Transportation along Langevin semigroups

Dan Mikulincer

MIT

Joint work with Yair Shenfeld

Transport maps

Let $X \sim \mu$ be a measure on $\mathbb{R}^d$ and let $G \sim \gamma$ stand for the standard Gaussian.
If φ is such that $\varphi(G) \stackrel{law}{=} X$, we call φ a *transport map*.

Transport maps

Let $X \sim \mu$ be a measure on $\mathbb{R}^d$ and let $G \sim \gamma$ stand for the standard Gaussian.
If φ is such that $\varphi(G) \stackrel{law}{=} X$, we call φ a *transport map*.

The existence and properties of such maps are useful for:

- Generative models and sampling algorithms.
- Understanding analytic properties of μ.

Optimal transport

Definition (Wasserstein distance between μ and γ)

$$\mathcal{W}_2(\mu, \gamma) := \inf_{\pi} \left\{ \mathbb{E}_{\pi} \left[||x - y||^2 \right] \right\}^{1/2}$$

where π ranges over all possible couplings of μ and γ.

Optimal transport

Definition (Wasserstein distance between μ and γ)

$$\mathcal{W}_2(\mu, \gamma) := \inf_{\pi} \left\{ \mathbb{E}_{\pi} \left[||x - y||^2 \right] \right\}^{1/2}$$

where π ranges over all possible couplings of μ and γ.

Brenier 87': There exists a transport map $\psi^{\mathrm{opt}} : \mathbb{R}^d \to \mathbb{R}^d$:

$$\mathbb{E} \left[\|\psi^{\mathrm{opt}}(G) - G\|^2 \right] = \mathcal{W}_2^2(\mu, \gamma).$$

Optimal transport

Definition (Wasserstein distance between μ and γ)

$$\mathcal{W}_2(\mu, \gamma) := \inf_{\pi} \left\{ \mathbb{E}_{\pi} \left[||x - y||^2 \right] \right\}^{1/2}$$

where π ranges over all possible couplings of μ and γ.

Brenier 87': There exists a transport map $\psi^{\mathrm{opt}} : \mathbb{R}^d \to \mathbb{R}^d$:

$$\mathbb{E}\left[\|\psi^{\mathrm{opt}}(G) - G\|^2\right] = \mathcal{W}_2^2(\mu, \gamma).$$

Caffarelli 00': If μ is more log-concave than γ_d, ψ^{opt} is 1-Lipschitz.

Optimal transport

Definition (Wasserstein distance between μ and γ)

$$\mathcal{W}_2(\mu,\gamma) := \inf_{\pi} \left\{ \mathbb{E}_\pi \left[||x-y||^2 \right] \right\}^{1/2}$$

where π ranges over all possible couplings of μ and γ.

Brenier 87': There exists a transport map $\psi^{\mathrm{opt}} : \mathbb{R}^d \to \mathbb{R}^d$:

$$\mathbb{E}\left[\|\psi^{\mathrm{opt}}(G) - G\|^2\right] = \mathcal{W}_2^2(\mu,\gamma).$$

Caffarelli 00': If μ is more log-concave than γ_d, ψ^{opt} is 1-Lipschitz.

(strong log-concavity: $-\nabla^2 \log\left(\frac{d\mu}{dx}(x)\right) \succeq \mathrm{Id}.$)

Poincaré inequalities

Gaussian Poincaré inequality: For any test function f,

$$\mathrm{Var}(f(G)) \leq \mathbb{E}\left[\|\nabla f(G)\|^2\right].$$

Poincaré inequalities

Gaussian Poincaré inequality: For any test function f,

$$\mathrm{Var}(f(G)) \leq \mathbb{E}\left[\|\nabla f(G)\|^2\right].$$

In general, $X \sim \mu$ satisfies a Poincaré inequality with constant $C_{\mathrm{p}}(\mu) > 0$, if,

$$\mathrm{Var}(f(X)) \leq C_{\mathrm{p}}(\mu)\mathbb{E}\left[\|\nabla f(X)\|^2\right].$$

An inequality of Brascamp and Lieb

Theorem (Brascamp-Lieb 76')

If μ is more log-concave than γ_d, then $C_{\mathrm{p}}(\mu) \leq 1$.

An inequality of Brascamp and Lieb

Theorem (Brascamp-Lieb 76')

If μ is more log-concave than γ_d, then $C_{\mathrm{p}}(\mu) \leq 1$.

Proof.

$$\begin{aligned}\mathrm{Var}_\mu(f) &= \mathrm{Var}_{\gamma_d}(f \circ \psi^{\mathrm{opt}}) \leq \mathbb{E}_{\gamma_d}\left[\|\nabla\left(f \circ \psi^{\mathrm{opt}}\right)\|^2\right] \\ &\leq \mathbb{E}_{\gamma_d}\left[\|\nabla \psi^{\mathrm{opt}}\|^2 \|\nabla f(\psi^{\mathrm{opt}})\|^2\right] = \mathbb{E}_\mu\left[\|\nabla f\|^2\right].\end{aligned}$$

□

Main question and motivation

Question

Can we find Lipschitz transport maps for target measures which are not strongly log-concave?

Main question and motivation

Question

Can we find Lipschitz transport maps for target measures which are not strongly log-concave?

A positive answer will go far beyond Poincaré inequalities:

1. Dimension-free Φ-Sobolev inequalities (generalizing both Poincaré and log-Sobolev).

Main question and motivation

Question

Can we find Lipschitz transport maps for target measures which are not strongly log-concave?

A positive answer will go far beyond Poincaré inequalities:

1. Dimension-free Φ-Sobolev inequalities (generalizing both Poincaré and log-Sobolev).
2. Bounds for higher eigenvalues of the weighted Laplacian.

Main question and motivation

Question

Can we find Lipschitz transport maps for target measures which are not strongly log-concave?

A positive answer will go far beyond Poincaré inequalities:

1. Dimension-free Φ-Sobolev inequalities (generalizing both Poincaré and log-Sobolev).
2. Bounds for higher eigenvalues of the weighted Laplacian.
3. Isoperimetric inequalities.

Main question and motivation

Question

Can we find Lipschitz transport maps for target measures which are not strongly log-concave?

A positive answer will go far beyond Poincaré inequalities:

1. Dimension-free Φ-Sobolev inequalities (generalizing both Poincaré and log-Sobolev).
2. Bounds for higher eigenvalues of the weighted Laplacian.
3. Isoperimetric inequalities.
4. Improved rates of convergence for the CLT.

How to transport μ to ν?

Let $(Y_t)_{t\geq 0}$ be the Langevin dynamics:

$$dY_t = \nabla \log\left(\frac{d\nu}{dx}\right)(Y_t)dt + \sqrt{2}dB_t, \quad Y_0 \sim \mu,$$

with $(B_t)_{t\geq 0}$ a Brownian motion in $\mathbb{R}^d$.

How to transport μ to ν?

Let $(Y_t)_{t\geq 0}$ be the Langevin dynamics:

$$dY_t = \nabla \log\left(\frac{d\nu}{dx}\right)(Y_t)dt + \sqrt{2}dB_t, \quad Y_0 \sim \mu,$$

with $(B_t)_{t\geq 0}$ a Brownian motion in $\mathbb{R}^d$. Let (Q_t) be the Langevin semigroup: $Q_t\eta(x) = E[\eta(Y_t)|Y_0 = x]$,

How to transport μ to ν?

Let $(Y_t)_{t\geq 0}$ be the Langevin dynamics:

$$dY_t = \nabla \log\left(\frac{d\nu}{dx}\right)(Y_t)dt + \sqrt{2}dB_t, \quad Y_0 \sim \mu,$$

with $(B_t)_{t\geq 0}$ a Brownian motion in $\mathbb{R}^d$. Let (Q_t) be the Langevin semigroup: $Q_t\eta(x) = E[\eta(Y_t)|Y_0 = x]$, and let $\rho_t := Q_t\left(\frac{d\mu}{d\nu}\right)d\nu = \text{Law}(Y_t)$ so that the path of measures $(\rho_t)_{t\geq 0}$ interpolates between $\rho_0 = \mu$ to $\rho_\infty = \nu$.

The continuity equation

The Langevin path $(\rho_t)_{t\geq 0}$ satisfies the continuity equation

$$\partial_t \rho_t + \nabla(V_t \rho_t) = 0,$$

where

$$V_t(x) = -\nabla \log\left(\frac{d\rho_t}{d\nu}\right)(x) = -\nabla \log Q_t\left(\frac{d\mu}{d\nu}\right)(x)$$

(because $\partial_t Q_t \eta = \Delta Q_t \eta + \langle \nabla Q_t \eta, \nabla \log\left(\frac{d\nu}{dx}\right)\rangle$).

Transportation along Langevin semigroups

Define the family of diffeomorphisms $S_t : \mathbb{R}^d \to \mathbb{R}^d$ by

$$\partial_t S_t(x) = V_t(S_t(x)), \quad S_0(x) = x.$$

Transportation along Langevin semigroups

Define the family of diffeomorphisms $S_t : \mathbb{R}^d \to \mathbb{R}^d$ by

$$\partial_t S_t(x) = V_t(S_t(x)), \quad S_0(x) = x.$$

S_t transports $\mu = \rho_0$ to ρ_t and $T_t := S_t^{-1}$ transports ρ_t to $\rho_0 = \mu$.

Transportation along Langevin semigroups

Define the family of diffeomorphisms $S_t : \mathbb{R}^d \to \mathbb{R}^d$ by

$$\partial_t S_t(x) = V_t(S_t(x)), \quad S_0(x) = x.$$

S_t transports $\mu = \rho_0$ to ρ_t and $T_t := S_t^{-1}$ transports ρ_t to $\rho_0 = \mu$.
The transport maps along Langevin semigroups are defined as

$$S_{\mathrm{LVN}} := \lim_{t\to\infty} S_t \quad \text{transports } \mu = \rho_0 \text{ to } \rho_\infty = \nu,$$

$$T_{\mathrm{LVN}} := \lim_{t\to\infty} T_t \quad \text{transports } \nu = \rho_\infty \text{ to } \rho_0 = \mu.$$

Warm-up

Theorem (M, Shenfeld)

- *If* $\nu = \gamma_d$ *and* $\mu = \kappa$*-log-concave (i.e.,* $-\nabla^2 \log\left(\frac{d\mu}{dx}(x)\right) \succeq \kappa I_d$*), for* $\kappa > 0$*, then* T_{LVN} *is* $\frac{1}{\sqrt{\kappa}}$*-Lipschitz. The result is sharp and already follows from results of Kim and E. Milman.*

Warm-up

Theorem (M, Shenfeld)

- *If* $\nu = \gamma_d$ *and* $\mu = \kappa$*-log-concave (i.e.,* $-\nabla^2 \log\left(\frac{d\mu}{dx}(x)\right) \succeq \kappa I_d$*), for* $\kappa > 0$*, then* T_{LVN} *is* $\frac{1}{\sqrt{\kappa}}$*-Lipschitz. The result is sharp and already follows from results of Kim and E. Milman.*

- *If* $\nu = \gamma_d$ *and* $\mu =$ *log-concave and* β*-log-convex (i.e.,* $-\nabla^2 \log\left(\frac{d\mu}{dx}(x)\right) \preceq \beta I_d$*), for* $\beta > 0$*, then* S_{LVN} *is* $\sqrt{\beta}$*-Lipschitz. The result is sharp.*

Warm-up

Theorem (M, Shenfeld)

- *If* $\nu = \gamma_d$ *and* $\mu = \kappa$*-log-concave (i.e.,* $-\nabla^2 \log\left(\frac{d\mu}{dx}(x)\right) \succeq \kappa I_d$*), for* $\kappa > 0$*, then* T_{LVN} *is* $\frac{1}{\sqrt{\kappa}}$*-Lipschitz. The result is sharp and already follows from results of Kim and E. Milman.*

- *If* $\nu = \gamma_d$ *and* $\mu =$ *log-concave and* β*-log-convex (i.e.,* $-\nabla^2 \log\left(\frac{d\mu}{dx}(x)\right) \preceq \beta I_d$*), for* $\beta > 0$*, then* S_{LVN} *is* $\sqrt{\beta}$*-Lipschitz. The result is sharp.*

The theorem parallels the analogous results for the optimal transport map.

Semi-log-concave measures with bounded support

Theorem (M, Shenfeld)

- *If* $\nu = \gamma_d$ *and* μ *is* κ*-log-concave with* $\mathrm{diam}(\mathrm{supp}(\mu)) \leq R$*, and* $\kappa R^2 < 1$*, then* T_{LVN} *is* $e^{\frac{1-\kappa R^2}{2}} R$*-Lipschitz.*

Semi-log-concave measures with bounded support

Theorem (M, Shenfeld)

- *If $\nu = \gamma_d$ and μ is κ-log-concave with $\operatorname{diam}(\operatorname{supp}(\mu)) \le R$, and $\kappa R^2 < 1$, then T_{LVN} is $e^{\frac{1-\kappa R^2}{2}} R$-Lipschitz.*
- *In particular, if μ is log-concave (so $\kappa = 0$) with $\operatorname{diam}(\operatorname{supp}(\mu)) \le R$, then T_{LVN} is $e^{1/2} R$-Lipschitz. The order of the Lipschitz constant is sharp.*

Semi-log-concave measures with bounded support

Theorem (M, Shenfeld)

- *If $\nu = \gamma_d$ and μ is κ-log-concave with $\mathrm{diam}(\mathrm{supp}(\mu)) \leq R$, and $\kappa R^2 < 1$, then T_{LVN} is $e^{\frac{1-\kappa R^2}{2}} R$-Lipschitz.*
- *In particular, if μ is log-concave (so $\kappa = 0$) with $\mathrm{diam}(\mathrm{supp}(\mu)) \leq R$, then T_{LVN} is $e^{1/2} R$-Lipschitz. The order of the Lipschitz constant is sharp.*

The question (due to Kolesnikov) of whether the optimal transport map from γ_d to μ which is log-concave with $\mathrm{diam}(\mathrm{supp}(\mu)) \leq R$ is $O(R)$-Lipschitz, is open.

Gaussian mixtures with bounded mixing measure

Theorem (M, Shenfeld)

If $\nu = \gamma_d$ and $\mu = \gamma_d \star m$ with $\mathrm{diam}(\mathrm{supp}(m)) \leq R$, then T_{LVN} is $e^{\frac{R^2}{2}}$-Lipschitz. The order of the Lipschitz constant is sharp.

Gaussian mixtures with bounded mixing measure

Theorem (M, Shenfeld)

If $\nu = \gamma_d$ and $\mu = \gamma_d \star m$ with $\mathrm{diam}(\mathrm{supp}(m)) \leq R$, then T_{LVN} is $e^{\frac{R^2}{2}}$-Lipschitz. The order of the Lipschitz constant is sharp.

The theorem leads to improved log-Sobolev inequalities for mixtures of Gaussians.

Isotropic log-concave measures - the Brownian transport map

There are adaptions of the technique to other settings.

Theorem (M, Shenfeld)

If $\nu = \gamma_\infty$ and μ is log-concave and isotropic. There exists map T, such that $T_*\nu = \mu$ and,

$$\mathbb{E}_\gamma \left[\|D\Phi\|^2\right] \leq \mathrm{polylog}(d).$$

Isotropic log-concave measures - the Brownian transport map

There are adaptions of the technique to other settings.

Theorem (M, Shenfeld)

If $\nu = \gamma_\infty$ and μ is log-concave and isotropic. There exists map T, such that $T_*\nu = \mu$ and,

$$\mathbb{E}_\gamma\left[\|D\Phi\|^2\right] \leq \mathrm{polylog}(d).$$

The result it tightly connected to the KLS conjecture and builds upon recent advances.

High level idea of proofs

Recall

$$\partial_t S_t(x) = V_t(S_t(x)), \quad S_0(x) = x,$$

High level idea of proofs

Recall

$$\partial_t S_t(x) = V_t(S_t(x)), \quad S_0(x) = x,$$

so

$$\partial_t \nabla S_t(x) = \nabla V_t(S_t(x)) \nabla S_t(x).$$

High level idea of proofs

Recall

$$\partial_t S_t(x) = V_t(S_t(x)), \quad S_0(x) = x,$$

so

$$\partial_t \nabla S_t(x) = \nabla V_t(S_t(x)) \nabla S_t(x).$$

Lemma

- *The Lipschitz constant of T_{LVN} is at most* $\exp\left(\int_0^\infty \sup_x \lambda_{\max}(-\nabla V_t(x))dt\right)$.

Proof idea

- The key is to control $\lambda_{\max}(-\nabla V_t(x))$.
- Recall $-\nabla V_t(x) = \nabla^2 \log Q_t\left(\frac{d\mu}{d\gamma_d}\right)(x)$.

Proof idea

- The key is to control $\lambda_{\max}(-\nabla V_t(x))$.
- Recall $-\nabla V_t(x) = \nabla^2 \log Q_t \left(\frac{d\mu}{d\gamma_d} \right)(x)$.
- We show that $\nabla V_t(x)$ can be represented as a covariance matrix of some measure μ_t.
- The measure μ_t turns out to be a Gaussian tilt of the measure μ.
- This allows to bound $\nabla V_t(x)$ using covariance inequalities such as Brascamp-Lieb or bounded support inequalities.

Thank You