Accelerated Flow for Probability Distributions

Thirty-sixth International Conference on Machine Learning, Long Beach, 2019

Amirhossein Taghvaei
Joint work with P. G. Mehta

Coordinated Science Laboratory
University of Illinois at Urbana-Champaign

June 13, 2019

ILLINOIS

Objective and main idea

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

Objective: Construct accelerated flows for probability distribution

Approach:

- (Wibisono, et. al. 2017) proposed a variational formulation to construct accelerated flows on Euclidean space
- Our approach is to extend the variational formulation for probability distributions

Objective and main idea

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

Objective: Construct accelerated flows for probability distribution

Approach:

- (Wibisono, et. al. 2017) proposed a variational formulation to construct accelerated flows on Euclidean space
- Our approach is to extend the variational formulation for probability distributions

Objective and main idea

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

Objective: Construct accelerated flows for probability distribution

Approach:

- (Wibisono, et. al. 2017) proposed a variational formulation to construct accelerated flows on Euclidean space
- Our approach is to extend the variational formulation for probability distributions

Variational formulation for Euclidean space

	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	$t^3(\frac{1}{2}\lvert\dot{x}_t\rvert^2 - f(x_t))$	?
Accelerated flow	$\ddot{x}_t = -\frac{3}{t}\dot{x}_t - \nabla f(x_t)$	?

- Accelerated flow is obtained by minimizing the action integral of the Lagrangian

Wasserstein gradient flow

	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	$\ddot{x}_t = -\frac{3}{t}\dot{x}_t - \nabla f(x_t)$	?

- The Wasserstein gradient flow with respect to relative entropy is the Fokker-Planck equation (Jordan, et. al. 1998)
- The Fokker-Planck equation is realized with the Langevin sde
- The goal is to obtain accelerated forms of the sde

Wasserstein gradient flow

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

- The Wasserstein gradient flow with respect to relative entropy is the Fokker-Planck equation (Jordan, et. al. 1998)
- The Fokker-Planck equation is realized with the Langevin sde
- The goal is to obtain accelerated forms of the sde

Wasserstein gradient flow

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

- The Wasserstein gradient flow with respect to relative entropy is the Fokker-Planck equation (Jordan, et. al. 1998)
- The Fokker-Planck equation is realized with the Langevin sde
- The goal is to obtain accelerated forms of the 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	$t^3(\frac{1}{2}\|\dot{x}_t\|^2 - f(x_t))$	$\mathsf{E}[t^3(\frac{1}{2}\|\dot{X}_t\|^2 - f(X_t) - \log(\rho(X_t)))]$
Accelerated flow	$\ddot{x}_t = -\frac{3}{t}\dot{x}_t - \nabla f(x_t)$	$\ddot{X}_t = -\frac{3}{t}\dot{X}_t - \nabla f(X_t) - \nabla \log(\rho_t(X_t))$

- The accelerated flow involves a mean-field term $\nabla \log \rho_t(X_t)$ which depends on the distribution of X_t
- The numerical algorithm involves a system of interacting particles
- The mean-field term is approximated in terms of particles

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	$t^3(\frac{1}{2}\|\dot{x}_t\|^2 - f(x_t))$	$\mathsf{E}[t^3(\frac{1}{2}\|\dot{X}_t\|^2 - f(X_t) - \log(\rho(X_t)))]$
Accelerated flow	$\ddot{x}_t = -\frac{3}{t}\dot{x}_t - \nabla f(x_t)$	$\ddot{X}_t = -\frac{3}{t}\dot{X}_t - \nabla f(X_t) - \nabla \log(\rho_t(X_t))$

- The accelerated flow involves a mean-field term $\nabla \log \rho_t(X_t)$ which depends on the distribution of X_t
- The numerical algorithm involves a system of interacting particles
- The mean-field term is approximated in terms of particles

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	$t^3(\frac{1}{2}\|\dot{x}_t\|^2 - f(x_t))$	$\mathsf{E}[t^3(\frac{1}{2}\|\dot{X}_t\|^2 - f(X_t) - \log(\rho(X_t)))]$
Accelerated flow	$\ddot{x}_t = -\frac{3}{t}\dot{x}_t - \nabla f(x_t)$	$\ddot{X}_t = -\frac{3}{t}\dot{X}_t - \nabla f(X_t) - \nabla \log(\rho_t(X_t))$

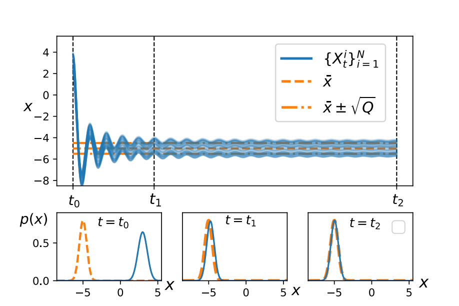
- The accelerated flow involves a mean-field term $\nabla \log \rho_t(X_t)$ which depends on the distribution of X_t
- The numerical algorithm involves a system of interacting particles
- The mean-field term is approximated in terms of particles

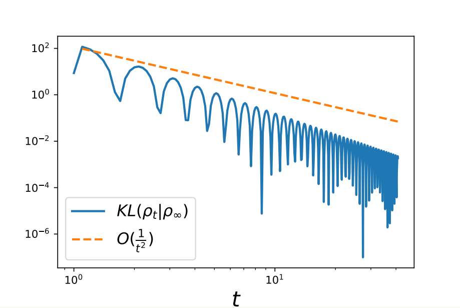
Summary

	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	$t^3(\frac{1}{2}\|\dot{x}_t\|^2 - f(x_t))$	$\mathsf{E}[t^3(\frac{1}{2}\|\dot{X}_t\|^2 - f(X_t) - \log(\rho(X_t)))]$
Accelerated flow	$\ddot{x}_t = -\frac{3}{t}\dot{x}_t - \nabla f(x_t)$	$\ddot{X}_t = -\frac{3}{t}\dot{X}_t - \nabla f(X_t) - \nabla \log(\rho_t(X_t))$

- The accelerated flow involves a mean-field term $\nabla \log \rho_t(X_t)$ which depends on the distribution of X_t
- The numerical algorithm involves a system of interacting particles
- The mean-field term is approximated in terms of particles

- The target distribution is Gaussian

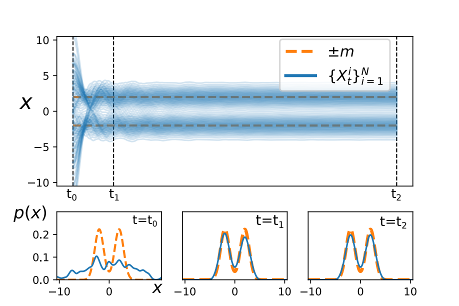


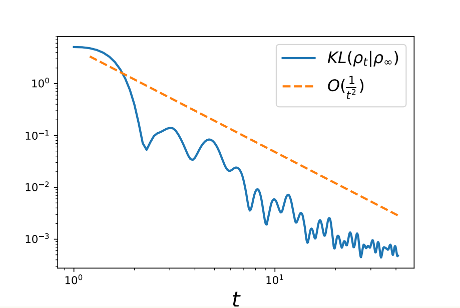


Numerical example

non-Gaussian

- The target distribution is mixture of two Gaussians

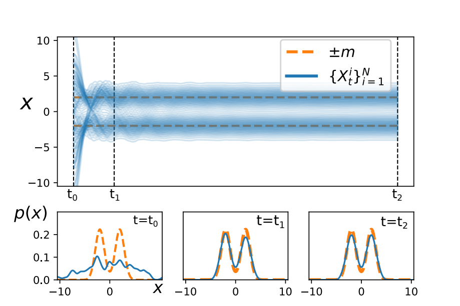


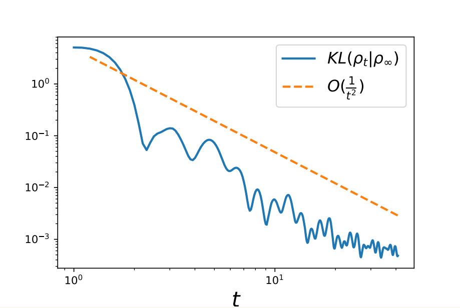


Thanks for your attention. For more details come to see poster #206

Numerical example
non-Gaussian

- The target distribution is mixture of two Gaussians





Thanks for your attention. For more details come to see poster #206