A Gradual, Semi-Discrete Approach to Generative Network Training via Explicit Wasserstein Minimization

Yucheng Chen[1] Matus Telgarsky[1] Chao Zhang[1]
Bolton Bailey[1] Daniel Hsu[2] Jian Peng[1]

[1]Department of Computer Science, UIUC, Urbana, IL

[2]Department of Computer Science, Columbia University, New York, NY

International Conference on Machine Learning
June 12, 2019

Explicit Wasserstein Minimization

- **Goal:** To train a generator network g minimizing the Wasserstein distance $W(g\#\mu, \nu)$ between the generated distribution $g\#\mu$ and the target distribution ν, where μ is a simple distribution such as uniform or Gaussian.
  - Indirectly pursued by WGAN (Arjovsky et al., 2017)
- **Motivation:** If the optimal transport plan between $g\#\mu$ and ν can be computed, why not use it to explicitly minimize $W(g\#\mu, \nu)$ without any adversarial procedure?

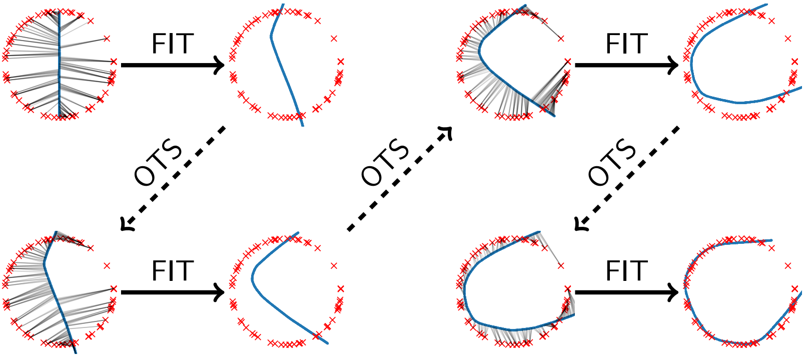
Key Observations

In the "semi-discrete setting", where $g\#\mu$ is continuous and ν is discrete (denoted as $\hat{\nu}$),

1. $W(g\#\mu, \hat{\nu})$ is realized by a **deterministic optimal transport mapping** T between $g\#\mu$ and $\hat{\nu}$, and
2. fitting the generated data $g\#\mu$ towards the corresponding target points $T\#g\#\mu$ may lead to a new generator g' with lower Wasserstein distance $W(g'\#\mu, \hat{\nu})$.

An algorithm iterating these two steps (called as "OTS" and "FIT") would explicitly minimize $W(g\#\mu, \hat{\nu})$.

A Synthetic Example



The Algorithm

- **OTS:** Compute the semi-discrete optimal transport between $g\#\mu$ and $\hat{\nu}$ by minimizing (Genevay et al., 2016)

$$-\int_X \min_i (c(x, y_i) - \hat{\psi}_i) \mathrm{d}g\#\mu(x) - \frac{1}{N}\sum_{i=1}^{N} \hat{\psi}_i.$$

  and the Monge OT plan is given by $T(x) := y_{\arg\min_i c(x,y_i) - \hat{\psi}_i}$.
- **FIT:** Find a new generator g' by minimizing

$$\int_z c(g'(z), T(g(z))) \mathrm{d}\mu(z).$$

- **Overall algorithm:** Iterate OTS and FIT.

Experimental Results

- MNIST: Better visual quality, better WD/IS/FID (even with small MLP architectures!)
- CelebA/CIFAR: Worse visual quality, but still lower WD
- Lower Wasserstein distance does not always lead to better visual quality: importance of regularizing discriminator in GANs (Huang et al., 2017; Bai et al., 2019).



References

Martín Arjovsky, Soumith Chintala, and Léon Bottou. Wasserstein generative adversarial networks. In *ICML*, 2017.

Yu Bai, Tengyu Ma, and Andrej Risteski. Approximability of discriminators implies diversity in GANs. In *ICLR*, 2019.

Aude Genevay, Marco Cuturi, Gabriel Peyré, and Francis R. Bach. Stochastic optimization for large-scale optimal transport. In *NIPS*, 2016.

Gabriel Huang, Gauthier Gidel, Hugo Berard, Ahmed Touati, and Simon Lacoste-Julien. Adversarial divergences are good task losses for generative modeling. 2017. `arXiv:1708.02511 [cs.LG]`.

Thank you!

Poster: Pacific Ballroom #4

6:30PM, Jun 12