Straight-Through Estimator as Projected Wasserstein Gradient Flow

Pengyu Cheng[1], Chang Liu[2], Chunyuan Li[3], Dinghan Shen[1], Ricardo Henao[1] and Lawrence Carin[1]

[1]Duke University, [2]Tsinghua University, [3]Microsoft Research

April 5, 2020

$\mathcal{M}$ is a d-dimensional discrete distribution family parameterized by $\boldsymbol{\theta}$.
Let $\tilde{\mathcal{M}}$ be the d-dimensional 2-Wasserstein space. f is a cost function.
We aim to minimize the expected cost

$$\min_{\boldsymbol{\theta}} \mathbb{E}_{\boldsymbol{z}\sim p_{\boldsymbol{\theta}}}[f(\boldsymbol{z})] = \min_{\mu\in\mathcal{M}} \mathbb{E}_{\boldsymbol{z}\sim\mu}[f(\boldsymbol{z})] =: \min_{\mu\in\mathcal{M}} F[\mu].$$

We propose a 3-step updating scheme: In k-th iteration,

- A: draw samples $\{\boldsymbol{z}_n\}$ from current distribution μ_k;
- B: update $\{\boldsymbol{z}_n\}$ to $\{\tilde{\boldsymbol{z}}_n\} \sim \tilde{\mu}_k$ via Wasserstein gradient flow in $\tilde{\mathcal{M}}$;
- C: project $\tilde{\mu}_k$ back to μ_{k+1} by minimizing Wasserstein distance $W(\mu, \tilde{\mu}_k)$.

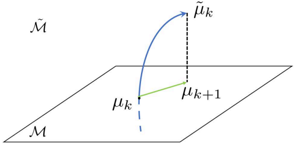


Figure: Updating scheme

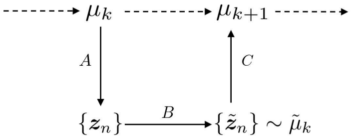


Figure: Algorithm outline

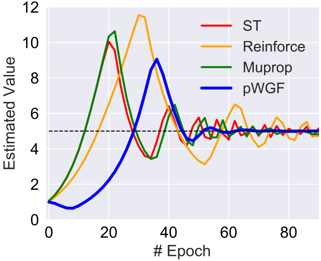
ST as pWGF and its improvement

Various algorithms can be derived from step C:

Algorithms	Approximation to Wasserstein distance
Straight-Through (ST)	**Difference in Means**
Proposed: pWGF-MMD	**Maximum Mean Discrepancy**

Experiments on Poisson inference task show improvement by pWGF:

- Real data $\{z_n\} \sim p(z) = \text{Poisson}(\lambda_0 = 5)$.
- Generate fake data $\{z'_n\} \sim q_\lambda(z) = \text{Poisson}(\lambda)$
- Discriminator $D_\omega(z)$ gives probability that z comes from real data.
- $\max_\lambda \min_\omega \{\mathbb{E}_{z\sim p}[\log D_\omega(z)] + \mathbb{E}_{z'\sim q_\lambda}[\log(1 - D_\omega(z'))]\}$



	Mean	Std
pWGF	5.0076	0.013
ST	5.1049	0.161
Muprop	5.0196	0.159
Reinforce	4.9452	0.173