Variational Wasserstein gradient flow

Jiaojiao Fan[1], Qinsheng Zhang[1], Amirhossein Taghvaei[2], and Yongxin Chen[1]

[1]Georgia Institute of Technology
[2]University of Washington, Seattle

ICML 2022

Optimization over distributional space

$$\min_{P} \ \mathcal{F}(P)$$

- Kullback-Leibler divergence w.r.t. target distribution Q

$$\mathcal{D}(P\|Q) := \int \log\left(\frac{dP}{dQ}\right) dP$$

- Generalized entropy

$$\mathcal{G}(P) := \frac{1}{m-1}\int P^m(x)dx, \quad m > 1$$

- Jensen-Shannon divergence

$$\mathrm{JSD}(P\|Q) := \mathcal{D}\left(P \,\middle\|\, \frac{P+Q}{2}\right) + \mathcal{D}\left(Q \,\middle\|\, \frac{P+Q}{2}\right)$$

Optimization over distributional space

- Kullback-Leibler divergence w.r.t. target distribution Q

$$\mathcal{D}(P||Q) := \int \log\left(\frac{dP}{dQ}\right) dP$$

- Generalized entropy

$$\mathcal{G}(P) := \frac{1}{m-1}\int P^m(x)dx, \quad m > 1$$

- Jensen-Shannon divergence

$$\mathrm{JSD}(P\|Q) := \mathcal{D}\left(P\,\middle\|\,\frac{P+Q}{2}\right) + \mathcal{D}\left(Q\,\middle\|\,\frac{P+Q}{2}\right)$$

$\Rightarrow$ f-divergence $\qquad D_f(P\|Q) = \mathbb{E}_Q\left[f\left(\frac{dP}{dQ}\right)\right]$

Wasserstein Gradient Flow (WGF)

$$\frac{\partial P}{\partial t} = \nabla \cdot \left(P \nabla \frac{\delta \mathcal{F}}{\delta P} \right)$$

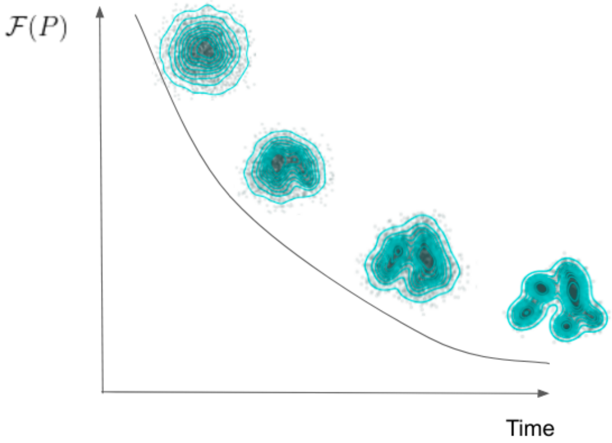
$\mathcal{F}(P)$	PDE $\partial P / \partial t =$	Class
$\int \log P dP + \int V dP$	$\nabla \cdot (P \nabla V) + \Delta P$	Fokker-Planck equation
$\frac{1}{m-1} \int P^m(x) dx \ (m > 1)$	ΔP^m	Porous media equation
$\int f\left(\frac{dP}{dQ}\right) dQ$	$\nabla \cdot \left(P \ \nabla f'\left(\frac{P}{Q}\right)\right)$	WGF of f-divergence

Proposition

If $\mathcal{F}(P)$ can be written as $D_f(P \| Q)$, then

$$\frac{d}{dt} \mathcal{F}(P_t) = -\mathbb{E}_{P_t}(\|\nabla f'(P_t/Q)\|^2)$$

Time discretization



Time discretization

JKO scheme (1998)

$$P_{k+1} = \arg\min_{P} \frac{1}{2a} W_2^2(P, P_k) + D_f(P\|Q)$$

(Step1) Brenier's Theorem:
if ν admits density, $\exists\ \nabla\varphi$ s.t.

$$W_2^2(\nu, \mu) = \int \|x - \nabla\varphi(x)\|^2 d\nu(x)$$

(Step2) Variational formula of $D_f(P\|Q)$:
if $P \ll Q$ and f is differentiable

$$D_f(P\|Q) = \sup_h \mathbb{E}_P[h(X)] - \mathbb{E}_Q[f^*(h(Y))]$$

$f^*(y) = \sup_{x\in\mathbb{R}}[xy - f(x)]$ is convex conjugate of f

Specialization of f-divergence

$$P_{k+1} = \nabla\varphi_k \sharp P_k$$

$$\nabla\varphi_k = \arg\min_{\varphi} \frac{1}{2a} \int \|x - \nabla\varphi(x)\|_2^2 dP_k(x) + \max_h \mathbb{E}_P[h(X)] - \mathbb{E}_Q[f^*(h(Y))]$$

$$X \sim P_k \Rightarrow \nabla\varphi_k(X) \sim P_{k+1}$$

$\mathcal{F}(P)$	$f(x)$	$f^*(y)$
KL divergence	$x \log x$	$\exp(y-1)$
Generalized entropy	$\frac{1}{m-1}(x^m - x)$	$\left(\frac{(m-1)y+1}{m}\right)^{\frac{m}{m-1}}$
JSD	$-(x+1)\log((1+x)/2) + x\log x$	$-\log(2-\exp(y))$

Specialization of f-divergence

$$P_{k+1} = T_k \sharp P_k$$

$$T_k = \arg\min_T \frac{1}{2a} \int \|x - T(x)\|_2^2 dP_k(x) + \max_h \mathbb{E}_P[h(X)] - \mathbb{E}_Q[f^*(h(Y))]$$

Computational complexity

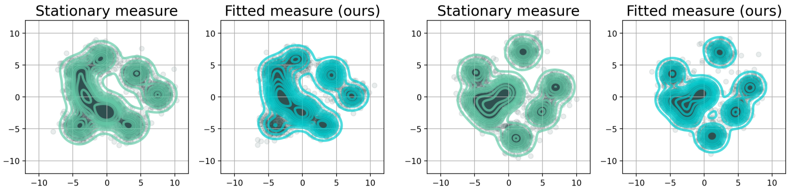
$n =$ dimension, $m = \#$ conjugate gradient steps,
$M =$ batch size, $H =$ neural network size

- Mokov et al. 2021, Alvarez-Melis et al. 2021
  - $O(n^3 + (k+n)MH)$ if exactly compute $\log |\nabla^2 \varphi|$
  - $O(mn^2 + (k+n)MH)$ if approximate $\log |\nabla^2 \varphi|$
- Ours: $O(kMH)$

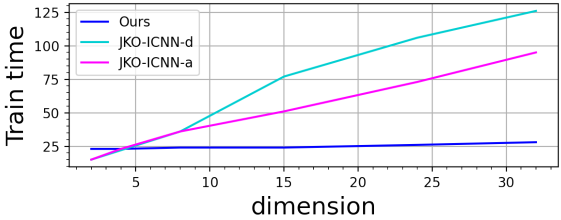
[Mokov et al.] "Large-Scale Wasserstein Gradient Flows" Neurips 2021

[Alvarez-Melis et al.] "Optimizing Functionals on the Space of Probabilities with Input Convex Neural Networks." Neurips OTML Workshop 2021

KL divergence: Sampling from target distribution Q



Dimension $n = 64$ Dimension $n = 128$

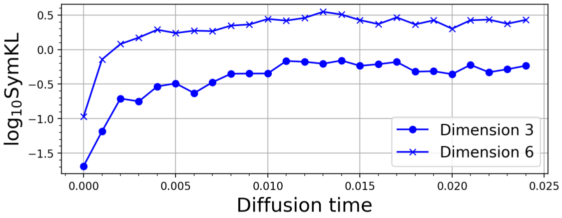


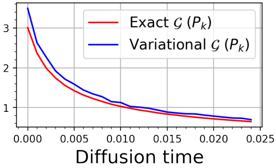
[JKO-ICNN-d] "Large-Scale Wasserstein Gradient Flows" Neurips 2021
[JKO-ICNN-a] "Optimizing Functionals on the Space of Probabilities with Input Convex Neural Networks." Neurips OTML Workshop 2021

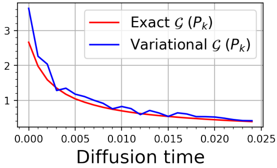
Generalized entropy: Porous media equation

Barenblatt profile

$$P(t,x) = (t+t_0)^{-\alpha}\left(C - \beta\|x-x_0\|^2 (t+t_0)^{\frac{-2\alpha}{n}}\right)_+^{\frac{1}{m-1}}$$

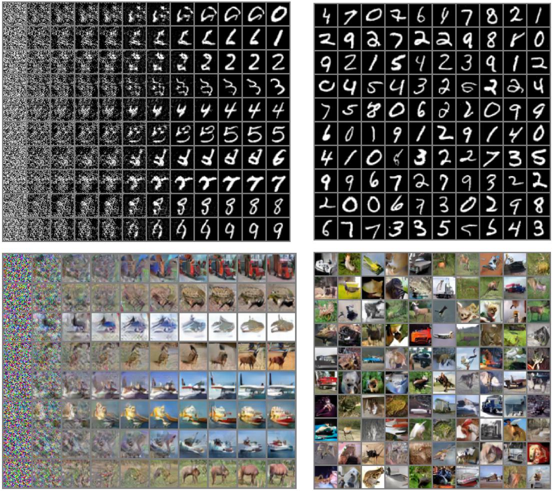






Dimension $n = 3$ Dimension $n = 6$

JSD: Gradient flow in pixel space



Trajectory

Uncurated samples

Thank you!

Contact:
jiaojiaofan@gatech.edu

arXiv:
https://arxiv.org/abs/2112.02424

code:
https://github.com/sbyebss/variational_wgf