Wasserstein of Wasserstein Loss for Learning Generative Models
ICML 2019

Yonatan Dukler*, Wuchen Li*, Alex Tong Lin*, Guido Montúfar*
{ydukler, wcl, atlin, montufar}@math.ucla.edu

June 11, 2019
Wasserstein GAN

- In the Wasserstein GAN framework \([2, 1]\), given \(P_{\text{data}}\), \(G(\Theta)\), and \(F(\Phi, x)\) trained as

\[
\min_{\Theta} \max_{\Phi} \mathbb{E}_{x \sim P_{G(\Theta)}} F(\Phi, x) - \mathbb{E}_{x \sim P_{\text{data}}} F(\Phi, x) \\
+ \lambda \mathbb{E}_{\tilde{x}} (\|\nabla_{x} F(\Phi, x)\|_{L^2} - 1)^2.
\]

  Lipschitz enforcing term

- Based on the Kanatrovich-Rubenstein duality

\[
\mathcal{W}_D(\mu, \nu) = \sup_{f \in \mathcal{C}} \left\{ \int f d\nu dx - \int f d\mu dx \right\}
\]

where \(\mathcal{C} := \{f : \text{Lip}_{d}(f) \leq 1\}\) and \(d(x, y)\) is a ground metric on samples (high dim. images) \([3]\).
Adding Prior Knowledge When Choosing $d(x, y)$

\[
C := \{ f : \text{Lip}_d(f) \leq 1 \} \text{ and } d(x, y)
\]

(A) Geometry of natural images: Lie in lower dimensional manifold than entire image space.

(B) Natural images satisfy well known symmetries and transformations cf. data-augmentation techniques.

A good metric should satisfy these priors.
Wasserstein metric on images

- Image as a distribution over pixels
- Pixel space is a graph
- Cost of transport based on graph’s weights

Pixel space graph

Raster image

Distribution
Wasserstein of Wasserstein GAN (WWGAN)
Wasserstein distance on images is computationally expansive

Instead utilize Wasserstein-2 metric’s Riemannian structure.

**WWGAN:** Replacing $L^2$ gradient with Wasserstein-2 gradient $\nabla_{x}^{W_2} F$ in WGAN

Penalty is computed efficiently via convolutions. (add’l. comp. cost $\sim 8\%$)
WWGAN is more robust to natural perturbations
References


Yonatan Dukler*, Wuchen Li*, Alex Tong Lin*, Guido Montúfar

Wasserstein of Wasserstein Loss for Learning Generative Models