Accelerated Gradient Flow for Probability Distributions

CSL Student Conference, UIUC, February, 2019

Amirhossein Taghvaei
Joint work with P. G. Mehta

Coordinated Science Laboratory
University of Illinois at Urbana-Champaign

Feb 7, 2019

ILLINOIS

Motivation and objective

- Many machine learning problems are modelled as an optimization problem on the space of probability distributions
  - Bayesian inference
  - Learning generative models
  - Policy optimization in reinforcement learning

- Solution approaches by constructing gradient flows for probability distributions
  - Liu & Wang, 2016. *"Stein variational gradient descent"*
  - Zhang, et. al. 2018. *"Policy optimization as wasserstein gradient flows"*
  - Frogner & Poggio, 2018. *"Approximate inference with wasserstein gradient flows"*
  - Chizat & Bach, 2018. *"On the global convergence of gradient descent for over-parameterized models using optimal transport"*

- **This talk:** Construct accelerated gradient flows for probability distribution

Motivation and objective

- Many machine learning problems are modelled as an optimization problem on the space of probability distributions
  - Bayesian inference
  - Learning generative models
  - Policy optimization in reinforcement learning
- Solution approaches by constructing gradient flows for probability distributions
  - Liu & Wang, 2016. *"Stein variational gradient descent"*
  - Zhang, et. al. 2018. *"Policy optimization as wasserstein gradient flows"*
  - Frogner & Poggio, 2018. *"Approximate inference with wasserstein gradient flows"*
  - Chizat & Bach, 2018. *"On the global convergence of gradient descent for over-parameterized models using optimal transport"*
- **This talk:** Construct accelerated gradient flows for probability distribution

Motivation and objective

- Many machine learning problems are modelled as an optimization problem on the space of probability distributions
  - Bayesian inference
  - Learning generative models
  - Policy optimization in reinforcement learning

- Solution approaches by constructing gradient flows for probability distributions
  - Liu & Wang, 2016. *"Stein variational gradient descent"*
  - Zhang, et. al. 2018. *"Policy optimization as wasserstein gradient flows"*
  - Frogner & Poggio, 2018. *"Approximate inference with wasserstein gradient flows"*
  - Chizat & Bach, 2018. *"On the global convergence of gradient descent for over-parameterized models using optimal transport"*

- **This talk:** Construct accelerated gradient flows for probability distribution

Our approach and main idea

Euclidean space	Space of probability distributions
Gradient descent	Wasserstein gradient flow
Accelerated methods	?

(1) Variational formulation on accelerated methods in optimization (Wibisono, et. al. 2017).

(2) Riemannian geometry for probability distributions from optimal transportation theory (Jordan, et. al. 1998) (Ambrosio, et. al. 2008)

(3) Extend (1) using (2) to formulate a variational from for probability distributions that produces accelerated flows

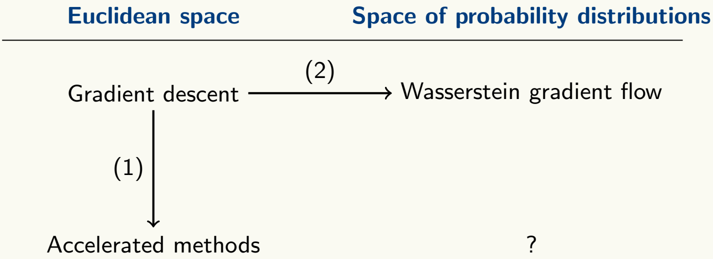
Our approach and main idea

Euclidean space	**Space of probability distributions**
Gradient descent	Wasserstein gradient flow
↓ (1)	
Accelerated methods	?

(1) Variational formulation on accelerated methods in optimization (Wibisono, et. al. 2017).

(2) Riemannian geometry for probability distributions from optimal transportation theory (Jordan, et. al. 1998) (Ambrosio, et. al. 2008)

(3) Extend (1) using (2) to formulate a variational from for probability distributions that produces accelerated flows

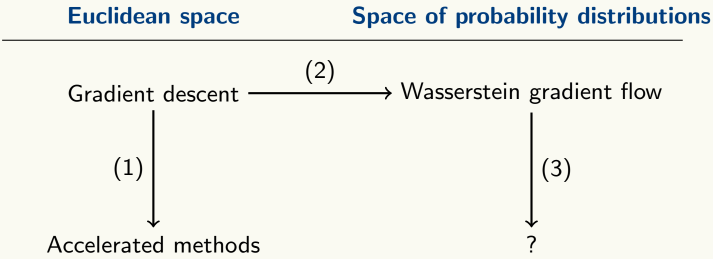


(1) Variational formulation on accelerated methods in optimization (Wibisono, et. al. 2017).

(2) Riemannian geometry for probability distributions from optimal transportation theory (Jordan, et. al. 1998) (Ambrosio, et. al. 2008)

(3) Extend (1) using (2) to formulate a variational from for probability distributions that produces accelerated flows

Our approach and main idea



(1) Variational formulation on accelerated methods in optimization (Wibisono, et. al. 2017).

(2) Riemannian geometry for probability distributions from optimal transportation theory (Jordan, et. al. 1998) (Ambrosio, et. al. 2008)

(3) Extend (1) using (2) to formulate a variational from for probability distributions that produces accelerated flows

Review of variational formulation of (Wibinoso, et. al. 2016)

- Optimization problem:

$$\min_{x\in\mathbb{R}^d} \; f(x) \quad \text{(Assume } f \text{ is convex)}$$

- Gradient flow:

$$\dot{x}_t = -\nabla f(x_t) \quad \Longrightarrow \quad f(x_t) - f(\bar{x}) \leq O(\frac{1}{t})$$

- Accelerated gradient flow (Su, et. al. 2014):

$$\ddot{x}_t = -\frac{3}{t}\dot{x}_t - \nabla f(x_t) \quad \Longrightarrow \quad f(x_t) - f(\bar{x}) \leq O(\frac{1}{t^2})$$

- $\{x_t\}$ is the solution to the following variational problem (Wibinoso, et. al. 2016):

$$\text{Minimize:} \quad \int_0^\infty t^3(\frac{1}{2}|u_t|^2 - f(x_t))\,\mathrm{d}t$$

$$\text{Subject to:} \quad \frac{\mathrm{d}x_t}{\mathrm{d}t} = u_t, \quad x_0 = x,\; \dot{x}_0 = v$$

Review of variational formulation of (Wibinoso, et. al. 2016)

- Optimization problem:

$$\min_{x \in \mathbb{R}^d} \; f(x) \quad \text{(Assume } f \text{ is convex)}$$

- Gradient flow:

$$\dot{x}_t = -\nabla f(x_t) \quad \Longrightarrow \quad f(x_t) - f(\bar{x}) \leq O(\frac{1}{t})$$

- Accelerated gradient flow (Su, et. al. 2014):

$$\ddot{x}_t = -\frac{3}{t}\dot{x}_t - \nabla f(x_t) \quad \Longrightarrow \quad f(x_t) - f(\bar{x}) \leq O(\frac{1}{t^2})$$

- $\{x_t\}$ is the solution to the following variational problem (Wibinoso, et. al. 2016):

$$\text{Minimize:} \quad \int_0^\infty t^3 (\frac{1}{2}|u_t|^2 - f(x_t))\,\mathrm{d}t$$

$$\text{Subject to:} \quad \frac{\mathrm{d}x_t}{\mathrm{d}t} = u_t, \quad x_0 = x, \; \dot{x}_0 = v$$

Review of variational formulation of (Wibinoso, et. al. 2016)

- Optimization problem:

$$\min_{x\in\mathbb{R}^d} \; f(x) \quad \text{(Assume } f \text{ is convex)}$$

- Gradient flow:

$$\dot{x}_t = -\nabla f(x_t) \quad \Longrightarrow \quad f(x_t) - f(\bar{x}) \leq O(\frac{1}{t})$$

- Accelerated gradient flow (Su, et. al. 2014):

$$\ddot{x}_t = -\frac{3}{t}\dot{x}_t - \nabla f(x_t) \quad \Longrightarrow \quad f(x_t) - f(\bar{x}) \leq O(\frac{1}{t^2})$$

- $\{x_t\}$ is the solution to the following variational problem (Wibinoso, et. al. 2016):

$$\text{Minimize:} \quad \int_0^\infty t^3(\frac{1}{2}|u_t|^2 - f(x_t))\,\mathrm{d}t$$

$$\text{Subject to:} \quad \frac{\mathrm{d}x_t}{\mathrm{d}t} = u_t, \quad x_0 = x,\; \dot{x}_0 = v$$

Review of variational formulation of (Wibinoso, et. al. 2016)

- Optimization problem:

$$\min_{x\in\mathbb{R}^d} \; f(x) \quad \text{(Assume } f \text{ is convex)}$$

- Gradient flow:

$$\dot{x}_t = -\nabla f(x_t) \quad \Longrightarrow \quad f(x_t) - f(\bar{x}) \leq O(\frac{1}{t})$$

- Accelerated gradient flow (Su, et. al. 2014):

$$\ddot{x}_t = -\frac{3}{t}\dot{x}_t - \nabla f(x_t) \quad \Longrightarrow \quad f(x_t) - f(\bar{x}) \leq O(\frac{1}{t^2})$$

- $\{x_t\}$ is the solution to the following variational problem (Wibinoso, et. al. 2016):

$$\text{Minimize:} \quad \int_0^\infty t^3 (\frac{1}{2}|u_t|^2 - f(x_t))\,\mathrm{d}t$$

$$\text{Subject to:} \quad \frac{\mathrm{d}x_t}{\mathrm{d}t} = u_t, \quad x_0 = x, \; \dot{x}_0 = v$$

Summary

	vector variables $\mathbb{R}^d$	probability distribution $\mathcal{P}_2(\mathbb{R}^d)$
Objective funct.	$f(x)$	?
Gradient flow	$\dot{x}_t = -\nabla f(x_t)$	?
Lagrangian	$t^3(\frac{1}{2}\lvert u_t\rvert^2 - f(x_t))$	?
Accelerated flow	$\ddot{x}_t = -\frac{3}{t}\dot{x}_t - \nabla f(x_t)$	?

Background: Wasserstein gradient

- Objective functional:

$$\mathsf{F} : \mathcal{P}_2(\mathbb{R}^d) \to \mathbb{R}$$

- Wasserstein gradient: $\nabla_W \mathsf{F}(\rho) : \mathbb{R}^d \to \mathbb{R}^d$ is a vector field that satisfies

$$\frac{\mathrm{d}}{\mathrm{d}t}\mathsf{F}(\rho_t)\Big|_{t=0} = \langle \nabla_W \mathsf{F}(\rho), u \rangle_{L^2(\rho)},$$

$$\text{for all path } \{\rho_t\} \text{ s.t } \quad \frac{\partial \rho_t}{\partial t} + \nabla \cdot (\rho_t u) = 0$$

- Example:

$$\mathsf{F}(\rho) = \mathsf{D}(\rho \| \rho_\infty) \quad \text{(relative entropy)}$$

$$\Longrightarrow \quad \nabla_W \mathsf{F}(\rho)(x) = \nabla \log(\rho(x)) + \nabla f(x)$$

where $f = -\log(\rho_\infty)$

Background: Wasserstein gradient

- Objective functional:

$$\mathsf{F} : \mathcal{P}_2(\mathbb{R}^d) \to \mathbb{R}$$

- Wasserstein gradient: $\nabla_W \mathsf{F}(\rho) : \mathbb{R}^d \to \mathbb{R}^d$ is a vector field that satisfies

$$\frac{\mathrm{d}}{\mathrm{d}t}\mathsf{F}(\rho_t)\Big|_{t=0} = \langle \nabla_W \mathsf{F}(\rho), u \rangle_{L^2(\rho)},$$

$$\text{for all path } \{\rho_t\} \text{ s.t } \quad \frac{\partial \rho_t}{\partial t} + \nabla \cdot (\rho_t u) = 0$$

- Example:

$$\mathsf{F}(\rho) = \mathsf{D}(\rho \| \rho_\infty) \quad \text{(relative entropy)}$$

$$\implies \quad \nabla_W \mathsf{F}(\rho)(x) = \nabla \log(\rho(x)) + \nabla f(x)$$

where $f = -\log(\rho_\infty)$

Background: Wasserstein gradient

- Objective functional:

$$\mathsf{F} : \mathcal{P}_2(\mathbb{R}^d) \to \mathbb{R}$$

- Wasserstein gradient: $\nabla_W \mathsf{F}(\rho) : \mathbb{R}^d \to \mathbb{R}^d$ is a vector field that satisfies

$$\frac{\mathrm{d}}{\mathrm{d}t}\mathsf{F}(\rho_t)\Big|_{t=0} = \langle \nabla_W \mathsf{F}(\rho), u \rangle_{L^2(\rho)},$$

$$\text{for all path } \{\rho_t\} \text{ s.t } \quad \frac{\partial \rho_t}{\partial t} + \nabla \cdot (\rho_t u) = 0$$

- Example:

$$\mathsf{F}(\rho) = \mathsf{D}(\rho \| \rho_\infty) \quad \text{(relative entropy)}$$

$$\implies \quad \nabla_W \mathsf{F}(\rho)(x) = \nabla \log(\rho(x)) + \nabla f(x)$$

where $f = -\log(\rho_\infty)$

Background: Wasserstein gradient flow

- Optimization problem:

$$\min_{\rho \in \mathcal{P}_2(\mathbb{R}^d)} \mathsf{F}(\rho)$$

- Wasserstein gradient flow:

$$\frac{\partial \rho_t}{\partial t} = \nabla \cdot (\rho_t \nabla_W \mathsf{F}(\rho_t))$$

- Example: $\mathsf{F}(\rho) = \mathsf{D}(\rho \| \rho_\infty)$, then (Jordan, et. al. 1998)

$$\frac{\partial \rho_t}{\partial t} = \nabla \cdot (\rho_t \nabla f) + \Delta \rho_t \quad \text{(Fokker-Planck eq.)}$$

- Probabilistic form:

$$\mathrm{d}X_t = -\nabla f(X_t)\,\mathrm{d}t + \sqrt{2}\,\mathrm{d}B_t, \quad \text{(Langevin sde)}$$

in the sense that $\rho_t = \mathsf{Law}(X_t)$.

- The goal is to build accelerated version of this sde

Background: Wasserstein gradient flow

- Optimization problem:

$$\min_{\rho \in \mathcal{P}_2(\mathbb{R}^d)} \mathsf{F}(\rho)$$

- Wasserstein gradient flow:

$$\frac{\partial \rho_t}{\partial t} = \nabla \cdot (\rho_t \nabla_W \mathsf{F}(\rho_t))$$

- Example: $\mathsf{F}(\rho) = \mathsf{D}(\rho \| \rho_\infty)$, then (Jordan, et. al. 1998)

$$\frac{\partial \rho_t}{\partial t} = \nabla \cdot (\rho_t \nabla f) + \Delta \rho_t \quad \text{(Fokker-Planck eq.)}$$

- Probabilistic form:

$$\mathrm{d}X_t = -\nabla f(X_t)\,\mathrm{d}t + \sqrt{2}\,\mathrm{d}B_t, \quad \text{(Langevin sde)}$$

  in the sense that $\rho_t = \mathsf{Law}(X_t)$.

- The goal is to build accelerated version of this sde

Background: Wasserstein gradient flow

- Optimization problem:

$$\min_{\rho \in \mathcal{P}_2(\mathbb{R}^d)} \mathsf{F}(\rho)$$

- Wasserstein gradient flow:

$$\frac{\partial \rho_t}{\partial t} = \nabla \cdot (\rho_t \nabla_W \mathsf{F}(\rho_t))$$

- Example: $\mathsf{F}(\rho) = \mathsf{D}(\rho \| \rho_\infty)$, then (Jordan, et. al. 1998)

$$\frac{\partial \rho_t}{\partial t} = \nabla \cdot (\rho_t \nabla f) + \Delta \rho_t \quad \text{(Fokker-Planck eq.)}$$

- Probabilistic form:

$$\mathrm{d}X_t = -\nabla f(X_t)\,\mathrm{d}t + \sqrt{2}\,\mathrm{d}B_t, \quad \text{(Langevin sde)}$$

in the sense that $\rho_t = \mathsf{Law}(X_t)$.

- The goal is to build accelerated version of this sde

Background: Wasserstein gradient flow

- Optimization problem:

$$\min_{\rho \in \mathcal{P}_2(\mathbb{R}^d)} \mathsf{F}(\rho)$$

- Wasserstein gradient flow:

$$\frac{\partial \rho_t}{\partial t} = \nabla \cdot (\rho_t \nabla_W \mathsf{F}(\rho_t))$$

- Example: $\mathsf{F}(\rho) = \mathsf{D}(\rho \| \rho_\infty)$, then (Jordan, et. al. 1998)

$$\frac{\partial \rho_t}{\partial t} = \nabla \cdot (\rho_t \nabla f) + \Delta \rho_t \quad \text{(Fokker-Planck eq.)}$$

- Probabilistic form:

$$\mathrm{d}X_t = -\nabla f(X_t)\,\mathrm{d}t + \sqrt{2}\,\mathrm{d}B_t, \quad \text{(Langevin sde)}$$

in the sense that $\rho_t = \mathsf{Law}(X_t)$.

- The goal is to build accelerated version of this sde

Background: Wasserstein gradient flow

- Optimization problem:

$$\min_{\rho \in \mathcal{P}_2(\mathbb{R}^d)} \; \mathsf{F}(\rho)$$

- Wasserstein gradient flow:

$$\frac{\partial \rho_t}{\partial t} = \nabla \cdot (\rho_t \nabla_W \mathsf{F}(\rho_t))$$

- Example: $\mathsf{F}(\rho) = \mathsf{D}(\rho \| \rho_\infty)$, then (Jordan, et. al. 1998)

$$\frac{\partial \rho_t}{\partial t} = \nabla \cdot (\rho_t \nabla f) + \Delta \rho_t \quad \text{(Fokker-Planck eq.)}$$

- Probabilistic form:

$$\mathrm{d}X_t = -\nabla f(X_t)\,\mathrm{d}t + \sqrt{2}\,\mathrm{d}B_t, \quad \text{(Langevin sde)}$$

  in the sense that $\rho_t = \mathsf{Law}(X_t)$.
- The goal is to build accelerated version of this sde

Summary

	vector variables $\mathbb{R}^d$	probability distribution $\mathcal{P}_2(\mathbb{R}^d)$
Objective funct.	$f(x)$	$\mathsf{F}(\rho) = \mathsf{D}(\rho \| \rho_\infty)$
Gradient flow	$\dot{x}_t = -\nabla f(x_t)$	$\mathrm{d}X_t = -\nabla f(X_t)\,\mathrm{d}t + \sqrt{2}\,\mathrm{d}B_t$
Lagrangian	$t^3(\frac{1}{2}\|u_t\|^2 - f(x_t))$	?
Accelerated flow	$\ddot{x}_t = -\frac{3}{t}\dot{x}_t - \nabla f(x_t)$	?

Proposed variational formulation

- Variational problem (probabilistic form):

$$\text{Minimize} \quad \mathsf{E}\left[\int_0^\infty t^3 \left(\frac{1}{2}|U_t|^2 - \tilde{\mathsf{F}}(\rho_t, X_t)\right) \mathrm{d}t\right]$$

$$\text{Subject to} \quad \frac{\mathrm{d}X_t}{\mathrm{d}t} = U_t, \quad X_0 \sim \rho_0,\ \dot{X}_0 \sim q_0$$

where $\rho_t = \mathsf{Law}(X_t)$.

- It is a mean-field optimal control problem (Bensoussan, et al. 2013, Carmona & Delarue, 2017)

Proposed variational formulation

- Variational problem (probabilistic form):

$$\text{Minimize} \quad \mathsf{E}\left[\int_0^\infty t^3\Big(\frac{1}{2}|U_t|^2 - \tilde{\mathsf{F}}(\rho_t, X_t)\Big)\,\mathrm{d}t\right]$$

$$\text{Subject to} \quad \frac{\mathrm{d}X_t}{\mathrm{d}t} = U_t, \quad X_0 \sim \rho_0,\ \dot{X}_0 \sim q_0$$

where $\rho_t = \mathsf{Law}(X_t)$.

- It is a mean-field optimal control problem (Bensoussan, et al. 2013, Carmona & Delarue, 2017)

Main result

Theorem

Consider the proposed variational problem. Then,

1. (Maximum principle) The optimal trajectory satisfies the second-order system:

$$\ddot{X}_t = -\frac{3}{t}\dot{X}_t - \nabla_W \mathsf{F}(\rho_t)(X_t), \quad X_0 \sim \rho_0$$

   where $\rho_t = \mathsf{Law}(X_t)$.

2. (Convergence) If the functional F is displacement convex, and the dimension $d = 1$. Then

$$\mathsf{F}(\rho_t) - \min_\rho \mathsf{F}(\rho) \leq O(\frac{1}{t^2})$$

We expect the dimension $d = 1$ assumption is not necessary

Main result

Theorem

Consider the proposed variational problem. Then,

1. (Maximum principle) The optimal trajectory satisfies the second-order system:

$$\ddot{X}_t = -\frac{3}{t}\dot{X}_t - \nabla_W \mathsf{F}(\rho_t)(X_t), \quad X_0 \sim \rho_0$$

   where $\rho_t = \mathsf{Law}(X_t)$.

2. (Convergence) If the functional F is displacement convex, and the dimension $d = 1$. Then

$$\mathsf{F}(\rho_t) - \min_{\rho} \mathsf{F}(\rho) \leq O(\frac{1}{t^2})$$

We expect the dimension $d = 1$ assumption is not necessary

Main result

Theorem

Consider the proposed variational problem. Then,

1. (Maximum principle) The optimal trajectory satisfies the second-order system:

$$\ddot{X}_t = -\frac{3}{t}\dot{X}_t - \nabla_W \mathsf{F}(\rho_t)(X_t), \quad X_0 \sim \rho_0$$

   where $\rho_t = \mathsf{Law}(X_t)$.

2. (Convergence) If the functional F is displacement convex, and the dimension $d = 1$. Then

$$\mathsf{F}(\rho_t) - \min_{\rho} \mathsf{F}(\rho) \leq O(\frac{1}{t^2})$$

We expect the dimension $d = 1$ assumption is not necessary

Minimizing relative entropy

- If $\mathsf{F}(\rho) = \mathsf{D}(\rho \| \rho_\infty)$ where $\rho_\infty = e^{-f}$. Then the accelerated flow is

$$\ddot{X}_t = -\frac{3}{t}\dot{X}_t - \nabla f(X_t) - \underbrace{\nabla \log(\rho_t(X_t))}_{\text{mean-field term}}, \quad X_0 \sim \rho_0$$

- $\mathsf{F}(\rho)$ is displacement convex iff $f(x)$ is convex
- If ρ_∞ is Gaussian, then X_t is also Gaussian and the mean evolves according to the accelerated gradient flow in Euclidean space

Minimizing relative entropy

- If $\mathsf{F}(\rho) = \mathsf{D}(\rho \| \rho_\infty)$ where $\rho_\infty = e^{-f}$. Then the accelerated flow is

$$\ddot{X}_t = -\frac{3}{t}\dot{X}_t - \nabla f(X_t) - \underbrace{\nabla \log(\rho_t(X_t))}_{\text{mean-field term}}, \quad X_0 \sim \rho_0$$

- $\mathsf{F}(\rho)$ is displacement convex iff $f(x)$ is convex
- If ρ_∞ is Gaussian, then X_t is also Gaussian and the mean evolves according to the accelerated gradient flow in Euclidean space

Minimizing relative entropy

- If $\mathsf{F}(\rho) = \mathsf{D}(\rho \| \rho_\infty)$ where $\rho_\infty = e^{-f}$. Then the accelerated flow is

$$\ddot{X}_t = -\frac{3}{t}\dot{X}_t - \nabla f(X_t) - \underbrace{\nabla \log(\rho_t(X_t))}_{\text{mean-field term}}, \quad X_0 \sim \rho_0$$

- $\mathsf{F}(\rho)$ is displacement convex iff $f(x)$ is convex
- If ρ_∞ is Gaussian, then X_t is also Gaussian and the mean evolves according to the accelerated gradient flow in Euclidean space

Numerical algorithm

- Accelerated flow for minimizing relative entropy

$$\ddot{X}_t = -\frac{3}{t}\dot{X}_t - \nabla f(X_t) - \underbrace{\nabla \log(\rho_t(X_t))}_{\text{mean-field term}}, \quad X_0 \sim \rho_0$$

- Realized with system of interacting particles $\{X_t^i\}_{i=1}^N$

$$\ddot{X}_t^i = -\frac{3}{t}\dot{X}_t^i - \nabla f(X_t^i) - \underbrace{I_t^{(N)}(X_t^i)}_{\text{interaction term}}, \quad X_0^i \overset{\text{i.i.d}}{\sim} \rho_0$$

- (parametric) Gaussian approximation
- (non-parametric) Diffusion-map approximation, density estimation
- Time discretization using the symplectic method

Numerical algorithm

- Accelerated flow for minimizing relative entropy

$$\ddot{X}_t = -\frac{3}{t}\dot{X}_t - \nabla f(X_t) - \underbrace{\nabla \log(\rho_t(X_t))}_{\text{mean-field term}}, \quad X_0 \sim \rho_0$$

- Realized with system of interacting particles $\{X_t^i\}_{i=1}^N$

$$\ddot{X}_t^i = -\frac{3}{t}\dot{X}_t^i - \nabla f(X_t^i) - \underbrace{I_t^{(N)}(X_t^i)}_{\text{interaction term}}, \quad X_0^i \overset{\text{i.i.d}}{\sim} \rho_0$$

- (parametric) Gaussian approximation
- (non-parametric) Diffusion-map approximation, density estimation
- Time discretization using the symplectic method

Numerical algorithm

- Accelerated flow for minimizing relative entropy

$$\ddot{X}_t = -\frac{3}{t}\dot{X}_t - \nabla f(X_t) - \underbrace{\nabla \log(\rho_t(X_t))}_{\text{mean-field term}}, \quad X_0 \sim \rho_0$$

- Realized with system of interacting particles $\{X_t^i\}_{i=1}^N$

$$\ddot{X}_t^i = -\frac{3}{t}\dot{X}_t^i - \nabla f(X_t^i) - \underbrace{I_t^{(N)}(X_t^i)}_{\text{interaction term}}, \quad X_0^i \overset{\text{i.i.d}}{\sim} \rho_0$$

- (parametric) Gaussian approximation
- (non-parametric) Diffusion-map approximation, density estimation
- Time discretization using the symplectic method

Numerical algorithm

- Accelerated flow for minimizing relative entropy

$$\ddot{X}_t = -\frac{3}{t}\dot{X}_t - \nabla f(X_t) - \underbrace{\nabla \log(\rho_t(X_t))}_{\text{mean-field term}}, \quad X_0 \sim \rho_0$$

- Realized with system of interacting particles $\{X_t^i\}_{i=1}^N$

$$\ddot{X}_t^i = -\frac{3}{t}\dot{X}_t^i - \nabla f(X_t^i) - \underbrace{I_t^{(N)}(X_t^i)}_{\text{interaction term}}, \quad X_0^i \overset{\text{i.i.d}}{\sim} \rho_0$$

- (parametric) Gaussian approximation
- (non-parametric) Diffusion-map approximation, density estimation
- Time discretization using the symplectic method

Numerical algorithm

- Accelerated flow for minimizing relative entropy

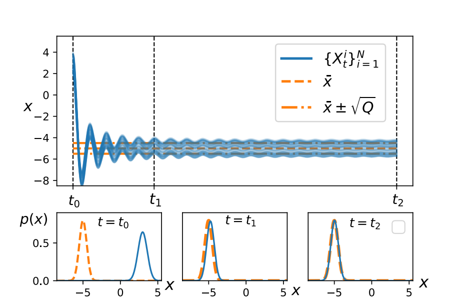
$$\ddot{X}_t = -\frac{3}{t}\dot{X}_t - \nabla f(X_t) - \underbrace{\nabla \log(\rho_t(X_t))}_{\text{mean-field term}}, \quad X_0 \sim \rho_0$$

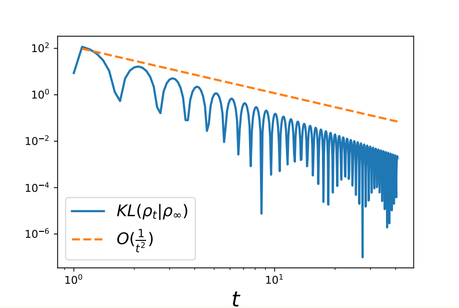
- Realized with system of interacting particles $\{X_t^i\}_{i=1}^N$

$$\ddot{X}_t^i = -\frac{3}{t}\dot{X}_t^i - \nabla f(X_t^i) - \underbrace{I_t^{(N)}(X_t^i)}_{\text{interaction term}}, \quad X_0^i \overset{\text{i.i.d}}{\sim} \rho_0$$

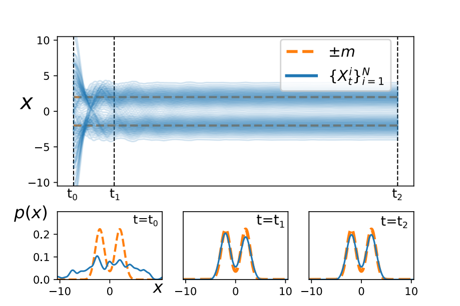
- (parametric) Gaussian approximation
- (non-parametric) Diffusion-map approximation, density estimation
- Time discretization using the symplectic method

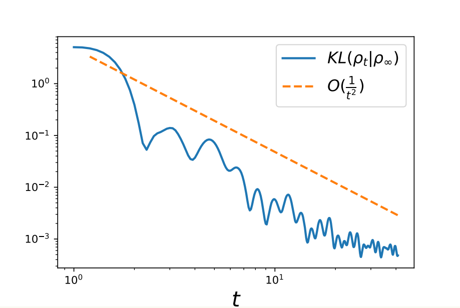
- The target distribution is Gaussian





- The target distribution is mixture of two Gaussians





Comparison to Hamiltonian MCMC

- Proposed accelerated flow:

$$\ddot{X}_t = -\frac{3}{t}\dot{X}_t - \nabla f(X_t) - \underbrace{\nabla \log(\rho_t(X_t))}_{\text{mean-field term}}$$

- Continuous-time limit of Hamiltonian MCMC (under-damped Langevin eq.)

$$\begin{aligned} \mathrm{d}X_t &= v_t \,\mathrm{d}t \\ \mathrm{d}v_t &= -\gamma v_t \,\mathrm{d}t - \nabla f(X_t)\,\mathrm{d}t + \underbrace{\sqrt{2}\,\mathrm{d}B_t}_{\text{stochastic term}} \end{aligned}$$

- Trade-off between computational efficiency and accuracy

Comparison to Hamiltonian MCMC

- Proposed accelerated flow:

$$\ddot{X}_t = -\frac{3}{t}\dot{X}_t - \nabla f(X_t) - \underbrace{\nabla \log(\rho_t(X_t))}_{\text{mean-field term}}$$

- Continuous-time limit of Hamiltonian MCMC (under-damped Langevin eq.)

$$\mathrm{d}X_t = v_t \,\mathrm{d}t$$
$$\mathrm{d}v_t = -\gamma v_t \,\mathrm{d}t - \nabla f(X_t)\,\mathrm{d}t + \underbrace{\sqrt{2}\,\mathrm{d}B_t}_{\text{stochastic term}}$$

- Trade-off between computational efficiency and accuracy

Comparison to Hamiltonian MCMC

- Proposed accelerated flow:

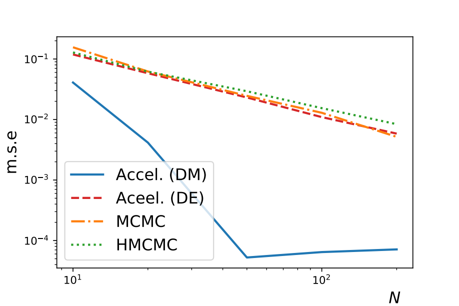
$$\ddot{X}_t = -\frac{3}{t}\dot{X}_t - \nabla f(X_t) - \underbrace{\nabla \log(\rho_t(X_t))}_{\text{mean-field term}}$$

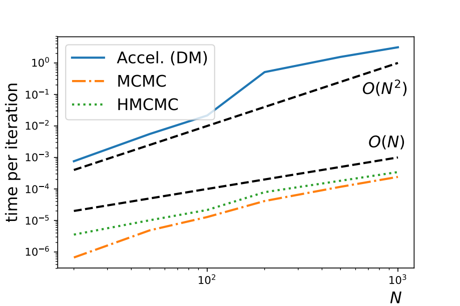
- Continuous-time limit of Hamiltonian MCMC (under-damped Langevin eq.)

$$\begin{aligned} \mathrm{d}X_t &= v_t \,\mathrm{d}t \\ \mathrm{d}v_t &= -\gamma v_t \,\mathrm{d}t - \nabla f(X_t)\,\mathrm{d}t + \underbrace{\sqrt{2}\,\mathrm{d}B_t}_{\text{stochastic term}} \end{aligned}$$

- Trade-off between computational efficiency and accuracy

Numerical example
comparison with MCMC and HMCMC





Summary

	vector variables $\mathbb{R}^d$	probability distribution $\mathcal{P}_2(\mathbb{R}^d)$
Objective funct.	$f(x)$	$\mathsf{F}(\rho) = \mathsf{D}(\rho \| \rho_\infty)$
Gradient flow	$\dot{x}_t = -\nabla f(x_t)$	$\mathrm{d}X_t = -\nabla f(X_t)\,\mathrm{d}t + \sqrt{2}\,\mathrm{d}B_t$
Lagrangian	$t^3(\frac{1}{2}\|u_t\|^2 - f(x_t))$	$\mathsf{E}[t^3(\frac{1}{2}\|U_t\|^2 - f(X_t) - \log(\rho(X_t)))]$
Accelerated flow	$\ddot{x}_t = -\frac{3}{t}\dot{x}_t - \nabla f(x_t)$	$\ddot{X}_t = -\frac{3}{t}\dot{X}_t - \nabla f(X_t) - \nabla \log(\rho_t(X_t))$

Future work:

- Removing the assumption $d = 1$
- Convergence analysis of the discretized algorithm

Summary

	vector variables $\mathbb{R}^d$	probability distribution $\mathcal{P}_2(\mathbb{R}^d)$
Objective funct.	$f(x)$	$\mathsf{F}(\rho) = \mathsf{D}(\rho \| \rho_\infty)$
Gradient flow	$\dot{x}_t = -\nabla f(x_t)$	$\mathrm{d}X_t = -\nabla f(X_t)\,\mathrm{d}t + \sqrt{2}\,\mathrm{d}B_t$
Lagrangian	$t^3(\frac{1}{2}\|u_t\|^2 - f(x_t))$	$\mathsf{E}[t^3(\frac{1}{2}\|U_t\|^2 - f(X_t) - \log(\rho(X_t)))]$
Accelerated flow	$\ddot{x}_t = -\frac{3}{t}\dot{x}_t - \nabla f(x_t)$	$\ddot{X}_t = -\frac{3}{t}\dot{X}_t - \nabla f(X_t) - \nabla \log(\rho_t(X_t))$

Future work:

- Removing the assumption $d = 1$
- Convergence analysis of the discretized algorithm