ELeGANt: Euler-Lagrange Analysis of Wasserstein Generative Adversarial Networks

DLDE-III —
NeurIPS 2023 Workshops on The
Symbiosis of Deep Learning and Differential Equations III

Siddarth Asokan
Robert Bosch Centre for Cyber Physical Systems, Indian Institute of Science
Bangalore
siddartha@iisc.ac.in

Chandra Sekhar Seelamantula
Department of Electrical Engineering
Indian Institute of Science
Bangalore
css@iisc.ac.in

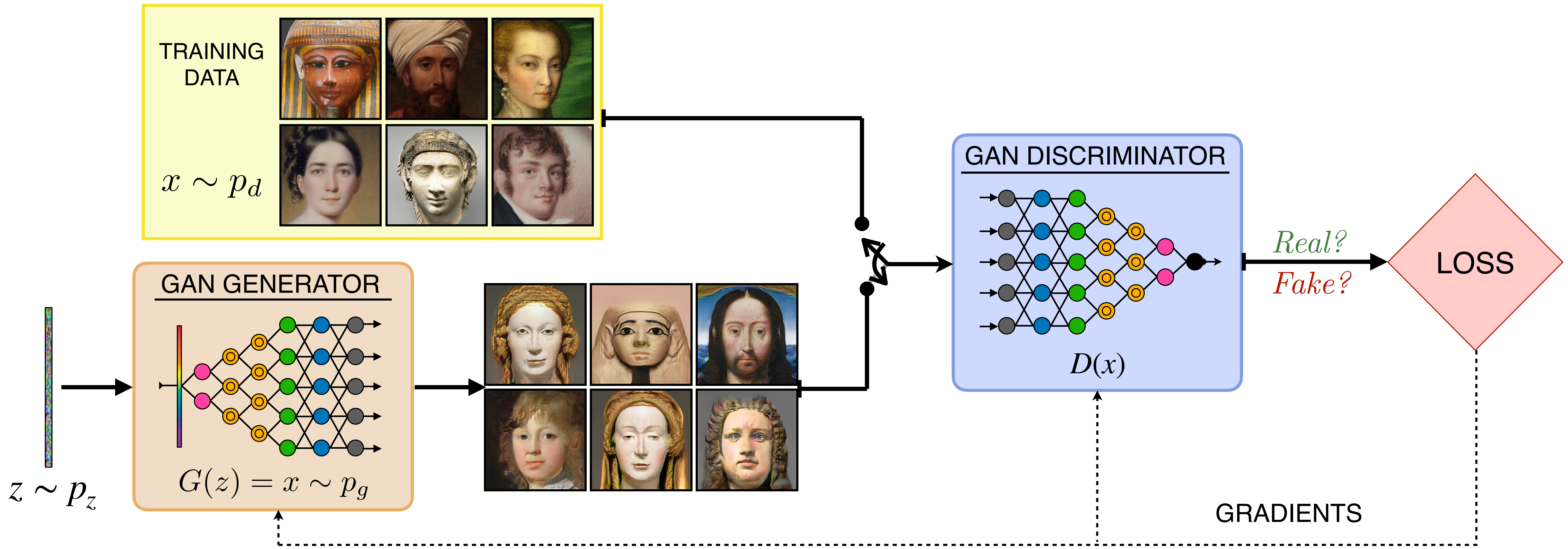
Robert Bosch Centre for CYBER PHYSICAL SYSTEMS

INDIAN INSTITUTE OF SCIENCE
भारतीय विज्ञान संस्थान

SPECTRUM LAB

Generative Adversarial Networks (GANs)

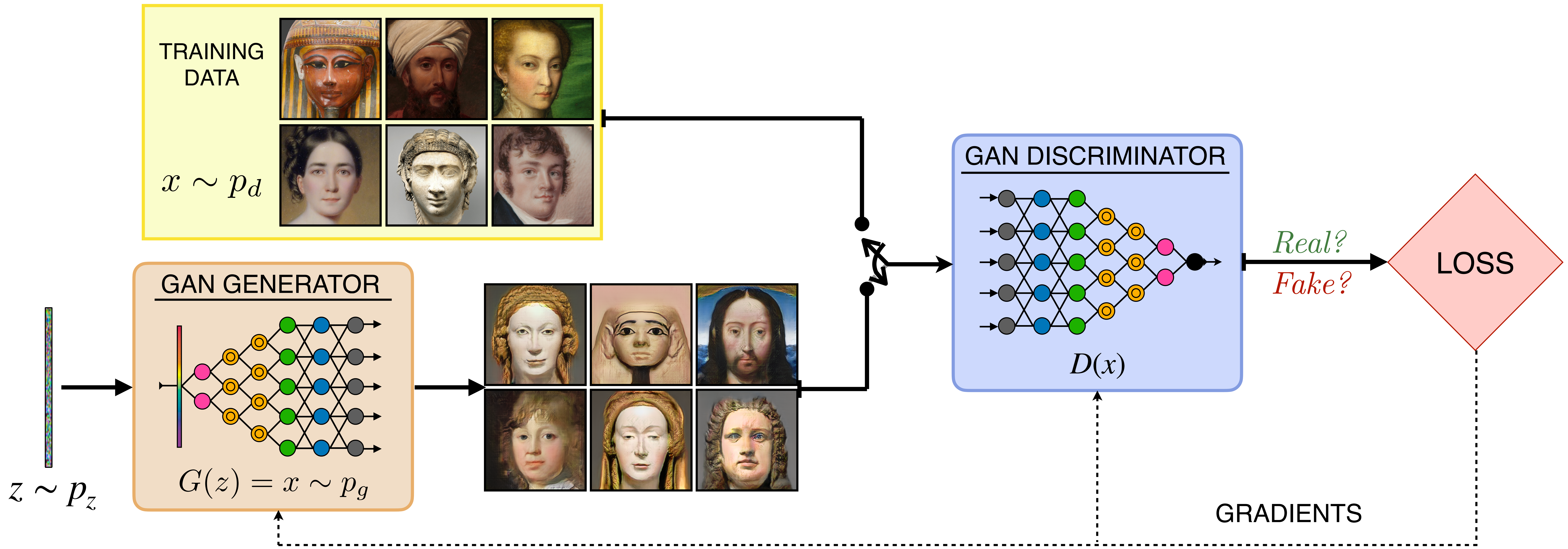
- A framework for training generative models to learn a desired data distribution.



[1] Goodfellow et al., "Generative Adversarial Networks," NeurIPS 2014

Generative Adversarial Networks (GANs)

- A framework for training generative models to learn a desired data distribution.

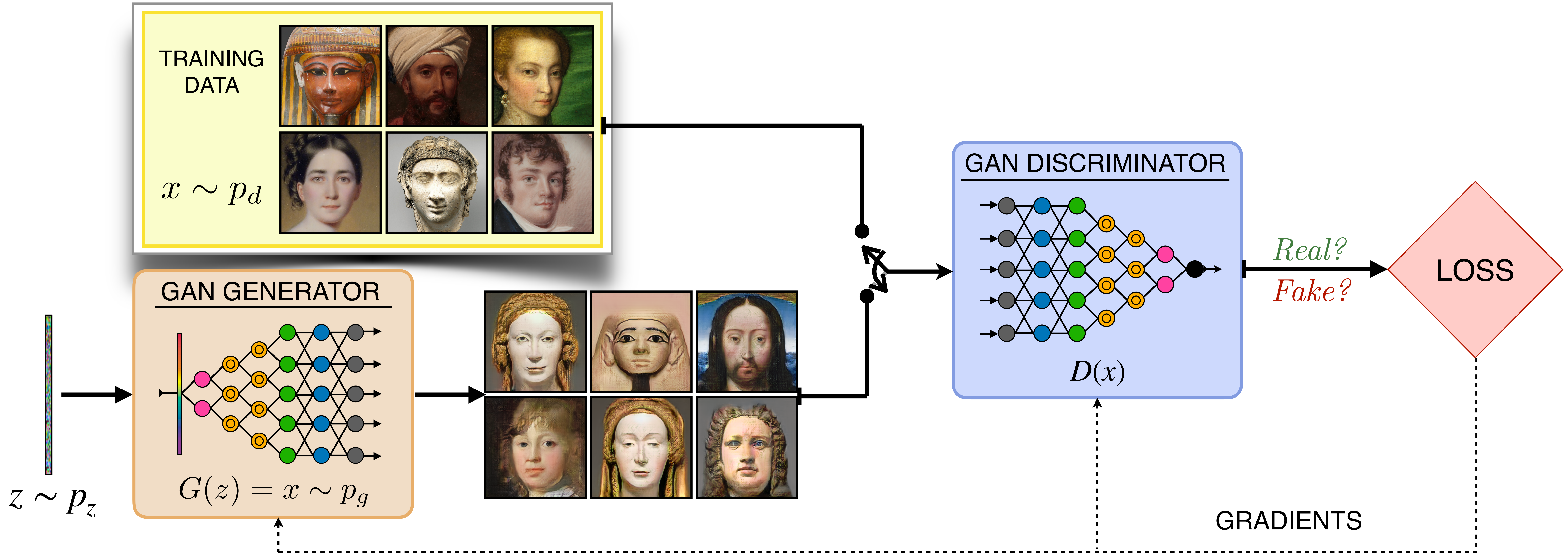


- A min-max game between two players.

[1] Goodfellow et al., "Generative Adversarial Networks," NeurIPS 2014

Generative Adversarial Networks (GANs)

- A framework for training generative models to learn a desired data distribution.

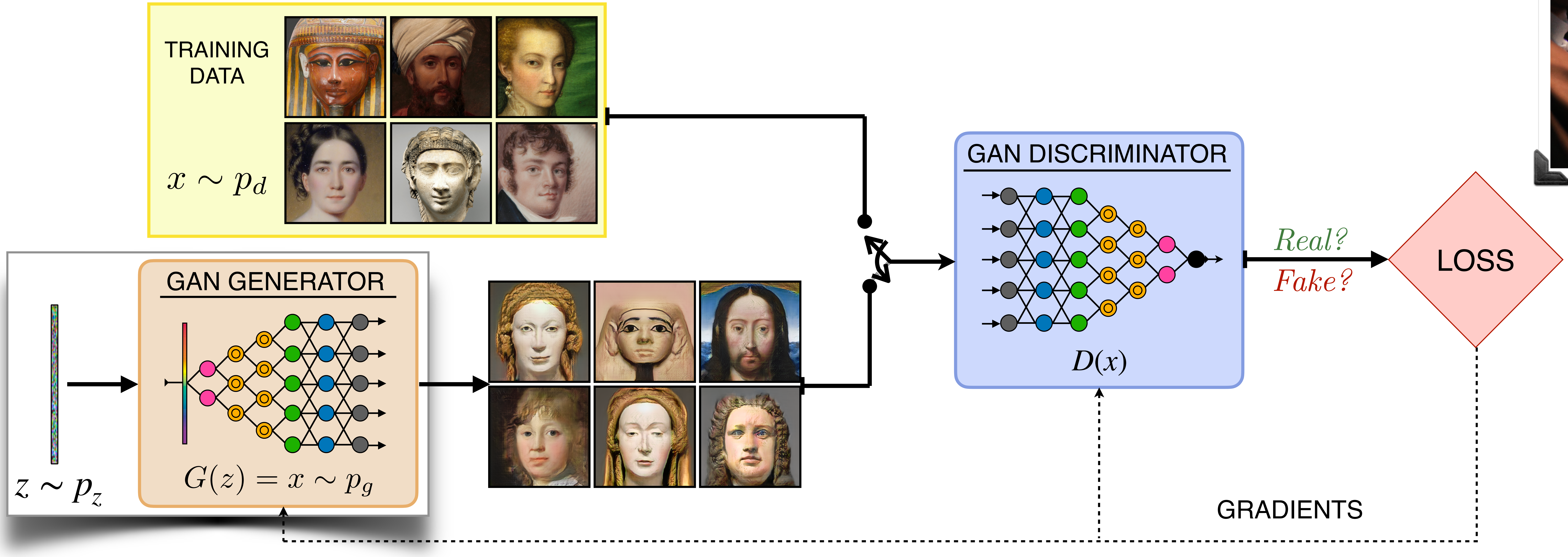


- A min-max game between two players.

[1] Goodfellow et al., "Generative Adversarial Networks," NeurIPS 2014

Generative Adversarial Networks (GANs)

- A framework for training generative models to learn a desired data distribution.

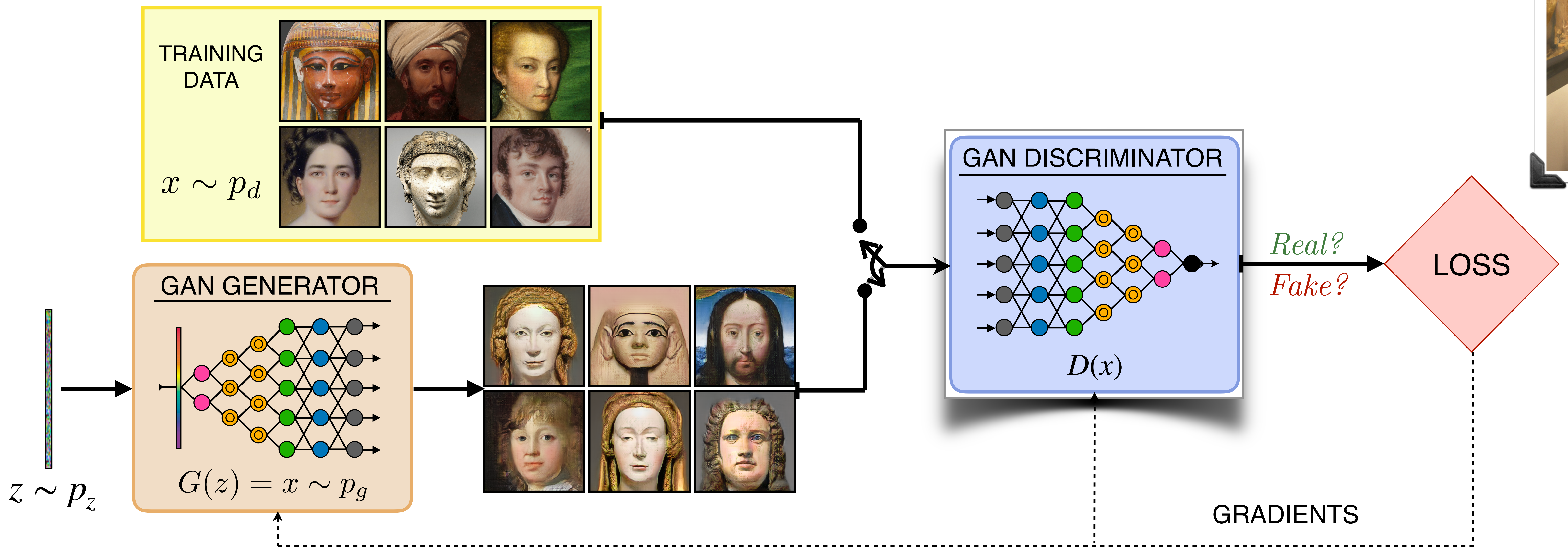


- A min-max game between two players:
  - The generator $G(z)$: which learns and creates samples from the distribution of the real data.

[1] Goodfellow et al., "Generative Adversarial Networks," NeurIPS 2014

Generative Adversarial Networks (GANs)

- A framework for training generative models to learn a desired data distribution.

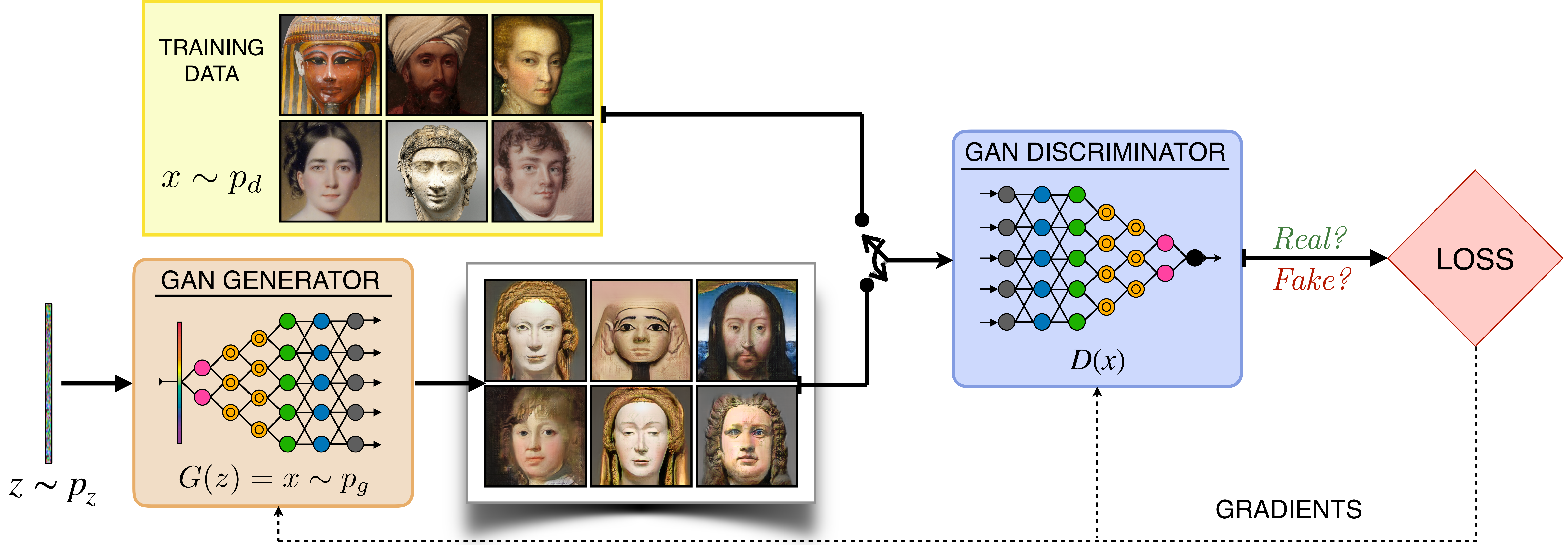


- A min-max game between two players:
  - The generator $G(z)$: which learns and creates samples from the distribution of the real data.
  - The discriminator $D(x)$: which differentiates between the real and fake data.

[1] Goodfellow et al., "Generative Adversarial Networks," NeurIPS 2014

Generative Adversarial Networks (GANs)

- A framework for training generative models to learn a desired data distribution.



- A min-max game between two players:
  - The generator $G(z)$: which learns and creates samples from the distribution of the real data.
  - The discriminator $D(x)$: which differentiates between the real and fake data.
  - The objective: learn the optimal $G(z)$

[1] Goodfellow et al., "Generative Adversarial Networks," NeurIPS 2014

Divergence-minimizing GANs

- The standard GAN (SGAN)[1], least-squares GAN (LSGAN)[2], etc. minimize various choices of $\boldsymbol{f}$**-divergence** between p_d and p_g.

[1]Goodfellow et al., NeurIPS, 2014; [2]Mao et al., ICCV, 2017

Divergence-minimizing GANs

- The standard GAN (SGAN)[1], least-squares GAN (LSGAN)[2], etc. minimize various choices of ***f*-divergence** between p_d and p_g.
  - SGAN: $\min_G \max_D \left\{ \underset{\boldsymbol{x}\sim p_d}{\mathbb{E}} [\ln(D(\boldsymbol{x}))] + \underset{\boldsymbol{x}\sim p_g}{\mathbb{E}} [\ln(1 - D(\boldsymbol{x}))] \right\}$ – Jensen-Shannon divergence.

[1]Goodfellow et al., NeurIPS, 2014; [2]Mao et al., ICCV, 2017

Divergence-minimizing GANs

- The standard GAN (SGAN)[1], least-squares GAN (LSGAN)[2], etc. minimize various choices of $\boldsymbol{f}$**-divergence** between p_d and p_g.
  - SGAN: $\min_G \max_D \left\{ \mathop{\mathbb{E}}_{\boldsymbol{x}\sim p_d} [\ln(D(\boldsymbol{x}))] + \mathop{\mathbb{E}}_{\boldsymbol{x}\sim p_g} [\ln(1 - D(\boldsymbol{x}))] \right\}$ – Jensen-Shannon divergence.
  - LSGAN: $\begin{matrix} \min_D \left\{ \mathop{\mathbb{E}}_{\boldsymbol{x}\sim p_d} \left[(D(\boldsymbol{x}) - b)^2 \right] + \mathop{\mathbb{E}}_{\boldsymbol{x}\sim p_g} \left[(D(\boldsymbol{x}) - a)^2 \right] \right\} \\ \min_G \left\{ \mathop{\mathbb{E}}_{\boldsymbol{x}\sim p_d} \left[(D(\boldsymbol{x}) - c)^2 \right] + \mathop{\mathbb{E}}_{\boldsymbol{x}\sim p_g} \left[(D(\boldsymbol{x}) - c)^2 \right] \right\} \end{matrix}$ – Pearson-χ^2 divergence.

[1]Goodfellow et al., NeurIPS, 2014; [2]Mao et al., ICCV, 2017

Divergence-minimizing GANs

- The standard GAN (SGAN)[1], least-squares GAN (LSGAN)[2], etc. minimize various choices of ***f*-divergence** between p_d and p_g.
  - SGAN: $\min_G \max_D \left\{ \mathbb{E}_{\boldsymbol{x}\sim p_d} [\ln(D(\boldsymbol{x}))] + \mathbb{E}_{\boldsymbol{x}\sim p_g} [\ln(1 - D(\boldsymbol{x}))] \right\}$ – Jensen-Shannon divergence.
  - LSGAN: $\begin{matrix} \min_D \left\{ \mathbb{E}_{\boldsymbol{x}\sim p_d} \left[(D(\boldsymbol{x}) - b)^2\right] + \mathbb{E}_{\boldsymbol{x}\sim p_g} \left[(D(\boldsymbol{x}) - a)^2\right] \right\} \\ \min_G \left\{ \mathbb{E}_{\boldsymbol{x}\sim p_d} \left[(D(\boldsymbol{x}) - c)^2\right] + \mathbb{E}_{\boldsymbol{x}\sim p_g} \left[(D(\boldsymbol{x}) - c)^2\right] \right\} \end{matrix}$ – Pearson-χ^2 divergence.
  - f-GANs: $\min_G \max_D \left\{ \mathbb{E}_{\boldsymbol{x}\sim p_d} [g(D(\boldsymbol{x}))] + \mathbb{E}_{\boldsymbol{x}\sim p_g} [f^c(g(D(\boldsymbol{x})))] \right\}$ – any f-divergence.

[1]Goodfellow et al., NeurIPS, 2014; [2]Mao et al., ICCV, 2017; [3]Nowozin et al., NeurIPS, 2016

Divergence-minimizing GANs

- The standard GAN (SGAN)[1], least-squares GAN (LSGAN)[2], etc. minimize various choices of ***f*-divergence** between p_d and p_g.
  - SGAN: $\min_G \max_D \left\{ \mathbb{E}_{\boldsymbol{x}\sim p_d} [\ln(D(\boldsymbol{x}))] + \mathbb{E}_{\boldsymbol{x}\sim p_g} [\ln(1 - D(\boldsymbol{x}))] \right\}$ – Jensen-Shannon divergence.
  - LSGAN: $\begin{matrix} \min_D \left\{ \mathbb{E}_{\boldsymbol{x}\sim p_d} \left[(D(\boldsymbol{x}) - b)^2 \right] + \mathbb{E}_{\boldsymbol{x}\sim p_g} \left[(D(\boldsymbol{x}) - a)^2 \right] \right\} \\ \min_G \left\{ \mathbb{E}_{\boldsymbol{x}\sim p_d} \left[(D(\boldsymbol{x}) - c)^2 \right] + \mathbb{E}_{\boldsymbol{x}\sim p_g} \left[(D(\boldsymbol{x}) - c)^2 \right] \right\} \end{matrix}$ – Pearson-χ^2 divergence.
  - f-GANs: $\min_G \max_D \left\{ \mathbb{E}_{\boldsymbol{x}\sim p_d} [g(D(\boldsymbol{x}))] + \mathbb{E}_{\boldsymbol{x}\sim p_g} [f^c(g(D(\boldsymbol{x})))] \right\}$ – any f-divergence.
- Training is unstable when the supports of p_d and p_g are non-overlapping.

[1]Goodfellow et al., NeurIPS, 2014; [2]Mao et al., ICCV, 2017; [3]Nowozin et al., NeurIPS, 2016

Integral Probability Metric GANs

- The GAN discriminator minimizes a chosen distance metric – Appropriate constraint space on the class of discriminator functions.

- Popular variants – Wasserstein GAN (WGAN)[4], Sobolev GAN[5], Fisher GAN[6]

[4] Arjovsky et al., ICML, 2017; [5] Mroueh et al., ICLR, 2018; [6] Mroueh et al., NeurIPS, 2017

Integral Probability Metric GANs

- The GAN discriminator minimizes a chosen distance metric – Appropriate constraint space on the class of discriminator functions.

- Popular variants – Wasserstein GAN (WGAN)[4], Sobolev GAN[5], Fisher GAN[6]

  – WGAN: $\min_{G} \max_{D} \left\{ \mathop{\mathbb{E}}_{\boldsymbol{x}\sim p_d} [D(\boldsymbol{x})] - \mathop{\mathbb{E}}_{\boldsymbol{x}\sim p_g} [D(\boldsymbol{x})] \right\}$ – Lipschitz-1 space.

[4] Arjovsky et al., ICML, 2017; [5] Mroueh et al., ICLR, 2018; [6] Mroueh et al., NeurIPS, 2017

Integral Probability Metric GANs

- The GAN discriminator minimizes a chosen distance metric – Appropriate constraint space on the class of discriminator functions.

- Popular variants – Wasserstein GAN (WGAN)[4], Sobolev GAN[5], Fisher GAN[6]
  - WGAN: $\min_G \max_D \left\{ \mathop{\mathbb{E}}_{\boldsymbol{x}\sim p_d} [D(\boldsymbol{x})] - \mathop{\mathbb{E}}_{\boldsymbol{x}\sim p_g} [D(\boldsymbol{x})] \right\}$ – Lipschitz-1 space.
  - Sobolev GAN: $\min_G \max_D \left\{ \mathop{\mathbb{E}}_{\boldsymbol{x}\sim p_d} [D(\boldsymbol{x})] - \mathop{\mathbb{E}}_{\boldsymbol{x}\sim p_g} [D(\boldsymbol{x})] + \mathop{\mathbb{E}}_{\boldsymbol{x}\sim \mu(\boldsymbol{x})} \left[\|\nabla D(\boldsymbol{x})\|_2^2 \right] \right\}$ – Sobolev spaces.

[4]Arjovsky et al., ICML, 2017; [5]Mroueh et al., ICLR, 2018; [6]Mroueh et al., NeurIPS, 2017

IPM-GANs and Gradient Penalties

- WGANs[4] consider a Lipschitz-1 constraint on the discriminator.

$$\min_G \max_D \left\{ \mathbb{E}_{\boldsymbol{x} \sim p_d} \left[D(\boldsymbol{x}) \right] - \mathbb{E}_{\boldsymbol{x} \sim p_g} \left[D(\boldsymbol{x}) \right] \right\}, \quad \text{s.t.} \quad \|D\|_{\mathrm{L}} \leq 1$$

[4] Arjovsky et al., ICML, 2017;

IPM-GANs and Gradient Penalties

- WGANs[4] consider a Lipschitz-1 constraint on the discriminator.

$$\min_G \max_D \left\{\mathbb{E}_{\boldsymbol{x}\sim p_d}\left[D(\boldsymbol{x})\right] - \mathbb{E}_{\boldsymbol{x}\sim p_g}\left[D(\boldsymbol{x})\right]\right\}, \quad \text{s.t.} \quad \|D\|_{\mathrm{L}} \leq 1$$

- WGAN-GP[7] employs a gradient penalty to approximate the Lipschitz-1 constraint.

$$\min_G \max_D \left\{\mathbb{E}_{\boldsymbol{x}\sim p_d}\left[D(\boldsymbol{x})\right] - \mathbb{E}_{\boldsymbol{x}\sim p_g}\left[D(\boldsymbol{x})\right] + \lambda\,\mathbb{E}_{\boldsymbol{x}\sim p_g,\,\hat{\boldsymbol{x}}\sim p_d}\left[\left(\|\nabla D(\alpha\boldsymbol{x} + (1-\alpha)\hat{\boldsymbol{x}})\|_2 - 1\right)^2\right]\right\}$$

[4]Arjovsky et al., ICML, 2017; [7]Gulrajani et al., NeurIPS, 2017

IPM-GANs and Gradient Penalties

- WGANs[4] consider a Lipschitz-1 constraint on the discriminator.

$$\min_G \max_D \left\{\mathbb{E}_{\boldsymbol{x}\sim p_d}\left[D(\boldsymbol{x})\right] - \mathbb{E}_{\boldsymbol{x}\sim p_g}\left[D(\boldsymbol{x})\right]\right\}, \quad \text{s.t.} \quad \|D\|_{\mathrm{L}} \leq 1$$

- WGAN-GP[7] employs a gradient penalty to approximate the Lipschitz-1 constraint.

$$\min_G \max_D \left\{\mathbb{E}_{\boldsymbol{x}\sim p_d}\left[D(\boldsymbol{x})\right] - \mathbb{E}_{\boldsymbol{x}\sim p_g}\left[D(\boldsymbol{x})\right] + \lambda\,\mathbb{E}_{\boldsymbol{x}\sim p_g,\,\hat{\boldsymbol{x}}\sim p_d}\left[\left(\|\nabla D(\alpha\boldsymbol{x} + (1-\alpha)\hat{\boldsymbol{x}})\|_2 - 1\right)^2\right]\right\}$$

- Variants such as WGAN-LP[8], WGAN-ALP[9], WGAN-R_d+R_g[10] also incorporate gradient penalties on WGAN to improve training stability.

[4]Arjovsky et al., ICML, 2017; [7]Gulrajani et al., NeurIPS, 2017; [8]Petzka et al., ICLR, 2018; [9]Terjék, ICLR, 2020; [10] Mescheder et al., ICML, 2018

IPM-GANs and Gradient Penalties

- Existing losses typically enforce the penalty over:
  - Interpolations between $\boldsymbol{x} \sim p_g$ and $\tilde{\boldsymbol{x}} \sim p_d$.
  - A generic reference measure $\nu_p(\boldsymbol{x})$,
  - Adversarially derived *worst-case* directions $\boldsymbol{r}_{adv}$.

WGAN flavor	Discriminator loss
WGAN	$\mathcal{L}_D^{\mathrm{W}} = -\mathop{\mathbb{E}}_{\boldsymbol{x}\sim p_d}[D(\boldsymbol{x})] + \mathop{\mathbb{E}}_{\boldsymbol{x}\sim p_g}[D(\boldsymbol{x})]$
WGAN-GP	$\mathcal{L}_D^{\mathrm{W}} + \lambda \mathop{\mathbb{E}}_{\boldsymbol{x}\sim \alpha p_g + (1-\alpha)p_d}\left[(\Vert\nabla D(\boldsymbol{x})\Vert_2 - 1)^2\right];\ 0 \leq \alpha \leq 1$
WGAN-$\mathrm{R}_d\mathrm{R}_g$	$\mathcal{L}_D^{\mathrm{W}} + \frac{\lambda_1}{2}\mathop{\mathbb{E}}_{\boldsymbol{x}\sim p_d}\left[\Vert\nabla D(\boldsymbol{x})\Vert_2^2\right] + \frac{\lambda_2}{2}\mathop{\mathbb{E}}_{\boldsymbol{x}\sim p_g}\left[\Vert\nabla D(\boldsymbol{x})\Vert_2^2\right]$
Sobolev GAN	$\mathcal{L}_D^{\mathrm{W}} + \lambda \mathop{\mathbb{E}}_{\boldsymbol{x}\sim \nu_p(\boldsymbol{x})}\left[\Vert\nabla D(\boldsymbol{x})\Vert_2^2\right]$, where $\nu_p(\boldsymbol{x}) \geq 0;\ \int_{\mathcal{X}} \nu_p(\boldsymbol{x})\mathrm{d}\boldsymbol{x} = 1$
WGAN-LP	$\mathcal{L}_D^{\mathrm{W}} + \lambda \mathop{\mathbb{E}}_{\boldsymbol{x}\sim \alpha p_g + (1-\alpha)p_d}\left[(\max(\Vert\nabla D(\boldsymbol{x})\Vert_2 - 1, 0))^2\right];\ 0 \leq \alpha \leq 1$
WGAN-ALP	$\mathcal{L}_D^{\mathrm{W}} + \lambda \mathop{\mathbb{E}}_{\boldsymbol{x}\sim p_d}\left[\left(\max\left(\frac{D(\boldsymbol{x}) - D(\boldsymbol{x}+\boldsymbol{r}_{adv})}{\Vert\boldsymbol{r}_{adv}\Vert_2} - 1, 0\right)\right)^2\right]$, where $\boldsymbol{r}_{adv} = \max_{\boldsymbol{r}:\Vert\boldsymbol{r}\Vert_2>0}\left\{\frac{D(\boldsymbol{x})-D(\boldsymbol{x}+\boldsymbol{r})}{\Vert\boldsymbol{r}\Vert_2}\right\}$

Our Results

- Analyzed the optimal discriminator in divergence-minimizing and IPM GANs leveraging results from *functional Calculus*.
- Showed that the discriminator in WGAN with a gradient-norm penalty (WGAN-GNP) is governed by a Poisson PDE.
- Fourier-series-based approximations to the optimal discriminator in WGAN-GNP.
- Optimal Lagrange multiplier servses to mesure generator convergence!
- Superior performance in latent-space matching with Wasserstein autoencoders.

Asokan and Seelamantula, *"Euler-Lagrange Analysis of Generative Adversarial Networks,"* JMLR, 2023

Functional Calculus

- Variational Calculus deals with **functional** optimization:

$$\min_{D\in\mathcal{G}} \left\{ \mathcal{L}(\boldsymbol{x}, D(\boldsymbol{x}), \nabla D(\boldsymbol{x})) = \int_{\mathcal{X}} \mathcal{F}(\boldsymbol{x}, D(\boldsymbol{x}), \nabla D(\boldsymbol{x}))\mathrm{d}\boldsymbol{x} \right\} \quad \Rightarrow \quad \delta\mathcal{L}\Big|_{D=D^*} = 0$$

- The optimality condition on the *first variation* ($\delta\mathcal{L}$) is the **Euler-Lagrange** condition:

$$\frac{\partial \mathcal{F}}{\partial D} - \sum_i \frac{\partial}{\partial x_i} \frac{\partial \mathcal{F}}{\partial (D_i')}\Bigg|_{D(\boldsymbol{x})=D^*(\boldsymbol{x})} = 0, \quad \text{where} \quad D_i' = \frac{\partial D}{\partial x_i}$$

- The check on the *second-variation* to identify a minimizer is $\delta^2\mathcal{L} \geq 0 \Rightarrow \mathbb{H}_{y,\mathcal{H}}\Big|_{y=y^*} \succ 0$,

$$\text{where} \quad \mathcal{H} = \sum_{i=1}^{n} \left(D_i' \frac{\partial \mathcal{F}}{\partial D_i'} \right) - \mathcal{F} \quad \text{and} \quad [\mathbb{H}_{D,\mathcal{H}}]_{i,j} = \frac{\partial^2 \mathcal{H}}{\partial D_i' \partial D_j'}.$$

WGAN with Gradient-norm Penalty (WGAN-GNP)

- The WGAN-GNP discriminator loss:

$$\mathcal{L}_D^{\mathrm{W}} = \mathop{\mathbb{E}}_{\boldsymbol{x}\sim p_d}[D(\boldsymbol{x})] - \mathop{\mathbb{E}}_{\boldsymbol{x}\sim p_g}[D(\boldsymbol{x})] + \boxed{\lambda_d \int_{\mathcal{X}} \left(\|\nabla D(\boldsymbol{x})\|_2^2 - 1\right) \mathrm{d}\boldsymbol{x}}$$

Asokan and Seelamantula, *"Euler-Lagrange Analysis of Generative Adversarial Networks,"* JMLR, 2023

WGAN with Gradient-norm Penalty (WGAN-GNP)

- The WGAN-GNP discriminator loss:

$$\begin{aligned}\mathcal{L}_D^{\mathrm{W}} &= \underset{\boldsymbol{x}\sim p_d}{\mathbb{E}}\left[D(\boldsymbol{x})\right] - \underset{\boldsymbol{x}\sim p_g}{\mathbb{E}}\left[D(\boldsymbol{x})\right] + \lambda_d \int_{\mathcal{X}} \left(\|\nabla D(\boldsymbol{x})\|_2^2 - 1\right)\,\mathrm{d}\boldsymbol{x} \\ &= \int_{\mathcal{X}} \Big(D(\boldsymbol{x})\left(p_d(\boldsymbol{x}) - p_g(\boldsymbol{x})\right) + \lambda_d \|\nabla D(\boldsymbol{x})\|_2^2\Big)\,\mathrm{d}\boldsymbol{x} - \lambda_d|\mathcal{X}|\end{aligned}$$

Asokan and Seelamantula, *"Euler-Lagrange Analysis of Generative Adversarial Networks,"* JMLR, 2023

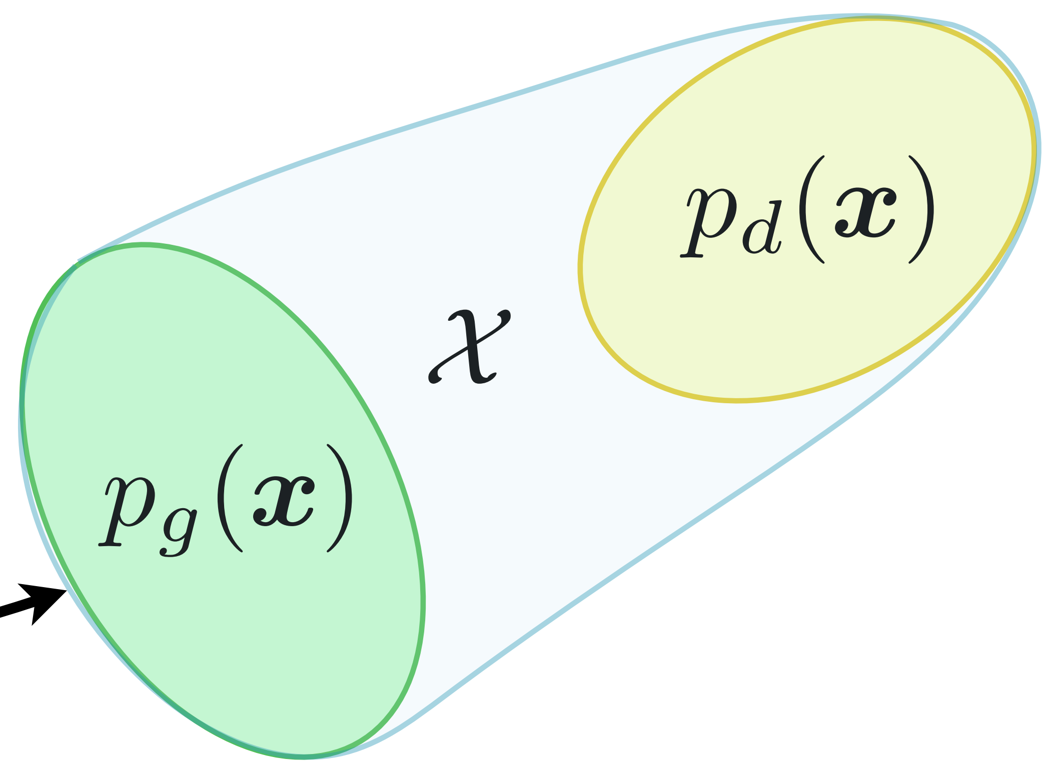
WGAN with Gradient-norm Penalty (WGAN-GNP)

- The WGAN-GNP discriminator loss:

$$\mathcal{L}_D^{\mathrm{W}} = \underset{\boldsymbol{x}\sim p_d}{\mathbb{E}}[D(\boldsymbol{x})] - \underset{\boldsymbol{x}\sim p_g}{\mathbb{E}}[D(\boldsymbol{x})] + \lambda_d \int_{\mathcal{X}} \left(\|\nabla D(\boldsymbol{x})\|_2^2 - 1\right) \mathrm{d}\boldsymbol{x}$$

$$= \int_{\mathcal{X}} \underbrace{\left(D(\boldsymbol{x})\left(p_d(\boldsymbol{x}) - p_g(\boldsymbol{x})\right) + \lambda_d \|\nabla D(\boldsymbol{x})\|_2^2\right)}_{\mathcal{F}(\boldsymbol{x}, D(\boldsymbol{x}), \nabla D(\boldsymbol{x}))} \mathrm{d}\boldsymbol{x} - \lambda_d |\mathcal{X}|$$

The convex hull of the supports of p_d and p_g.



Asokan and Seelamantula, *"Euler-Lagrange Analysis of Generative Adversarial Networks,"* JMLR, 2023

WGAN with Gradient-norm Penalty (WGAN-GNP)

- The WGAN-GNP discriminator loss:

$$\begin{aligned}\mathcal{L}_D^{\mathrm{W}} &= \mathop{\mathbb{E}}_{\boldsymbol{x}\sim p_d}[D(\boldsymbol{x})] - \mathop{\mathbb{E}}_{\boldsymbol{x}\sim p_g}[D(\boldsymbol{x})] + \lambda_d \int_{\mathcal{X}} \left(\|\nabla D(\boldsymbol{x})\|_2^2 - 1\right) \mathrm{d}\boldsymbol{x} \\ &= \int_{\mathcal{X}} \underbrace{\left(D(\boldsymbol{x})\left(p_d(\boldsymbol{x}) - p_g(\boldsymbol{x})\right) + \lambda_d \|\nabla D(\boldsymbol{x})\|_2^2\right)}_{\mathcal{F}(\boldsymbol{x}, D(\boldsymbol{x}), \nabla D(\boldsymbol{x}))} \mathrm{d}\boldsymbol{x} - \lambda_d |\mathcal{X}|\end{aligned}$$

- Applying Euler-Lagrange (EL) condition to $\mathcal{F}$, we have:

$$\frac{\partial \mathcal{F}}{\partial D} = p_d(\boldsymbol{x}) - p_g(\boldsymbol{x}), \quad \text{and} \quad \sum_i \frac{\partial}{\partial x_i} \frac{\partial \mathcal{F}}{\partial D_i'} = \sum_i \frac{\partial^2 D}{\partial x_i^2} = \Delta D(\boldsymbol{x})$$

$$\Rightarrow \boxed{\Delta D(\boldsymbol{x})\big|_{D=D^*} = \frac{p_d(\boldsymbol{x}) - p_g(\boldsymbol{x})}{2\lambda_d},} \ \text{where } \Delta = \nabla.\nabla = \left(\partial_{x_1}^2 + \partial_{x_2}^2 + \ldots + \partial_{x_n}^2\right)$$

Asokan and Seelamantula, *"Euler-Lagrange Analysis of Generative Adversarial Networks,"* JMLR, 2023

Green's Function

- An approach for obtaining the *particular solution* to PDEs

- Given a differential operator $\mathfrak{L}f(\boldsymbol{x}) = g(\boldsymbol{x})$, find a function $\phi(\boldsymbol{x})$ such that $\mathfrak{L}\phi(\boldsymbol{x}) = \delta(\boldsymbol{x})$.

Also called fundamental solution

Green's Function

- An approach for obtaining the *particular solution* to PDEs
- Given a differential operator $\mathfrak{L}f(\boldsymbol{x}) = g(\boldsymbol{x})$, find a function $\phi(\boldsymbol{x})$ such that $\mathfrak{L}\phi(\boldsymbol{x}) = \delta(\boldsymbol{x})$.

  Also called fundamental solution
- Convolving both sides with $\phi(\boldsymbol{x})$ yields

$$\mathfrak{L}(f * \phi)(\boldsymbol{x}) = (g * \phi)(\boldsymbol{x})$$

$$(f * \mathfrak{L}\phi)(\boldsymbol{x}) = (g * \phi)(\boldsymbol{x})$$

$$\Rightarrow f(\boldsymbol{x}) = (g * \phi)(\boldsymbol{x})$$

Fundamental Solution to WGAN-GNP

- The WGAN-GNP discriminator PDE

$$\Delta D(\boldsymbol{x})\big|_{D=D^*} = \frac{p_g(\boldsymbol{x}) - p_d(\boldsymbol{x})}{2\lambda_d}, \text{ where } \Delta = \nabla.\nabla = \left(\partial_{x_1}^2 + \partial_{x_2}^2 + \ldots + \partial_{x_n}^2\right)$$

- The *particular solution* to the discriminator PDE:

$$D_p^*(\boldsymbol{x}) = \frac{1}{2\lambda_d}\left((p_g - p_d) * \phi\right)(\boldsymbol{x}), \quad \text{where} \quad \phi(\boldsymbol{x}) = \begin{cases} -\dfrac{1}{2\pi}\ln(\|\boldsymbol{x}\|), & \text{for } n = 2, \text{ and} \\ \dfrac{\Gamma\left(\frac{n}{2}+1\right)}{n(n-2)\pi^{\frac{n}{2}}}\dfrac{1}{\|\boldsymbol{x}\|^{n-2}}, & \text{for } n \geq 3, \end{cases}$$

Fundamental Solution to WGAN-GNP

- The WGAN-GNP discriminator PDE

$$\Delta D\left(\boldsymbol{x}\right)\Big|_{D=D^*} = \frac{p_g(\boldsymbol{x}) - p_d(\boldsymbol{x})}{2\lambda_d}, \text{ where } \Delta = \nabla.\nabla = \left(\partial^2_{x_1} + \partial^2_{x_2} + \ldots + \partial^2_{x_n}\right)$$

- The *particular solution* to the discriminator PDE:

$$D^*_p(\boldsymbol{x}) = \frac{1}{2\lambda_d}\left(\left(p_g - p_d\right) * \phi\right)(\boldsymbol{x}), \quad \text{where} \quad \phi(\boldsymbol{x}) = \begin{cases} -\dfrac{1}{2\pi}\ln(\|\boldsymbol{x}\|), & \text{for } n = 2, \text{ and} \\ \dfrac{\Gamma\left(\frac{n}{2}+1\right)}{n(n-2)\pi^{\frac{n}{2}}}\dfrac{1}{\|\boldsymbol{x}\|^{n-2}}, & \text{for } n \geq 3, \end{cases}$$

- The solution $D^*(\boldsymbol{x})$ must also include the *homogeneous component* $D_h(\boldsymbol{x})$ such that $\Delta D_h(\boldsymbol{x}) = 0$.

Fundamental Solution to WGAN-GNP

- The WGAN-GNP discriminator PDE

$$\Delta D(\boldsymbol{x})\Big|_{D=D^*} = \frac{p_g(\boldsymbol{x}) - p_d(\boldsymbol{x})}{2\lambda_d}, \text{ where } \Delta = \nabla.\nabla = \left(\partial^2_{x_1} + \partial^2_{x_2} + \ldots + \partial^2_{x_n}\right)$$

- The *particular solution* to the discriminator PDE:

$$D_p^*(\boldsymbol{x}) = \frac{1}{2\lambda_d}\left((p_g - p_d) * \phi\right)(\boldsymbol{x}), \quad \text{where} \quad \phi(\boldsymbol{x}) = \begin{cases} -\frac{1}{2\pi}\ln(\|\boldsymbol{x}\|), & \text{for } n = 2, \text{ and} \\ \frac{\Gamma\left(\frac{n}{2}+1\right)}{n(n-2)\pi^{\frac{n}{2}}} \frac{1}{\|\boldsymbol{x}\|^{n-2}}, & \text{for } n \geq 3, \end{cases}$$

- The solution $D^*(\boldsymbol{x})$ must also include the *homogeneous component* $D_h(\boldsymbol{x})$ such that $\Delta D_h(\boldsymbol{x}) = 0$.

Family of all affine functions: $\langle \boldsymbol{a}, \boldsymbol{x} \rangle + a_0$

Fundamental Solution to WGAN-GNP

- The WGAN-GNP discriminator PDE

$$\Delta D(\boldsymbol{x})\big|_{D=D^*} = \frac{p_g(\boldsymbol{x}) - p_d(\boldsymbol{x})}{2\lambda_d}, \quad \text{where } \Delta = \nabla.\nabla = \left(\partial^2_{x_1} + \partial^2_{x_2} + \ldots + \partial^2_{x_n}\right)$$

- The *particular solution* to the discriminator PDE:

$$D_p^*(\boldsymbol{x}) = \frac{1}{2\lambda_d}\left((p_g - p_d) * \phi\right)(\boldsymbol{x}), \quad \text{where} \quad \phi(\boldsymbol{x}) = \begin{cases} -\frac{1}{2\pi}\ln(\|\boldsymbol{x}\|), & \text{for } n = 2, \text{ and} \\ \frac{\Gamma\left(\frac{n}{2}+1\right)}{n(n-2)\pi^{\frac{n}{2}}}\frac{1}{\|\boldsymbol{x}\|^{n-2}}, & \text{for } n \geq 3, \end{cases}$$

- The solution $D^*(\boldsymbol{x})$ must also include the *homogeneous component* $D_h(\boldsymbol{x})$ such that $\Delta D_h(\boldsymbol{x}) = 0$.

- **Theorem:** The generator convergence $p_g^*(\boldsymbol{x}) = p_d(\boldsymbol{x})$ is independent of the choice of $D_h(\boldsymbol{x})$

Implementing WGAN-GNP

- Two alternatives for implementing the optimal discriminator.

$$\Delta D(\boldsymbol{x})\big|_{D=D^*} = \frac{p_g(\boldsymbol{x}) - p_d(\boldsymbol{x})}{2\lambda_d},$$

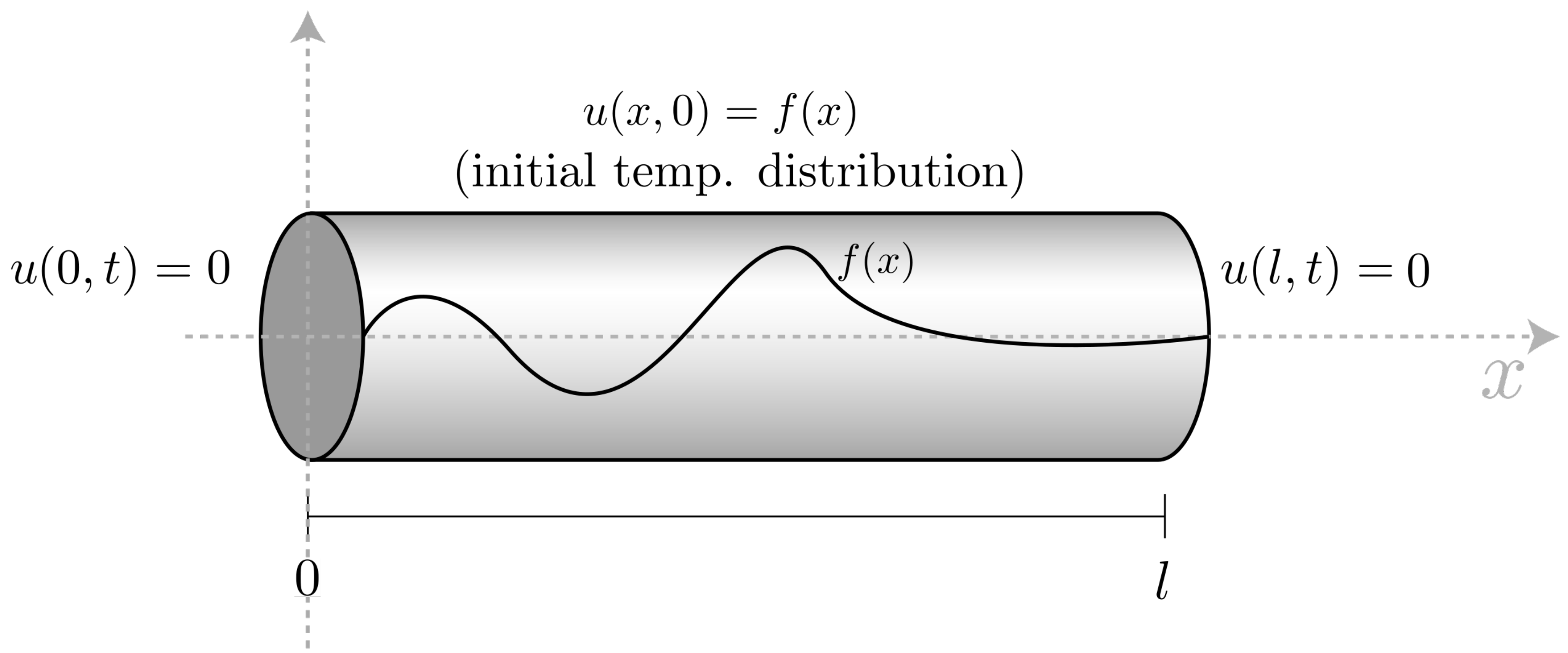
Leverage the form of the PDE to derive an implementable form of the discriminator.

$$D_p^*(\boldsymbol{x}) = \frac{1}{2\lambda_d}\left((p_g - p_d) * \phi\right)(\boldsymbol{x}),$$

Develop a neural-network approximation of the convolution-based solution.

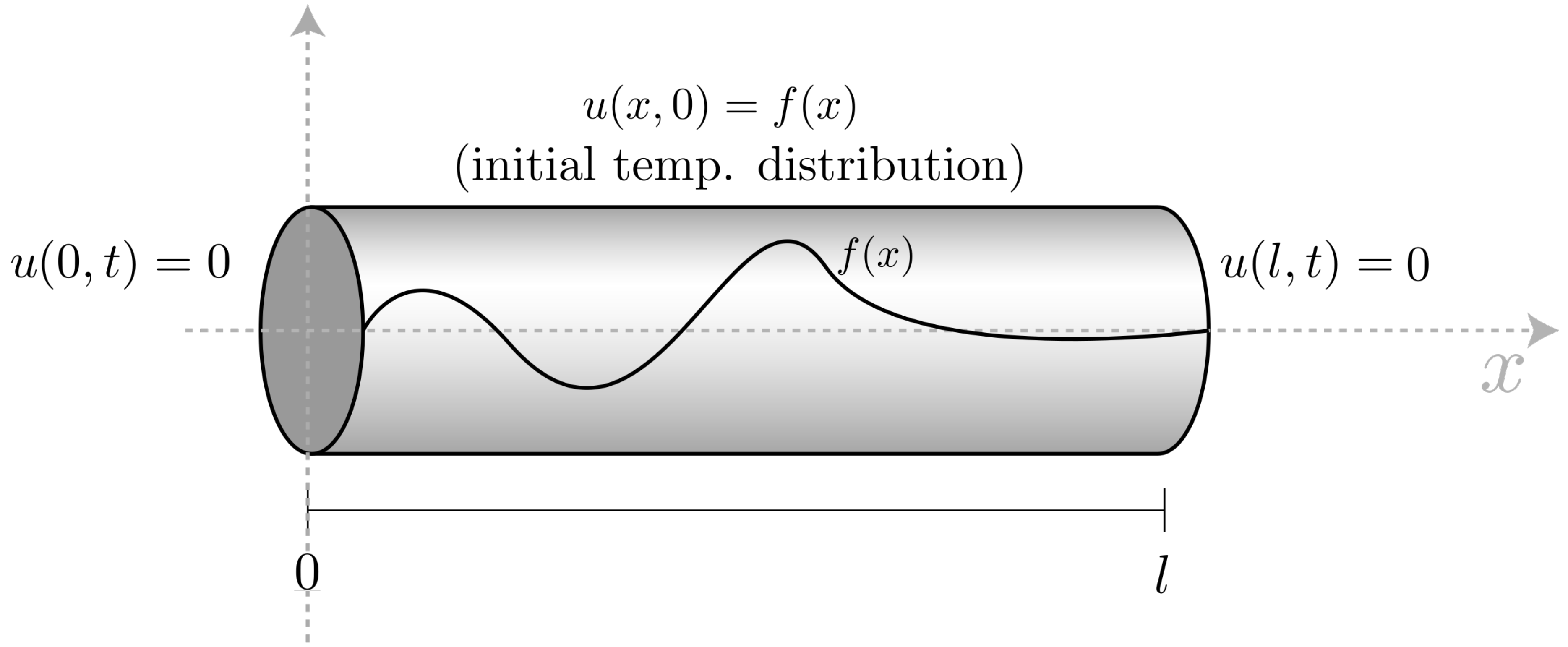
The Fourier-series Discriminator

- Fourier series – Originally proposed to solve the *heat equation* by Fourier in 1822.



The Fourier-series Discriminator

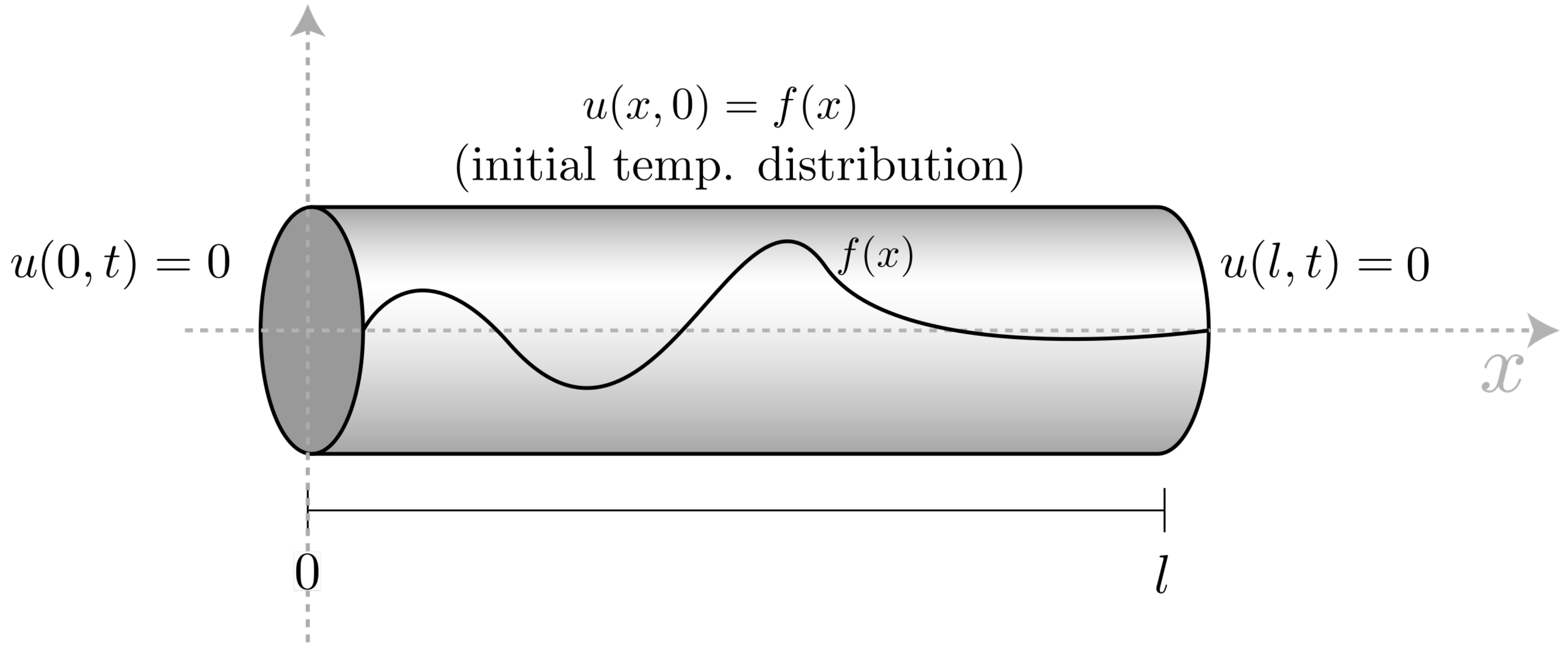
- Fourier series – Originally proposed to solve the *heat equation* by Fourier in 1822.



- Derive Fourier-series expansions of the data and generator distributions, and the discriminator

$$p_d(\boldsymbol{x}) = \sum_{\boldsymbol{m} \in \mathbb{Z}^n} \alpha_{\boldsymbol{m}} e^{\mathrm{j}\langle \boldsymbol{\omega_m}, \boldsymbol{x} \rangle}, \quad p_g(\boldsymbol{x}) = \sum_{\boldsymbol{m} \in \mathbb{Z}^n} \beta_{\boldsymbol{m}} e^{\mathrm{j}\langle \boldsymbol{\omega_m}, \boldsymbol{x} \rangle}, \quad \text{and} \quad D_{FS}(\boldsymbol{x}) = \frac{1}{\lambda_d} \sum_{\boldsymbol{m} \in \mathbb{Z}^n} \gamma_{\boldsymbol{m}} e^{\mathrm{j}\langle \boldsymbol{\omega_m}, \boldsymbol{x} \rangle}$$

The Fourier-series Discriminator



- Fourier series – Originally proposed to solve the *heat equation* by Fourier in 1822.

- Derive Fourier-series expansions of the data and generator distributions, and the discriminator

$$p_d(\boldsymbol{x}) = \sum_{\boldsymbol{m}\in\mathbb{Z}^n} \alpha_{\boldsymbol{m}} e^{\mathrm{j}\langle \boldsymbol{\omega}_{\boldsymbol{m}}, \boldsymbol{x}\rangle}, \quad p_g(\boldsymbol{x}) = \sum_{\boldsymbol{m}\in\mathbb{Z}^n} \beta_{\boldsymbol{m}} e^{\mathrm{j}\langle \boldsymbol{\omega}_{\boldsymbol{m}}, \boldsymbol{x}\rangle}, \quad \text{and} \quad D_{FS}(\boldsymbol{x}) = \frac{1}{\lambda_d} \sum_{\boldsymbol{m}\in\mathbb{Z}^n} \gamma_{\boldsymbol{m}} e^{\mathrm{j}\langle \boldsymbol{\omega}_{\boldsymbol{m}}, \boldsymbol{x}\rangle}$$

- Coefficients $\alpha_{\boldsymbol{m}}$ and $\beta_{\boldsymbol{m}}$ can be computed via the characteristic function of the *p.d.f.*

$$\alpha_{\boldsymbol{m}} = \frac{1}{T}\int_0^T p_d(\boldsymbol{x}) e^{\mathrm{j}\langle \boldsymbol{\omega}_{\boldsymbol{m}}, \boldsymbol{x}\rangle}\, \mathrm{d}\boldsymbol{x} = \frac{1}{T} \mathop{\mathbb{E}}_{x\sim p_d}\left[e^{\mathrm{j}\langle \boldsymbol{\omega}_{\boldsymbol{m}}, \boldsymbol{x}\rangle}\right] = \frac{1}{T}\varphi^*_{p_d}(\boldsymbol{\omega}_{\boldsymbol{m}}) \approx \frac{1}{NT}\sum_{k=1}^{N} e^{\mathrm{j}\langle \boldsymbol{\omega}_{\boldsymbol{m}}, \boldsymbol{x}_k\rangle}$$

The Fourier-series Discriminator

- Derive Fourier-series expansions of the data and generator distributions, and the discriminator

$$p_d(\boldsymbol{x}) = \sum_{\boldsymbol{m}\in\mathbb{Z}^n} \alpha_{\boldsymbol{m}} e^{\mathrm{j}\langle \boldsymbol{\omega}_{\boldsymbol{m}}, \boldsymbol{x}\rangle}, \quad p_g(\boldsymbol{x}) = \sum_{\boldsymbol{m}\in\mathbb{Z}^n} \beta_{\boldsymbol{m}} e^{\mathrm{j}\langle \boldsymbol{\omega}_{\boldsymbol{m}}, \boldsymbol{x}\rangle}, \quad \text{and} \quad D_{FS}(\boldsymbol{x}) = \frac{1}{\lambda_d}\sum_{\boldsymbol{m}\in\mathbb{Z}^n} \gamma_{\boldsymbol{m}} e^{\mathrm{j}\langle \boldsymbol{\omega}_{\boldsymbol{m}}, \boldsymbol{x}\rangle}$$

- Coefficients $\alpha_{\boldsymbol{m}}$ and $\beta_{\boldsymbol{m}}$ can be computed via the characteristic function of the *p.d.f.*

$$\alpha_{\boldsymbol{m}} = \frac{1}{T}\int_0^T p_d(\boldsymbol{x}) e^{\mathrm{j}\langle \boldsymbol{\omega}_{\boldsymbol{m}}, \boldsymbol{x}\rangle}\,\mathrm{d}\boldsymbol{x} = \frac{1}{T}\mathop{\mathbb{E}}_{x\sim p_d}\left[e^{\mathrm{j}\langle \boldsymbol{\omega}_{\boldsymbol{m}}, \boldsymbol{x}\rangle}\right] = \frac{1}{T}\varphi^*_{p_d}(\boldsymbol{\omega}_{\boldsymbol{m}}) \approx \frac{1}{NT}\sum_{k=1}^{N} e^{\mathrm{j}\langle \boldsymbol{\omega}_{\boldsymbol{m}}, \boldsymbol{x}_k\rangle}$$

- The discriminator coefficients can be computed in closed form:

$$\gamma_{\boldsymbol{m}} = \frac{1}{2}\left(\frac{\alpha_{\boldsymbol{m}} - \beta_{\boldsymbol{m}}}{\|\boldsymbol{\omega}_{\boldsymbol{m}}\|^2}\right), \quad \boldsymbol{m} \in \mathbb{Z}^n - \{\mathbf{0}\}, \quad \text{and} \quad \gamma_{\mathbf{0}} = 0$$

Asokan and Seelamantula, *"Euler-Lagrange Analysis of Generative Adversarial Networks,"* JMLR, 2023

WGAN with Fourier-series Discriminator

- Derive Fourier-series expansions of the data and generator distributions, and the discriminator

$$p_d(\boldsymbol{x}) = \sum_{\boldsymbol{m}\in\mathbb{Z}^n} \alpha_{\boldsymbol{m}} e^{\mathrm{j}\langle \boldsymbol{\omega}_{\boldsymbol{m}}, \boldsymbol{x}\rangle}, \quad p_g(\boldsymbol{x}) = \sum_{\boldsymbol{m}\in\mathbb{Z}^n} \beta_{\boldsymbol{m}} e^{\mathrm{j}\langle \boldsymbol{\omega}_{\boldsymbol{m}}, \boldsymbol{x}\rangle}, \quad \text{and} \quad D_{FS}(\boldsymbol{x}) = \frac{1}{\lambda_d} \sum_{\boldsymbol{m}\in\mathbb{Z}^n} \gamma_{\boldsymbol{m}} e^{\mathrm{j}\langle \boldsymbol{\omega}_{\boldsymbol{m}}, \boldsymbol{x}\rangle}.$$

- Coefficients can be computed in closed form:

$$\alpha_{\boldsymbol{m}} = \frac{1}{T}\varphi^*_{p_d}(\boldsymbol{\omega}_{\boldsymbol{m}}), \quad \beta_{\boldsymbol{m}} = \frac{1}{T}\varphi^*_{p_g}(\boldsymbol{\omega}_{\boldsymbol{m}}), \quad \text{and} \quad \gamma_{\boldsymbol{m}} = \frac{1}{2}\left(\frac{\alpha_{\boldsymbol{m}} - \beta_{\boldsymbol{m}}}{\|\boldsymbol{\omega}_{\boldsymbol{m}}\|^2}\right), \quad \boldsymbol{m} \neq 0;\ \gamma_{\mathbf{0}} = 0$$

WGAN with Fourier-series Discriminator

- Derive Fourier-series expansions of the data and generator distributions, and the discriminator

$$p_d(\boldsymbol{x}) = \sum_{\boldsymbol{m}\in\mathbb{Z}^n} \alpha_{\boldsymbol{m}} e^{\mathrm{j}\langle \boldsymbol{\omega}_{\boldsymbol{m}},\boldsymbol{x}\rangle}, \quad p_g(\boldsymbol{x}) = \sum_{\boldsymbol{m}\in\mathbb{Z}^n} \beta_{\boldsymbol{m}} e^{\mathrm{j}\langle \boldsymbol{\omega}_{\boldsymbol{m}},\boldsymbol{x}\rangle}, \quad \text{and} \quad D_{FS}(\boldsymbol{x}) = \frac{1}{\lambda_d} \sum_{\boldsymbol{m}\in\mathbb{Z}^n} \gamma_{\boldsymbol{m}} e^{\mathrm{j}\langle \boldsymbol{\omega}_{\boldsymbol{m}},\boldsymbol{x}\rangle}.$$

- Coefficients can be computed in closed form:

$$\alpha_{\boldsymbol{m}} = \frac{1}{T}\varphi^*_{p_d}(\boldsymbol{\omega}_{\boldsymbol{m}}), \quad \beta_{\boldsymbol{m}} = \frac{1}{T}\varphi^*_{p_g}(\boldsymbol{\omega}_{\boldsymbol{m}}), \quad \text{and} \quad \gamma_{\boldsymbol{m}} = \frac{1}{2}\left(\frac{\alpha_{\boldsymbol{m}} - \beta_{\boldsymbol{m}}}{\|\boldsymbol{\omega}_{\boldsymbol{m}}\|^2}\right), \quad \boldsymbol{m} \neq 0; \; \gamma_{\boldsymbol{0}} = 0$$

- A truncated Fourier-series is implementable.

$$\tilde{p}_d(\boldsymbol{x}) = \sum_{\boldsymbol{m}\in[M]^n} \alpha_{\boldsymbol{m}} e^{\mathrm{j}\omega_o\langle \boldsymbol{m},\boldsymbol{x}\rangle}, \; \tilde{p}_g(\boldsymbol{x}) = \sum_{\boldsymbol{m}\in[M]^n} \beta_{\boldsymbol{m}} e^{\mathrm{j}\omega_o\langle \boldsymbol{m},\boldsymbol{x}\rangle}, \; \text{and} \; \tilde{D}_{FS}(\boldsymbol{x}) = \frac{1}{\lambda_d} \sum_{\boldsymbol{m}\in[M]^n} \gamma_{\boldsymbol{m}} e^{\mathrm{j}\omega_o\langle \boldsymbol{m},\boldsymbol{x}\rangle},$$

WGAN with Fourier-series Discriminator (WGAN-FS)

- Complex Fourier series is challenging to implement

$$\tilde{p}_d(\boldsymbol{x}) = \sum_{\boldsymbol{m}\in[M]^n} \alpha_{\boldsymbol{m}} e^{\mathrm{j}\omega_o\langle \boldsymbol{m},\boldsymbol{x}\rangle}, \quad \tilde{p}_g(\boldsymbol{x}) = \sum_{\boldsymbol{m}\in[M]^n} \beta_{\boldsymbol{m}} e^{\mathrm{j}\omega_o\langle \boldsymbol{m},\boldsymbol{x}\rangle}, \quad \text{and} \quad \tilde{D}_{FS}(\boldsymbol{x}) = \frac{1}{\lambda_d}\sum_{\boldsymbol{m}\in[M]^n} \gamma_{\boldsymbol{m}} e^{\mathrm{j}\omega_o\langle \boldsymbol{m},\boldsymbol{x}\rangle},$$

WGAN with Fourier-series Discriminator (WGAN-FS)

- Complex Fourier series is challenging to implement

$$\tilde{p}_d(\boldsymbol{x}) = \sum_{\boldsymbol{m}\in[M]^n} \alpha_{\boldsymbol{m}} e^{\mathrm{j}\omega_o\langle \boldsymbol{m},\boldsymbol{x}\rangle}, \quad \tilde{p}_g(\boldsymbol{x}) = \sum_{\boldsymbol{m}\in[M]^n} \beta_{\boldsymbol{m}} e^{\mathrm{j}\omega_o\langle \boldsymbol{m},\boldsymbol{x}\rangle}, \quad \text{and} \quad \tilde{D}_{FS}(\boldsymbol{x}) = \frac{1}{\lambda_d} \sum_{\boldsymbol{m}\in[M]^n} \gamma_{\boldsymbol{m}} e^{\mathrm{j}\omega_o\langle \boldsymbol{m},\boldsymbol{x}\rangle},$$

- In practice, employ the trigonometric Fourier series

$$D^*_{FS}(\boldsymbol{x}) \approx \frac{1}{\lambda^*_{FS}}\left(\frac{\gamma_{\boldsymbol{0}}}{2} + \sum_{\boldsymbol{m}\in\mathcal{M}} \gamma^r_{\boldsymbol{m}} \cos(\omega_o\langle \boldsymbol{m},\boldsymbol{x}\rangle) + \sum_{\boldsymbol{m}\in\mathcal{M}} \gamma^i_{\boldsymbol{m}} \sin(\omega_o\langle \boldsymbol{m},\boldsymbol{x}\rangle)\right), \quad \text{where}$$

$$\bar{\alpha}^r_{\boldsymbol{m}} \approx \frac{1}{NT} \sum_{\substack{k=1 \\ \boldsymbol{x}_k\sim p_d}}^{N} \cos(\omega_o\langle \boldsymbol{m},\boldsymbol{x}_k\rangle), \qquad \bar{\alpha}^i_{\boldsymbol{m}} \approx \frac{1}{NT} \sum_{\substack{k=1 \\ \boldsymbol{x}_k\sim p_d}}^{N} \sin(\omega_o\langle \boldsymbol{m},\boldsymbol{x}_k\rangle),$$

$$\bar{\beta}^r_{\boldsymbol{m}} \approx \frac{1}{NT} \sum_{\substack{k=1 \\ \boldsymbol{x}_k\sim p_g}}^{N} \cos(\omega_o\langle \boldsymbol{m},\boldsymbol{x}_k\rangle), \quad \text{and} \quad \bar{\beta}^i_{\boldsymbol{m}} \approx \frac{1}{NT} \sum_{\substack{k=1 \\ \boldsymbol{x}_k\sim p_g}}^{N} \sin(\omega_o\langle \boldsymbol{m},\boldsymbol{x}_k\rangle).$$

The Optimal Lagrange Multiplier

- WGAN-FS discriminator based on the trigonometric Fourier series

$$D^*_{FS}(\boldsymbol{x}) \approx \frac{1}{\lambda^*_{FS}}\left(\frac{\gamma_{\mathbf{0}}}{2} + \sum_{\boldsymbol{m}\in\mathcal{M}} \gamma^r_{\boldsymbol{m}} \cos(\omega_o\langle \boldsymbol{m}, \boldsymbol{x}\rangle) + \sum_{\boldsymbol{m}\in\mathcal{M}} \gamma^i_{\boldsymbol{m}} \sin(\omega_o\langle \boldsymbol{m}, \boldsymbol{x}\rangle)\right)$$

The Optimal Lagrange Multiplier

- WGAN-FS discriminator based on the trigonometric Fourier series

$$D^*_{FS}(\boldsymbol{x}) \approx \frac{1}{\lambda^*_{FS}}\left(\frac{\gamma_{\mathbf{0}}}{2} + \sum_{\boldsymbol{m}\in\mathcal{M}} \gamma^r_{\boldsymbol{m}} \cos(\omega_o\langle\boldsymbol{m},\boldsymbol{x}\rangle) + \sum_{\boldsymbol{m}\in\mathcal{M}} \gamma^i_{\boldsymbol{m}} \sin(\omega_o\langle\boldsymbol{m},\boldsymbol{x}\rangle)\right)$$

- The optimal Lagrange multiplier can be found via dual optimization

$$\lambda^*_d \text{ is such that } \int_{\mathcal{X}} \left(\|\nabla D^*(\boldsymbol{x})\|_2^2 - 1\right) \, \mathrm{d}\boldsymbol{x} = 0, \text{ where } D^*(\boldsymbol{x}) = \frac{1}{2\lambda_d}\left((p_g - p_d) * \phi\right)(\boldsymbol{x}) + D^*_h(\boldsymbol{x})$$

$$\lambda^*_{FS} \text{ is such that } \int_{\mathcal{X}} \left(\|\nabla D^*_{FS}(\boldsymbol{x})\|_2^2 - 1\right) \, \mathrm{d}\boldsymbol{x} = 0.$$

The Optimal Lagrange Multiplier

- WGAN-FS discriminator based on the trigonometric Fourier series

$$D^*_{FS}(\boldsymbol{x}) \approx \frac{1}{\lambda^*_{FS}} \left(\frac{\gamma_{\mathbf{0}}}{2} + \sum_{\boldsymbol{m} \in \mathcal{M}} \gamma^r_{\boldsymbol{m}} \cos(\omega_o \langle \boldsymbol{m}, \boldsymbol{x} \rangle) + \sum_{\boldsymbol{m} \in \mathcal{M}} \gamma^i_{\boldsymbol{m}} \sin(\omega_o \langle \boldsymbol{m}, \boldsymbol{x} \rangle) \right)$$

- The optimal Lagrange multiplier can be found via dual optimization

$$\lambda^*_d \text{ is such that} \int_{\mathcal{X}} \left(\|\nabla D^*(\boldsymbol{x})\|_2^2 - 1 \right) \, \mathrm{d}\boldsymbol{x} = 0, \text{ where } D^*(\boldsymbol{x}) = \frac{1}{2\lambda_d} \left((p_g - p_d) * \phi \right)(\boldsymbol{x}) + D^*_h(\boldsymbol{x})$$

$$\lambda^*_{FS} \text{ is such that} \int_{\mathcal{X}} \left(\|\nabla D^*_{FS}(\boldsymbol{x})\|_2^2 - 1 \right) \, \mathrm{d}\boldsymbol{x} = 0.$$

- An estimate of λ^*_{FS}:

$$\lambda^*_{FS} \approx \sqrt{(2|\mathcal{M}|+1) \left(\sum_{\boldsymbol{m} \in \mathcal{M}} (\tau^i_{\boldsymbol{m}} + \tau^r_{\boldsymbol{m}}) + \frac{1}{N} \sum_{k=1}^{N} \sum_{\boldsymbol{m} \in \mathcal{M}} (\tau^i_{\boldsymbol{m}} - \tau^r_{\boldsymbol{m}}) \cos(2\omega_o \langle \boldsymbol{m}, \boldsymbol{x}_k \rangle) \right)},$$

$$\text{where} \quad \tau^r_{\boldsymbol{m}} = \frac{1}{2} (\gamma^r_{\boldsymbol{m}})^2 \omega_o^2 \|\boldsymbol{m}\|^2, \quad \text{and} \quad \tau^i_{\boldsymbol{m}} = \frac{1}{2} (\gamma^i_{\boldsymbol{m}})^2 \omega_o^2 \|\boldsymbol{m}\|^2.$$

The Optimal Lagrange Multiplier

- WGAN-FS discriminator based on the trigonometric Fourier series

$$D^*_{FS}(\boldsymbol{x}) \approx \frac{1}{\lambda^*_{FS}} \left(\frac{\gamma_{\mathbf{0}}}{2} + \sum_{\boldsymbol{m} \in \mathcal{M}} \gamma^r_{\boldsymbol{m}} \cos(\omega_o \langle \boldsymbol{m}, \boldsymbol{x} \rangle) + \sum_{\boldsymbol{m} \in \mathcal{M}} \gamma^i_{\boldsymbol{m}} \sin(\omega_o \langle \boldsymbol{m}, \boldsymbol{x} \rangle) \right)$$

- The optimal Lagrange multiplier can be found via dual optimization

$$\lambda^*_d \text{ is such that } \int_{\mathcal{X}} \left(\|\nabla D^*(\boldsymbol{x})\|_2^2 - 1 \right) \, \mathrm{d}\boldsymbol{x} = 0, \text{ where } D^*(\boldsymbol{x}) = \frac{1}{2\lambda_d} \left((p_g - p_d) * \phi \right)(\boldsymbol{x}) + D^*_h(\boldsymbol{x})$$

- The value of λ^*_d can be computed in closed form:

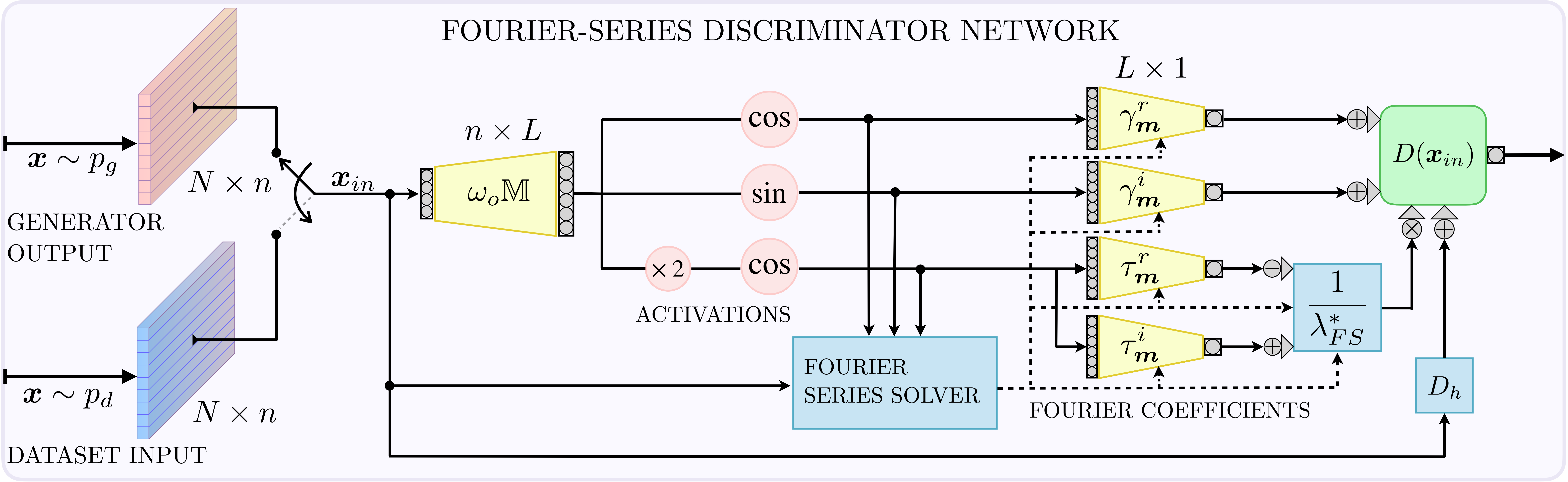
$$\lambda^*_d = \sqrt{\frac{1}{|\mathcal{X}|} \int_{\mathcal{X}} \sum_{i=1}^{n} \left(\left(K^\lambda_{n,i} * (p_g - p_d) \right)(\boldsymbol{x}) + a_i \right)^2 \mathrm{d}\boldsymbol{x}}, \quad \text{where}$$

$$K^\lambda_{n,i}(\boldsymbol{x}) = \frac{\partial \phi(\boldsymbol{x})}{\partial x_i} = \begin{cases} \frac{2}{\kappa_2} \left(\frac{x_i}{\|\boldsymbol{x}\|} \right), & \text{for } n = 2, \text{ and} \\ \frac{2-n}{\kappa_n} \left(\frac{x_i}{\|\boldsymbol{x}\|^n} \right), & \text{for } n \geq 3. \end{cases}$$

The WGAN-FS Discriminator Network

- WGAN-FS discriminator based on the trigonometric Fourier series

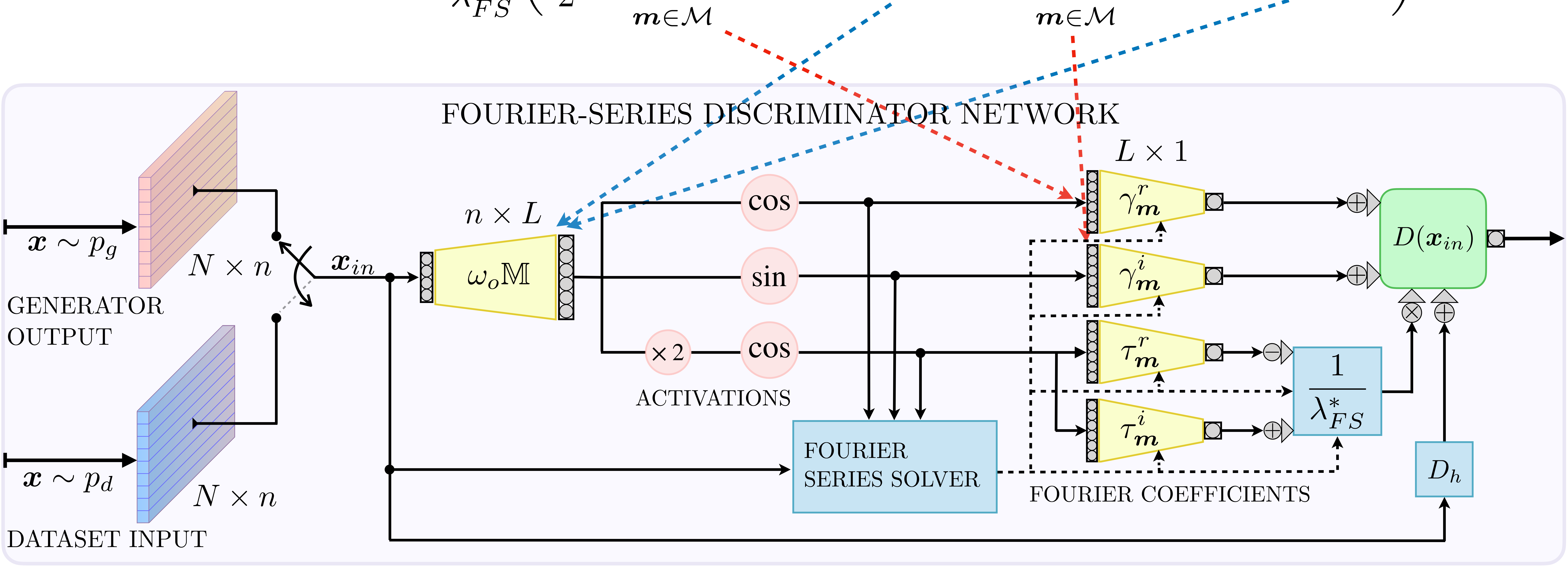
$$D^*_{FS}(\boldsymbol{x}) \approx \frac{1}{\lambda^*_{FS}}\left(\frac{\gamma_{\mathbf{0}}}{2} + \sum_{\boldsymbol{m}\in\mathcal{M}} \gamma^r_{\boldsymbol{m}} \cos(\omega_o\langle \boldsymbol{m}, \boldsymbol{x}\rangle) + \sum_{\boldsymbol{m}\in\mathcal{M}} \gamma^i_{\boldsymbol{m}} \sin(\omega_o\langle \boldsymbol{m}, \boldsymbol{x}\rangle)\right)$$



The WGAN-FS Discriminator Network

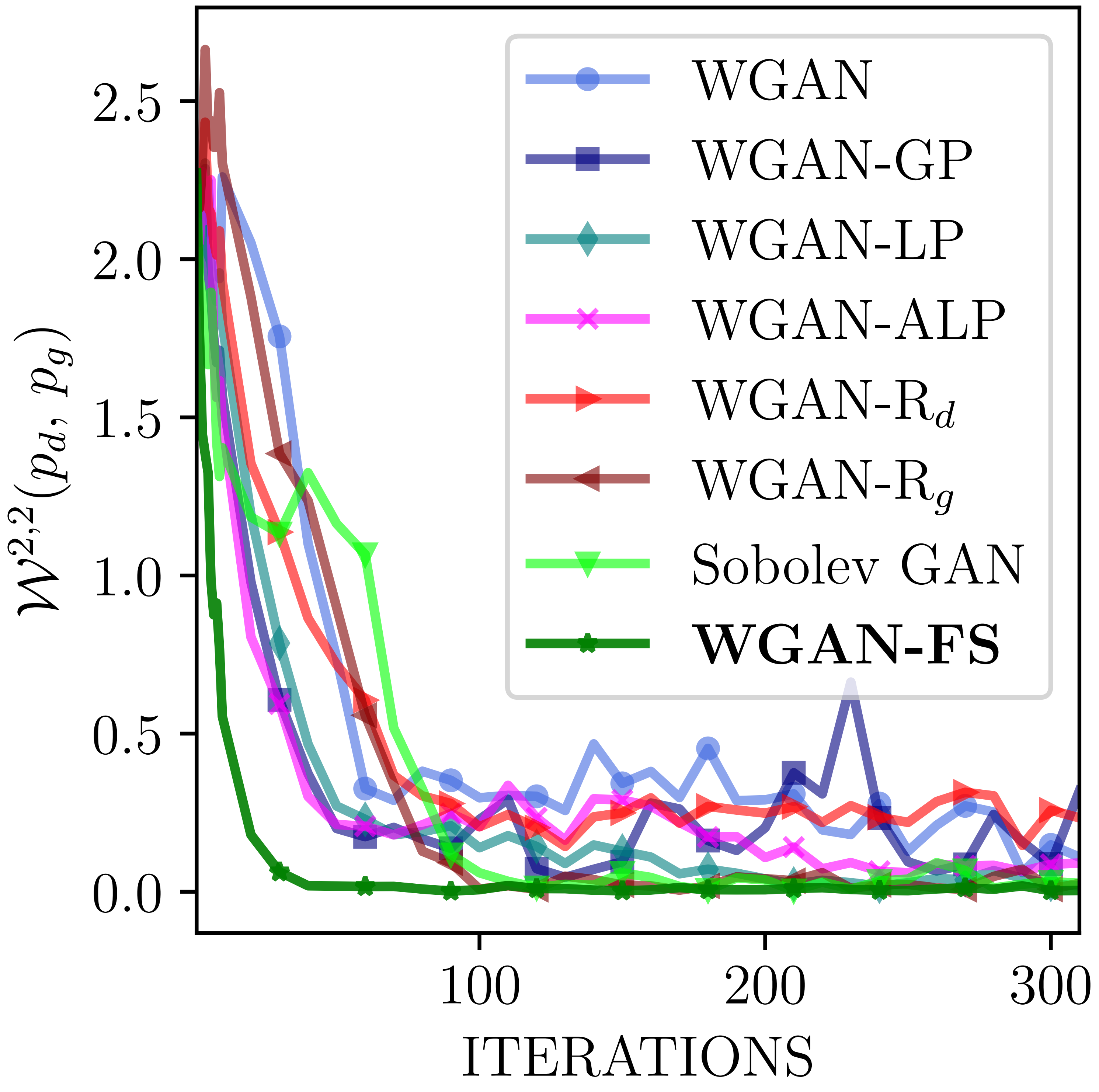
- WGAN-FS discriminator based on the trigonometric Fourier series

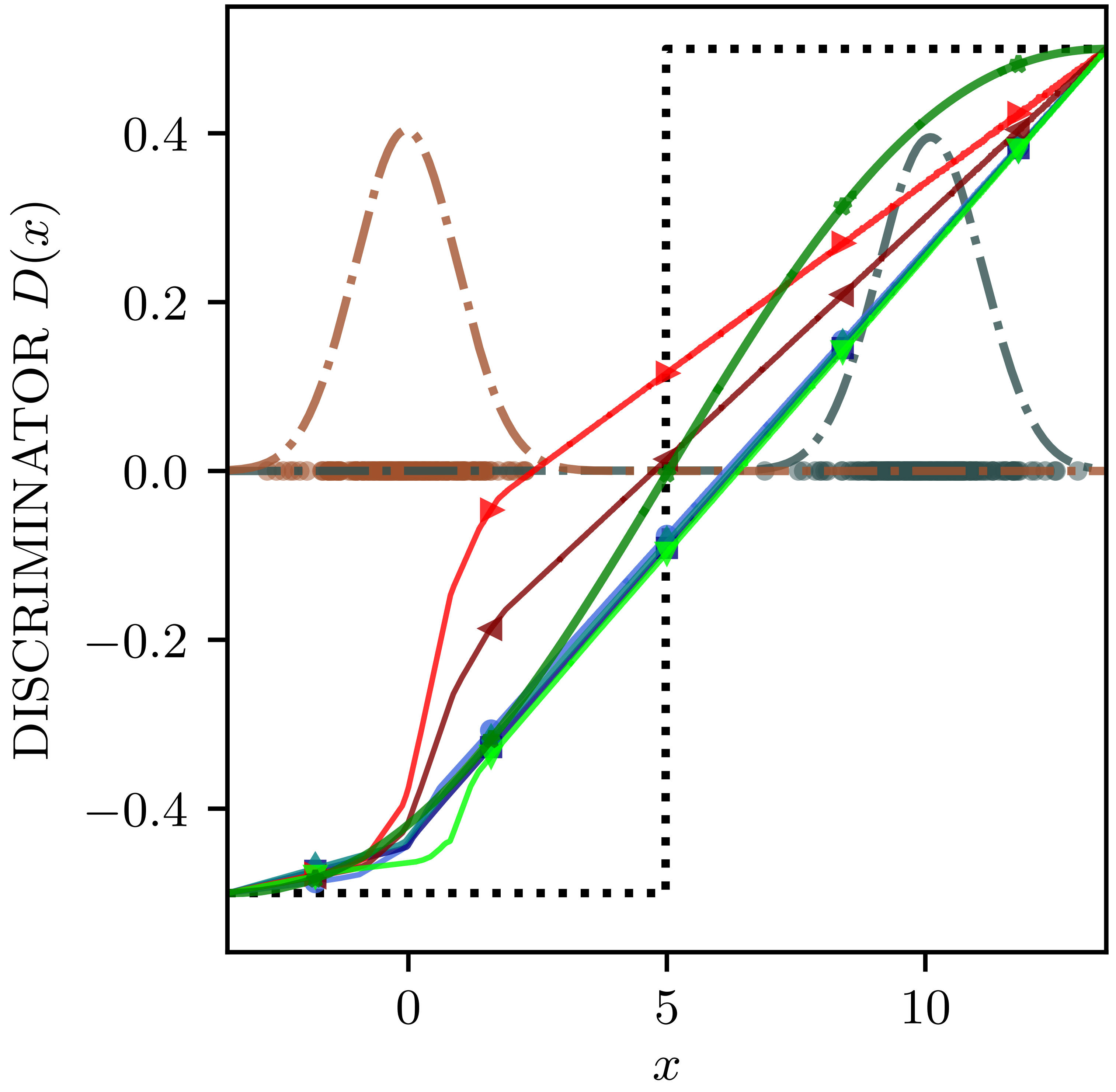
$$D^*_{FS}(\boldsymbol{x}) \approx \frac{1}{\lambda^*_{FS}}\left(\frac{\gamma_{\mathbf{0}}}{2} + \sum_{\boldsymbol{m}\in\mathcal{M}} \gamma^r_{\boldsymbol{m}} \cos(\omega_o\langle \boldsymbol{m}, \boldsymbol{x}\rangle) + \sum_{\boldsymbol{m}\in\mathcal{M}} \gamma^i_{\boldsymbol{m}} \sin(\omega_o\langle \boldsymbol{m}, \boldsymbol{x}\rangle)\right)$$

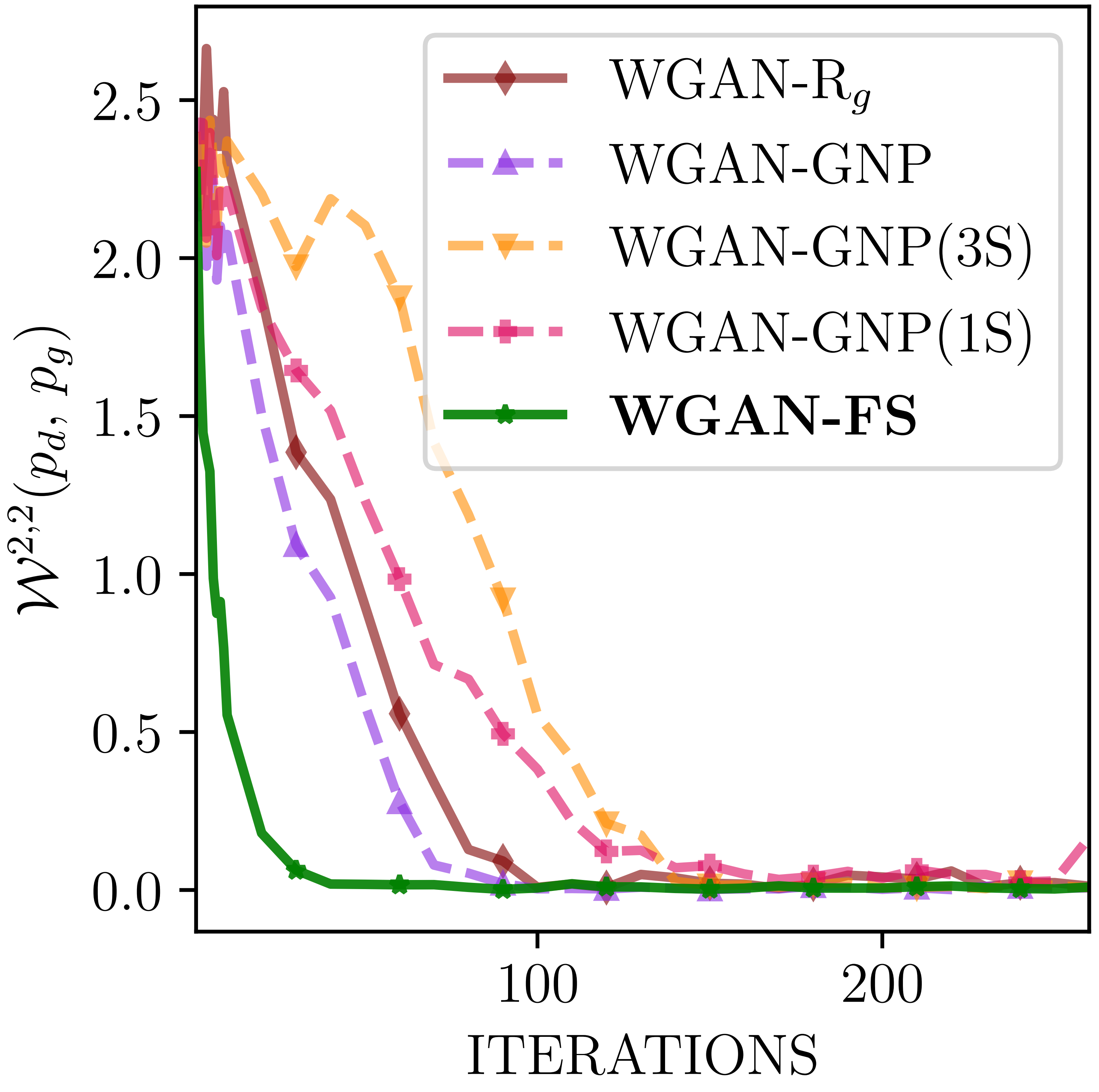


Experiments on Gaussian Learning

- WGAN-FS outperforms WGANs with a trainable discriminator, and trainable versions of WGAN-GNP.







Experiments on Gaussian Learning

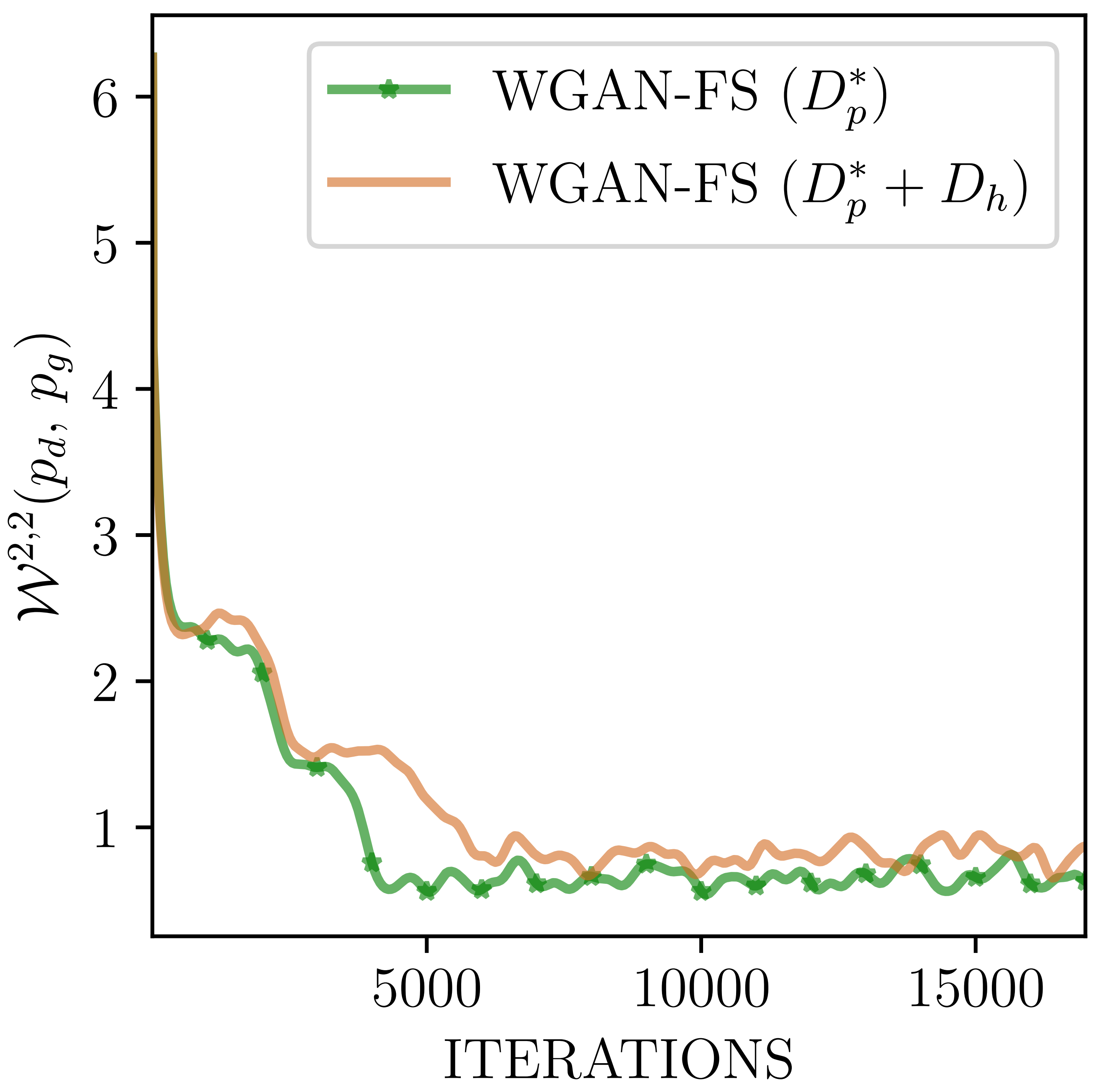


Fig. Impact of the homogeneous component.

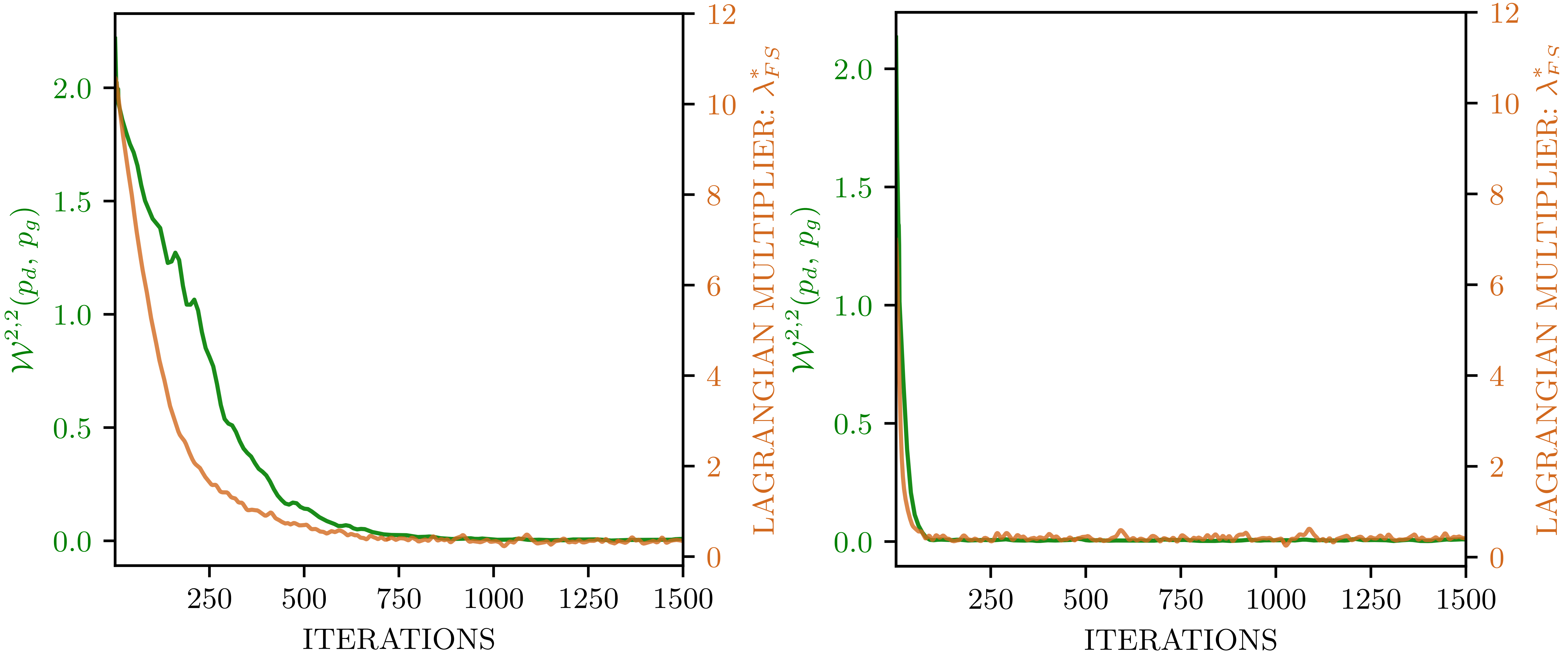
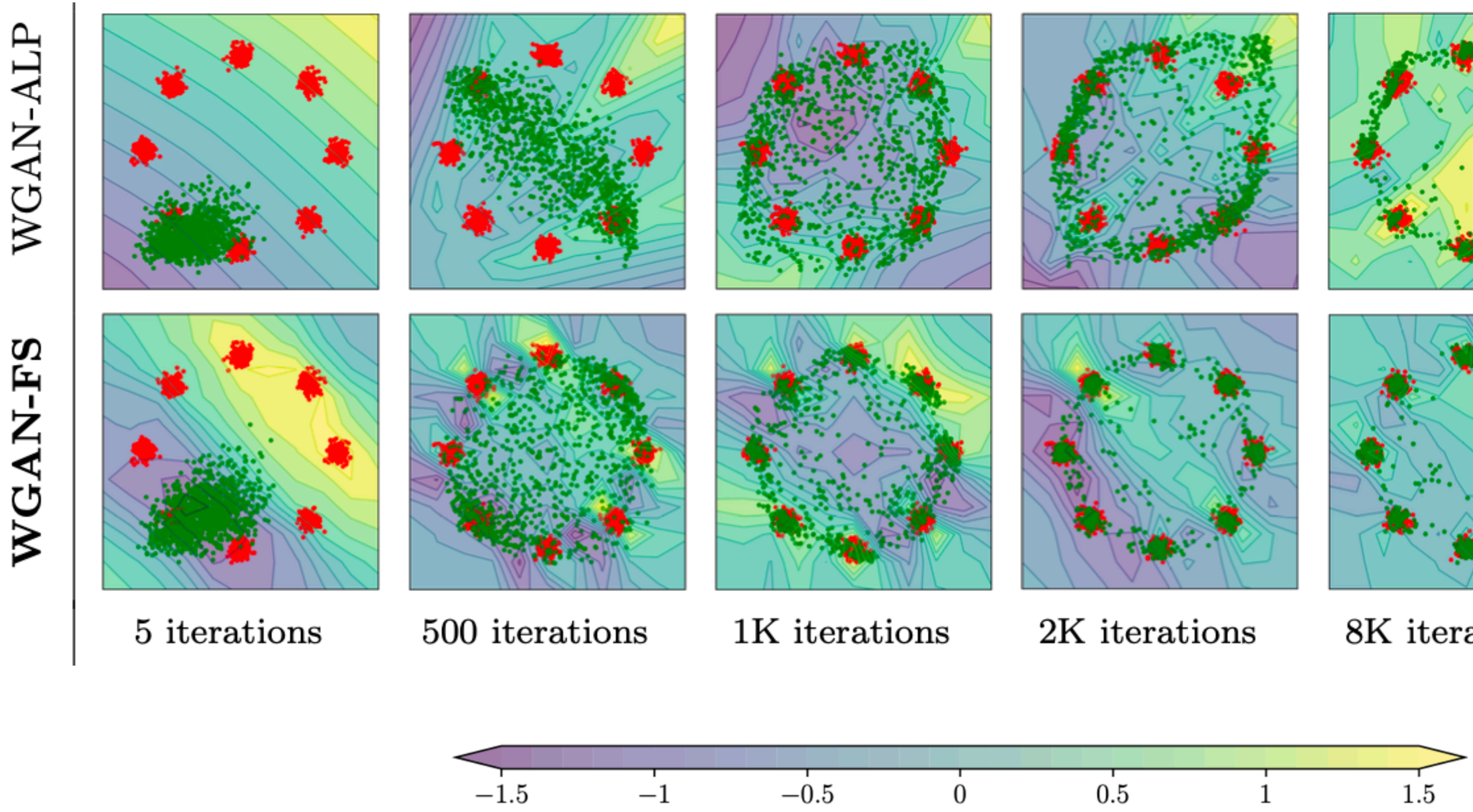


Fig. The Lagrange multiplier as an proxy for tracking model convergence.

Experiments on Gaussian Learning



WGAN-FS in Higher-dimensions

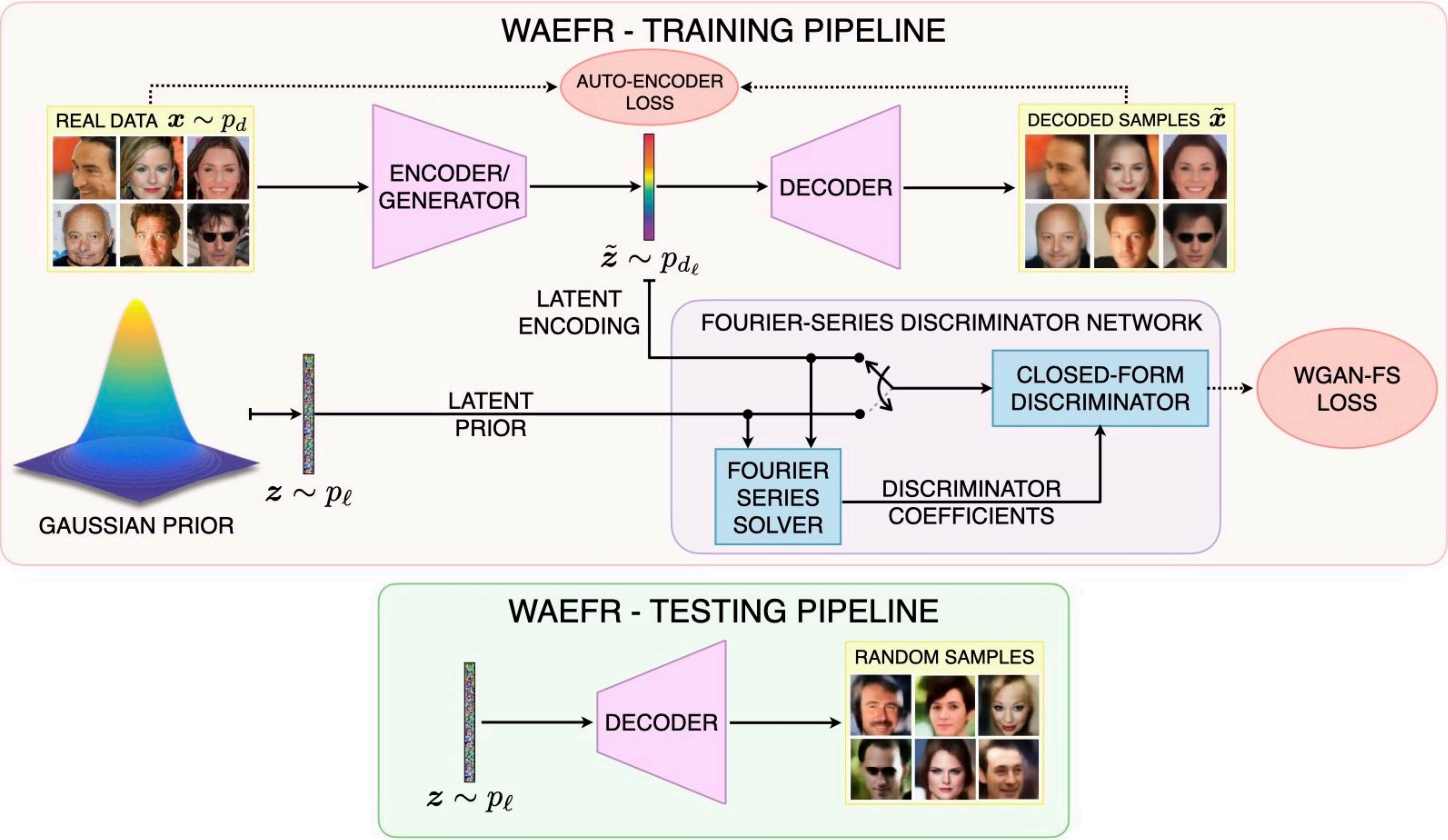
- The number of coefficients for a truncation harmonic M grows as n^M in $\mathbb{R}^n$.
- Consider MNIST, $\boldsymbol{x} \in \mathbb{R}^{784}$ — for $M = 50$, we have $\mathcal{O}(10^{144})$ terms!!

WGAN-FS in Higher-dimensions

- The number of coefficients for a truncation harmonic M grows as n^M in $\mathbb{R}^n$.
- Consider MNIST, $\boldsymbol{x} \in \mathbb{R}^{784}$ — for $M = 50$, we have $\mathcal{O}(10^{144})$ terms!!
- Explore applications to latent-space learning with Wasserstein Autoencoders[16].
- The WGAN-FS discriminator is used to match the latent-space distribution of a data to a target Gaussian – Wasserstein autoencoders (WAEs).

[16] Tolstikhin *et al.*, ICLR 2018

WAE + Fourier-series Discriminator (WAEFR)



WAE + Fourier-series Discriminator (WAEFR)

- Compare WAEs[16] with the Fourier-represented discriminator (WAEFR) against baseline WAEs and adversarial autoencoders (AAEs)

- WAEFR generates images of superior visual quality, with improved convergence.

[16]Tolstikhin *et al.*, ICLR 2018

WAE + Fourier-series Discriminator (WAEFR)

	GAN flavor	MNIST	SVHN	CelebA	Ukiyo-E	CIFAR-10 (Averaged)	CIFAR-10
FID ↓	WAE	21.676	46.083	42.943	204.446	124.165	123.88
	WAE-KL	26.231	59.717	59.223	215.013	112.650	115.96
	WAAE-LP	20.240	47.332	43.509	195.133	108.512	108.95
	WAAE-ALP	22.306	48.128	45.628	200.330	**107.509**	110.223
	CWAE	22.125	46.757	47.963	207.350	114.689	102.062
	WAEFR	**19.793**	**44.811**	**37.195**	**192.049**	108.804	**100.754**
$\langle RE \rangle$ ↓	WAE	0.0827	0.0425	0.0939	0.0520	0.1786	0.125
	WAE-KL	0.0707	0.0380	0.0776	0.0421	0.1254	0.116
	WAAE-LP	0.0747	0.0353	0.0868	0.0429	0.1382	0.117
	WAAE-ALP	0.0836	0.0377	0.0956	0.0479	0.1294	0.119
	CWAE	0.0735	0.0478	0.0852	0.0831	0.1729	0.112
	WAEFR	**0.0693**	**0.0310**	**0.0762**	**0.0417**	**0.1227**	**0.107**

		GAN flavor	MNIST	SVHN	CelebA	Ukiyo-E	CIFAR-10 (Averaged)	CIFAR-10
Sharpness	Random	WAE	0.1567	0.0018	0.0015	0.1210	0.0625	0.0011
		WAE-KL	0.1317	0.0014	0.0018	0.1255	0.0039	0.0032
		WAAE-LP	0.1520	0.0017	0.0044	**0.1566**	0.0155	0.0029
		WAAE-ALP	0.1609	0.0017	0.0035	0.1441	0.0150	0.0039
		CWAE	0.1703	0.0019	0.0036	0.0821	0.0158	0.0086
		WAEFR	**0.1717**	**0.0028**	**0.0084**	0.2275	**0.0194**	**0.0110**
	Interpolation	WAE	0.1681	0.0022	0.0032	0.0270	0.0035	0.0027
		WAE-KL	0.1629	0.0022	0.0044	0.0229	0.0053	0.0054
		WAAE-LP	0.1706	0.0024	0.0044	**0.0383**	0.0036	0.0041
		WAAE-ALP	0.1031	0.0019	0.0038	0.0345	0.0061	0.0031
		CWAE	0.1387	0.0028	0.0034	0.0136	0.0061	0.0045
		WAEFR	**0.1746**	**0.0029**	**0.0077**	0.0330	**0.0068**	**0.0064**
		Benchmark	0.1950	0.0051	0.0318	0.1805	0.0278	0.0361

- WAEFR achieves superior FID, and average reconstruction error (of the autoencoder sub-block), while generating sharper images when compared against the baselines.

Conclusions

- Derived the optimal discriminator PDE in WGANs with gradient-based regularization.
- Developed a novel Fourier-series approach to implementing the optimal discriminator.
- **Future Scope:** Analyzing other gradient-regularized GAN discriminator variants.
- **Future Scope:** Exploring alternatives to the Fourier-series approximation.

References

[01] Goodfellow et al., “Generative adversarial networks,” NeurIPS 2014

[02] Mao et al., “Least-squares generative adversarial networks,” ICCV 2017

[03] Nowozin et al.,“f-GAN: Training Generative Neural Samplers using Variational Divergence Minimization,” NeurIPS 2016

[04] Arjovsky et al.,“ Wasserstein GANs,” ICML 2017

[05] Mroueh et al., “Sobolev GAN,”ICLR 2018

[06] Mroueh et al., “Fisher GAN,”NeurIPS 2017

[07] Gulrajani et al., “Improved training of Wasserstein GANs,” NeurIPS 2017

[08] Petzka et al., “On the regularization of Wasserstein GANs,” ICLR 2018

[09] Terjék, “Adversarial Lipschitz regularization,” ICLR 2020

[10] Mescheder et al., “Which training methods for GANs do actually converge?” ICML 2018

[11] Tolstikhin et al., “Wasserstein auto-encoders,” ICLR, 2018

Thank You!

ADDITIONAL SLIDES

WGAN with Fourier-series Discriminator

- Given truncated Fourier-series is implementable.

$$\tilde{p}_d(\boldsymbol{x}) = \sum_{\boldsymbol{m}\in[M]^n} \alpha_{\boldsymbol{m}} e^{\mathrm{j}\omega_o\langle \boldsymbol{m},\boldsymbol{x}\rangle}, \quad \alpha_{\boldsymbol{m}} \approx \frac{1}{NT}\sum_{k=1}^{N} e^{\mathrm{j}\langle \boldsymbol{\omega}_{\boldsymbol{m}},\boldsymbol{x}_k\rangle}, \quad \text{and} \quad \tilde{D}_{FS}(\boldsymbol{x}) = \frac{1}{\lambda_d}\sum_{\boldsymbol{m}\in[M]^n} \gamma_{\boldsymbol{m}} e^{\mathrm{j}\omega_o\langle \boldsymbol{m},\boldsymbol{x}\rangle},$$

- We derived error bounds on the Fourier-series truncation and sample-estimation of the coefficients.

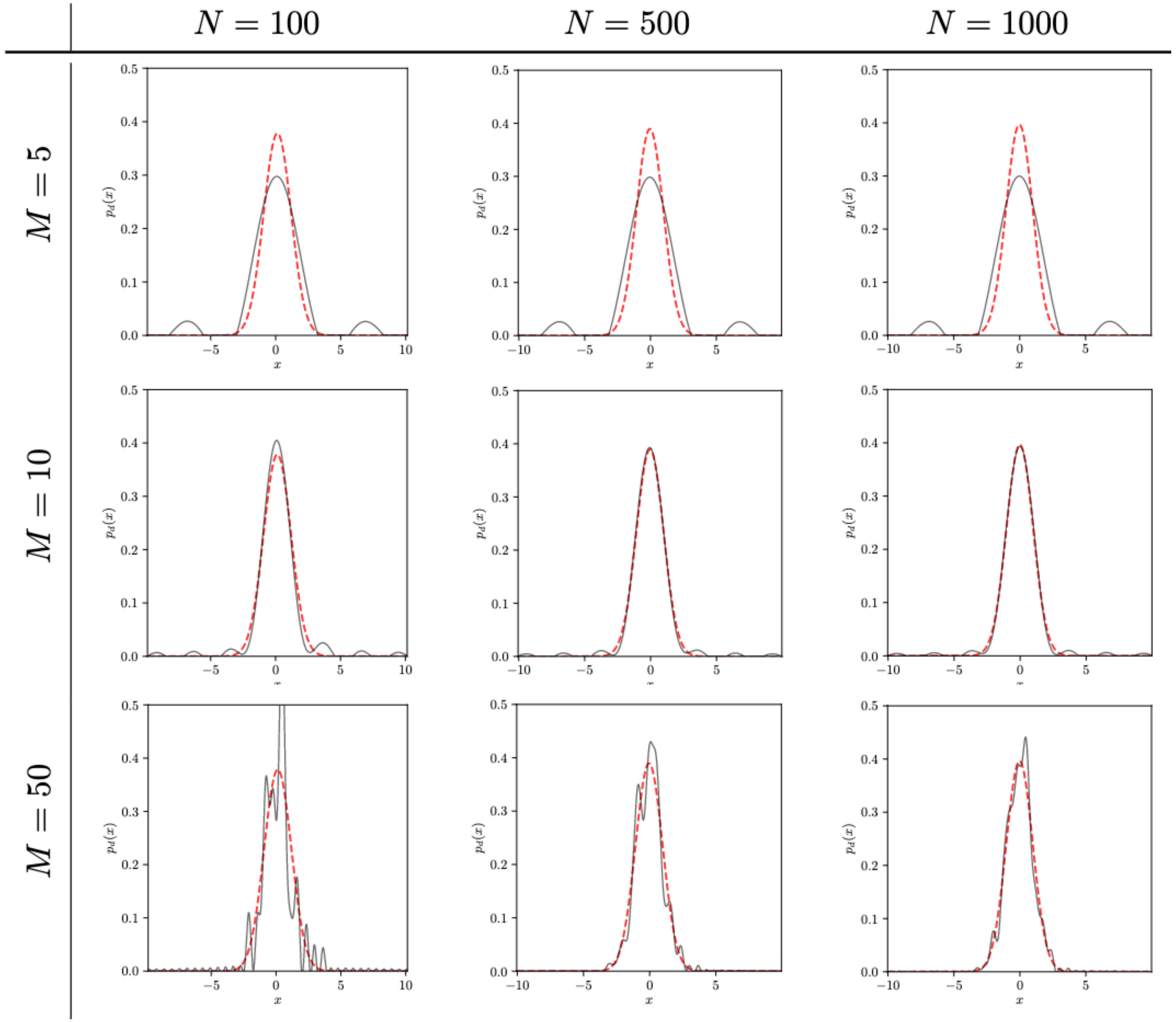
$$\epsilon_D^2 = \|D_{FS}(\boldsymbol{x}) - \tilde{D}_{FS}(\boldsymbol{x})\|_2^2 \leq \mathfrak{C}_{n,T}\left(\frac{(M^2 n)^{-(k-2)}}{k-2}\right), \quad \text{for} \quad p_d, p_g \in \mathcal{C}^k(\mathcal{X}) \bigcap W^{k,2}(\mathcal{X})$$

$$\mathbb{E}_{\boldsymbol{x}}\left[\epsilon_{p_d}^2\right] \leq \underbrace{\frac{M^n}{N}\left(1 - \frac{\mathfrak{m}_{p_d}}{n^{k+\frac{n+1}{2}}}\right)}_{\varepsilon_{\text{stat}}} + \underbrace{\frac{\mathfrak{M}_{p_d}\mathfrak{C}_n}{(2k+1)}\left(\frac{1}{M^{2k+1}}\right)}_{\varepsilon_{\text{trunc}}} \text{for } 0 < k < \infty,$$

WGAN with Fourier-series Discriminator

- We derived error bounds on the Fourier-series truncation and sample-estimation of the coefficients.

$$\mathbb{E}_{\boldsymbol{x}}\left[\epsilon_{p_d}^2\right] \leq \underbrace{\frac{M^n}{N}\left(1-\frac{\mathfrak{m}_{p_d}}{n^{k+\frac{n+1}{2}}}\right)}_{\varepsilon_{\text{stat}}} + \underbrace{\frac{\mathfrak{M}_{p_d}\mathfrak{C}_n}{(2k+1)}\left(\frac{1}{M^{2k+1}}\right)}_{\varepsilon_{\text{trunc}}},$$

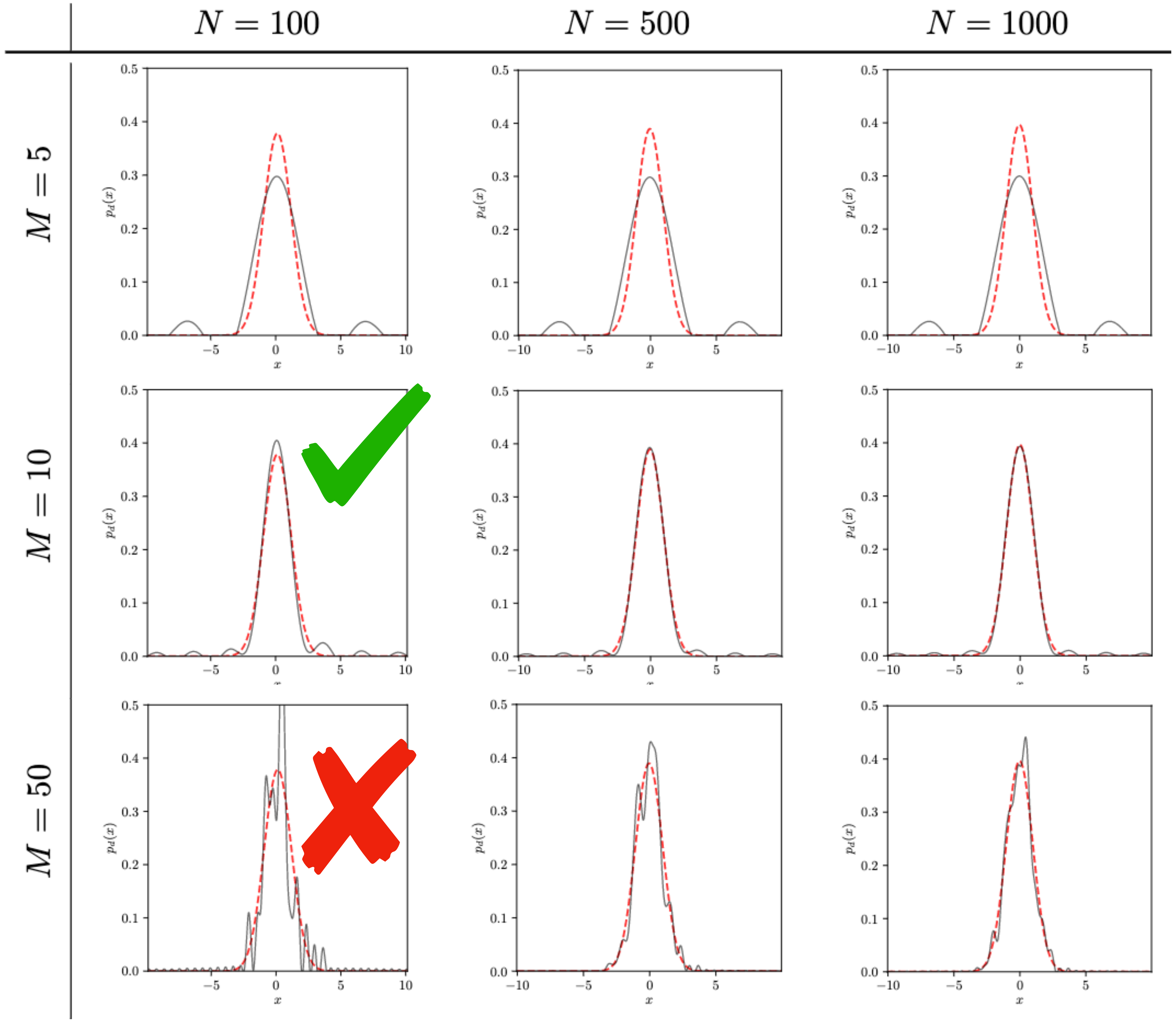


WGAN with Fourier-series Discriminator

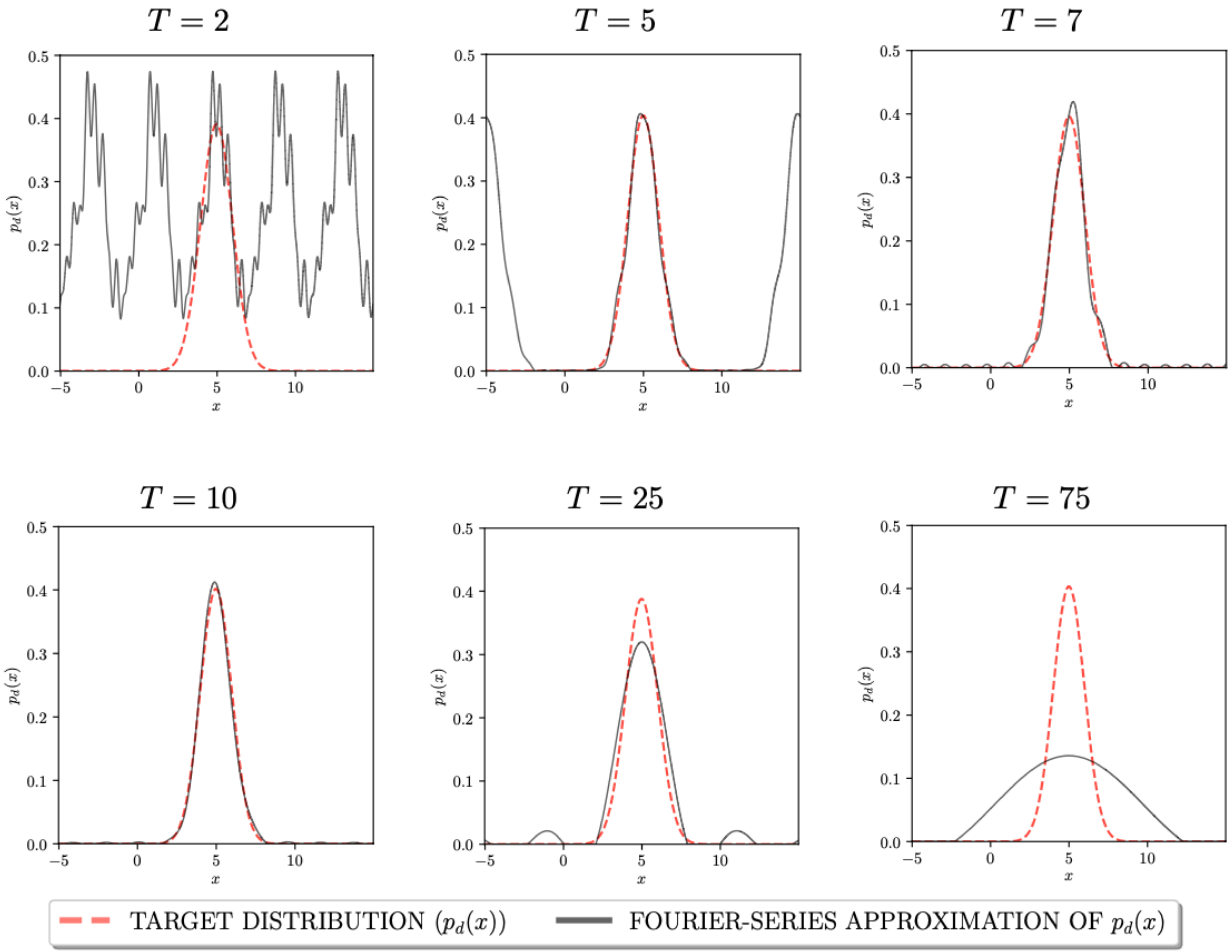
- We derived error bounds on the Fourier-series truncation and sample-estimation of the coefficients.

$$\mathop{\mathbb{E}}_{\boldsymbol{x}}\left[\epsilon_{p_d}^2\right] \leq \underbrace{\frac{M^n}{N}\left(1-\frac{\mathfrak{m}_{p_d}}{n^{k+\frac{n+1}{2}}}\right)}_{\varepsilon_{\text{stat}}} + \underbrace{\frac{\mathfrak{M}_{p_d}\mathfrak{C}_n}{(2k+1)}\left(\frac{1}{M^{2k+1}}\right)}_{\varepsilon_{\text{trunc}}},$$

- Discarding higher-frequency coefficients is better than approximating them with too few samples!



WGAN with Fourier-series Discriminator



Thank You