates ?
- diffusion matrix: isotropic diffusions to tackle anisotropic targets

Underdamped Langevin dynamics

- No variational framework
- optimization of constant diffusion (optimal friction matrix)[9]
  $\rightarrow$ Generalized HMC

Thank you !

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

Proof of exponential convergence

Since $\pi_t = \mathrm{e}^{t\mathcal{L}_\mathcal{D}}\pi_0$ and $\mathcal{L}_\mathcal{D} \leqslant -\Lambda(\mathcal{D})\mathrm{id}$ on $L^2(\mu)$,

$$\begin{aligned}\frac{d}{dt}\left\|\frac{\pi_t}{\pi} - 1\right\|^2_{L^2(\pi)} &= 2\left\langle \frac{\pi_t}{\pi} - 1, \mathcal{L}_\mathcal{D}\left(\frac{\pi_t}{\pi} - 1\right)\right\rangle_{L^2(\pi)} \\ &\leqslant -2\Lambda(\mathcal{D})\left\|\frac{\pi_t}{\pi} - 1\right\|^2_{L^2(\pi)}\end{aligned}$$

We then apply Gronwall's lemma.

Settings: $p \in (1, +\infty)$, $|\cdot|_{\mathrm{F}} \equiv$ Frobenius norm, $a = 0$ and $b > 0$ 'small enough', continuity assumption for $\mathcal{D}^\star$ for $d = 1$

Claim If $\mathcal{D}^\star$ is uniformly definite positive, then the eigenvalue $\Lambda(\mathcal{D}^\star)$ is degenerate

Proof If not, write the Euler–Lagrange equation (perturbation theory)

$$\int_{\mathbb{T}^d} \delta\mathcal{D} : (\nabla u_{\mathcal{D}^\star} \otimes \nabla u_{\mathcal{D}^\star})\,\mathrm{d}\pi = p\gamma \int_{\mathbb{T}^d} |\mathcal{D}^\star|_{\mathrm{F}}^{p-2}\,\mathcal{D}^\star : \delta\mathcal{D}\,\mathrm{e}^{-pV}\,\mathrm{d}q$$

where $-\mathcal{L}_{\mathcal{D}^\star} u_{\mathcal{D}^\star} = \lambda(\mathcal{D}^\star) u_{\mathcal{D}^\star}$ so that
$\mathcal{D}^\star = \alpha\,|\mathcal{D}^\star|^{2-p}\,\mathrm{e}^{(p-1)V}\nabla u_{\mathcal{D}^\star} \otimes \nabla u_{\mathcal{D}^\star}$ $\rightarrow$ contradiction !

Smooth-maximum approach: $\sup_{\mathcal{D}\in\mathfrak{D}_p^{a,b}} f_\alpha(\mathcal{D})$ and let $\alpha \to +\infty$

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

- Euler–Lagrange equation for f_α

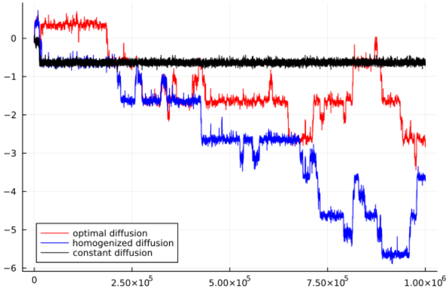
$$\mathcal{D}^{\star,\alpha} = \gamma_\alpha \left|\mathcal{D}^{\star,\alpha}\right|_{\mathrm{F}}^{2-p}\mathrm{e}^{(p-1)V}\sum_{k\geqslant 2}\left[\frac{G_\alpha(1+\alpha\lambda_{k,\alpha})-\alpha H_\alpha}{G_\alpha^2}\mathrm{e}^{\alpha\lambda_{k,\alpha}}\right]\nabla e_{k,\alpha}\otimes\nabla e_{k,\alpha}$$

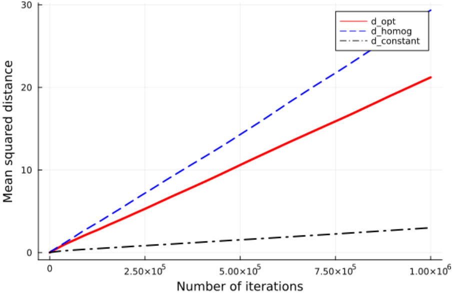
where $G_\alpha = \sum_{i\geqslant 2}\mathrm{e}^{\alpha\lambda_{i,\alpha}}$, $H_\alpha = \sum_{i\geqslant 2}\lambda_{i,\alpha}\mathrm{e}^{\alpha\lambda_{i,\alpha}}$

- The limit depends on $\lim_{\alpha\to+\infty}\alpha(\lambda_{j,\alpha}-\lambda_{2,\alpha}) =: \eta$
- Typical example: $d=1$, degeneracy of order 2 for $\lambda_{2,\infty}$

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

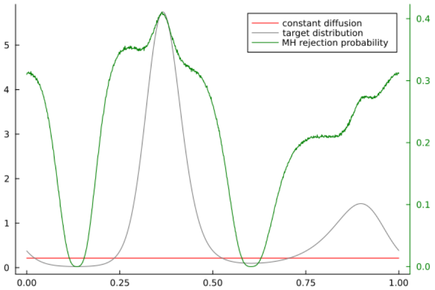
$$V(q) = \sin(4\pi q)(2 + \sin(2\pi q))$$

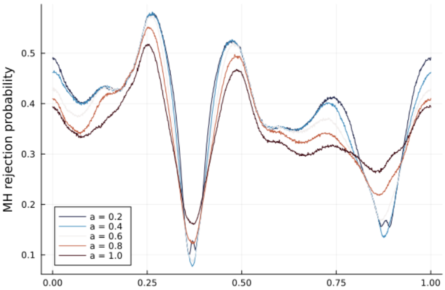




Mean square distance

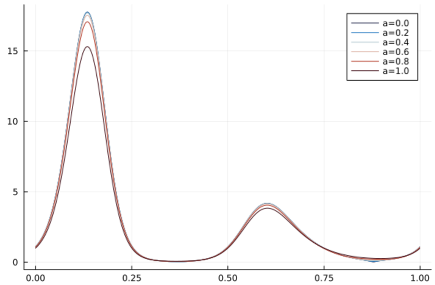
Spectral gaps: 0.81 (constant), 10.6 (homogenized), 11.2 (optimal)



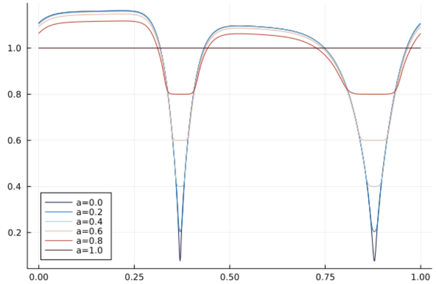


Rejection probabilities for constant diffusion mostly where V is maximal

Rejection probabilities for optimized diffusion mostly where V is minimal



Optimal diffusions

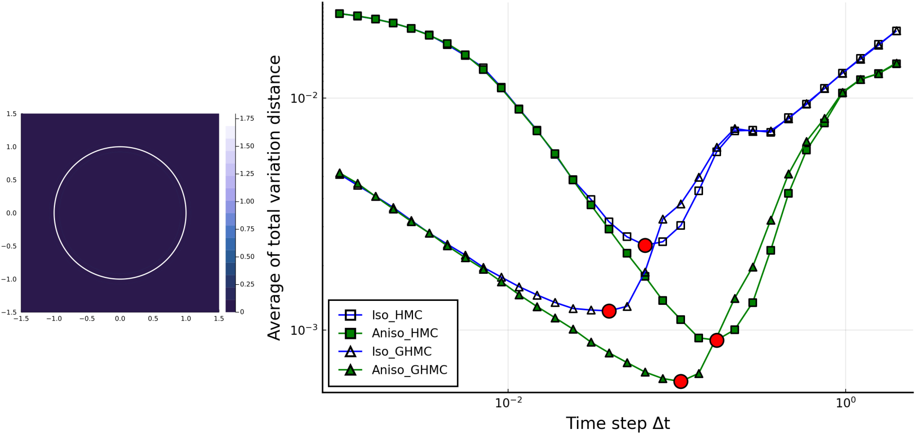


Normalized by $\mathcal{D}^{\star}_{\mathrm{Hom}}$

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

- Anisotropic diffusion coefficient $\mathcal{D}_{\mathsf{Tan}}(q) = \varepsilon \mathrm{I}_2 + \tilde{q}\tilde{q}^\mathsf{T}/\|q\|^2,\ \tilde{q} = (-y\ x)^\mathsf{T}$
- Isotropic diffusion coefficient $\mathcal{D}_{\mathsf{One}} \equiv (1+\varepsilon)\mathrm{I}_2, \quad \varepsilon = 0.1$

Computing: after fixed number of iterations, distance to the invariant measure of the angle distribution (uniform on $[0, 2\pi]$)



$\Rightarrow$ Compromise: small/large time steps (exploration vs rejection)

Definition of H-convergence

A sequence $(\mathcal{A}^k)_{k\geqslant 1} \subset L^\infty(\mathbb{T}^d, \mathcal{M}_{a,b})$ H-converges to $\overline{\mathcal{A}} \in L^\infty(\mathbb{T}^d, \mathcal{M}_{a,b})$ if, for any $f \in H^{-1}(\mathbb{T}^d)$ such that $\langle f, \mathbf{1}\rangle_{H^{-1},H^1} = 0$, the sequence $(u^k)_{k\geqslant 1} \subset H^1(\mathbb{T}^d)$ of solutions to

$$\begin{cases} -\operatorname{div}\left(\mathcal{A}^k \nabla u^k\right) = f \quad \text{on } \mathbb{T}^d, \\ \displaystyle\int_{\mathbb{T}^d} u^k(q)\mathrm{d}q = 0 \end{cases}$$

satisfies in the limit $k \to +\infty$,

$$\begin{cases} u^k \rightharpoonup u & \text{weakly in } H^1(\mathbb{T}^d), \\ \mathcal{A}^k \nabla u^k \rightharpoonup \overline{\mathcal{A}} \nabla u & \text{weakly in } L^2(\mathbb{T}^d)^d, \end{cases}$$

where $u \in H^1(\mathbb{T}^d)$ is the solution of the homogenized problem

$$\begin{cases} -\operatorname{div}\left(\overline{\mathcal{A}} \nabla u\right) = f \quad \text{on } \mathbb{T}^d, \\ \displaystyle\int_{\mathbb{T}^d} u(q)\mathrm{d}q = 0 \end{cases}$$

Correctors

If $\mathcal{A} = \mathcal{D}\exp(-V)$, $(w_i)_{1\leqslant i\leqslant d} \subset H^1(\mathbb{T}^d)$ is the family of unique solutions to the problem

$$\begin{cases} -\operatorname{div}(\mathcal{A}(e_i + \nabla w_i)) = 0, \\ \displaystyle\int_{\mathbb{T}^d} w = 0 \end{cases}$$

Then for any $\xi \in \mathbb{R}^d$,

$$\xi^{\mathsf{T}}\overline{D}\xi = \xi^{\mathsf{T}}\left(\int_{\mathbb{T}^d}\mathcal{D}(q)\mathrm{d}\pi\right)\xi - \int_{\mathbb{T}^d}\nabla w_\xi^{\mathsf{T}}\mathcal{D}\nabla w_\xi \mathrm{d}\pi.$$