Lipschitz Generative Adversarial Nets

Zhiming Zhou[1] Jiadong Liang[2] Lantao Yu[3] Yong Yu[1] Zhihua Zhang[2]

[1]Shanghai Jiao Tong University

[2]Peking University

[2]Stanford University

ICML, 2019

Generalized Formulation for GANs

$$\min_{f \in \mathcal{F}} \mathbb{E}_{z \sim \mathbb{P}_z}[\phi(f(g(z)))] + \mathbb{E}_{x \sim \mathbb{P}_r}[\psi(f(x))],$$
$$\min_{g \in \mathcal{G}} \mathbb{E}_{z \sim \mathbb{P}_z}[\varphi(f(g(z)))], \quad (1)$$

where

- $\mathbb{P}_z$: the source distribution in $\mathbb{R}^{n_z}$;
- $\mathbb{P}_r$: the target (real) distribution in $\mathbb{R}^{n_r}$;
- g: the generative function $\mathbb{R}^{n_z} \to \mathbb{R}^{n_r}$;
- f: the discriminative function $\mathbb{R}^{n_r} \to \mathbb{R}$;
- $\mathcal{G}$: the generative function space;
- $\mathcal{F}$: the discriminative function space;

ϕ, ψ, φ: $\mathbb{R} \to \mathbb{R}$ are loss metrics.

GAN	:	$\phi(x) = \psi(-x) = -\log(\sigma(-x))$
WGAN	:	$\phi(x) = \psi(-x) = x$
LSGAN	:	$\phi(x) = \psi(-x) = (x + \alpha)^2$

We denote the generation distribution by $\mathbb{P}_g$.

The Gradient Uninformativeness

The problem that the gradient from the discriminator does not contain any informative about the real distribution.

A new perspective for the training instability / convergence issue of GANs.

For GANs with unrestricted $\mathcal{F}$:

$$f^*(x) = \underset{f(x)\in\mathbb{R}}{\arg\min}\ \mathbb{P}_g(x)\phi(f(x)) + \mathbb{P}_r(x)\psi(f(x)),\ \forall x. \tag{2}$$

- $f^*(x)$ is independently defined and only reflects the local densities $\mathbb{P}_r(x)$ and $\mathbb{P}_g(x)$;
- $\nabla_x f^*(x)$ does not reflect any information about the other distribution, if the supports of two distributions are disjoint.

The Gradient Uninformativeness

Unrestricted GANs	MUST suffer from this problem
Restricted GANs	May suffer from this problem
GANs with W-Distance	May suffer from this problem
Lipschitz GANs	DO NOT suffer from this problem

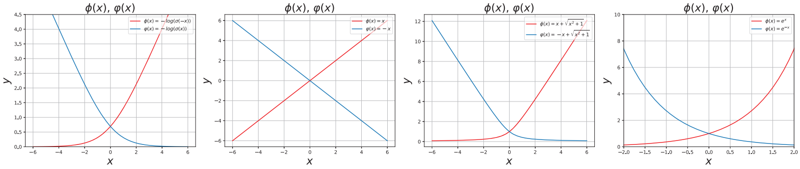
Lipschitz Generative Adversarial Nets (LGANs)

$$\min_{f \in \mathcal{F}} \mathbb{E}_{z \sim \mathbb{P}_z}[\phi(f(g(z)))] + \mathbb{E}_{x \sim \mathbb{P}_r}[\psi(f(x))] + \boxed{\lambda \cdot k(f)^2},$$
$$\min_{g \in \mathcal{G}} \mathbb{E}_{z \sim \mathbb{P}_z}[\varphi(f(g(z)))]. \qquad \boxed{k(f)\text{: Lipschitz constant of } f} \tag{3}$$

We require ϕ and ψ to satisfy:

$$\begin{cases} \phi'(x) > 0, \\ \phi''(x) \geq 0, \end{cases} \quad \textit{and} \quad \phi(x) = \psi(-x). \tag{4}$$

Any increasing function with non-decreasing derivative.



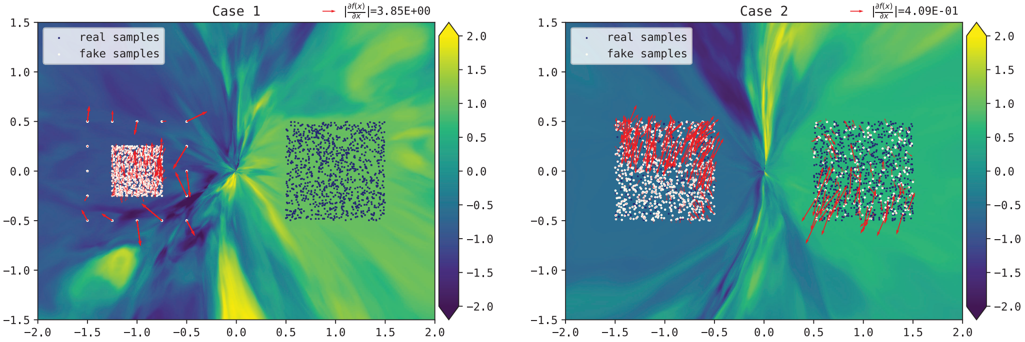
Lipschitz Generative Adversarial Nets (LGANs)

Theoretically guaranteed properties:

- The optimal discriminative function f^* exists;
- If ϕ is strictly convex, then f^* is unique;
- There exists a unique Nash equilibrium where $\mathbb{P}_r = \mathbb{P}_g$ and $k(f^*) = 0$;
- Do not suffer from gradient uninformativeness;
- For each generated sample, the gradient directly points towards a real sample.

Experiments: Gradient Uninformativeness

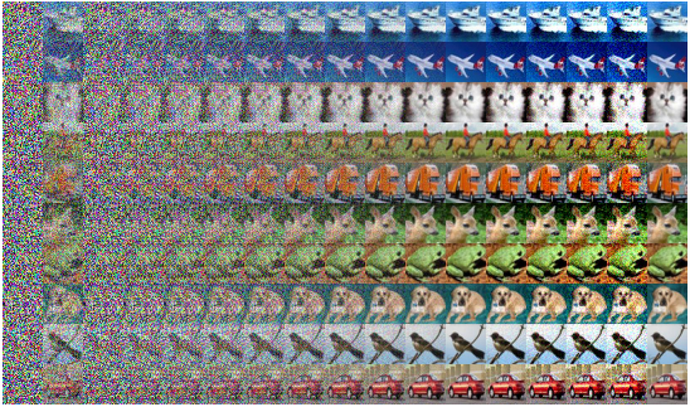
Gradient uninformativeness practically behaviors as noisy gradient.



(a) Disjoint Case (b) Overlapping Case

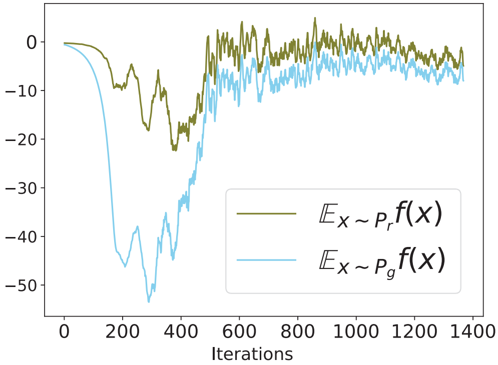
Experiments: $\nabla_x f^*(x)$ in LGANs

$\nabla_x f^*(x)$ directly point towards real samples.

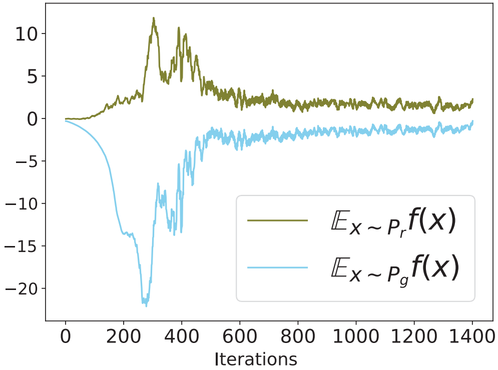


Experiments: f^* is Unique

Uniqueness of f^* leads to stabilized discriminative functions.



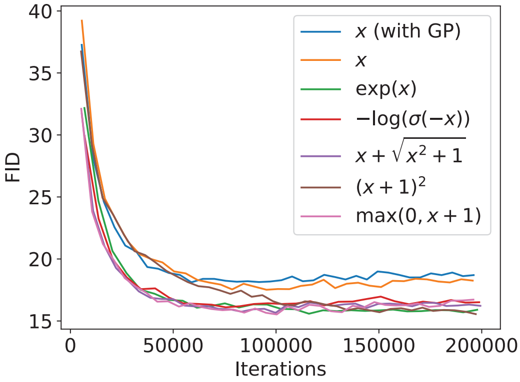
(a) $f(x)$ in WGAN.



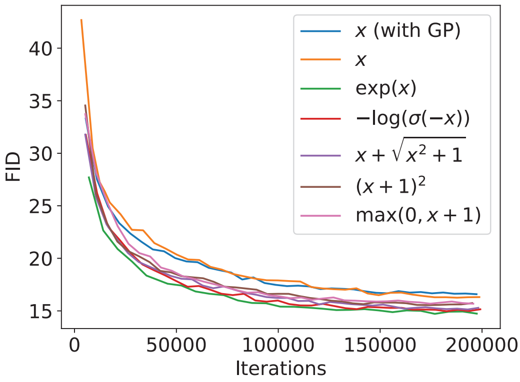
(b) $f(x)$ in LGANs.

Experiments: Unsupervised Image Generation

LGANs, with different choices of $\phi(x)$, consistently outperform WGAN.



(a) Training curves on CIFAR.



(b) Training curves on Tiny.

Summary

Poster: this evening, #19

- Gradient uninformativeness

Unrestricted GANs	MUST suffer from this problem	
Restricted GANs	May suffer from this problem	(5)
GANs with W-Distance	May suffer from this problem	
Lipschitz GANs	DO NOT suffer from this problem	

- Lipschitz GANs:
  - Penalize the Lipschitz constant of f;
  - Set $\phi(x)$ to be any increasing function with non-decreasing derivative;
  - If ϕ is strictly convex, then f^* is unique;
  - The gradients directly point towards real samples.