On Characterizing GAN Convergence Through Proximal Duality Gap

Paper ID: 4023

Sahil Sidheekh

Aroof Aimen

Dr. Narayanan C. K

Laboratory of Statistical Artificial Intelligence & Machine Learning

Indian Institute of Technology Ropar

भारतीय प्रौद्योगिकी संस्थान रोपड़

ICML

International Conference On Machine Learning

18-2 July, 2021

Overview

Objective : Quantifying and understanding GAN convergence

Questions We Address :

1. How to quantitatively identify if a GAN has converged and learned the real data distribution ?
2. How do GAN game configurations relate to the nature of the learned data distribution ?

GAN Formulation

A zero sum min-max game

$$\min_{\theta_g \in \Theta_G} \max_{\theta_d \in \Theta_D} V(D_{\theta_d}, G_{\theta_g}),$$

Classic GAN	$V = V_c = \mathbb{E}_{\mathrm{x}\sim P_r}[\log D(\mathrm{x})] + \mathbb{E}_{\mathrm{x}\sim P_{\theta_g}}[\log(1 - D(\mathrm{x}))]$
$f-$GAN	$V = V_f = \mathbb{E}_{\mathrm{x}\sim P_r}[D(\mathrm{x})] - \mathbb{E}_{\mathrm{x}\sim P_{\theta_g}}[f^*(D(\mathrm{x}))],$
WGAN	$V = V_w = \mathbb{E}_{\mathrm{x}\sim P_r}[D(\mathrm{x})] - \mathbb{E}_{\mathrm{x}\sim P_{\theta_g}}[D^c(\mathrm{x})],$

P_r : real data distribution

P_{θ_g}: generated data distribution

Each formulation minimises a particular divergence between P_r and P_{θ_g}

GAN Optimality

What is GAN convergence ?

An adversarial game converges to an equilibrium

Classical notion of GAN convergence - *Nash Equilibrium*: (θ_d^*, θ_g^*)

$$V(D_{\theta_d}, G_{\theta_g^*}) \leq V(D_{\theta_d^*}, G_{\theta_g^*}) \leq V(D_{\theta_d^*}, G_{\theta_g}) \;; \forall\ \theta_d, \theta_g$$

Need not always exist

A more generic notion of GAN convergence - *Proximal Equilibrium*: (θ_d^*, θ_g^*)

$$V(D_{\theta_d}, G_{\theta_g^*}) \leq V(D_{\theta_d^*}, G_{\theta_g^*}) \leq V^\lambda(D_{\theta_d^*}, G_{\theta_g});\ \forall\ \theta_d, \theta_g$$

Guaranteed to exist

where, $V^\lambda(D_{\theta_d}, G_{\theta_g}) = \max_{\tilde{\theta}_d \in \Theta_D} V(D_{\tilde{\theta}_d}, G_{\theta_g}) - \lambda || D_{\tilde{\theta}_d} - D_{\theta_d} ||^2$

Covers a spectrum of equilibria through λ

Characterizing GAN Convergence

Quantify GAN convergence as attaining the game attaining a proximal equilibrium

Introducing Proximal Duality Gap

For a GAN configuration (θ_d, θ_g), we define proximal duality gap (DG^{λ}) as :

$$DG^{\lambda}(\theta_d, \theta_g) = V_{D_w}(\theta_g) - V^{\lambda}_{G_w}(\theta_d) \,,$$

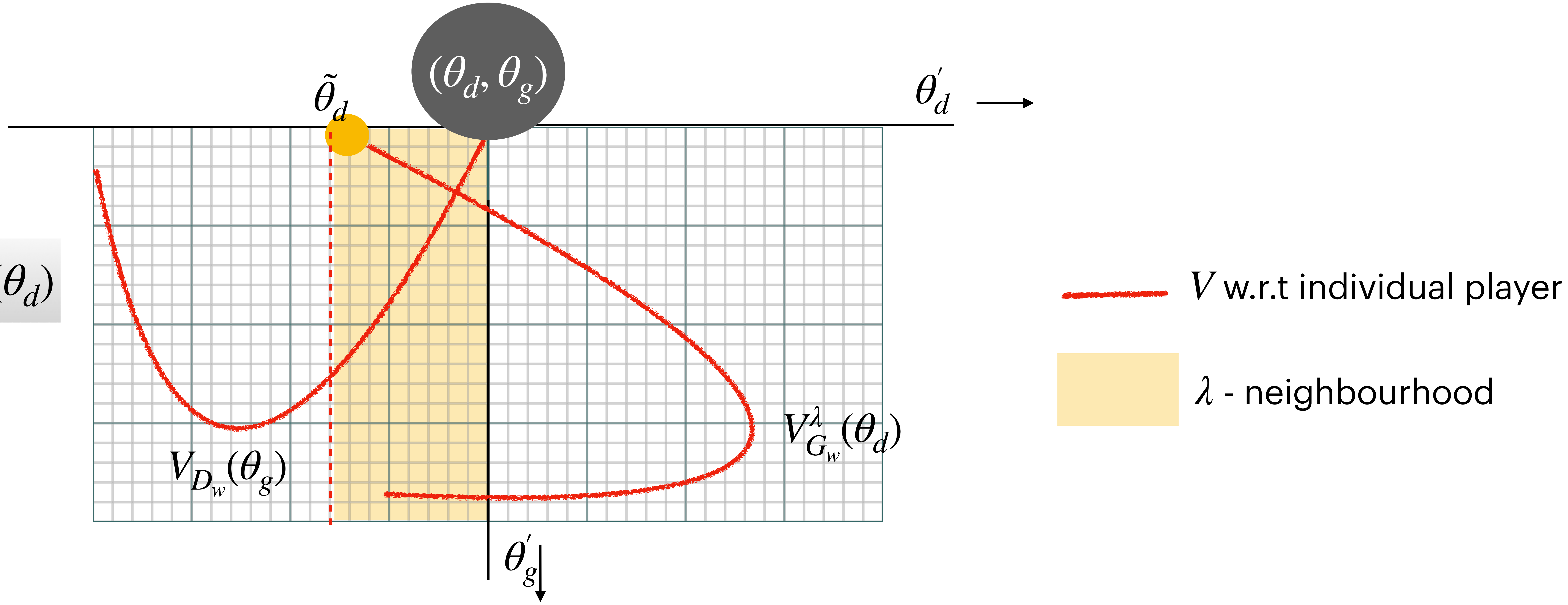
where , $$V_{D_w}(\theta_g) = \max_{\theta_d^{'} \in \Theta_D} V(D_{\theta_d^{'}}, G_{\theta_g})$$

$$V^{\lambda}_{G_w}(\theta_d) = \min_{\theta_g^{'} \in \Theta_G} V^{\lambda}(D_{\theta_d}, G_{\theta_g^{'}})$$

Measure the ability of the players to deviate from a given configuration w.r.t the proximal objective (V^{λ})

Proximal Duality Gap for GANs

$$DG^{\lambda}(\theta_d, \theta_g) = V_{D_w}(\theta_g) - V^{\lambda}_{G_w}(\theta_d)$$



At a $\lambda-$proximal equilibrium $(\theta_d{}^*, \theta_g{}^*)$, $V_{D_w}(\theta_g^*) = V^{\lambda}_{G_w}(\theta_d^*) = V(\theta_d^*, \theta_g^*)$

$DG^{\lambda}(\theta_d^*, \theta_g^*) = 0$ → Quantifies GAN Convergence!

Proximal Duality Gap

What does proximal duality gap tell us about the nature of the learned data distribution ?

DG^{λ} is lower bounded closely by the divergence between the real and generated data distributions.

$$DG^{\lambda}(\theta_d, \theta_g) \geq DIV(P_{\theta_g} || P_r) - \kappa$$

Where κ (≥ 0) denotes the minimum divergence that the considered class of generator functions can achieve with the real data distribution.

$DG^{\lambda} \rightarrow 0$ not only implies that the GAN has reached an equilibrium, but also $P_r \approx P_g$

Proximal Duality Gap

Implications of Proximal Duality Gap : Better Understanding GAN optimality

The generator attains the minimum divergence with the real data distribution at a proximal equilibrium.

Not all Proximal equilibria are Nash equilibria

↓

GANs can learn / attain the minimum divergence with P_r at non-Nash points as well

But then, can $P_{\theta_g} = P_r$ at non-proximal equilibria ?

Proximal Duality Gap

Implications of Proximal Duality Gap : Better Understanding GAN optimality

DG^{λ} at a configuration (θ_d^*, θ_g^*) for the GAN game is equal to zero for $\lambda = 0$, when the generator learns the real data distribution.

$$P_{\theta_g^*} = P_r \implies DG^{\lambda=0}(\theta_d^*, \theta_g^*) = 0$$

GANs can capture P_r at a game configuration if and only if it corresponds to a Stackelberg Equilibrium.

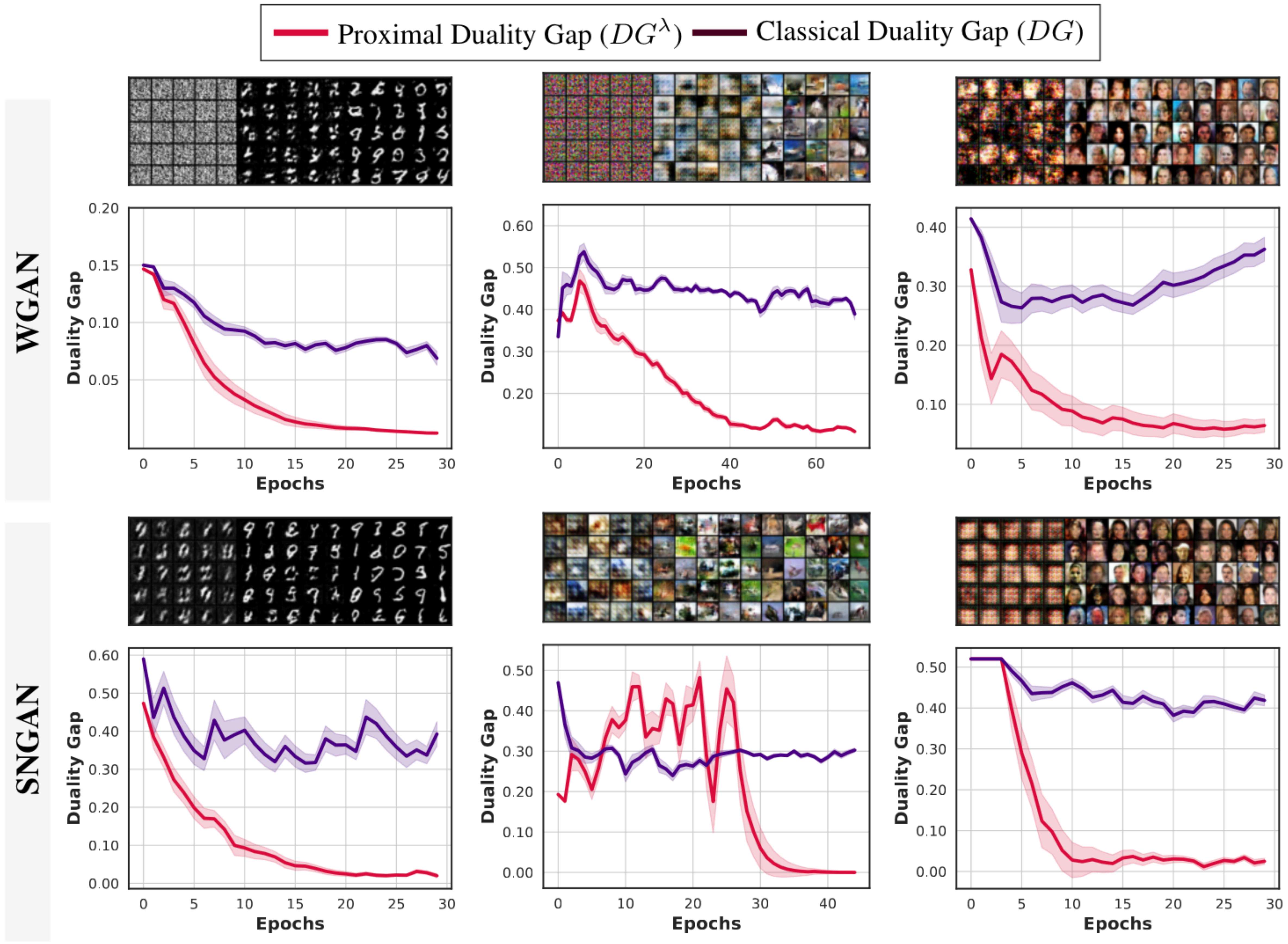
DG^{λ} is sufficient to quantify GAN convergence in the wild

Experiments and Results

Simulate GAN convergence & non-convergence

Monitor GAN training using DG^{λ}

DG^{λ} tends to zero when GAN converges



Thank You

Visit the link below to have a look at our paper !

Link: https://arxiv.org/abs/2105.04801

Laboratory of Statistical Artificial Intelligence & Machine Learning

Indian Institute of Technology Ropar

भारतीय प्रौद्योगिकी संस्थान रोपड़

ICML

International Conference On Machine Learning