# Linearized Optimal Transport on Particle Systems and Related Applications

Nicholas Karris
nkarris@ucsd.edu

Candidacy Defense
December 6, 2024

# Overview

# General Setting

Consider a particle system governed by some microscale simulator $S_h$ that depends on a small time step $h > 0$.

# General Setting

Consider a particle system governed by some microscale simulator $S_h$ that depends on a small time step $h > 0$.

In our case, $S_h$ is

# General Setting

Consider a particle system governed by some microscale simulator $S_h$ that depends on a small time step $h > 0$.

In our case, $S_h$ is

1. random (e.g., $S_h$ simulates a stochastic process)

# General Setting

Consider a particle system governed by some microscale simulator $S_h$ that depends on a small time step $h > 0$.

In our case, $S_h$ is

1. random (e.g., $S_h$ simulates a stochastic process)
2. computationally expensive

# General Setting

Consider a particle system governed by some microscale simulator $S_h$ that depends on a small time step $h > 0$.

In our case, $S_h$ is

1. random (e.g., $S_h$ simulates a stochastic process)
2. computationally expensive

If we want to simulate the system until time $T$, we can iteratively compute $S_h$ for $\frac{T}{h}$ steps.

# General Setting

Consider a particle system governed by some microscale simulator $S_h$ that depends on a small time step $h > 0$.

In our case, $S_h$ is

1. random (e.g., $S_h$ simulates a stochastic process)
2. computationally expensive

If we want to simulate the system until time $T$, we can iteratively compute $S_h$ for $\frac{T}{h}$ steps.

Motivating Question: Is there a way to skip some of these expensive microscale steps?

# Motivating Example

# Motivating Example

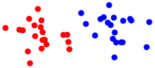
Can we use two snapshots at nearby times to predict where the particles will be at some later time?

# Motivating Example

Can we use two snapshots at nearby times to predict where the particles will be at some later time?

# Motivating Example

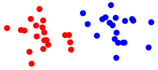
Can we use two snapshots at nearby times to predict where the particles will be at some later time?



# Motivating Example

Can we use two snapshots at nearby times to predict where the particles will be at some later time?

If we can use this data to predict the evolution, we could skip the interim microscale steps.

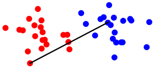


# Motivating Example

Can we use two snapshots at nearby times to predict where the particles will be at some later time?

If we can use this data to predict the evolution, we could skip the interim microscale steps.

One idea: Predict each particle's location.

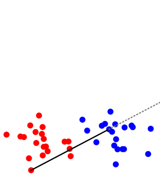


# Motivating Example

Can we use two snapshots at nearby times to predict where the particles will be at some later time?

If we can use this data to predict the evolution, we could skip the interim microscale steps.

One idea: Predict each particle's location.

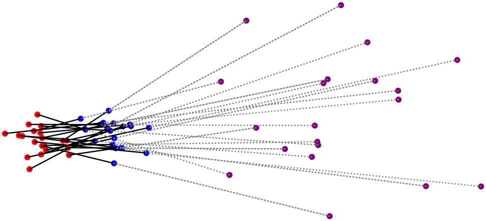


# Motivating Example

Can we use two snapshots at nearby times to predict where the particles will be at some later time?

If we can use this data to predict the evolution, we could skip the interim microscale steps.

One idea: Predict each particle's location.

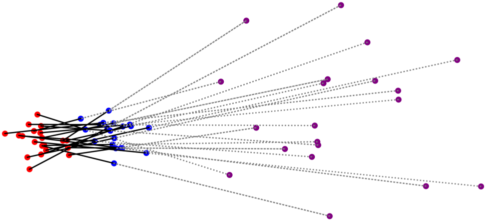


# Motivating Example

Can we use two snapshots at nearby times to predict where the particles will be at some later time?

If we can use this data to predict the evolution, we could skip the interim microscale steps.

One idea: Predict each particle's location. — Bad!



# Key Insight

# Key Insight

Many of these systems exhibit behavior on multiple time scales:

# Key Insight

Many of these systems exhibit behavior on multiple time scales:

- The individual particles move chaotically on short time scales

# Key Insight

Many of these systems exhibit behavior on multiple time scales:

- The individual particles move chaotically on short time scales
- The distribution of particles evolves much more slowly and predictably over long time scales.

# Key Insight

Many of these systems exhibit behavior on multiple time scales:

- The individual particles move chaotically on short time scales
- The distribution of particles evolves much more slowly and predictably over long time scales.

Key Idea: Use optimal transport to approximate the evolution of the distribution rather than the individual particles.

# Overview

# Optimal Transport Maps

Monge Problem: Given two probability distributions $\sigma, \mu$ on $\mathbb{R}^d$, define the optimal transport map from $\sigma$ to $\mu$ to be

$$T_\sigma^\mu = \underset{T_\# \sigma = \mu}{\operatorname{argmin}} \int_{\mathbb{R}^d} \|x - T(x)\|_2^2 \, d\sigma(x).$$

# Optimal Transport Maps

Monge Problem: Given two probability distributions $\sigma, \mu$ on $\mathbb{R}^d$, define the optimal transport map from $\sigma$ to $\mu$ to be

$$T_\sigma^\mu = \underset{T_\# \sigma = \mu}{\operatorname{argmin}} \int_{\mathbb{R}^d} \|x - T(x)\|_2^2 \, d\sigma(x).$$

**Theorem (Brenier)**

*A unique minimizer exists when one of $\sigma$ or $\mu$ has a density.*

# Optimal Transport Maps

Monge Problem: Given two probability distributions $\sigma, \mu$ on $\mathbb{R}^d$, define the optimal transport map from $\sigma$ to $\mu$ to be

$$T_\sigma^\mu = \underset{T_\# \sigma = \mu}{\operatorname{argmin}} \int_{\mathbb{R}^d} \|x - T(x)\|_2^2 \, d\sigma(x).$$

## Theorem (Brenier)

*A unique minimizer exists when one of $\sigma$ or $\mu$ has a density.*

## Proposition ([1])

*If $N \in \mathbb{N}$ is fixed, and $\sigma = \frac{1}{N}\sum_{i=1}^N \delta_{x_i}$ and $\mu = \frac{1}{N}\sum_{i=1}^N \delta_{y_i}$ are uniform measures on $N$ points in $\mathbb{R}^d$, then there exists an optimal transport map $T_\sigma^\mu$ that solves the Monge problem, and $T_\sigma^\mu(x_i) = y_{\tau(i)}$ for some permutation $\tau \in S_N$.*

[1] Peyré and Cuturi. Computational Optimal Transport (2019)

# Velocity Flow Fields

**Theorem ([2])**

# Velocity Flow Fields

## Theorem ([2])

*Let* $t \mapsto \mu_t \in \mathcal{P}(\mathbb{R}^d)$ *be absolutely continuous in the Wasserstein-2 distance.*

# Velocity Flow Fields

## Theorem ([2])

*Let* $t \mapsto \mu_t \in \mathcal{P}(\mathbb{R}^d)$ *be absolutely continuous in the Wasserstein-2 distance. Then there is a unique minimal-energy vector field* $\mathbf{v}_t : \mathbb{R}^d \to \mathbb{R}^d$ *such that the continuity equation*

$$\partial_t \mu_t + \nabla \cdot (\mathbf{v}_t \mu_t) = 0$$

*holds in the sense of distributions.*

# Velocity Flow Fields

## Theorem ([2])

*Let $t \mapsto \mu_t \in \mathcal{P}(\mathbb{R}^d)$ be absolutely continuous in the Wasserstein-2 distance. Then there is a unique minimal-energy vector field $\mathbf{v}_t : \mathbb{R}^d \to \mathbb{R}^d$ such that the continuity equation*

$$\partial_t \mu_t + \nabla \cdot (\mathbf{v}_t \mu_t) = 0$$

*holds in the sense of distributions. Moreover, for $\mathcal{L}^1$-a.e. $t$,*

$$\lim_{H \to 0} \frac{W_2(\mu_{t+H}, (\mathrm{id} + H\mathbf{v}_t)_{\#}\mu_t)}{|H|} = 0$$

# Velocity Flow Fields

## Theorem ([2])

*Let* $t \mapsto \mu_t \in \mathcal{P}(\mathbb{R}^d)$ *be absolutely continuous in the Wasserstein-2 distance. Then there is a unique minimal-energy vector field* $\mathbf{v}_t : \mathbb{R}^d \to \mathbb{R}^d$ *such that the continuity equation*

$$\partial_t \mu_t + \nabla \cdot (\mathbf{v}_t \mu_t) = 0$$

*holds in the sense of distributions. Moreover, for* $\mathcal{L}^1$*-a.e.* $t$*,*

$$\lim_{H \to 0} \frac{W_2(\mu_{t+H}, (\mathrm{id} + H\mathbf{v}_t)_{\#}\mu_t)}{|H|} = 0$$

*and, assuming* $\mu_t \ll \mathcal{L}^d$*,*

$$\lim_{h \to 0} \frac{1}{h}(T_{\mu_t}^{\mu_{t+h}} - \mathrm{id}) = \mathbf{v}_t \qquad \textit{in } L^2(\mu_t).$$

[2] Ambrosio, Gigli, Savaré. Gradient Flows in Metric Spaces and in the Space of Probability Measures (2008)

# Differential Geometry Perspective

# Differential Geometry Perspective

On the Wasserstein Manifold $\mathbb{W}_2 = (\mathcal{P}_2(\mathbb{R}^d), W_2)$, the tangent space at $\mu \in \mathcal{P}_2(\mathbb{R}^d)$ can be identified with the space of vector fields in $L^2(\mu)$.

# Differential Geometry Perspective

On the Wasserstein Manifold $\mathbb{W}_2 = (\mathcal{P}_2(\mathbb{R}^d), W_2)$, the tangent space at $\mu \in \mathcal{P}_2(\mathbb{R}^d)$ can be identified with the space of vector fields in $L^2(\mu)$. The exponential map at $\mu$ is

$$\exp_\mu^{\mathbb{W}_2} : L^2(\mu) \to \mathcal{P}_2(\mathbb{R}^d) \qquad \exp_\mu^{\mathbb{W}_2}(\mathbf{v}) = (\mathrm{id} + \mathbf{v})_{\#}\mu,$$

and the curve $H \mapsto \exp_\mu^{\mathbb{W}_2}(H\mathbf{v})$ is a constant-speed geodesic.

# Differential Geometry Perspective

On the Wasserstein Manifold $\mathbb{W}_2 = (\mathcal{P}_2(\mathbb{R}^d), W_2)$, the tangent space at $\mu \in \mathcal{P}_2(\mathbb{R}^d)$ can be identified with the space of vector fields in $L^2(\mu)$. The exponential map at $\mu$ is

$$\exp_\mu^{\mathbb{W}_2} : L^2(\mu) \to \mathcal{P}_2(\mathbb{R}^d) \qquad \exp_\mu^{\mathbb{W}_2}(\mathbf{v}) = (\mathrm{id} + \mathbf{v})_\# \mu,$$

and the curve $H \mapsto \exp_\mu^{\mathbb{W}_2}(H\mathbf{v})$ is a constant-speed geodesic.

Thus, the earlier theorem can be interpreted as saying

$$\mu_{t+H} \approx (\mathrm{id} + H\mathbf{v}_t)_\# \mu_t = \exp_{\mu_t}^{\mathbb{W}_2}(H\mathbf{v}_t).$$

# Differential Geometry Perspective

On the Wasserstein Manifold $\mathbb{W}_2 = (\mathcal{P}_2(\mathbb{R}^d), W_2)$, the tangent space at $\mu \in \mathcal{P}_2(\mathbb{R}^d)$ can be identified with the space of vector fields in $L^2(\mu)$. The exponential map at $\mu$ is

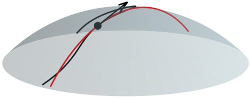
$$\exp_\mu^{\mathbb{W}_2} : L^2(\mu) \to \mathcal{P}_2(\mathbb{R}^d) \qquad \exp_\mu^{\mathbb{W}_2}(\mathbf{v}) = (\mathrm{id} + \mathbf{v})_{\#}\mu,$$

and the curve $H \mapsto \exp_\mu^{\mathbb{W}_2}(H\mathbf{v})$ is a constant-speed geodesic.

Thus, the earlier theorem can be interpreted as saying

$$\mu_{t+H} \approx (\mathrm{id} + H\mathbf{v}_t)_{\#}\mu_t = \exp_{\mu_t}^{\mathbb{W}_2}(H\mathbf{v}_t).$$

I.e., we approximate $\mu_{t+H}$ by the geodesic through $\mu_t$ in the direction of the "tangent vector" $\mathbf{v}_t$.



# Euler's Method on Wasserstein Manifold

This motivates the following as the analog of Euler's method on $\mathbb{W}_2$:

# Euler's Method on Wasserstein Manifold

This motivates the following as the analog of Euler's method on $\mathbb{W}_2$:

Suppose $F(t, \nu) \in T_\nu \mathbb{W}_2$ for all $t \in [0, T]$ and all $\nu \in \mathcal{P}_2(\mathbb{R}^d)$,

# Euler's Method on Wasserstein Manifold

This motivates the following as the analog of Euler's method on $\mathbb{W}_2$:

Suppose $F(t, \nu) \in T_\nu \mathbb{W}_2$ for all $t \in [0, T]$ and all $\nu \in \mathcal{P}_2(\mathbb{R}^d)$, and suppose $(\mu_t)_{t \in [0,T]}$ is such that $(\mu_t, \mathbf{v}_t)$ satisfies the continuity equation with

$$\mathbf{v}_t = F(t, \mu_t).$$

# Euler's Method on Wasserstein Manifold

This motivates the following as the analog of Euler's method on $\mathbb{W}_2$:

Suppose $F(t,\nu) \in T_\nu \mathbb{W}_2$ for all $t \in [0,T]$ and all $\nu \in \mathcal{P}_2(\mathbb{R}^d)$, and suppose $(\mu_t)_{t\in[0,T]}$ is such that $(\mu_t, \mathbf{v}_t)$ satisfies the continuity equation with

$$\mathbf{v}_t = F(t, \mu_t).$$

Then for an initial measure $\mu_0$ and discrete times $t_n = nH$, we set

$$\mu_{(0)} = \mu_0, \qquad \mu_{(n+1)} = (\mathrm{id} + HF(t_n, \mu_{(n)}))_{\#}\mu_{(n)}.$$

# Euler's Method on Wasserstein Manifold

This motivates the following as the analog of Euler's method on $\mathbb{W}_2$:

Suppose $F(t, \nu) \in T_\nu \mathbb{W}_2$ for all $t \in [0, T]$ and all $\nu \in \mathcal{P}_2(\mathbb{R}^d)$, and suppose $(\mu_t)_{t \in [0,T]}$ is such that $(\mu_t, \mathbf{v}_t)$ satisfies the continuity equation with

$$\mathbf{v}_t = F(t, \mu_t).$$

Then for an initial measure $\mu_0$ and discrete times $t_n = nH$, we set

$$\mu_{(0)} = \mu_0, \qquad \mu_{(n+1)} = (\mathrm{id} + HF(t_n, \mu_{(n)}))_{\#}\mu_{(n)}.$$

The idea is that $\mu_{(n)} \approx \mu_{t_n}$.

# Local Truncation Error

**Theorem (K., Nikitopoulos, Kevrekidis, Lee, Cloninger)**

*(Informal) Let* $(\mu_t)$ *be absolutely continuous, and let* $\mathbf{v}_t = F(t, \mu_t)$ *be its velocity flow field.*

# Local Truncation Error

## Theorem (K., Nikitopoulos, Kevrekidis, Lee, Cloninger)

*(Informal) Let* $(\mu_t)$ *be absolutely continuous, and let* $\mathbf{v}_t = F(t, \mu_t)$ *be its velocity flow field. Assume the characteristic ODE*

$$\gamma_x(t) = x, \qquad \frac{d}{ds}\gamma_x(s) = \mathbf{v}_s(\gamma_x(s))$$

*admits a globally defined solution that is* $C^2(t, t+H)$ *for* $\mu_t$*-a.e.* $x \in \mathbb{R}^d$.

# Local Truncation Error

**Theorem (K., Nikitopoulos, Kevrekidis, Lee, Cloninger)**

*(Informal) Let $(\mu_t)$ be absolutely continuous, and let $\mathbf{v}_t = F(t, \mu_t)$ be its velocity flow field. Assume the characteristic ODE*

$$\gamma_x(t) = x, \qquad \frac{d}{ds}\gamma_x(s) = \mathbf{v}_s(\gamma_x(s))$$

*admits a globally defined solution that is $C^2(t, t+H)$ for $\mu_t$-a.e. $x \in \mathbb{R}^d$. Then*

$$W_2(\mu_{t+H}, (\mathrm{id} + HF(t, \mu_t))_\# \mu_t) \leq \frac{H^2}{2} \left( \int_{\mathbb{R}^d} \max_{r \in [t, t+H]} \|\gamma_x''(r)\|_2^2 \, d\mu_t(x) \right)^{1/2}$$

This says that the local truncation error of this method is $\mathcal{O}(H^2)$.

# First-Order Global Error Bound

**Theorem (K., Nikitopoulos, Kevrekidis, Lee, Clonigner)**

*(Informal)*

# First-Order Global Error Bound

**Theorem (K., Nikitopoulos, Kevrekidis, Lee, Clonigner)**

*(Informal) Suppose*

$$M = \left( \int_{\mathbb{R}^d} \max_{r \in [0,T]} \|\gamma_x''(r)\|_2^2 \, d\mu_0(x) \right)^{1/2} < \infty.$$

# First-Order Global Error Bound

## Theorem (K., Nikitopoulos, Kevrekidis, Lee, Clonigner)

*(Informal) Suppose*

$$M = \left( \int_{\mathbb{R}^d} \max_{r \in [0,T]} \|\gamma_x''(r)\|_2^2 \, d\mu_0(x) \right)^{1/2} < \infty.$$

*Then under sufficient Lipschitz conditions on $F$, the Euler's method approximations $\mu_{(n)}$ satisfy*

$$W_2(\mu_{(n)}, \mu_{t_n}) \leq \frac{HM}{2(L_1 + L_2)} \left( e^{(t_n - t_0)(L_1 + L_2)} - 1 \right),$$

*where $L_1, L_2$ are Lipschitz constants.*

# First-Order Global Error Bound

**Theorem (K., Nikitopoulos, Kevrekidis, Lee, Clonigner)**

*(Informal) Suppose*

$$M = \left( \int_{\mathbb{R}^d} \max_{r \in [0,T]} \|\gamma_x''(r)\|_2^2 \, d\mu_0(x) \right)^{1/2} < \infty.$$

*Then under sufficient Lipschitz conditions on $F$, the Euler's method approximations $\mu_{(n)}$ satisfy*

$$W_2(\mu_{(n)}, \mu_{t_n}) \leq \frac{HM}{2(L_1 + L_2)} \left( e^{(t_n - t_0)(L_1 + L_2)} - 1 \right),$$

*where $L_1, L_2$ are Lipschitz constants.*

The proof is very similar to the analogous bound for Euler's method in $\mathbb{R}^d$.

# Overview

# Discrete Algorithm

In the applications of interest, we do not have explicit access to the function $F$, so we must somehow approximate the vector field $F(t_n, \mu_{(n)})$.

# Discrete Algorithm

In the applications of interest, we do not have explicit access to the function $F$, so we must somehow approximate the vector field $F(t_n, \mu_{(n)})$.

Let $\mu_t^N$ be the uniform distribution on the locations of the particles at time $t$, and $\mu_{t+h}^N = S_h(\mu_t^N)$.

# Discrete Algorithm

In the applications of interest, we do not have explicit access to the function $F$, so we must somehow approximate the vector field $F(t_n, \mu_{(n)})$.

Let $\mu_t^N$ be the uniform distribution on the locations of the particles at time $t$, and $\mu_{t+h}^N = S_h(\mu_t^N)$. We assume $(\mu_t^N)$ is an empirical estimate of some absolutely continuous curve $(\mu_t)$.

# Discrete Algorithm

In the applications of interest, we do not have explicit access to the function $F$, so we must somehow approximate the vector field $F(t_n, \mu_{(n)})$.

Let $\mu_t^N$ be the uniform distribution on the locations of the particles at time $t$, and $\mu_{t+h}^N = S_h(\mu_t^N)$. We assume $(\mu_t^N)$ is an empirical estimate of some absolutely continuous curve $(\mu_t)$.

To approximate $\mathbf{v}_t$, we can compute

$$\mathbf{v}_t^{h,N}(x_i) = \frac{T_{\mu_t^N}^{\mu_{t+h}^N}(x_i) - x_i}{h}$$

for each $x_i \in \mathrm{supp}(\mu_t^N)$.

# Discrete Algorithm

In the applications of interest, we do not have explicit access to the function $F$, so we must somehow approximate the vector field $F(t_n, \mu_{(n)})$.

Let $\mu_t^N$ be the uniform distribution on the locations of the particles at time $t$, and $\mu_{t+h}^N = S_h(\mu_t^N)$. We assume $(\mu_t^N)$ is an empirical estimate of some absolutely continuous curve $(\mu_t)$.

To approximate $\mathbf{v}_t$, we can compute

$$\mathbf{v}_t^{h,N}(x_i) = \frac{T_{\mu_t^N}^{\mu_{t+h}^N}(x_i) - x_i}{h}$$

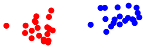
for each $x_i \in \mathrm{supp}(\mu_t^N)$. We can then approximate

$$\mu_{t+H}^N \approx (\mathrm{id} + H\mathbf{v}_t^{h,N})_{\#}\mu_t^N.$$

# Discrete Algorithm Example

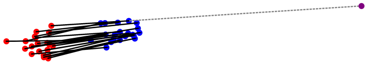
# Discrete Algorithm Example

# Discrete Algorithm Example

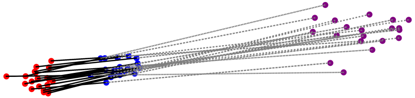


# Discrete Algorithm Example

# Discrete Algorithm Example

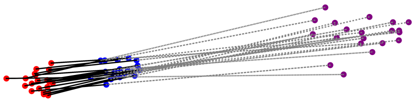


# Discrete Algorithm Example



# Discrete Algorithm Example

Since we have a microscale simulator, we can reinitialize the simulation and repeat.



# Discrete Algorithm Example

Since we have a microscale simulator, we can reinitialize the simulation and repeat.

# Discrete Algorithm Example

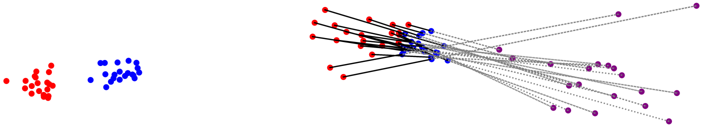
Since we have a microscale simulator, we can reinitialize the simulation and repeat.

# Discrete Algorithm Example

Since we have a microscale simulator, we can reinitialize the simulation and repeat.

# Discrete Algorithm Example

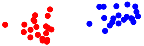
Since we have a microscale simulator, we can reinitialize the simulation and repeat.

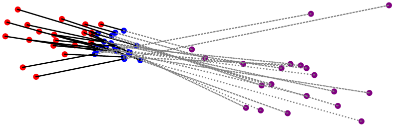


# Discrete Algorithm Example

Since we have a microscale simulator, we can reinitialize the simulation and repeat.

But we must “burn-in” first!





# Discrete Algorithm Example

Since we have a microscale simulator, we can reinitialize the simulation and repeat.

But we must “burn-in” first!

# Discrete Algorithm Example

Since we have a microscale simulator, we can reinitialize the simulation and repeat.

But we must “burn-in” first!

# Discrete Algorithm Example

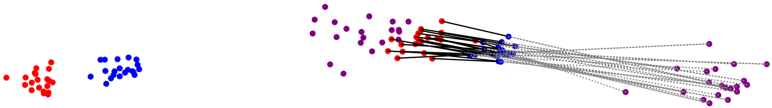
Since we have a microscale simulator, we can reinitialize the simulation and repeat.

But we must “burn-in” first!

# Discrete Algorithm Example

Since we have a microscale simulator, we can reinitialize the simulation and repeat.

But we must “burn-in” first!



# Experimental Results — 2D Diffusion

We consider the toy example where $S_h$ is an Euler-Maruyama numerical simulation of Brownian motion:

$$x_i(t_{j+1}) = x_i(t_j) + \lambda Z_i(t_j)$$

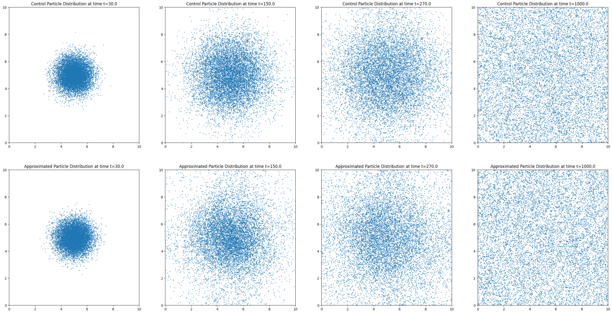
where $Z_i(t_j) \sim \mathcal{N}(0, hI_2)$ are i.i.d.

# Experimental Results — 2D Diffusion

We consider the toy example where $S_h$ is an Euler-Maruyama numerical simulation of Brownian motion:

$$x_i(t_{j+1}) = x_i(t_j) + \lambda Z_i(t_j)$$

where $Z_i(t_j) \sim \mathcal{N}(0, hI_2)$ are i.i.d.



# PCA of Approximations

If we fix $\sigma \in \mathcal{P}^N(\mathbb{R}^d)$, the LOT embedding

$$\mu_t^N \mapsto T_\sigma^{\mu_t^N}$$

embeds the data into $\mathbb{R}^{N \times d}$.

# PCA of Approximations

If we fix $\sigma \in \mathcal{P}^N(\mathbb{R}^d)$, the LOT embedding

$$\mu_t^N \mapsto T_\sigma^{\mu_t^N}$$

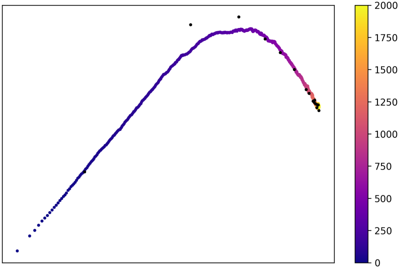
embeds the data into $\mathbb{R}^{N \times d}$. We can then perform PCA on the embedded data to visualize the distribution's evolution, and we can compare our approximations to the control data.

# PCA of Approximations

If we fix $\sigma \in \mathcal{P}^N(\mathbb{R}^d)$, the LOT embedding

$$\mu_t^N \mapsto T_\sigma^{\mu_t^N}$$

embeds the data into $\mathbb{R}^{N\times d}$. We can then perform PCA on the embedded data to visualize the distribution's evolution, and we can compare our approximations to the control data.



# Experimental Results — Cell Chemotaxis

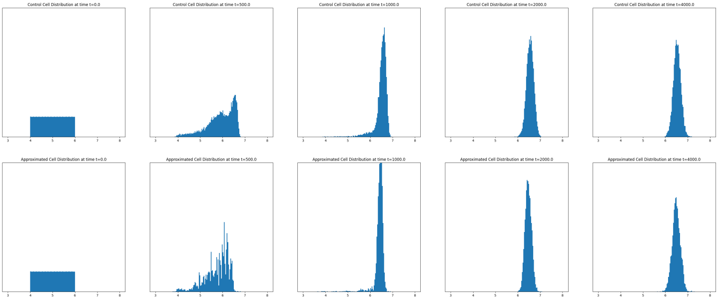
We perform the same experiments with $S_h$ a cell chemotaxis model, described in [3] (computationally expensive).

[3] Setayeshgar, Gear, Othmer, Kevrekidis. Application of Coarse Integration to Bacterial Chemotaxis (2005)

# Experimental Results — Cell Chemotaxis

We perform the same experiments with $S_h$ a cell chemotaxis model, described in [3] (computationally expensive).

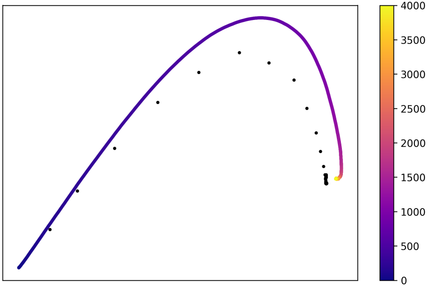
[3] Setayeshgar, Gear, Othmer, Kevrekidis. Application of Coarse Integration to Bacterial Chemotaxis (2005)



# Experimental Results — Cell Chemotaxis

We perform the same experiments with $S_h$ a cell chemotaxis model, described in [3] (computationally expensive).

[3] Setayeshgar, Gear, Othmer, Kevrekidis. Application of Coarse Integration to Bacterial Chemotaxis (2005)



# Overview

# Choosing an optimal $h$

# Choosing an optimal $h$

We want

$$\mathbf{v}_t \approx \mathbf{v}_t^{h,N} = \frac{T_{\mu_t^N}^{\mu_{t+h}^N} - \mathrm{id}}{h}.$$

# Choosing an optimal $h$

We want

$$\mathbf{v}_t \approx \mathbf{v}_t^{h,N} = \frac{T_{\mu_t^N}^{\mu_{t+h}^N} - \mathrm{id}}{h}.$$

Why not take $h \to 0$?

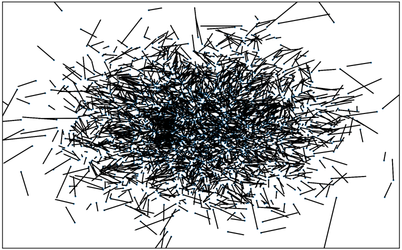
# Choosing an optimal $h$

We want

$$\mathbf{v}_t \approx \mathbf{v}_t^{h,N} = \frac{T_{\mu_t^N}^{\mu_{t+h}^N} - \mathrm{id}}{h}.$$

Why not take $h \to 0$?

h too small



# Velocity Flow Field for Gaussian Diffusion

For standard Brownian motion $dX_t = dW_t$, we have

$$\mathrm{Law}(X_t) = \mu_t = \mathcal{N}(0, tI),$$

# Velocity Flow Field for Gaussian Diffusion

For standard Brownian motion $dX_t = dW_t$, we have

$$\mathrm{Law}(X_t) = \mu_t = \mathcal{N}(0, tI),$$

and we can compute the velocity field explicitly (optimal transport between Gaussians has a closed form)

$$\mathbf{v}_t(x) = \frac{x}{2t}.$$

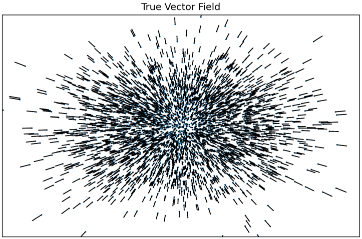
# Velocity Flow Field for Gaussian Diffusion

For standard Brownian motion $dX_t = dW_t$, we have

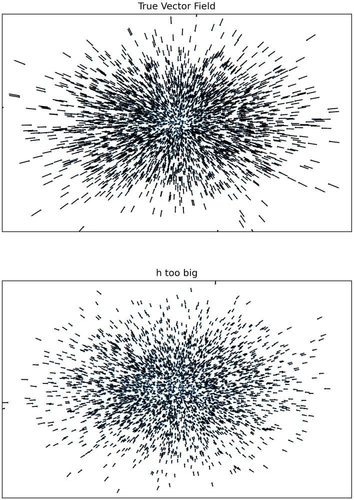
$$\text{Law}(X_t) = \mu_t = \mathcal{N}(0, tI),$$

and we can compute the velocity field explicitly (optimal transport between Gaussians has a closed form)

$$\mathbf{v}_t(x) = \frac{x}{2t}.$$



# Choosing an optimal $h$



# Choosing an optimal $h$

For different values of $h, N$, we can compute

$$\|\mathbf{v}_t^{h,N} - \mathbf{v}_t\|^2_{L^2(\mu_t^N)} = \frac{1}{N}\sum_{i=1}^{N} \left\|\mathbf{v}_t^{h,N}(x_i) - \mathbf{v}_t(x_i)\right\|_2^2 = \frac{1}{N}\sum_{i=1}^{N} \left|\mathbf{v}_t^{h,N}(x_i) - \frac{x_i}{2t}\right|^2.$$

# Choosing an optimal $h$

For different values of $h, N$, we can compute

$$\|\mathbf{v}_t^{h,N} - \mathbf{v}_t\|^2_{L^2(\mu_t^N)} = \frac{1}{N}\sum_{i=1}^{N}\left\|\mathbf{v}_t^{h,N}(x_i) - \mathbf{v}_t(x_i)\right\|_2^2 = \frac{1}{N}\sum_{i=1}^{N}\left|\mathbf{v}_t^{h,N}(x_i) - \frac{x_i}{2t}\right|^2.$$

We can also compute the true finite difference

$$\mathbf{v}_t^h(x) = \frac{T_{\mu_t}^{\mu_{t+h}}(x) - x}{h} = \frac{x(\sqrt{\frac{t+h}{t}} - 1)}{h}$$

# Choosing an optimal $h$

For different values of $h, N$, we can compute

$$\|\mathbf{v}_t^{h,N} - \mathbf{v}_t\|^2_{L^2(\mu_t^N)} = \frac{1}{N}\sum_{i=1}^{N}\left\|\mathbf{v}_t^{h,N}(x_i) - \mathbf{v}_t(x_i)\right\|_2^2 = \frac{1}{N}\sum_{i=1}^{N}\left|\mathbf{v}_t^{h,N}(x_i) - \frac{x_i}{2t}\right|^2.$$

We can also compute the true finite difference

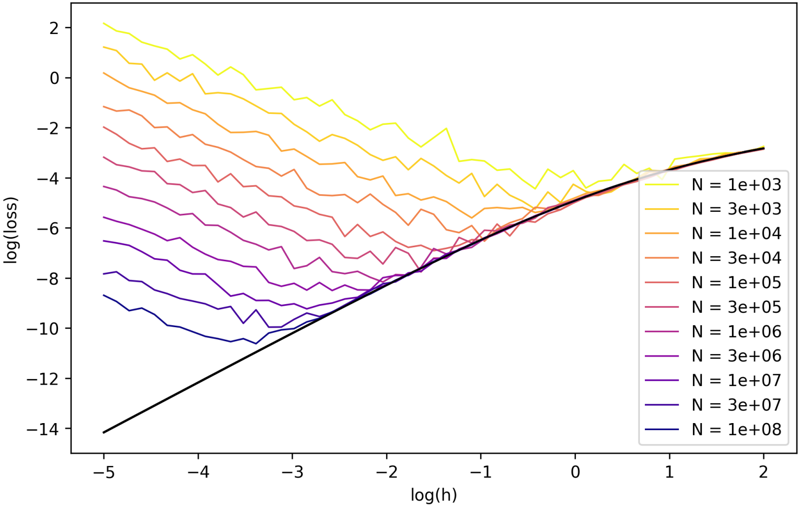
$$\mathbf{v}_t^h(x) = \frac{T_{\mu_t}^{\mu_{t+h}}(x) - x}{h} = \frac{x(\sqrt{\frac{t+h}{t}} - 1)}{h}$$

and then

$$\|\mathbf{v}_t^h - \mathbf{v}_t\|^2_{L^2(\mu_t^N)} = \frac{1}{N}\sum_{i=1}^{N}\left\|\frac{x_i(\sqrt{\frac{t+h}{t}} - 1)}{h} - \frac{x_i}{2t}\right\|_2^2$$

so that we have a benchmark for how well our discrete approximations can hope to do.

# Choosing an optimal $h$



# Estimating the Vector Field Error

Of course, we don't often know the true vector field $\mathbf{v}_t$ in practice.

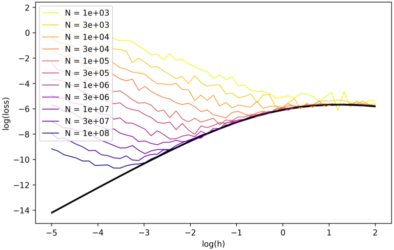
# Estimating the Vector Field Error

Of course, we don't often know the true vector field $\mathbf{v}_t$ in practice. How do we estimate $\|\mathbf{v}_t^{h,N} - \mathbf{v}_t\|^2_{L^2(\mu_t^N)}$ without it?

# Estimating the Vector Field Error

Of course, we don't often know the true vector field $\mathbf{v}_t$ in practice. How do we estimate $\|\mathbf{v}_t^{h,N} - \mathbf{v}_t\|^2_{L^2(\mu_t^N)}$ without it?

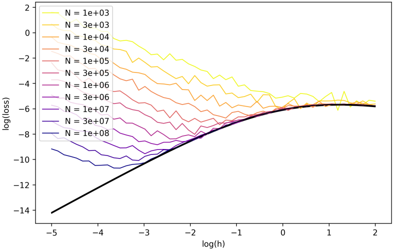
Idea: compute $\|\mathbf{v}_t^{2h,N} - \mathbf{v}_t^{h,N}\|^2_{L^2(\mu_t^N)}$.

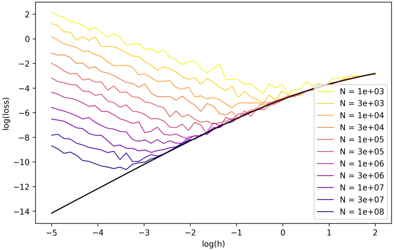


# Estimating the Vector Field Error

Of course, we don't often know the true vector field $\mathbf{v}_t$ in practice. How do we estimate $\|\mathbf{v}_t^{h,N} - \mathbf{v}_t\|^2_{L^2(\mu_t^N)}$ without it?

Idea: compute $\|\mathbf{v}_t^{2h,N} - \mathbf{v}_t^{h,N}\|^2_{L^2(\mu_t^N)}$.



# Discrete Convergence Analog

We want some theoretical result that guarantees accuracy of the approximation algorithm.

# Discrete Convergence Analog

We want some theoretical result that guarantees accuracy of the approximation algorithm.

There are results that show $T_{\sigma^N}^{\mu^N} \to T_{\sigma}^{\mu}$ as $N \to \infty$, where $\mu^N, \sigma^N$ are independent samples of $\mu, \sigma$ respectively (e.g., [4]).

# Discrete Convergence Analog

We want some theoretical result that guarantees accuracy of the approximation algorithm.

There are results that show $T_{\sigma^N}^{\mu^N} \to T_\sigma^\mu$ as $N \to \infty$, where $\mu^N, \sigma^N$ are independent samples of $\mu, \sigma$ respectively (e.g., [4]).

But there are two theoretical hurdles for us:

# Discrete Convergence Analog

We want some theoretical result that guarantees accuracy of the approximation algorithm.

There are results that show $T_{\sigma^N}^{\mu^N} \to T_{\sigma}^{\mu}$ as $N \to \infty$, where $\mu^N, \sigma^N$ are independent samples of $\mu, \sigma$ respectively (e.g., [4]).

But there are two theoretical hurdles for us:

1. $\mu_t^N$ and $\mu_{t+h}^N$ are not independent

# Discrete Convergence Analog

We want some theoretical result that guarantees accuracy of the approximation algorithm.

There are results that show $T_{\sigma^N}^{\mu^N} \to T_{\sigma}^{\mu}$ as $N \to \infty$, where $\mu^N, \sigma^N$ are independent samples of $\mu, \sigma$ respectively (e.g., [4]).

But there are two theoretical hurdles for us:

1. $\mu_t^N$ and $\mu_{t+h}^N$ are not independent
2. The behavior on the last slide happens even for independent samples of $\mu_t$ and $\mu_{t+h}$.

[4] Seguy, et al. Large-Scale Optimal Transport and Mapping Estimation

# Discrete Convergence Analog

Loose idea:

$$\|\mathbf{v}_t^{h,N} - \mathbf{v}_t\|_{L^2(\mu_t^N)} \leq \|\mathbf{v}_t^{h,N} - \mathbf{v}_t^h\|_{L^2(\mu_t^N)} + \|\mathbf{v}_t^h - \mathbf{v}_t\|_{L_2(\mu_t^N)}$$

and $\|\mathbf{v}_t^h - \mathbf{v}_t\|_{L_2(\mu_t^N)} = o(h)$,

# Discrete Convergence Analog

Loose idea:

$$\|\mathbf{v}_t^{h,N} - \mathbf{v}_t\|_{L^2(\mu_t^N)} \leq \|\mathbf{v}_t^{h,N} - \mathbf{v}_t^h\|_{L^2(\mu_t^N)} + \|\mathbf{v}_t^h - \mathbf{v}_t\|_{L_2(\mu_t^N)}$$

and $\|\mathbf{v}_t^h - \mathbf{v}_t\|_{L_2(\mu_t^N)} = o(h)$, so error depends on

$$\|\mathbf{v}_t^{h,N} - \mathbf{v}_t^h\|_{L^2(\mu_t^N)} = \frac{1}{h}\| T_{\mu_t^N}^{\mu_{t+h}^N} - T_{\mu_t}^{\mu_{t+h}}\|_{L^2(\mu_t^N)},$$

so need to understand how $T_{\sigma^N}^{\mu^N} \to T_\sigma^\mu$ as $N \to \infty$ and $\mu \to \sigma$ together.

# Higher Order Approximations

The approximation

$$\mu_{t+H} \approx (\mathrm{id} + H\mathbf{v}_t)_{\#}\mu_t$$

is the analog of Euler's method in $\mathbb{R}^d$.

# Higher Order Approximations

The approximation

$$\mu_{t+H} \approx (\mathrm{id} + H\mathbf{v}_t)_{\#}\mu_t$$

is the analog of Euler's method in $\mathbb{R}^d$.

Q: What is the second-order (and higher-orders) equivalent on the Wasserstein manifold?

# Higher Order Approximations

The approximation

$$\mu_{t+H} \approx (\mathrm{id} + H\mathbf{v}_t)_{\#}\mu_t$$

is the analog of Euler's method in $\mathbb{R}^d$.

Q: What is the second-order (and higher-orders) equivalent on the Wasserstein manifold?

Naturally leads to a question about a general Taylor's Theorem on the Wasserstein manifold.

# Newton's Method for Finding Steady States

If we want to find the steady state distribution, then we want to find a measure $\mu^*$ such that the governing vector field

$$F(t, \mu^*) = \mathbf{0}.$$

# Newton's Method for Finding Steady States

If we want to find the steady state distribution, then we want to find a measure $\mu^*$ such that the governing vector field

$$F(t, \mu^*) = \mathbf{0}.$$

For simplicity, assume the system is autonomous, i.e.,

$$F(t, \mu) = F(\mu).$$

# Newton's Method for Finding Steady States

If we want to find the steady state distribution, then we want to find a measure $\mu^*$ such that the governing vector field

$$F(t, \mu^*) = \mathbf{0}.$$

For simplicity, assume the system is autonomous, i.e.,

$$F(t, \mu) = F(\mu).$$

With a Taylor's Theorem, we could develop an analog of Newton's method to approximate the solution to $F(\mu) = \mathbf{0}$.

# Newton's Method for Finding Steady States

If we want to find the steady state distribution, then we want to find a measure $\mu^*$ such that the governing vector field

$$F(t, \mu^*) = \mathbf{0}.$$

For simplicity, assume the system is autonomous, i.e.,

$$F(t, \mu) = F(\mu).$$

With a Taylor's Theorem, we could develop an analog of Newton's method to approximate the solution to $F(\mu) = \mathbf{0}$.

How can we use this to quickly find steady states of particle systems governed by a microscale simulator?

# Newton's Method for Finding Steady States

If we want to find the steady state distribution, then we want to find a measure $\mu^*$ such that the governing vector field

$$F(t, \mu^*) = \mathbf{0}.$$

For simplicity, assume the system is autonomous, i.e.,

$$F(t, \mu) = F(\mu).$$

With a Taylor's Theorem, we could develop an analog of Newton's method to approximate the solution to $F(\mu) = \mathbf{0}$.

How can we use this to quickly find steady states of particle systems governed by a microscale simulator?

[5] Kevrekidis, et al. Equation-Free, Coarse-Grained Multiscale Computation: Enabling Microscopic Simulators to Perform System-Level Analysis (2003)
[6] Kevrekidis, et al. Equation-free: The computer-aided analysis of complex multiscale systems (2004)

# Diffusion Models

Forward Diffusion Process:

$$q(x_{t+1} \mid x_t) \sim \mathcal{N}(\sqrt{\alpha_{t+1}}x_t, \sqrt{1-\alpha_{t+1}}I).$$

# Diffusion Models

Forward Diffusion Process:

$$q(x_{t+1} \mid x_t) \sim \mathcal{N}(\sqrt{\alpha_{t+1}}x_t, \sqrt{1-\alpha_{t+1}}I).$$

Reverse Diffusion Process: want to approximate $q(x_t \mid x_{t+1})$.

# Diffusion Models

Forward Diffusion Process:

$$q(x_{t+1} \mid x_t) \sim \mathcal{N}(\sqrt{\alpha_{t+1}}x_t, \sqrt{1-\alpha_{t+1}}I).$$

Reverse Diffusion Process: want to approximate $q(x_t \mid x_{t+1})$.

LOT gives a way to approximate the vector field describing the evolution of the data in the forward process, and this can easily be reversed by negating the vector field.

# Weighted Wasserstein Barycenters

Motivated by applications to hyperspectral imaging, we are interested in the following problem:

# Weighted Wasserstein Barycenters

Motivated by applications to hyperspectral imaging, we are interested in the following problem:

Given distributions $\{\mu_k\}_{k=1}^{K}$ and some distribution $\nu$, we want to find weights $w_k$ that sum to 1 such that the $w$-weighted Wasserstein barycenter best approximates $\nu$,

# Weighted Wasserstein Barycenters

Motivated by applications to hyperspectral imaging, we are interested in the following problem:

Given distributions $\{\mu_k\}_{k=1}^{K}$ and some distribution $\nu$, we want to find weights $w_k$ that sum to 1 such that the $w$-weighted Wasserstein barycenter best approximates $\nu$, i.e., we want

$$\nu \approx \operatorname*{argmin}_{\mu} \sum_{k=1}^{K} w_k W_2(\mu_k, \mu)^2.$$

# Weighted Wasserstein Barycenters

Motivated by applications to hyperspectral imaging, we are interested in the following problem:

Given distributions $\{\mu_k\}_{k=1}^K$ and some distribution $\nu$, we want to find weights $w_k$ that sum to 1 such that the $w$-weighted Wasserstein barycenter best approximates $\nu$, i.e., we want

$$\nu \approx \operatorname*{argmin}_{\mu} \sum_{k=1}^{K} w_k W_2(\mu_k, \mu)^2.$$

Wasserstein barycenters are hard to compute, but LOT barycenters are easy to compute: if $\sigma$ is fixed, then the $w$-weighted LOT barycenter is

$$\left( \sum_{k=1}^{K} w_k T_{\sigma}^{\mu_k} \right)_{\#} \sigma.$$

# Thanks!

Questions?

# Precise Statement of Global Error

Suppose $F : \mathbb{R} \times \mathcal{P}_2(\mathbb{R}^d) \to (\mathbb{R}^d)^{\mathbb{R}^d}$ is a map such that, for all times $t \in \mathbb{R}$ and all measures $\nu \in \mathcal{P}_2(\mathbb{R}^d)$, the vector field $F(t, \nu)$ is in $T_\nu \mathbb{W}_2$. Fix a measure $\mu_0 \in \mathcal{P}_2(\mathbb{R}^d)$, and suppose there is a narrowly continuous curve $\mu : [0, T] \to \mathcal{P}_2(\mathbb{R}^d)$ given by $t \mapsto \mu_t$ that satisfies the continuity equation

$$\partial_t \mu_t + \nabla \cdot (\mu_t \mathbf{v}_t) = 0$$

in the sense of distributions, where $\mathbf{v}_t = F(t, \mu_t)$. Moreover, suppose $\mathbf{v}(t, x) = F(t, \mu_t)(x)$ is a $C^1$ vector field such that, for $\mu_0$-a.e. $x \in \mathbb{R}^d$, the ODE

$$\gamma_x(0) = x, \qquad \frac{d}{dt}\gamma_x(t) = \mathbf{v}_t(\gamma_x(t))$$

admits a unique solution on $[0, T]$.

# Precise Statement of Global Error (ctd)

Suppose

$$M = \left( \int_{\mathbb{R}^d} \max_{r \in [0,T]} \|\gamma_x''(r)\|_2^2 \, d\mu_0(x) \right)^{1/2} < \infty.$$

Also suppose $F(t,\nu)$ is Lipschitz for all $t \in [0,T], \nu \in \mathcal{P}_2(\mathbb{R}^d)$, and that $L_2$ is an upper bound for all such Lipschitz constants, i.e.,

$$\|F(t,\nu)(x_1) - F(t,\nu)(x_2)\|_2 \leq L_2 \|x_1 - x_2\|_2$$

for all $t \in [0,T], \nu \in \mathcal{P}_2(\mathbb{R}^d)$, and $x_1, x_2 \in \mathbb{R}^d$. Finally, suppose $F$ is itself Lipschitz with constant $L_1$ in the sense that

$$\|F(t,\nu_1) - F(t,\nu_2)\|_{L^2(\nu_1)} \leq L_1 W_2(\nu_1,\nu_2)$$

for all $\nu_1, \nu_2 \in \mathcal{P}_2(\mathbb{R}^d)$. Then for a time step $H > 0$ and discrete times $t_n = nH$, the sequence of Euler approximations iteratively defined by

$$\mu_{(0)} = \mu_0, \qquad \mu_{(n+1)} = (\mathrm{id} + HF(t_n, \mu_{(n)}))_{\#}\mu_{(n)}$$

satisfies

$$W_2(\mu_{(n)}, \mu_{t_n}) \leq \frac{HM}{2(L_1+L_2)} \left( e^{(t_n - t_0)(L_1+L_2)} - 1 \right).$$