Non-reversible lifts of reversible diffusions

joint work with Francis Lörler

Banff, September 2024

1.Introduction

$\mu(dx) \propto e^{-U(x)}dx$ on $S = \mathbb{R}^d$ (or Riemannian manifold)

Sampling from μ

① Reversible Markov Processes

Brownian motion

$$dX_t = -\frac{1}{2}\nabla U(X_t)\,dt + dB_t$$

Overdamped Langevin dynamics

$\rightsquigarrow$ Unadjusted Langevin algorithm (= Euler scheme), Metropolis adjusted LA, Random Walk Metropolis, ...

$$Lg = \frac{1}{2}\Delta g - \frac{1}{2}\nabla U \cdot \nabla g$$

$$\mathcal{E}(f,g) = -(f, Lg)_{L^2(\mu)} = \frac{1}{2}\int \nabla f \cdot \nabla g \, d\mu$$

Relaxation time

$$t_{rel} = 1/\mathrm{gap}(L)$$

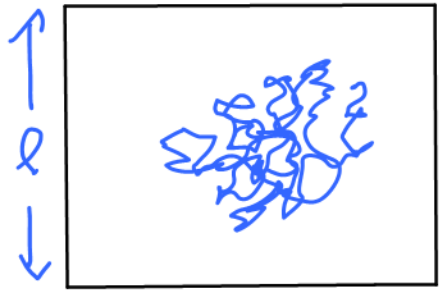
Spectral gap

$$\mathrm{gap}(L) = \inf_{f \neq 0} \frac{\mathcal{E}(f,f)}{\mathrm{Var}_\mu(f)}$$

EXAMPLE

1) $S = (\mathbb{R}/\ell\mathbb{Z})^d$, U uniformly bounded $\Rightarrow$ $t_{rel} \sim \ell^2$

Slow diffusive motion



2) $S = \mathbb{R}^d$, $\nabla^2 U \geq m I_d$, $m > 0$

$\Rightarrow$ $t_{rel} \leq \frac{2}{m}$, sharp in Gaussian case

Condition number dependence $\Omega(\kappa)$ for discretizations

② Non-reversible Markov Processes

$\hat{S} = \mathbb{R}^d \times \mathbb{R}^d = \{(x,v) : x, v \in \mathbb{R}^d\}$ or tangent bundle of Riem. manifold

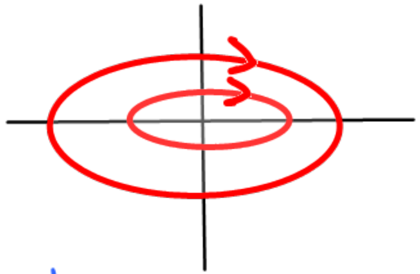
$\hat{\mu} = \mu \otimes \kappa$ $\quad \kappa = N(0, I_d)$ $\quad d\hat{\mu} \propto e^{-H(x,v)}\, dx\, dv$ $\quad H(x,v) = U(x) + \frac{1}{2}|v|^2$

Hamiltonian dynamics

$$\hat{L} = v\cdot\nabla_x - \nabla U(x)\cdot\nabla_v$$

$$dX_t = V_t\,dt$$
$$dV_t = -\nabla U(X_t)\,dt$$

$\rightsquigarrow$ MD



not ergodic $\rightarrow$ add noise in v-variable

Randomised Hamiltonian Monte Carlo

$\gamma > 0$ refreshment rate

$$\hat{L}_\gamma = \hat{L} + \gamma(\Pi_v - I)$$

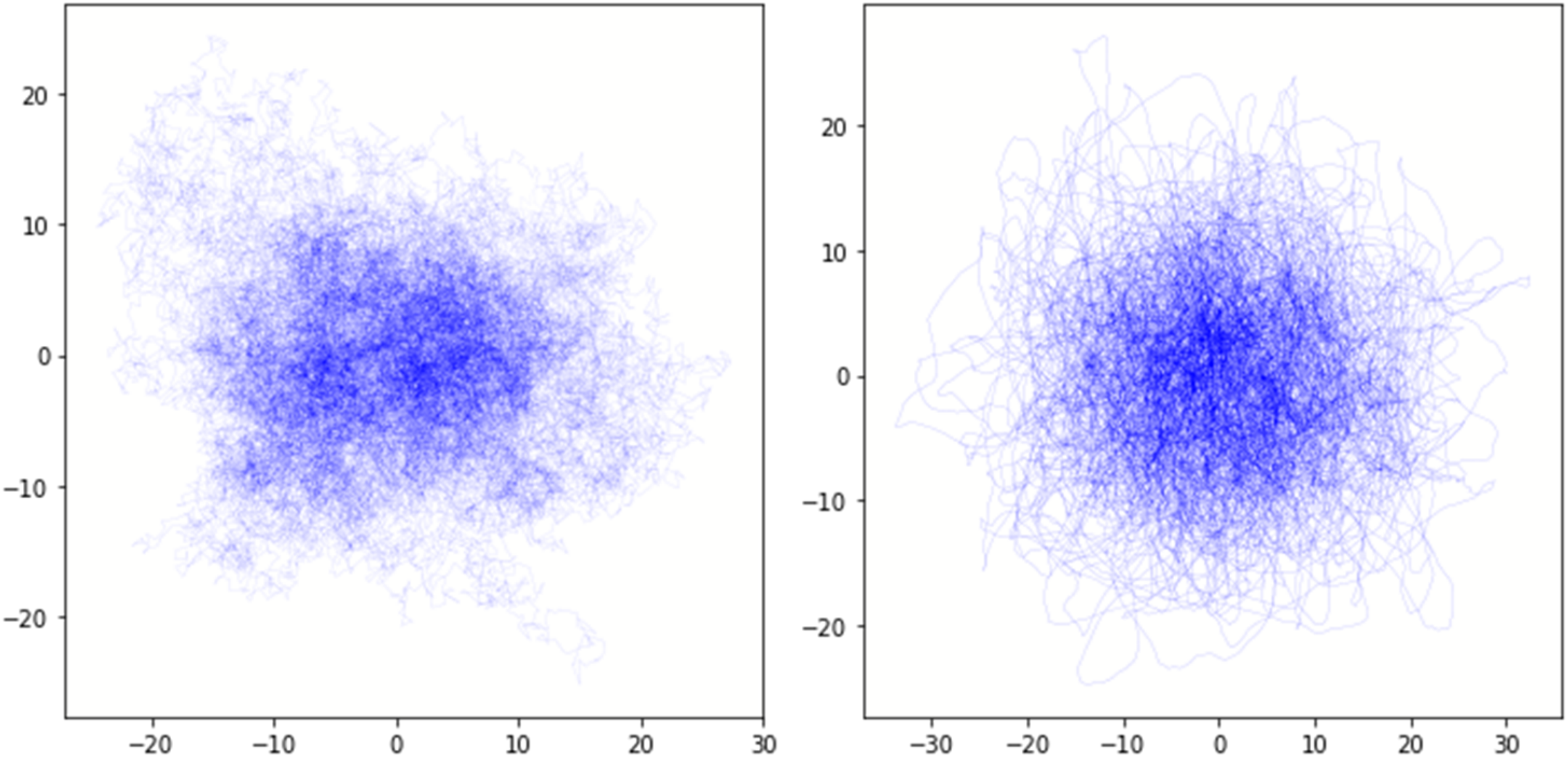
$$(\Pi_v f)(x,v) = \int f(x,w)\,\kappa(dw)$$

Langevin dynamics (LD)

$$\hat{L}_\gamma = \hat{L} + \gamma L_{ou}$$

$\hat{L} = \Delta_v - v\cdot\nabla_v$

$$dX_t = V_t\,dt$$
$$dV_t = -\nabla U(X_t)\,dt - \gamma V_t\,dt + \sqrt{2\gamma}\,dB_t$$

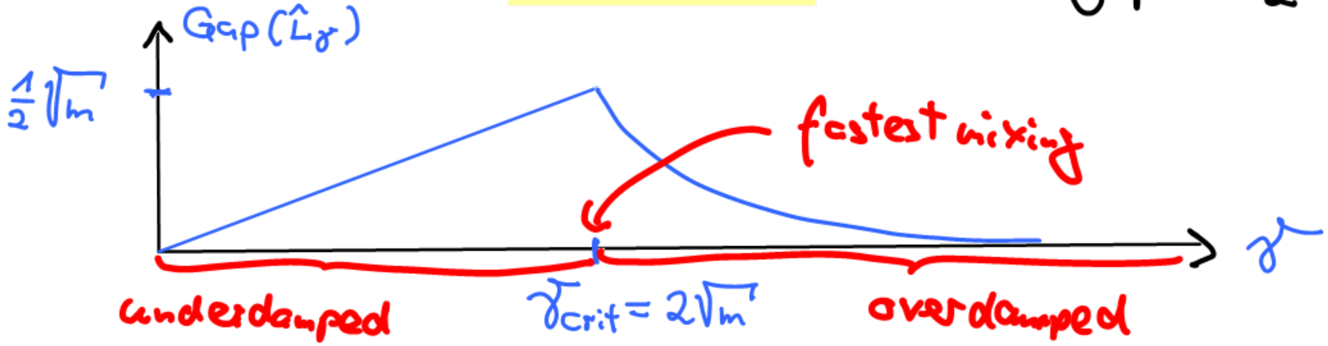


Overdamped Langevin dynamics vs. critical Langevin dynamics in a quadratic potential. Each plot shows a single trajectory up to time t = 20000.

- The sample path of critical LD changes its direction only slowly, and its empirical distribution gives a reasonable approximation of the Gaussian invariant measure.

- Conversely, due to random walk like behaviour, the empirical distribution of the sample path of OLD over the same time interval is still patchy and asymmetric.

EXAMPLE: Gaussian case $\quad U(x) = m x^2/2 \quad$ Spectral gap $= \dfrac{\gamma}{2} - \sqrt{\left(\dfrac{\gamma^2}{4} - m\right)^+}$



Diffusive to ballistic speed-up for $\gamma \propto \sqrt{m}$

QUESTIONS

1) Relation between ① and ② ?

$$\text{OLD} \underset{\text{collapse}}{\overset{\text{lift}}{\rightleftarrows}} \text{Ham. dynamics, LD, RHMC}$$

2) Consequence for relaxation times?

lower bound for relaxation times of lifts

3) Optimal lifts? Maximal speed-up?

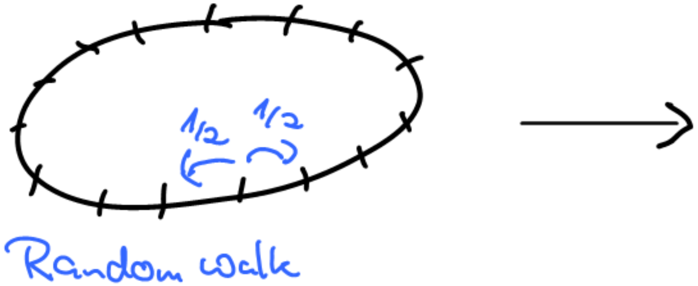
upper bound for relaxation times

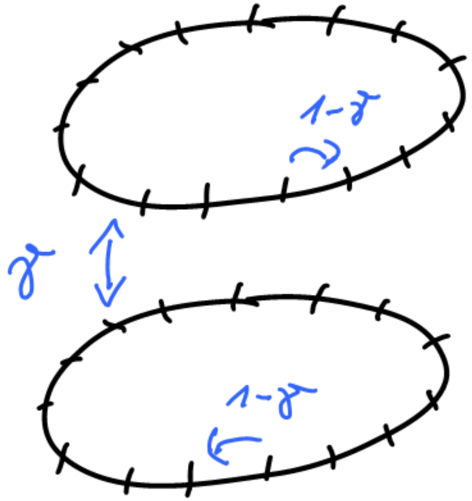
2. Lifts: From Markov chains to diffusions

a) Lifts of Markov chains

[Diaconis, Holmes, Neal 2000], [Chen, Lovasz, Pak 1999]

EXAMPLE





$\hat{S} = S \times V$

$\hat{\mu}(dx\, dv) = \mu(dx)\, K(x, dv)$

p transition kernel on S with invariant measure μ

$\hat{p}$ transition kernel on $\hat{S}$ with invariant measure $\hat{\mu}$

DEFINITION $\hat{p}$ is a lift of p iff

$$(*) \quad \int \hat{p}\big((x,v), A \times V\big)\, K(x, dv) = p(x, A) \quad \forall x \in S,\ A \subseteq V \text{ measurable}$$

REMARK This does $\underline{\underline{\text{not}}}$ imply a corresponding relation between p^n and $\hat{p}^n$.

$\rightsquigarrow$ lift is infinitesimal property

Equivalent formulations of lift property

$\pi(x,v) = x$

$$(**) \quad \int \hat{p}(f\circ\pi)(x,v)\, k(x,dv) = (pf)(x) \qquad \forall x \in S,\ f: S \to \mathbb{R} \text{ meas.+bounded}$$

$$(***) \quad \int \hat{p}(f\circ\pi)\, g\circ\pi \, d\hat{\mu} = \int pf\, g\, d\mu \qquad \forall f, g \in L^2(\mu)$$

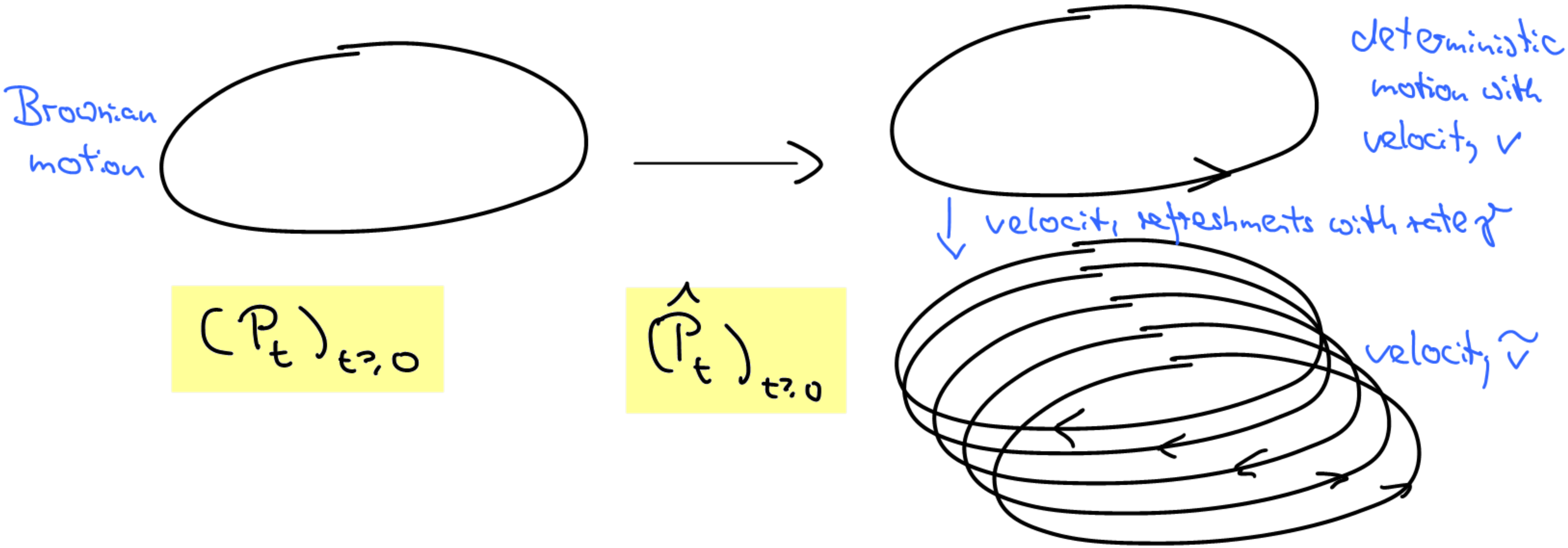
$$(****) \quad \int \hat{L}(f\circ\pi)\, g\circ\pi \, d\hat{\mu} = \underbrace{\int Lf\, g\, d\mu}_{= -\mathcal{E}(f,g)} \qquad \forall f, g \in L^2(\mu)$$

$\hat{L} = \hat{p} - I$, $L = p - I$

"First order lift"

b) From discrete to continuous time

EXAMPLE $\quad S=S^1, \quad V=\mathbb{R}, \quad \mu=\mathrm{Unif}(S), \quad \kappa(dv)=N(0,1)(dv)$



$$\int \hat{P}_t(f\circ\pi)(x,v)\,\kappa(dv) = \mathbb{E}_{\delta_x\otimes\kappa}\big[f(\underbrace{\hat{X}_t}_{\approx N(x,t^2)})\big]$$

Second order lift

$$\overset{f\in \mathrm{Dom}(L)}{=} \mathbb{E}_{\delta_x}\big[f(X_{t^2})\big] + o(t^2) = (P_{t^2}f)(x) + o(t^2) \quad \text{as } t\downarrow 0$$

Equivalent formulation via generators:

$$(**) \quad \lim_{t\downarrow 0} \int \underbrace{\frac{\hat{P}_t(f\circ\pi) - f\circ\pi}{t^2}}_{= \frac{1}{t}\hat{L}(f\circ\pi) + \frac{1}{2}\hat{L}^2(f\circ\pi) + o(1)} (x,v)\, \kappa(dv) = (Lf)(x) \quad \forall f \in \mathrm{Dom}(L)$$

$$(***) \quad \int \hat{L}(f\circ\pi)(x,v)\, \kappa(dv) = 0 \quad \text{and} \quad \frac{1}{2}\int \hat{L}^2(f\circ\pi)(x,v)\, \kappa(dv) = Lf(x)$$

c) Lifts of reversible diffusions

$(P_t)_{t\geq 0}$ Symmetric Markov semigroup on $L^2(\mu)$, generator L

$(\hat{P}_t)_{t\geq 0}$ Markov semigroup on $L^2(\hat{\mu})$, generator $\hat{L}$

DEFINITION [A.E., F. Lörler 2024] $(\hat{P}_t)_{t\geq 0}$ is a 2^{nd} order lift of $(P_t)_{t\geq 0}$ iff

(i) $f \in \mathrm{Dom}(L) \Rightarrow f\circ\pi \in \mathrm{Dom}(\hat{L})$

(ii) $\int \hat{L}(f\circ\pi)\, g\circ\pi \, d\hat{\mu} = 0 \quad \forall f,g \in \mathrm{Dom}(L)$

(iii) $\frac{1}{2}\int \hat{L}(f\circ\pi)\, \hat{L}(g\circ\pi)\, d\hat{\mu} = -\int f\, Lg \, d\mu = \mathcal{E}(f,g) \quad \forall f,g \in \mathrm{Dom}(L)$

DEFINITION [A.E., F. Lörler 2024] $(\hat{P}_t)_{t\geq 0}$ is a **2nd order lift** of $(P_t)_{t\geq 0}$ iff

(i) $f \in \mathrm{Dom}(L) \Rightarrow f\circ\pi \in \mathrm{Dom}(\hat{L})$

(ii) $\int \hat{L}(f\circ\pi)\, g\circ\pi \, d\hat{\mu} = 0 \quad \forall f,g \in \mathrm{Dom}(L)$

(iii) $\frac{1}{2}\int \hat{L}(f\circ\pi)\,\hat{L}(g\circ\pi)\, d\hat{\mu} = -\int f\, Lg \, d\mu = \mathcal{E}(f,g) \quad \forall f,g\in \mathrm{Dom}(L)$

EXAMPLES The following processes are <u>all</u> 2nd order lifts of OLD:

1) Hamiltonian dynamics
2) Langevin dynamics for any $\gamma \geq 0$
3) Randomized HMC for any $\gamma \geq 0$
4) Bouncy particle sampler

etc.

3. Bounds for relaxation times

$$t_{rel}(\hat{P}) := \inf\left\{ t \geq 0 : \|\hat{P}_t f\|_{L^2(\hat{\mu})} \leq \frac{1}{e} \|f\|_{L^2(\hat{\mu})} \ \forall f \in L^2_0(\hat{\mu}) \right\}$$

non-asymptotic L^2 relaxation time

WARNING In general $t_{rel}^{-1} \neq$ asymptotic decay rate,

$t_{rel} \neq$ asymptotic decorrelation time, $t_{rel} \neq$ inverse spectral gap !

EXAMPLE $\|\hat{P}_t\|_{L^2_0(\hat{\mu}) \to L^2_0(\hat{\mu})} \leq C e^{-\lambda t} \Rightarrow t_{rel} \leq \frac{1}{\lambda} \log C$

a) General lower bound for relaxation times of lifts

THEOREM [AE, F. Lörler] If $(\hat{P}_t)$ is a 2^{nd} order lift of (P_t) then

$$t_{rel}(\hat{P}) \geq \frac{1}{2\sqrt{2}} \sqrt{t_{rel}(P)}$$

THEOREM [AE, F. Lörler] If $(\hat{P}_t)$ is a 2^{nd} order lift of (P_t) then

$$t_{rel}(\hat{P}) \geq \frac{1}{2\sqrt{2}} \sqrt{t_{rel}(P)}$$

REMARK Similar result for lifts of Markov chains by Chen, Lovasz, Pak (1999), but with "set time" instead of relaxation time.
Proof via conductance.

PROOF $\lambda > gap(L) \Rightarrow \exists f \in L^2_0(\mu) : \mathcal{E}(f,f) \leq \lambda \|f\|^2_{L^2(\mu)}$

$$\Rightarrow \|\hat{L}(f\circ\pi)\|^2_{L^2(\hat{\mu})} \overset{\text{lift}}{=} 2\,\mathcal{E}(f,f) \leq 2\lambda \|f\|^2_{L^2(\mu)} = 2\lambda \|f\circ\pi\|^2_{L^2(\hat{\mu})}$$

$$\Rightarrow \underbrace{s(\hat{L})}_{} := \inf_{g \in L^2_0(\hat{\mu}) \cap Dom(\hat{L})} \frac{\|\hat{L}g\|_{L^2(\hat{\mu})}}{\|g\|_{L^2(\hat{\mu})}} \leq \sqrt{2\,gap(L)}$$

singular value gap, see also Chatterjee (2023)

One can show that $t_{rel} \geq \frac{1}{2s(\hat{L})}$ $(\dots)$ $\square$

b) Optimal lifts

DEFINITION Let $C \in [1,\infty)$. A lift is called **C-optimal** iff

$$t_{rel}(\hat{P}) \le \frac{C}{2\sqrt{2}} \sqrt{t_{rel}(P)} \,.$$

maximal acceleration up to factor C

EXAMPLE Gaussian case: Critically damped LD is 5.46 optimal lift of OLD.

THEOREM Suppose that

(i) μ satisfies Poincaré inequality with constant $m \in (0,\infty)$.

(ii) $\nabla^2 U \geq -cm$, $\quad c \in [0,\infty)$.

(iii) $\frac{1}{A} \le \frac{\gamma}{\sqrt{m}} \le A$, $\quad A \in [1,\infty)$.

Then RHMC is a C-optimal lift of OLD with $C = 2\sqrt{2}\left(482 \cdot (6 + \frac{5c}{7}) \cdot A + 3\right)$.

Similar results hold for Langevin dynamics, on convex domains in $\mathbb{R}^n$ (with reflection at boundary), and on Riemannian manifolds.

[AE, F. Lörler; work in progress]

PROOF Adaptation of Lu, Wang (2022) and Cao, Lu, Wang (2023).

Argument based on space-time Poincaré inequalities and divergence lemma, following approach in Albritton, Armstrong, Mourrat, Novack (2023).

Simplifications by lift-framework.