Transportation along Langevin semigroups

Dan Mikulincer

MIT

Joint work with Yair Shenfeld
Let $X \sim \mu$ be a measure on $\mathbb{R}^d$ and let $G \sim \gamma$ stand for the standard Gaussian.

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

The existence and properties of such maps are useful for:

- Generative models and sampling algorithms.
- Understanding analytic properties of $\mu$. 
Definition (Wasserstein distance between $\mu$ and $\gamma$)

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

where $\pi$ ranges over all possible couplings of $\mu$ and $\gamma$. 
Definition (Wasserstein distance between $\mu$ and $\gamma$)

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

where $\pi$ ranges over all possible couplings of $\mu$ and $\gamma$.

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

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

Definition (Wasserstein distance between $\mu$ and $\gamma$)

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

where $\pi$ ranges over all possible couplings of $\mu$ and $\gamma$.

Brenier 87’: There exists a transport map $\psi^{\text{opt}} : \mathbb{R}^d \rightarrow \mathbb{R}^d$:

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

Caffarelli 00’: If $\mu$ is more log-concave than $\gamma_d$, $\psi^{\text{opt}}$ is 1-Lipschitz.
Optimal transport

**Definition (Wasserstein distance between $\mu$ and $\gamma$)**

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

where $\pi$ ranges over all possible couplings of $\mu$ and $\gamma$.

**Brenier 87’**: There exists a transport map $\psi^{\text{opt}} : \mathbb{R}^d \to \mathbb{R}^d$:

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

**Caffarelli 00’**: If $\mu$ is more log-concave than $\gamma_d$, $\psi^{\text{opt}}$ is 1-Lipschitz.

(strong log-concavity: $-\nabla^2 \log \left( \frac{d\mu}{dx}(x) \right) \succeq \text{Id.}$)
Gaussian Poincaré inequality: For any test function \( f \),

\[
\text{Var}(f(G)) \leq \mathbb{E} \left[ \| \nabla f(G) \|^2 \right].
\]
Gaussian Poincaré inequality: For any test function $f$,

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

In general, $X \sim \mu$ satisfies a Poincaré inequality with constant $C_p(\mu) > 0$, if,

$$\text{Var}(f(X)) \leq C_p(\mu) \mathbb{E} \left[ \| \nabla f(X) \|^2 \right].$$
An inequality of Brascamp and Lieb

**Theorem (Brascamp-Lieb 76’)**

*If* \( \mu \) *is more log-concave than* \( \gamma_d \), *then* \( C_p(\mu) \leq 1 \).
An inequality of Brascamp and Lieb

**Theorem (Brascamp-Lieb 76’)**

If $\mu$ is more log-concave than $\gamma_d$, then $C_p(\mu) \leq 1$.

**Proof.**

$$\text{Var}_\mu(f) = \text{Var}_{\gamma_d}(f \circ \psi^{\text{opt}}) \leq \mathbb{E}_{\gamma_d} \left[ \| \nabla (f \circ \psi^{\text{opt}}) \|^2 \right]$$

$$\leq \mathbb{E}_{\gamma_d} \left[ \| \nabla \psi^{\text{opt}} \|^2 \| \nabla f(\psi^{\text{opt}}) \|^2 \right] = \mathbb{E}_\mu \left[ \| \nabla f \|^2 \right].$$
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 $\Phi$-Sobolev inequalities (generalizing both Poincaré and log-Sobolev).
2. Bounds for higher eigenvalues of the weighted Laplacian.
**Main question and motivation**

<table>
<thead>
<tr>
<th>Question</th>
</tr>
</thead>
<tbody>
<tr>
<td>Can we find Lipschitz transport maps for target measures which are not strongly log-concave?</td>
</tr>
</tbody>
</table>

A positive answer will go far beyond Poincaré inequalities:

1. Dimension-free $\Phi$-Sobolev inequalities (generalizing both Poincaré and log-Sobolev).
2. Bounds for higher eigenvalues of the weighted Laplacian.
3. Isoperimetric inequalities.
<table>
<thead>
<tr>
<th>Question</th>
</tr>
</thead>
<tbody>
<tr>
<td>Can we find Lipschitz transport maps for target measures which are not strongly log-concave?</td>
</tr>
</tbody>
</table>

A positive answer will go far beyond Poincaré inequalities:

1. Dimension-free $\Phi$-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 $\mu$ to $\nu$?

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 $\mu$ to $\nu$?

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: $Q_t\eta(x) = E[\eta(Y_t)|Y_0 = x]$,
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 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 \(\partial_t Q_t \eta = \Delta Q_t \eta + \langle \nabla Q_t \eta, \nabla \log \left( \frac{d\nu}{dx} \right) \rangle \)).
Define the family of diffeomorphisms $S_t : \mathbb{R}^d \to \mathbb{R}^d$ by

$$\partial_t S_t(x) = V_t(S_t(x)), \quad S_0(x) = x.$$
Define the family of diffeomorphisms $S_t : \mathbb{R}^d \rightarrow \mathbb{R}^d$ by

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

$S_t$ transports $\mu = \rho_0$ to $\rho_t$ and $T_t := S_t^{-1}$ transports $\rho_t$ to $\rho_0 = \mu$. 
Define the family of diffeomorphisms $S_t : \mathbb{R}^d \rightarrow \mathbb{R}^d$ by

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

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

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

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

- If $\nu = \gamma_d$ and $\mu$ = $\beta$-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.
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 $\beta$-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.,
  \( -\nabla^2 \log \left( \frac{d\mu}{dx}(x) \right) \geq \kappa I_d \)), 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\text{-concave and } \beta\text{-log-convex (i.e.,}
  \( -\nabla^2 \log \left( \frac{d\mu}{dx}(x) \right) \leq \beta I_d \)), 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.
Theorem (M, Shenfeld)

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

In particular, if $\mu$ is log-concave (so $\kappa = 0$) with $\text{diam}(\text{supp}(\mu)) \leq R$, then $T_{LVN}$ is $\frac{e}{2} R$-Lipschitz. The order of the Lipschitz constant is sharp.

The question (due to Kolesnikov) of whether the optimal transport map from $\gamma d$ to $\mu$ which is log-concave with $\text{diam}(\text{supp}(\mu)) \leq R$ is $O(R)$-Lipschitz, is open.
Theorem (M, Shenfeld)

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

- In particular, if $\mu$ is log-concave (so $\kappa = 0$) with $\text{diam}(\text{supp}(\mu)) \leq R$, then $T_{LVN}$ is $e^{1/2} R$-Lipschitz. The order of the Lipschitz constant is sharp.
Theorem (M, Shenfeld)

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

- In particular, if \( \mu \) is log-concave (so \( \kappa = 0 \)) with \( \text{diam}(\text{supp}(\mu)) \leq R \), 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 \( \gamma_d \) to \( \mu \) which is log-concave with \( \text{diam}(\text{supp}(\mu)) \leq R \) is \( O(R) \)-Lipschitz, is open.
Theorem (M, Shenfeld)

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

If \( \nu = \gamma_d \) and \( \mu = \gamma_d \ast m \) with \( \text{diam}(\text{supp}(m)) \leq R \), 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 adaptations of the technique to other settings.

**Theorem (M, Shenfeld)**

If \( \nu = \gamma_\infty \) and \( \mu \) is log-concave and isotropic. There exists map \( T \), such that \( T_\ast \nu = \mu \) and,

\[
\mathbb{E}_\gamma \left[ \| D\Phi \|^2 \right] \leq \text{polylog}(d).
\]
There are adaptations of the technique to other settings.

**Theorem (M, Shenfeld)**

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

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

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

\[ \partial_t S_t(x) = V_t(S_t(x)), \quad S_0(x) = x, \]

Lemma

The Lipschitz constant of \( T \) is at most \( \exp \int_0^\infty \sup_x \lambda_{\max}(-\nabla V_t(x)) \, dt \).
Recall

\[ \partial_t S_t(x) = \nabla V_t(S_t(x)), \quad S_0(x) = x, \]

so

\[ \partial_t \nabla S_t(x) = \nabla \nabla V_t(S_t(x)) \nabla S_t(x). \]
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

$$\exp\left( \int_0^\infty \sup_x \lambda_{\text{max}}(-\nabla V_t(x)) dt \right).$$
Proof idea

- The key is to control $\lambda_{\text{max}}(-\nabla V_t(x))$.
- 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 $\lambda_{\text{max}}(-\nabla V_t(x))$.
- 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 $\mu_t$.
- The measure $\mu_t$ turns out to be a Gaussian tilt of the measure $\mu$.
- This allows to bound $\nabla V_t(x)$ using covariance inequalities such as Brascamp-Lieb or bounded support inequalities.
Thank You