

# Generative Adversarial Networks (GANs)

Emanuele Sansone

23 febbraio 2017

# Goals

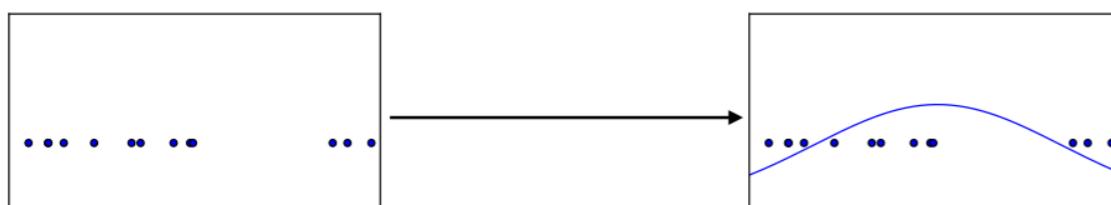
- Feedbacks for upcoming seminar
- Stimulate discussion



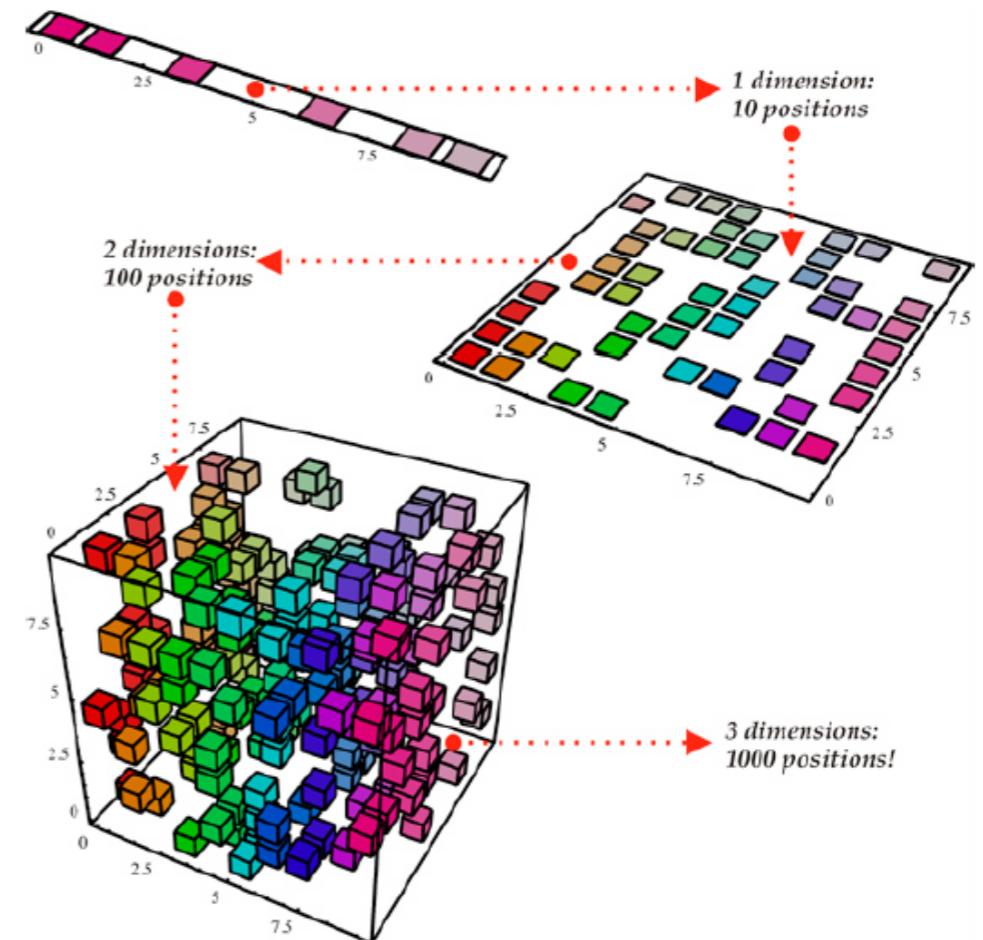
1. Why not using this model?
2. What don't you like in this model?
3. Why not improving it?
4. ...

# Sketch of GANs

Density estimation in high-dimensional data (manifold assumption)

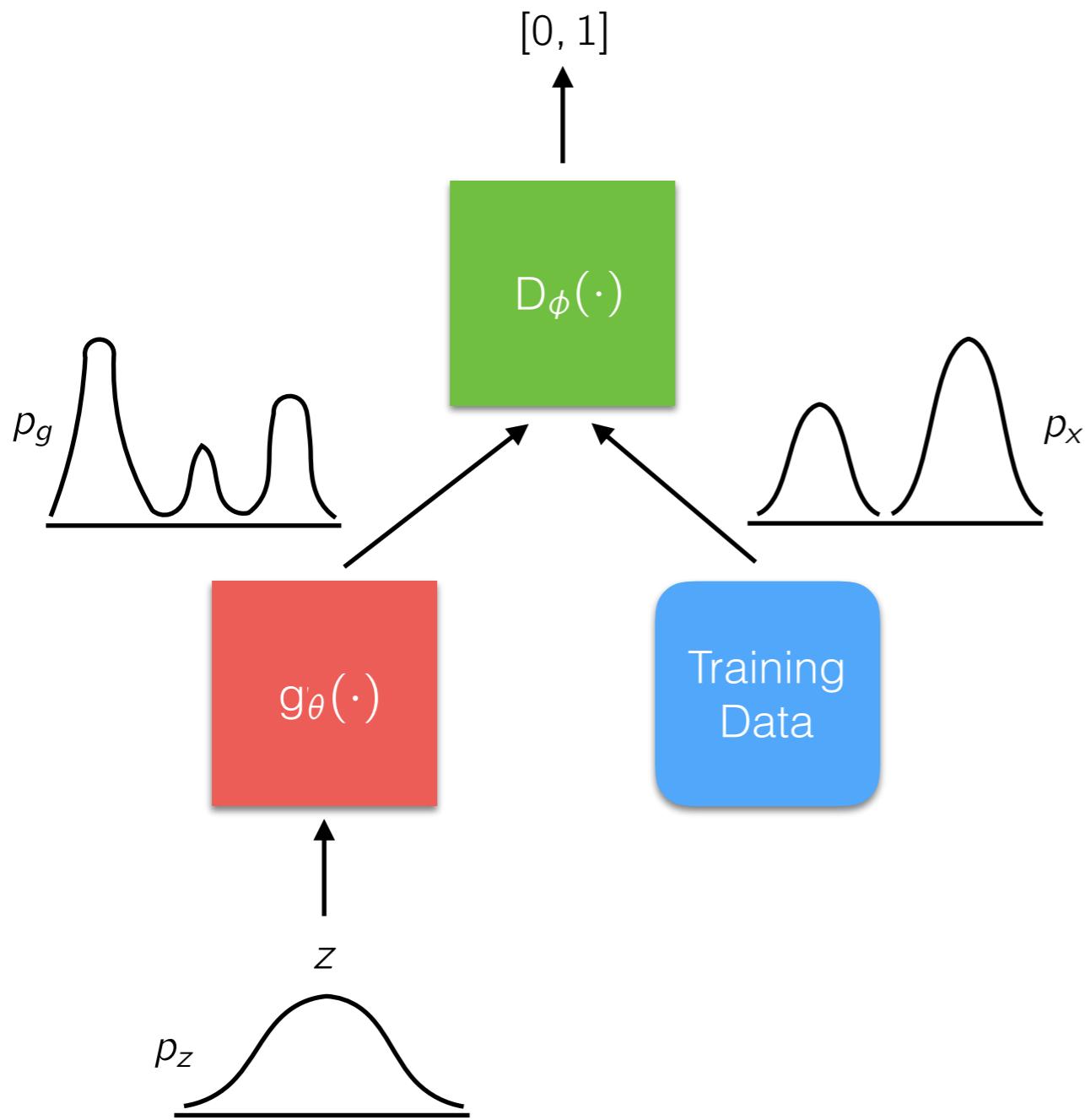


Ian Goodfellow, NIPS 2016 tutorial on GANs



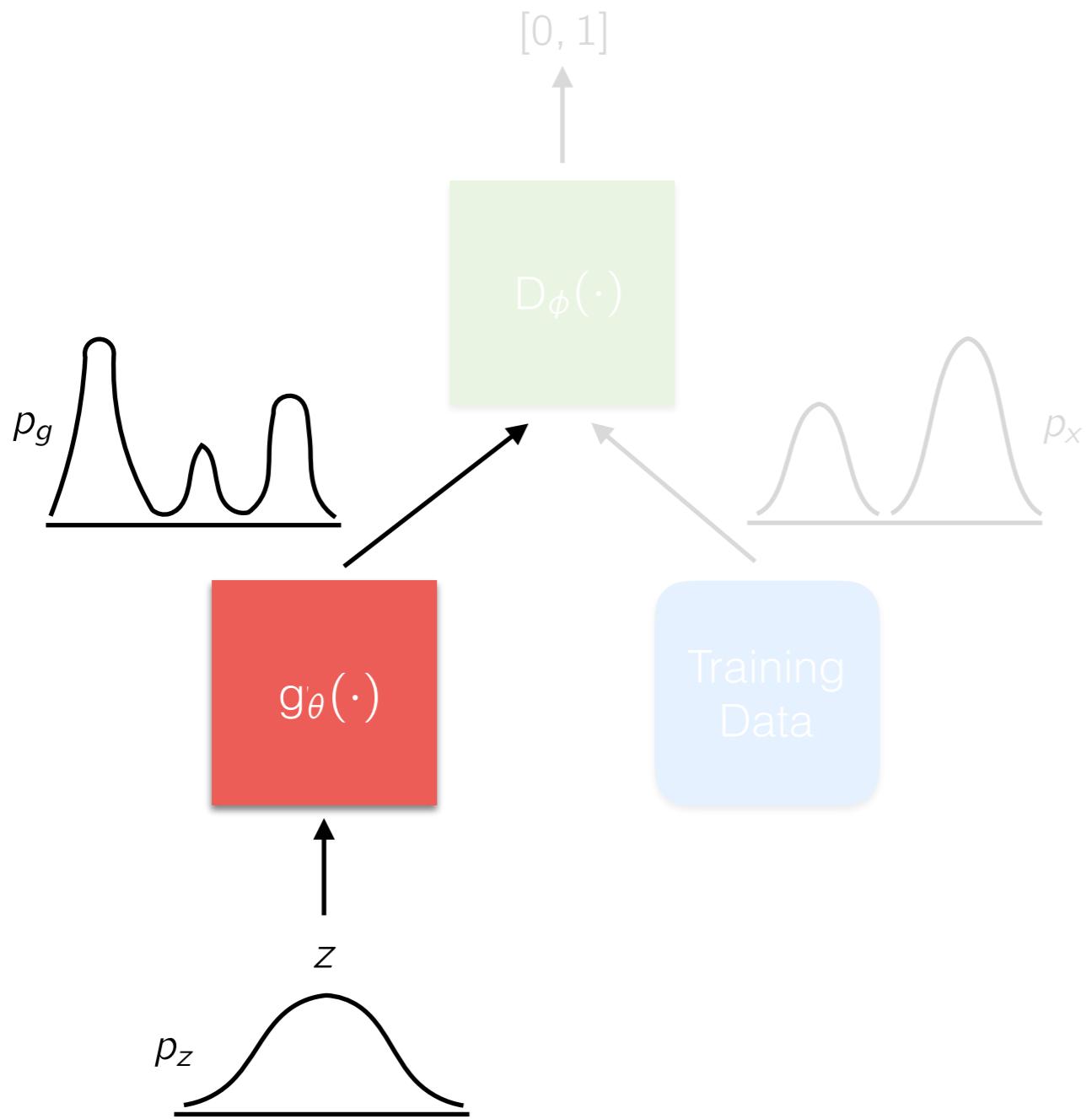
Yoshua Bengio, MLSS 2015 Austin TX

# Sketch of GANs



"Generative Adversarial Networks"  
ICLR 2014 Ian Goodfellow

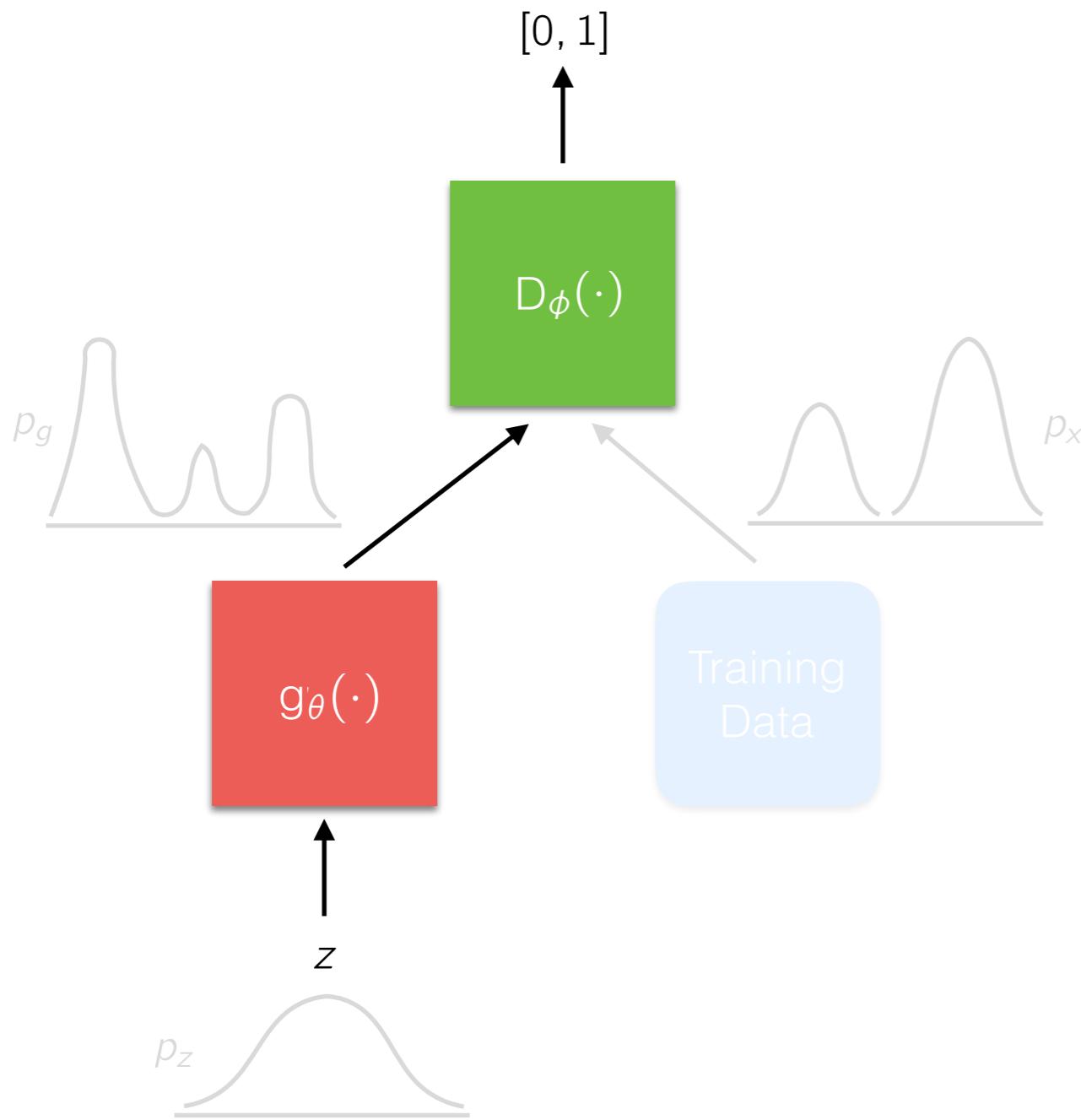
# Sketch of GANs



## GENERATIVE

1. Fast sample generation - discriminative function (no Markov chains)
2. Implicit definition of density families

# Sketch of GANs

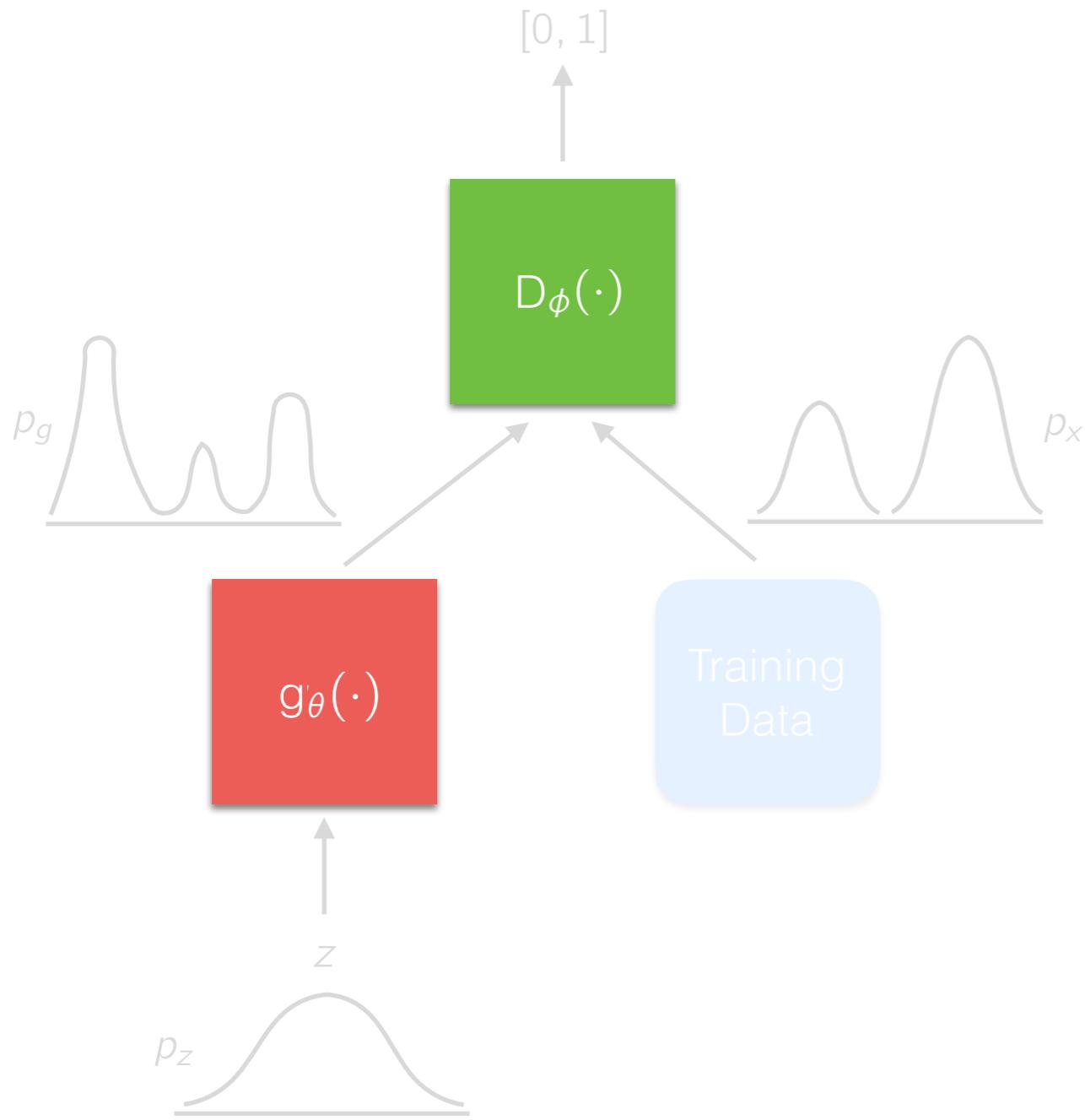


**ADVERSARIAL**

Game between two adversaries  
(no likelihood maximization)

SEE LATER

# Sketch of GANs

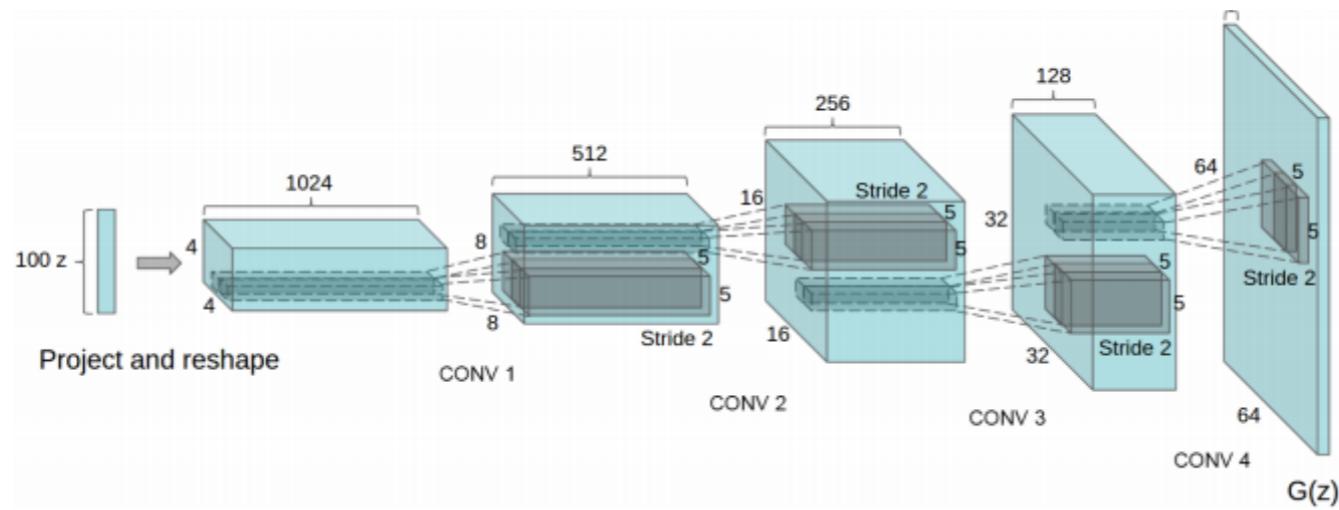
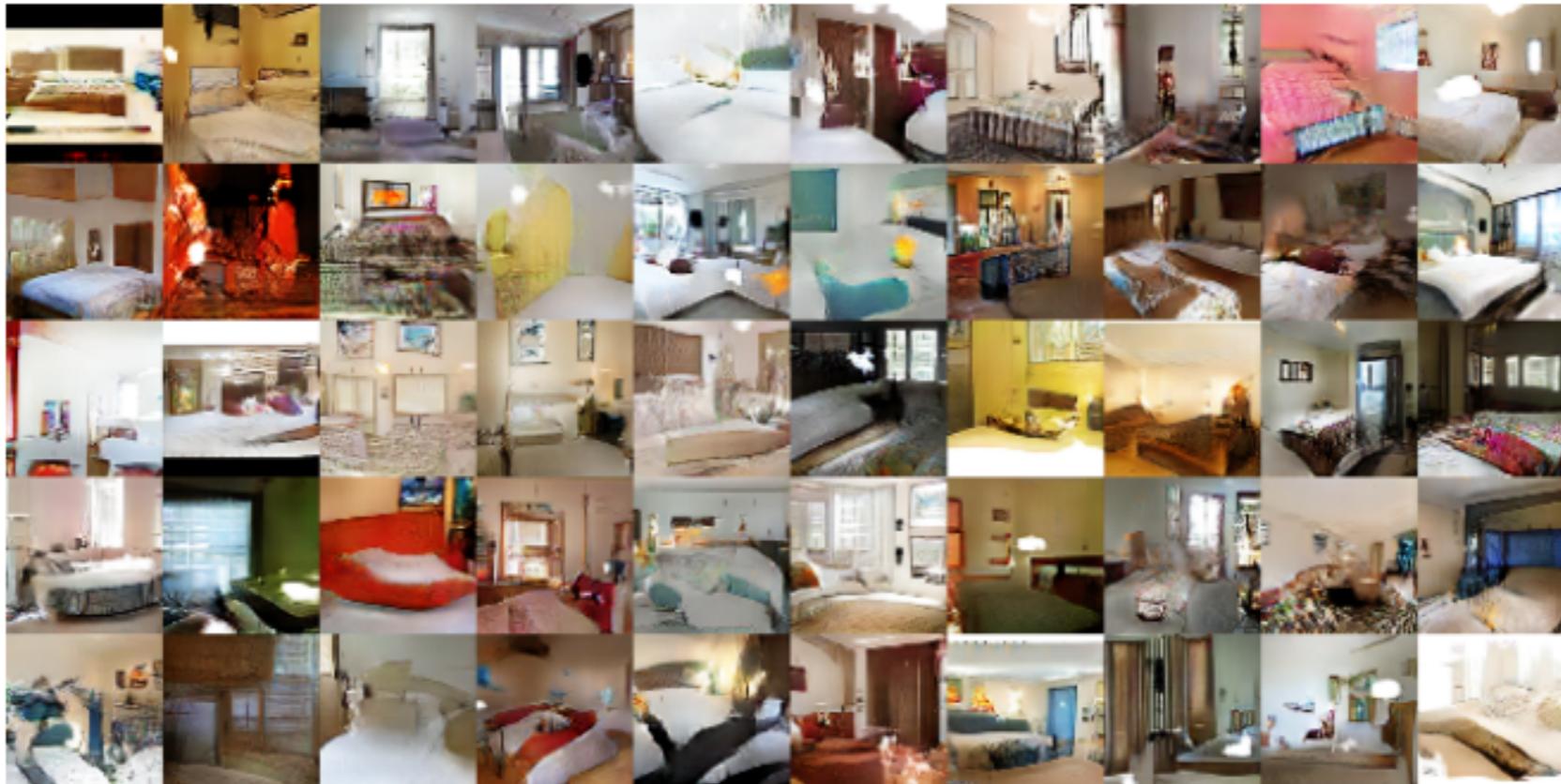


## NETWORKS

Exploitation of deep neural  
architectures  
(High capacity and efficient  
training...)

“Learning Deep Architectures for AI”  
Joshua Bengio 2009

# Application - Image synthesis



“Unsupervised Representation Learning with Deep Convolutional Generative Adversarial Networks”  
arXiv 2015 Alec Radford, Luke Metz, Soumith Chintala

# Application - Video synthesis

Hallucinated videos

Beach



Golf



Train



Baby



Conditional generation of videos

Input



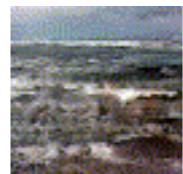
Output



Input



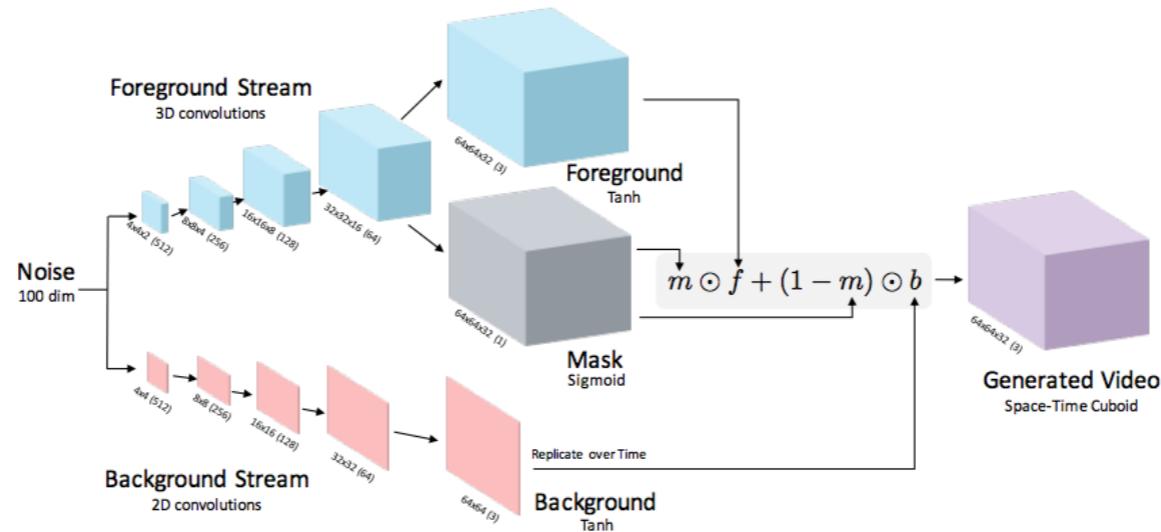
Output



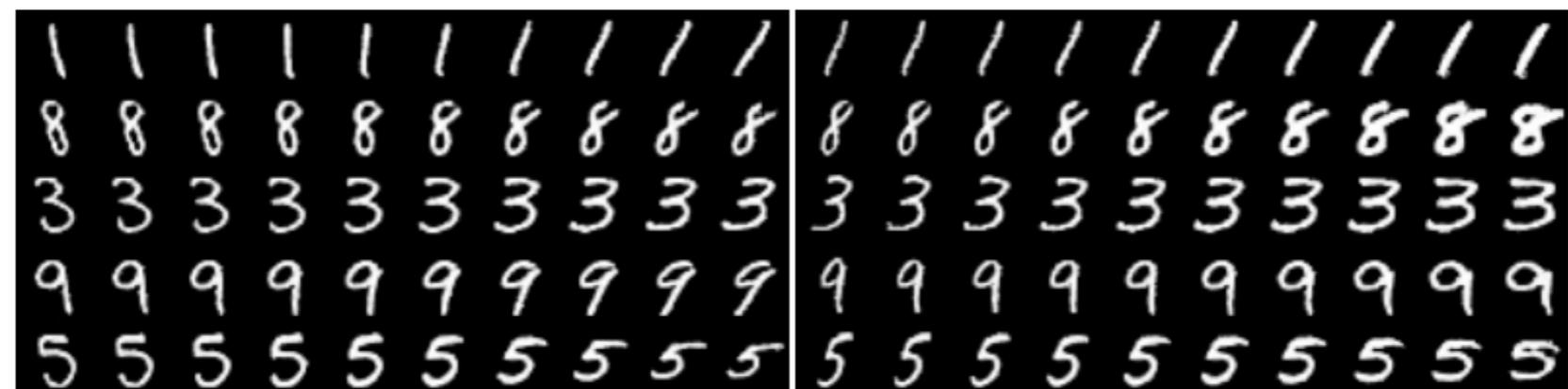
Input



Output



# Application - Representation learning 1



(c) Varying  $c_2$  from  $-2$  to  $2$  on InfoGAN (Rotation)

(d) Varying  $c_3$  from  $-2$  to  $2$  on InfoGAN (Width)

“InfoGAN: Interpretable Representation Learning by Information Maximizing Generative Adversarial Nets”  
NIPS 2016 Xi Chen, Yan Duan, Rein Houthooft, John Schulman, Ilya Sutskever, Pieter Abbeel

# Application - Representation Learning 2



(a) Azimuth (pose)

(b) Elevation



(c) Lighting

(d) Wide or Narrow

“InfoGAN: Interpretable Representation Learning by Information Maximizing Generative Adversarial Nets”  
NIPS 2016 Xi Chen, Yan Duan, Rein Houthooft, John Schulman, Ilya Sutskever, Pieter Abbeel

PAPERs about GANs 2014-2016	
Generative Adversarial Networks	ICLR 2014
Deep Generative Image Models Using a Laplacian Pyramid of Adversarial Networks (UNSUPERVISED)	NIPS 2015
Draw: A Recurrent Neural Network for Image Generation (UNSUPERVISED)	ICML 2015
Unsupervised Representation Learning with Deep Convolutional Generative Adversarial Networks	arXiv 2015
InfoGAN: Interpretable Representation Learning by Information Maximizing Generative Adversarial Nets	NIPS 2016
Towards Principled Unsupervised Learning	ICLR 2016 (workshop)
Improved Techniques for Training GANs	NIPS 2016
Unsupervised and Semi-Supervised Learning with Categorical Generative Adversarial Networks	ICLR 2016
Generating Videos with Scene Dynamics	NIPS 2016
NIPS 2016 Tutorial: Generative Adversarial Networks <b>PAPER+VIDEO</b>	NIPS 2016
PAPERs about GANs accepted to ICLR 2017	
Towards Principled Methods for Training Generative Adversarial Networks ( <b>FIRST THEORETICAL PAPER</b> )	ICLR 2017
Adversarially Learned Inference ( <b>SoA in SEMI-SUPERVISED LEARNING</b> )	ICLR 2017
Learning to Generate Samples from Noise Through Infusion Training	ICLR 2017
Improving Generative Adversarial Networks with Denoising Feature Matching	ICLR 2017
LR-GAN: Layered Recursive Generative Adversarial Networks for Image Generation	ICLR 2017
Mode Regularized Generative Adversarial Networks	ICLR 2017
Generative Models and Model Criticism via Optimized Maximum Mean Discrepancy	ICLR 2017
Calibrating Energy-Based Generative Adversarial Networks	ICLR 2017
Unrolled Generative Adversarial Networks	ICLR 2017
Generative Multi-Adversarial Networks	ICLR 2017
Energy-Based Generative Adversarial Networks	ICLR 2017

Probably hot topic?

# Generative Adversarial Networks (GANs)

Emanuele Sansone

23 febbraio 2017

# Problem formulation

$$\min_{\theta} \max_{\phi} \left\{ E_{x \sim p_x} \left\{ \log(D_\phi(x)) \right\} + E_{z \sim p_z} \left\{ \log(1 - D_\phi(g_\theta(z))) \right\} \right\}$$

# Problem formulation

$$\min_{\theta} \max_{\phi} \left\{ E_{x \sim p_x} \left\{ \log(D_\phi(x)) \right\} + E_{z \sim p_z} \left\{ \log(1 - D_\phi(g_\theta(z))) \right\} \right\}$$

Training Data    Generated Data

# Problem formulation

$$\min_{\theta} \max_{\phi} \left\{ E_{x \sim p_x} \left\{ \log(D_\phi(x)) \right\} + E_{z \sim p_z} \left\{ \log(1 - D_\phi(g_\theta(z))) \right\} \right\}$$

## **Optimal Discriminator (Proof sketch)**

$$\min_{\theta} \max_{\phi} \left\{ E_{x \sim p_x} \left\{ \log(D_\phi(x)) \right\} + E_{x \sim p_g} \left\{ \log(1 - D_\phi(x)) \right\} \right\}$$

# Problem formulation

$$\min_{\theta} \max_{\phi} \left\{ E_{x \sim p_x} \left\{ \log(D_\phi(x)) \right\} + E_{z \sim p_z} \left\{ \log(1 - D_\phi(g_\theta(z))) \right\} \right\}$$

## Optimal Discriminator (Proof sketch)

$$\begin{aligned} & \min_{\theta} \max_{\phi} \left\{ E_{x \sim p_x} \left\{ \log(D_\phi(x)) \right\} + E_{x \sim p_g} \left\{ \log(1 - D_\phi(x)) \right\} \right\} \\ & \int p_x(x) \log(D_\phi(x)) dx + \int p_g(x) \log(1 - D_\phi(x)) dx \end{aligned}$$

# Problem formulation

$$\min_{\theta} \max_{\phi} \left\{ E_{x \sim p_x} \left\{ \log(D_\phi(x)) \right\} + E_{z \sim p_z} \left\{ \log(1 - D_\phi(g_\theta(z))) \right\} \right\}$$

## Optimal Discriminator (Proof sketch)

$$\begin{aligned} & \min_{\theta} \max_{\phi} \left\{ E_{x \sim p_x} \left\{ \log(D_\phi(x)) \right\} + E_{x \sim p_g} \left\{ \log(1 - D_\phi(x)) \right\} \right\} \\ & \int p_x(x) \log(D_\phi(x)) dx + \int p_g(x) \log(1 - D_\phi(x)) dx \\ & \int \left\{ p_x(x) \log(D_\phi(x)) + p_g(x) \log(1 - D_\phi(x)) \right\} dx \end{aligned}$$

# Problem formulation

$$\min_{\theta} \max_{\phi} \left\{ E_{x \sim p_x} \left\{ \log(D_\phi(x)) \right\} + E_{z \sim p_z} \left\{ \log(1 - D_\phi(g_\theta(z))) \right\} \right\}$$

## Optimal Discriminator (Proof sketch)

$$\begin{aligned} & \min_{\theta} \max_{\phi} \left\{ E_{x \sim p_x} \left\{ \log(D_\phi(x)) \right\} + E_{x \sim p_g} \left\{ \log(1 - D_\phi(x)) \right\} \right\} \\ & \int p_x(x) \log(D_\phi(x)) dx + \int p_g(x) \log(1 - D_\phi(x)) dx \\ & \underbrace{\int \left\{ p_x(x) \log(D_\phi(x)) + p_g(x) \log(1 - D_\phi(x)) \right\} dx}_{\downarrow} \\ & a \log(y) + b \log(1 - y) \end{aligned}$$

# Problem formulation

$$\min_{\theta} \max_{\phi} \left\{ E_{x \sim p_x} \left\{ \log(D_\phi(x)) \right\} + E_{z \sim p_z} \left\{ \log(1 - D_\phi(g_\theta(z))) \right\} \right\}$$

## Optimal Discriminator (Proof sketch)

$$\begin{aligned} & \min_{\theta} \max_{\phi} \left\{ E_{x \sim p_x} \left\{ \log(D_\phi(x)) \right\} + E_{x \sim p_g} \left\{ \log(1 - D_\phi(x)) \right\} \right\} \\ & \int p_x(x) \log(D_\phi(x)) dx + \int p_g(x) \log(1 - D_\phi(x)) dx \\ & \underbrace{\int \left\{ p_x(x) \log(D_\phi(x)) + p_g(x) \log(1 - D_\phi(x)) \right\} dx}_{\downarrow} \\ & a \log(y) + b \log(1 - y) \end{aligned}$$

$$\forall (a, b) \in \mathbb{R}^2 \setminus \{0, 0\} \quad \frac{\partial a \log(y) + b \log(1 - y)}{\partial y} = 0 \iff y^* = \frac{a}{a + b}$$

# Problem formulation

$$\min_{\theta} \max_{\phi} \left\{ E_{x \sim p_x} \left\{ \log(D_\phi(x)) \right\} + E_{z \sim p_z} \left\{ \log(1 - D_\phi(g_\theta(z))) \right\} \right\}$$

## Optimal Discriminator (Proof sketch)

$$\begin{aligned} & \min_{\theta} \max_{\phi} \left\{ E_{x \sim p_x} \left\{ \log(D_\phi(x)) \right\} + E_{x \sim p_g} \left\{ \log(1 - D_\phi(x)) \right\} \right\} \\ & \int p_x(x) \log(D_\phi(x)) dx + \int p_g(x) \log(1 - D_\phi(x)) dx \\ & \underbrace{\int \left\{ p_x(x) \log(D_\phi(x)) + p_g(x) \log(1 - D_\phi(x)) \right\} dx}_{\downarrow} \\ & a \log(y) + b \log(1 - y) \end{aligned}$$

$$\forall (a, b) \in \mathbb{R}^2 \setminus \{0, 0\} \quad \frac{\partial a \log(y) + b \log(1 - y)}{\partial y} = 0 \iff y^* = \frac{a}{a + b}$$

$$D_\phi^*(x) = \frac{p_x(x)}{p_x(x) + p_g(x)}$$

# Problem formulation

$$\min_{\theta} \max_{\phi} \left\{ E_{x \sim p_x} \left\{ \log(D_\phi(x)) \right\} + E_{z \sim p_z} \left\{ \log(1 - D_\phi(g_\theta(z))) \right\} \right\}$$

**Optimal Generator (Proof sketch)**

$$\begin{aligned} & \min_{\theta} \max_{\phi} \left\{ E_{x \sim p_x} \left\{ \log(D_\phi(x)) \right\} + E_{x \sim p_g} \left\{ \log(1 - D_\phi(x)) \right\} \right\} \\ & \int p_x(x) \log(D_\phi(x)) dx + \int p_g(x) \log(1 - D_\phi(x)) dx \end{aligned}$$

# Problem formulation

$$\min_{\theta} \max_{\phi} \left\{ E_{x \sim p_x} \left\{ \log(D_\phi(x)) \right\} + E_{z \sim p_z} \left\{ \log(1 - D_\phi(g_\theta(z))) \right\} \right\}$$

**Optimal Generator (Proof sketch)**

$$\begin{aligned} & \min_{\theta} \max_{\phi} \left\{ E_{x \sim p_x} \left\{ \log(D_\phi(x)) \right\} + E_{x \sim p_g} \left\{ \log(1 - D_\phi(x)) \right\} \right\} \\ & \int p_x(x) \log(D_\phi(x)) dx + \int p_g(x) \log(1 - D_\phi(x)) dx \\ & \quad \downarrow \\ & D_\phi^*(x) = \frac{p_x(x)}{p_x(x) + p_g(x)} \\ & \quad \downarrow \\ & \int p_x(x) \log \left( \frac{p_x(x)}{p_x(x) + p_g(x)} \right) dx + \int p_g(x) \log \left( 1 - \frac{p_x(x)}{p_x(x) + p_g(x)} \right) dx \end{aligned}$$

# Problem formulation

$$\min_{\theta} \max_{\phi} \left\{ E_{x \sim p_x} \left\{ \log(D_{\phi}(x)) \right\} + E_{z \sim p_z} \left\{ \log(1 - D_{\phi}(g_{\theta}(z))) \right\} \right\}$$

## Optimal Generator (Proof sketch)

$$\begin{aligned} & \min_{\theta} \max_{\phi} \left\{ E_{x \sim p_x} \left\{ \log(D_{\phi}(x)) \right\} + E_{x \sim p_g} \left\{ \log(1 - D_{\phi}(x)) \right\} \right\} \\ & \int p_x(x) \log(D_{\phi}(x)) dx + \int p_g(x) \log(1 - D_{\phi}(x)) dx \\ & \quad \downarrow \\ & D_{\phi}^*(x) = \frac{p_x(x)}{p_x(x) + p_g(x)} \\ & \quad \downarrow \\ & \int p_x(x) \log \left( \frac{p_x(x)}{p_x(x) + p_g(x)} \right) dx + \int p_g(x) \log \left( 1 - \frac{p_x(x)}{p_x(x) + p_g(x)} \right) dx \\ & \int p_x(x) \log \left( \frac{p_x(x)}{p_x(x) + p_g(x)} \right) dx + \int p_g(x) \log \left( \frac{p_g(x)}{p_x(x) + p_g(x)} \right) dx \end{aligned}$$

# Problem formulation

$$\min_{\theta} \max_{\phi} \left\{ E_{x \sim p_x} \left\{ \log(D_\phi(x)) \right\} + E_{z \sim p_z} \left\{ \log(1 - D_\phi(g_\theta(z))) \right\} \right\}$$

## Optimal Generator (Proof sketch)

$$\begin{aligned} & \min_{\theta} \max_{\phi} \left\{ E_{x \sim p_x} \left\{ \log(D_\phi(x)) \right\} + E_{x \sim p_g} \left\{ \log(1 - D_\phi(x)) \right\} \right\} \\ & \int p_x(x) \log(D_\phi(x)) dx + \int p_g(x) \log(1 - D_\phi(x)) dx \\ & \quad \downarrow \\ & D_\phi^*(x) = \frac{p_x(x)}{p_x(x) + p_g(x)} \\ & \quad \downarrow \\ & \int p_x(x) \log \left( \frac{p_x(x)}{p_x(x) + p_g(x)} \right) dx + \int p_g(x) \log \left( 1 - \frac{p_x(x)}{p_x(x) + p_g(x)} \right) dx \\ & \quad \int p_x(x) \log \left( \frac{p_x(x)}{p_x(x) + p_g(x)} \right) dx + \int p_g(x) \log \left( \frac{p_g(x)}{p_x(x) + p_g(x)} \right) dx \\ & \quad \int p_x(x) \log \left( \frac{p_x(x)}{\frac{p_x(x) + p_g(x)}{2}} \right) dx + \int p_g(x) \log \left( \frac{p_g(x)}{\frac{p_x(x) + p_g(x)}{2}} \right) dx - 2 \log(2) \end{aligned}$$

# Problem formulation

$$\min_{\theta} \max_{\phi} \left\{ E_{x \sim p_x} \left\{ \log(D_\phi(x)) \right\} + E_{z \sim p_z} \left\{ \log(1 - D_\phi(g_\theta(z))) \right\} \right\}$$

## Optimal Generator (Proof sketch)

$$\begin{aligned}
 & \min_{\theta} \max_{\phi} \left\{ E_{x \sim p_x} \left\{ \log(D_\phi(x)) \right\} + E_{x \sim p_g} \left\{ \log(1 - D_\phi(x)) \right\} \right\} \\
 & \int p_x(x) \log(D_\phi(x)) dx + \int p_g(x) \log(1 - D_\phi(x)) dx \\
 & \quad \downarrow \\
 & D_\phi^*(x) = \frac{p_x(x)}{p_x(x) + p_g(x)} \\
 & \quad \downarrow \\
 & \int p_x(x) \log \left( \frac{p_x(x)}{p_x(x) + p_g(x)} \right) dx + \int p_g(x) \log \left( 1 - \frac{p_x(x)}{p_x(x) + p_g(x)} \right) dx \\
 & \quad \int p_x(x) \log \left( \frac{p_x(x)}{p_x(x) + p_g(x)} \right) dx + \int p_g(x) \log \left( \frac{p_g(x)}{p_x(x) + p_g(x)} \right) dx \\
 & \quad \underline{\int p_x(x) \log \left( \frac{p_x(x)}{\frac{p_x(x)+p_g(x)}{2}} \right) dx + \int p_g(x) \log \left( \frac{p_g(x)}{\frac{p_x(x)+p_g(x)}{2}} \right) dx - 2 \log(2)} \\
 & \quad \downarrow \\
 & 2JSD(p_x, p_g)
 \end{aligned}$$

# Problem formulation

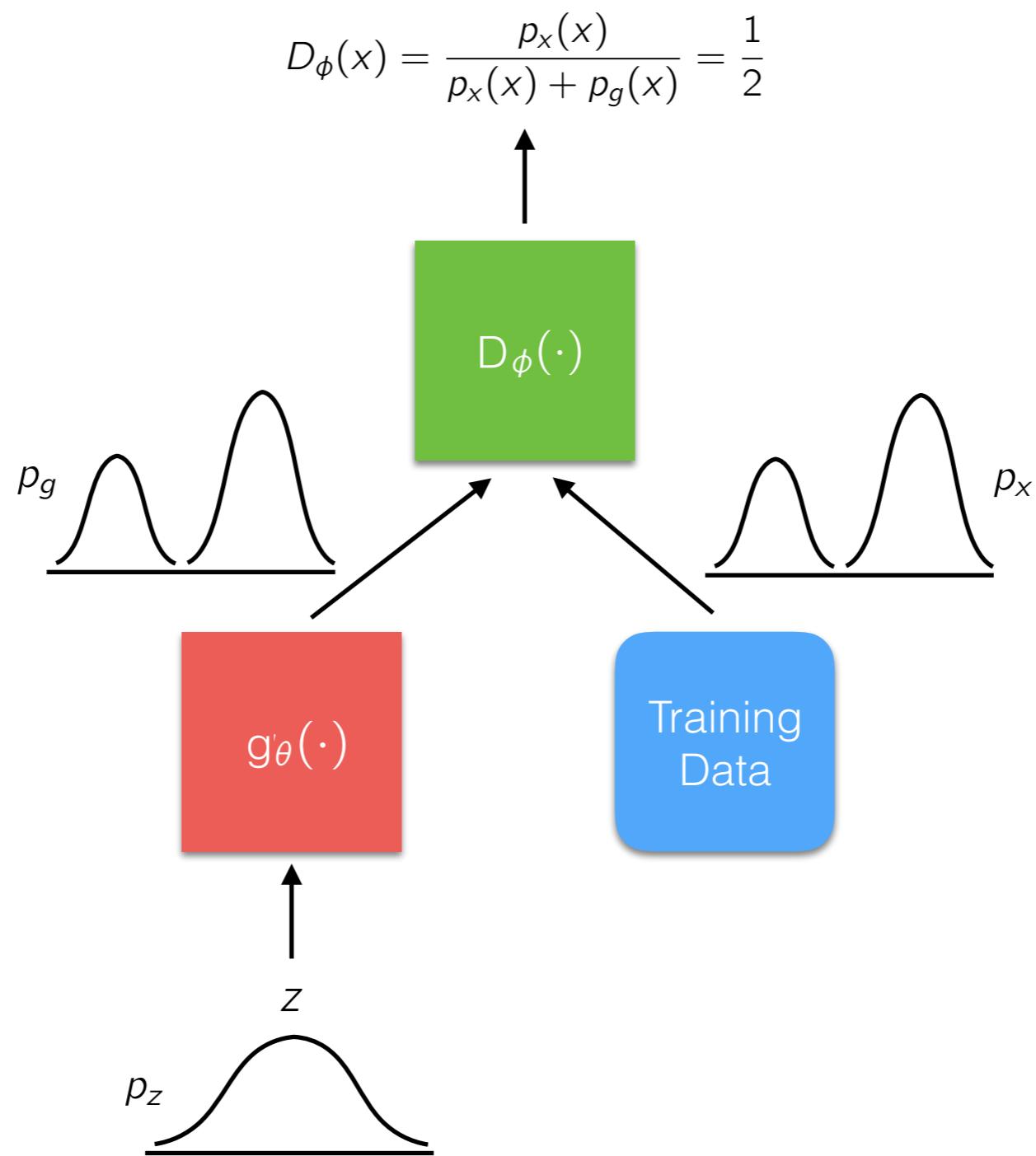
$$\min_{\theta} \max_{\phi} \left\{ E_{x \sim p_x} \left\{ \log(D_\phi(x)) \right\} + E_{z \sim p_z} \left\{ \log(1 - D_\phi(g_\theta(z))) \right\} \right\}$$

## Optimal Generator (Proof sketch)

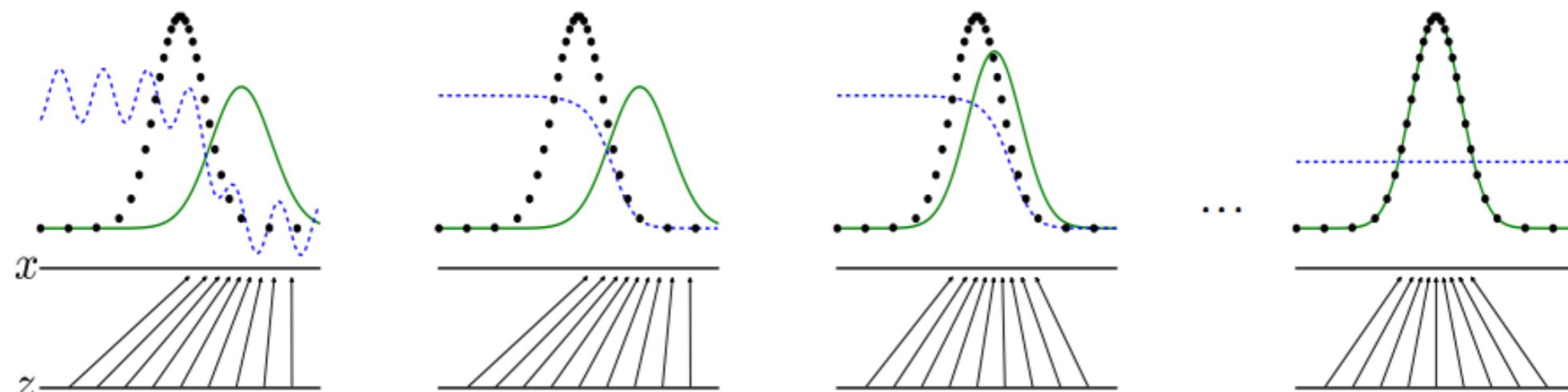
$$\begin{aligned}
 & \min_{\theta} \max_{\phi} \left\{ E_{x \sim p_x} \left\{ \log(D_\phi(x)) \right\} + E_{x \sim p_g} \left\{ \log(1 - D_\phi(x)) \right\} \right\} \\
 & \int p_x(x) \log(D_\phi(x)) dx + \int p_g(x) \log(1 - D_\phi(x)) dx \\
 & \quad \downarrow \\
 & D_\phi^*(x) = \frac{p_x(x)}{p_x(x) + p_g(x)} \\
 & \quad \downarrow \\
 & \int p_x(x) \log \left( \frac{p_x(x)}{p_x(x) + p_g(x)} \right) dx + \int p_g(x) \log \left( 1 - \frac{p_x(x)}{p_x(x) + p_g(x)} \right) dx \\
 & \quad \int p_x(x) \log \left( \frac{p_x(x)}{p_x(x) + p_g(x)} \right) dx + \int p_g(x) \log \left( \frac{p_g(x)}{p_x(x) + p_g(x)} \right) dx \\
 & \quad \underline{\int p_x(x) \log \left( \frac{p_x(x)}{\frac{p_x(x)+p_g(x)}{2}} \right) dx + \int p_g(x) \log \left( \frac{p_g(x)}{\frac{p_x(x)+p_g(x)}{2}} \right) dx - 2 \log(2)} \\
 & \quad \downarrow \\
 & 2JSD(p_x, p_g)
 \end{aligned}$$

The minimum value is  $-2\log(2)$  when  $JSD(p_x, p_g) = 0 \iff p_g(x) = p_x(x), \forall x$

# In practice

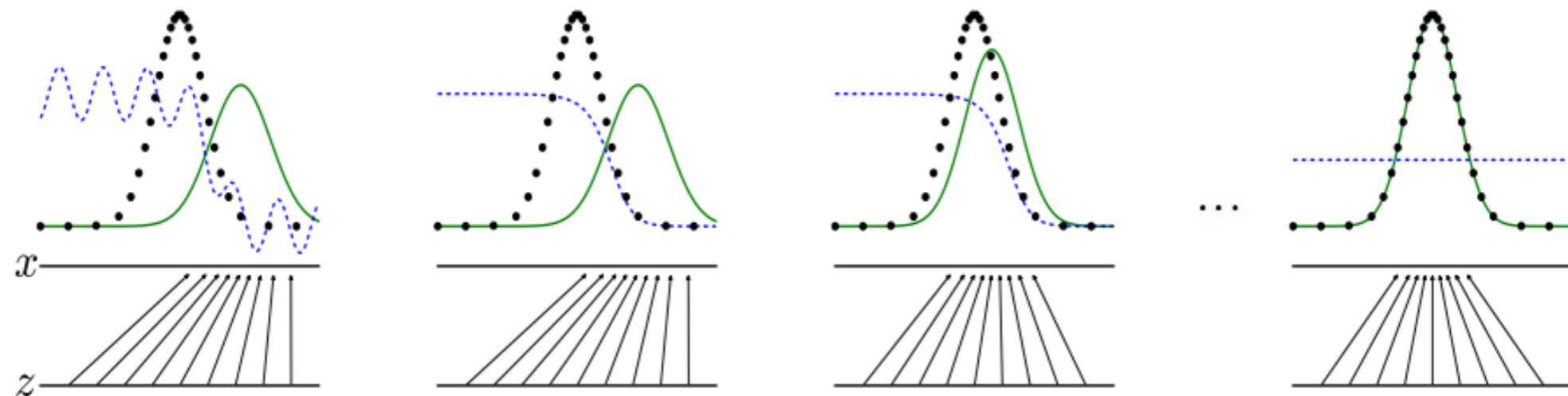


# In practice



“Generative Adversarial Networks”  
ICLR 2014 Ian Goodfellow

# In practice



“Generative Adversarial Networks”  
ICLR 2014 Ian Goodfellow

DON’T WORRY, THERE IS A DEMO

Why is it better than traditional  
generative approaches?

# Traditional generative approaches

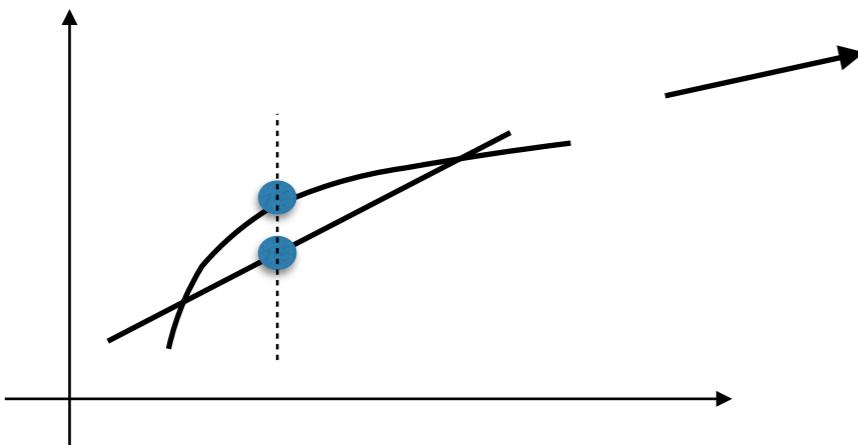
$$\ln p(\text{DATA}) = \ln \int p(\text{DATA}, \theta) d\theta$$

# Traditional generative approaches

$$\begin{aligned}\ln p(\text{DATA}) &= \ln \int p(\text{DATA}, \theta) d\theta \\ &= \ln \int q(\theta) \frac{p(\text{DATA}, \theta)}{q(\theta)} d\theta\end{aligned}$$

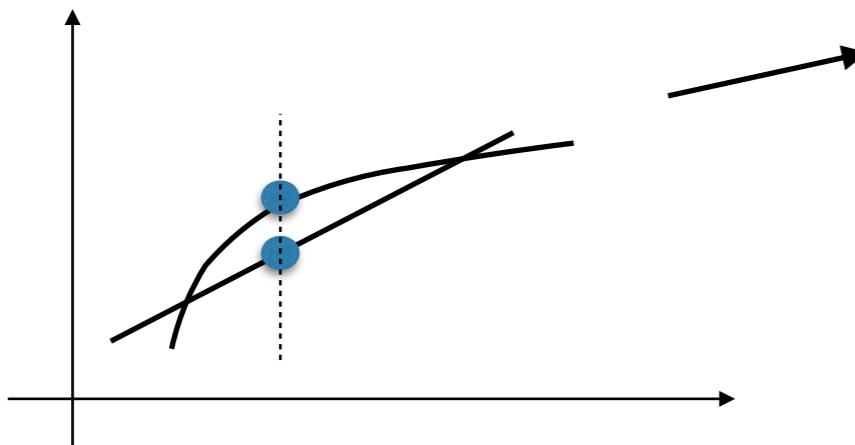
# Traditional generative approaches

$$\begin{aligned}\ln p(\text{DATA}) &= \ln \int p(\text{DATA}, \theta) d\theta \\ &= \ln \int q(\theta) \frac{p(\text{DATA}, \theta)}{q(\theta)} d\theta \\ &\geq \int q(\theta) \ln \frac{p(\text{DATA}, \theta)}{q(\theta)} d\theta\end{aligned}$$



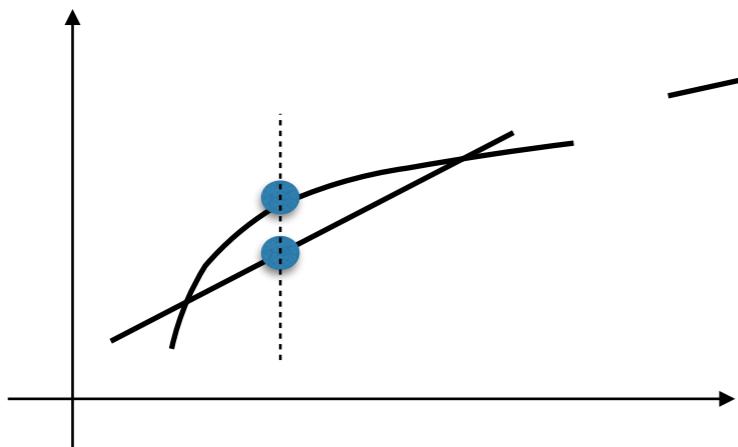
# Traditional generative approaches

$$\begin{aligned}\ln p(\text{DATA}) &= \ln \int p(\text{DATA}, \theta) d\theta \\ &= \ln \int q(\theta) \frac{p(\text{DATA}, \theta)}{q(\theta)} d\theta \\ &\geq \int q(\theta) \ln \frac{p(\text{DATA}, \theta)}{q(\theta)} d\theta \\ &\doteq -KL(q(\theta), p(\text{DATA}))\end{aligned}$$



# Traditional generative approaches

$$\begin{aligned}\ln p(\text{DATA}) &= \ln \int p(\text{DATA}, \theta) d\theta \\ &= \ln \int q(\theta) \frac{p(\text{DATA}, \theta)}{q(\theta)} d\theta \\ &\geq \int q(\theta) \ln \frac{p(\text{DATA}, \theta)}{q(\theta)} d\theta \\ &\doteq -KL(q(\theta), p(\text{DATA}))\end{aligned}$$



The maximum is achieved for  $q(\theta) = p(\theta|DATA)$

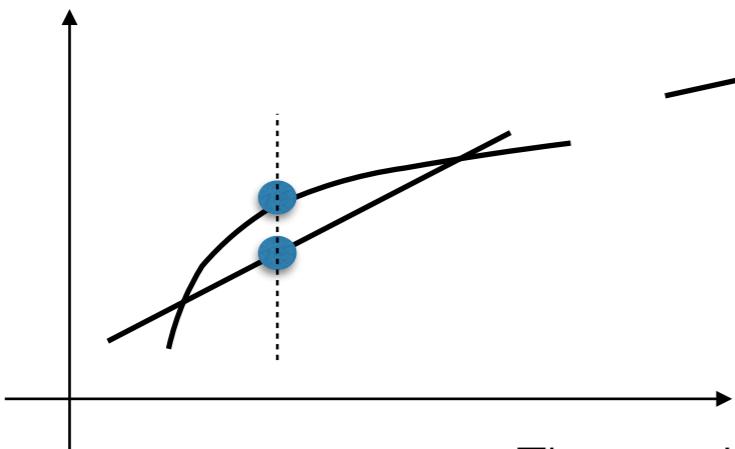


$q(\theta)$  can be regarded as  $p_g$

$p(\theta, DATA)$  can be regarded as  $p_x$

# Traditional generative approaches

$$\begin{aligned}\ln p(\text{DATA}) &= \ln \int p(\text{DATA}, \theta) d\theta \\ &= \ln \int q(\theta) \frac{p(\text{DATA}, \theta)}{q(\theta)} d\theta \\ &\geq \int q(\theta) \ln \frac{p(\text{DATA}, \theta)}{q(\theta)} d\theta \\ &\doteq -KL(q(\theta), p(\text{DATA}))\end{aligned}$$



The maximum is achieved for  $q(\theta) = p(\theta|DATA)$



$q(\theta)$  can be regarded as  $p_g$

$p(\theta, \text{DATA})$  can be regarded as  $p_x$



$\boxed{\min KL(p_g, p_x)}$

# Traditional generative approaches vs. GANS

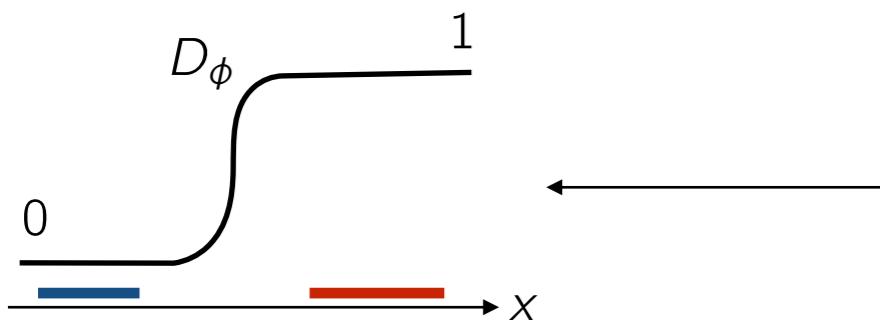
	Traditional Generative Approaches	GANs
Optimization	$\min KL(p_g, p_x)$	$\min JSD(p_g, p_x)$
	0 (Low cost for mode dropping)	$\log(2)$
	Infinity (High cost for fake data)	$\log(2)$
Minimum	0	0

Is there any issue?

# Problem of vanishing gradients

**Theorem 2.4:** if the discriminator is close to optimality, namely

$$D_\phi(x) \simeq \frac{p_x(x)}{p_x(x) + p_g(x)}$$



In other words,

$$D_\phi(x) \simeq 1, \forall x \in \text{supp}(p_x) \quad \text{---}$$
$$D_\phi(x) \simeq 0, \forall x \in \text{supp}(p_g) \quad \text{---}$$

(which is a very common case, i.e. Theorem 2.1-2.2)  
and the Jacobian of the generator is bounded  
(by any scalar  $M$ )

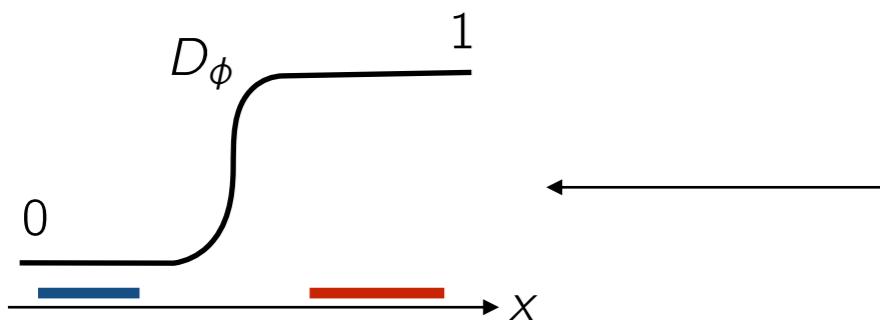


$$\|\nabla_\theta E_{z \sim p_z} \left\{ \log(1 - D_\phi(g_\theta(z))) \right\}\|_2 \leq M \frac{\epsilon}{1 - \epsilon}$$

# Problem of vanishing gradients

**Theorem 2.4:** if the discriminator is close to optimality, namely

$$D_\phi(x) \simeq \frac{p_x(x)}{p_x(x) + p_g(x)}$$



In other words,

$$D_\phi(x) \simeq 1, \forall x \in \text{supp}(p_x)$$

$$D_\phi(x) \simeq 0, \forall x \in \text{supp}(p_g)$$

(which is a very common case, i.e. Theorem 2.1-2.2)  
and the Jacobian of the generator is bounded  
(by any scalar  $M$ )

⇓

$$\|\nabla_\theta E_{z \sim p_z} \left\{ \log(1 - D_\phi(g_\theta(z))) \right\}\|_2 \leq M \frac{\epsilon}{1 - \epsilon}$$

This happens when the discriminator is “too strong”. For example, when it is updated more frequently than the generator

# Problem of vanishing gradients

$$\begin{aligned}
 \textbf{Proof: } & \| \nabla_{\theta} E_{z \sim p_z} \left\{ \log(1 - D_{\phi}(g_{\theta}(z))) \right\} \|_2^2 = \| E_{z \sim p_z} \left\{ \nabla_{\theta} \log(1 - D_{\phi}(g_{\theta}(z))) \right\} \|_2^2 \\
 & = \| E_{z \sim p_z} \left\{ - \frac{\nabla_{\theta} D_{\phi}(g_{\theta}(z))}{1 - D_{\phi}(g_{\theta}(z))} \right\} \|_2^2 \\
 \text{Jensen's inequality} \longrightarrow & \leq E_{z \sim p_z} \left\{ \left\| \frac{\nabla_{\theta} D_{\phi}(g_{\theta}(z))}{1 - D_{\phi}(g_{\theta}(z))} \right\|_2^2 \right\} \\
 & = E_{z \sim p_z} \left\{ \frac{\| \nabla_{\theta} D_{\phi}(g_{\theta}(z)) \|_2^2}{|1 - D_{\phi}(g_{\theta}(z))|^2} \right\} \\
 & = E_{z \sim p_z} \left\{ \frac{\| \nabla_x D_{\phi}(g_{\theta}(z)) J(g_{\theta}(z)) \|_2^2}{|1 - D_{\phi}(g_{\theta}(z))|^2} \right\} \\
 \text{Cauchy-Schwartz inequality} \longrightarrow & \leq E_{z \sim p_z} \left\{ \frac{\| \nabla_x D_{\phi}(g_{\theta}(z)) \|_2^2 \| J(g_{\theta}(z)) \|_2^2}{|1 - D_{\phi}(g_{\theta}(z))|^2} \right\}
 \end{aligned}$$

Since the discriminator is close to optimality, namely

$$\|D_{\phi} - D_{\phi}^*\| \doteq \sup_{x \in \mathcal{X}} \left\{ |D_{\phi} - D_{\phi}^*| + \|\nabla_x D_{\phi} - \nabla_x D_{\phi}^*\|_2 \right\} < \epsilon$$

(values and gradients are both similar) and since

$$\|\nabla_x D_{\phi} - \nabla_x D_{\phi}^*\|_2^2 \geq \|\nabla_x D_{\phi}\|_2^2 - \|\nabla_x D_{\phi}^*\|_2^2$$

$$\|\nabla_x D_{\phi}\|_2^2 < \|\nabla_x D_{\phi}^*\|_2^2 + \epsilon^2 \quad \textbf{Cont.}$$

# Problem of vanishing gradients

**Proof:**  $E_{z \sim p_z} \left\{ \frac{\|\nabla_x D_\phi(g_\theta(z))\|_2^2 \|J(g_\theta(z))\|_2^2}{|1 - D_\phi(g_\theta(z))|^2} \right\} < E_{z \sim p_z} \left\{ \frac{(\|\nabla_x D_\phi^*(g_\theta(z))\|_2^2 \|J(g_\theta(z))\|_2^2 + \epsilon^2) |1 - D_\phi(g_\theta(z))|^2}{|1 - D_\phi(g_\theta(z))|^2} \right\}$

$$\begin{aligned} |D_\phi - D_\phi^*| &< \epsilon \\ |1 - D_\phi + 1 - D_\phi^*| &< \epsilon \\ |1 - D_\phi^* - (1 - D_\phi)| &< \epsilon \\ |1 - D_\phi^*| - |(1 - D_\phi)| &\leq |1 - D_\phi^* - (1 - D_\phi)| \\ |1 - D_\phi^*| - |(1 - D_\phi)| &< \epsilon \end{aligned} \quad \xrightarrow{\quad} \quad E_{z \sim p_z} \left\{ \frac{(\|\nabla_x D_\phi^*(g_\theta(z))\|_2^2 \|J(g_\theta(z))\|_2^2 + \epsilon^2) |1 - D_\phi^*(g_\theta(z))| - \epsilon)^2}{(|1 - D_\phi^*(g_\theta(z))| - \epsilon)^2} \right\}$$

At optimality

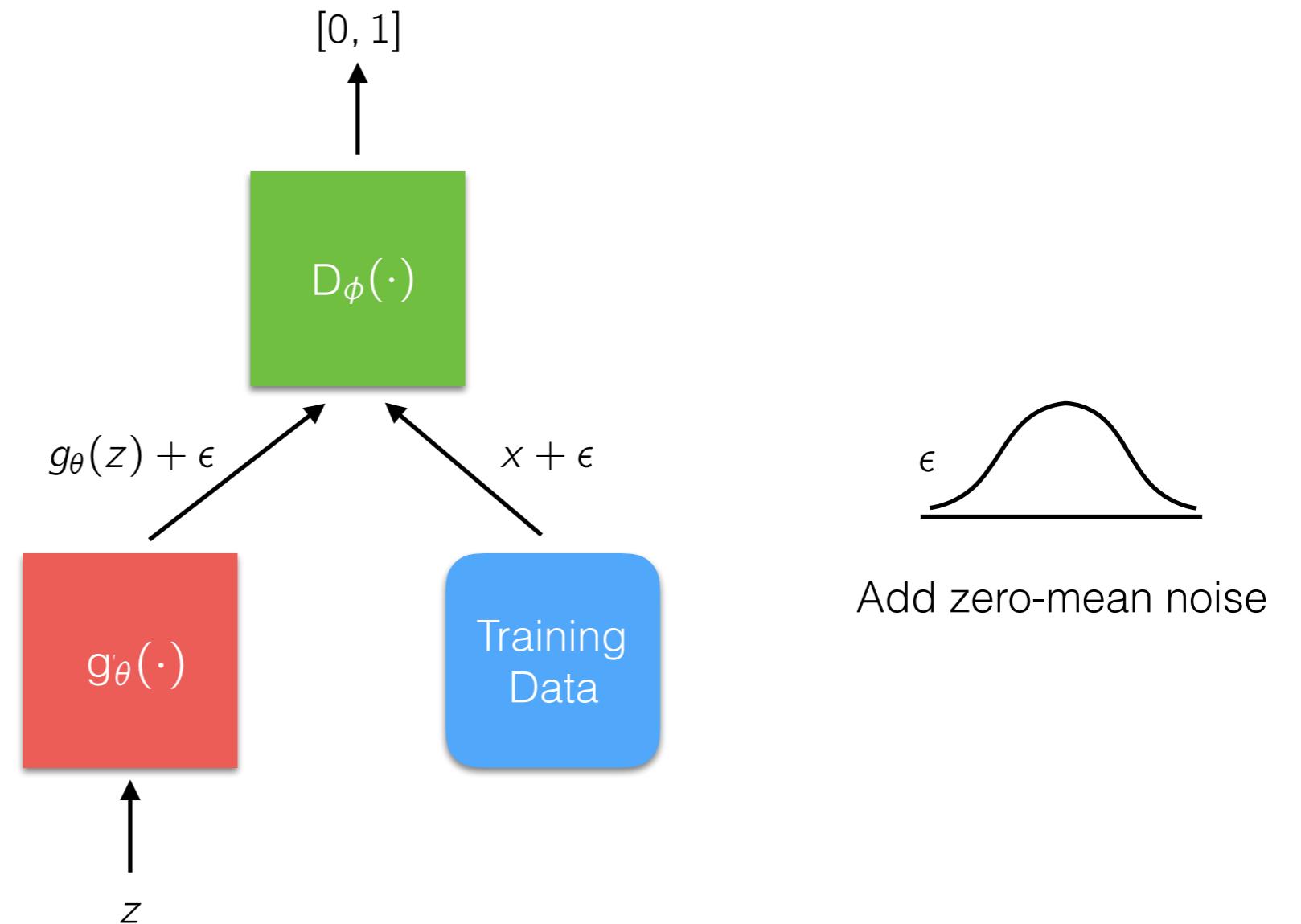
$$\begin{aligned} \nabla_x D_\phi^*(x) &= 0 \\ D_\phi^*(x) &= 0, \forall x \in \text{supp}(p_g) \setminus \mathcal{L} \end{aligned}$$

Therefore

$$\begin{aligned} \|\nabla_\theta E_{z \sim p_z} \left\{ \log(1 - D_\phi(g_\theta(z))) \right\}\|_2^2 &< E_{z \sim p_z} \left\{ \frac{\epsilon^2 \|J(g_\theta(z))\|_2^2}{(1 - \epsilon)^2} \right\} \\ &\leq M^2 \frac{\epsilon^2}{(1 - \epsilon)^2} \quad \mathbf{QED} \end{aligned}$$

How do you solve it?

# Recent Solution



“Towards Principled Methods for Training Generative Adversarial Networks”  
ICLR 2017 Martin Arjovsky, Leon Bottou

# Recent Solution

**Theorem 3.2:**

$$D_{\phi}^*(x) = \frac{p_{x+\epsilon}(x)}{p_{x+\epsilon}(x) + p_{g+\epsilon}(x)}$$

$$\epsilon \sim \mathcal{N}(0, \sigma^2 I)$$

# Recent Solution

**Theorem 3.2:**

$$D_\phi^*(x) = \frac{p_{x+\epsilon}(x)}{p_{x+\epsilon}(x) + p_{g+\epsilon}(x)}$$

$$\epsilon \sim \mathcal{N}(0, \sigma^2 I)$$

⇓

$$\begin{aligned} E_{z \sim p_z} \left\{ \nabla_\theta \log(1 - D_\phi^*(g_\theta(z))) \right\} &= E_{z \sim p_z} \left\{ a(z) \int p_\epsilon(g_\theta(z) - y) \nabla_\theta \|g_\theta(z) - y\|^2 p_x(y) dy \right. \\ &\quad \left. - b(z) \int p_\epsilon(g_\theta(z) - y) \nabla_\theta \|g_\theta(z) - y\|^2 p_g(y) dy \right\} \end{aligned}$$

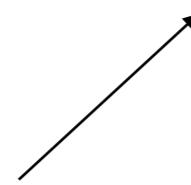
where

$$b(z) = a(z) \frac{p_{x+\epsilon}}{p_{g+\epsilon}}$$

# Recent Solution

**Proof:**

$$\begin{aligned}
 E_{z \sim p_z} \left\{ \nabla_{\theta} \log(1 - D_{\phi}^*(g_{\theta}(z))) \right\} &= E_{z \sim p_z} \left\{ \nabla_{\theta} \log \frac{p_{g+\epsilon}(g_{\theta}(z))}{p_{x+\epsilon}(g_{\theta}(z)) + p_{g+\epsilon}(g_{\theta}(z))} \right\} \\
 &= E_{z \sim p_z} \left\{ \nabla_{\theta} \log(p_{g+\epsilon}(g_{\theta}(z))) - \nabla_{\theta} \log(p_{x+\epsilon}(g_{\theta}(z)) + p_{g+\epsilon}(g_{\theta}(z))) \right\} \\
 &= E_{z \sim p_z} \left\{ \frac{\nabla_{\theta} p_{g+\epsilon}(g_{\theta}(z))}{p_{g+\epsilon}(g_{\theta}(z))} - \frac{\nabla_{\theta} p_{x+\epsilon}(g_{\theta}(z)) + \nabla_{\theta} p_{g+\epsilon}(g_{\theta}(z))}{p_{x+\epsilon}(g_{\theta}(z)) + p_{g+\epsilon}(g_{\theta}(z))} \right\} \\
 &= E_{z \sim p_z} \left\{ \nabla_{\theta} \{-p_{x+\epsilon}(g_{\theta}(z))\} - \right. \\
 &\quad \left. \frac{1}{p_{x+\epsilon}(g_{\theta}(z)) + p_{g+\epsilon}(g_{\theta}(z))} \frac{p_{x+\epsilon}(g_{\theta}(z))}{p_{g+\epsilon}(g_{\theta}(z))} \nabla_{\theta} \{-p_{g+\epsilon}(g_{\theta}(z))\} \right\} \\
 &= E_{z \sim p_z} \left\{ 2\sigma^2 a(z) \nabla_{\theta} \{-p_{x+\epsilon}(g_{\theta}(z))\} - 2\sigma^2 b(z) \nabla_{\theta} \{-p_{g+\epsilon}(g_{\theta}(z))\} \right\}
 \end{aligned}$$



$$a(z) \doteq \frac{1}{2\sigma^2} \frac{1}{p_{x+\epsilon}(g_{\theta}(z)) + p_{g+\epsilon}(g_{\theta}(z))}$$

$$b(z) \doteq \frac{1}{2\sigma^2} \frac{1}{p_{x+\epsilon}(g_{\theta}(z)) + p_{g+\epsilon}(g_{\theta}(z))} \frac{p_{x+\epsilon}(g_{\theta}(z))}{p_{g+\epsilon}(g_{\theta}(z))}$$

**Cont.**

# Recent Solution

**Proof:** Recall that adding two independent random variables produces a random variable with density obtained by convolving the two original densities

$$\begin{aligned}
 E_{z \sim p_z} \left\{ \nabla_\theta \log(1 - D_\phi^*(g_\theta(z))) \right\} &= E_{z \sim p_z} \left\{ -2\sigma^2 a(z) \nabla_\theta \int p_\epsilon(g_\theta(z) - y) p_x(y) dy + \right. \\
 &\quad \left. 2\sigma^2 b(z) \nabla_\theta \int p_\epsilon(g_\theta(z) - y) p_g(y) dy \right\} \\
 &= E_{z \sim p_z} \left\{ -2\sigma^2 a(z) \nabla_\theta \int \frac{1}{Z} e^{-\frac{\|g_\theta(z)-y\|^2}{2\sigma^2}} p_x(y) dy + \right. \\
 &\quad \left. 2\sigma^2 b(z) \nabla_\theta \int \frac{1}{Z} e^{-\frac{\|g_\theta(z)-y\|^2}{2\sigma^2}} p_g(y) dy \right\} \\
 &= E_{z \sim p_z} \left\{ -2\sigma^2 a(z) \int \nabla_\theta \frac{1}{Z} e^{-\frac{\|g_\theta(z)-y\|^2}{2\sigma^2}} p_x(y) dy + \right. \\
 &\quad \left. 2\sigma^2 b(z) \int \nabla_\theta \frac{1}{Z} e^{-\frac{\|g_\theta(z)-y\|^2}{2\sigma^2}} p_g(y) dy \right\} \\
 &= E_{z \sim p_z} \left\{ a(z) \int \frac{1}{Z} e^{-\frac{\|g_\theta(z)-y\|^2}{2\sigma^2}} \nabla_\theta \|g_\theta(z) - y\|^2 p_x(y) dy - \right. \\
 &\quad \left. b(z) \int \frac{1}{Z} e^{-\frac{\|g_\theta(z)-y\|^2}{2\sigma^2}} \nabla_\theta \|g_\theta(z) - y\|^2 p_g(y) dy \right\} \\
 &= E_{z \sim p_z} \left\{ a(z) \int p_\epsilon(g_\theta(z) - y) \nabla_\theta \|g_\theta(z) - y\|^2 p_x(y) dy - \right. \\
 &\quad \left. b(z) \int p_\epsilon(g_\theta(z) - y) \nabla_\theta \|g_\theta(z) - y\|^2 p_g(y) dy \right\} \quad \text{QED}
 \end{aligned}$$

**MISTAKE**  
**(it should be derived)!**  
 It comes from the  
 optimal discriminator  
 formula which is  
 assumed fixed (and  
 therefore not  
 dependent on \theta)

Why does it work?

# Interpretation

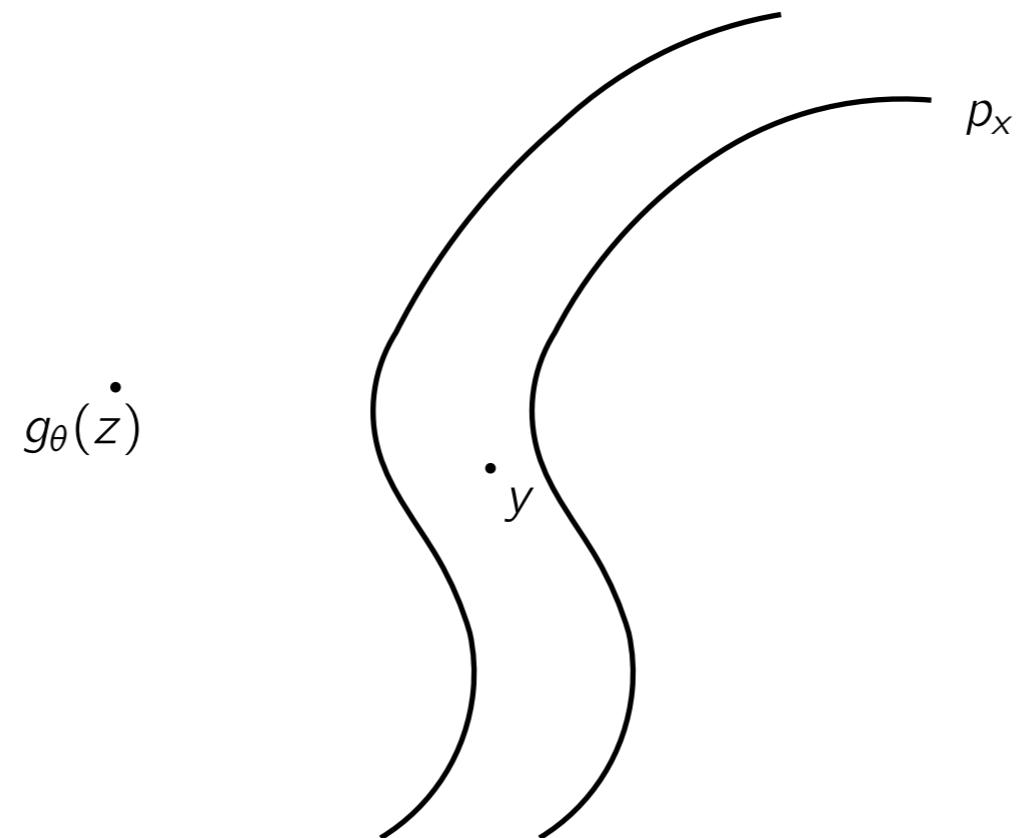
$$\begin{aligned} E_{z \sim p_z} \left\{ \nabla_{\theta} \log(1 - D_{\phi}^*(g_{\theta}(z))) \right\} &= E_{z \sim p_z} \left\{ a(z) \int p_{\epsilon}(g_{\theta}(z) - y) \nabla_{\theta} \|g_{\theta}(z) - y\|^2 p_x(y) dy \right. \\ &\quad \left. - b(z) \int p_{\epsilon}(g_{\theta}(z) - y) \nabla_{\theta} \|g_{\theta}(z) - y\|^2 p_g(y) dy \right\} \end{aligned}$$

# Interpretation

$$\begin{aligned} E_{z \sim p_z} \left\{ \nabla_{\theta} \log(1 - D_{\phi}^*(g_{\theta}(z))) \right\} &= E_{z \sim p_z} \left\{ a(z) \int p_{\epsilon}(g_{\theta}(z) - y) \nabla_{\theta} \|g_{\theta}(z) - y\|^2 p_x(y) dy \right. \\ &\quad \left. - b(z) \int p_{\epsilon}(g_{\theta}(z) - y) \nabla_{\theta} \|g_{\theta}(z) - y\|^2 p_g(y) dy \right\} \end{aligned}$$

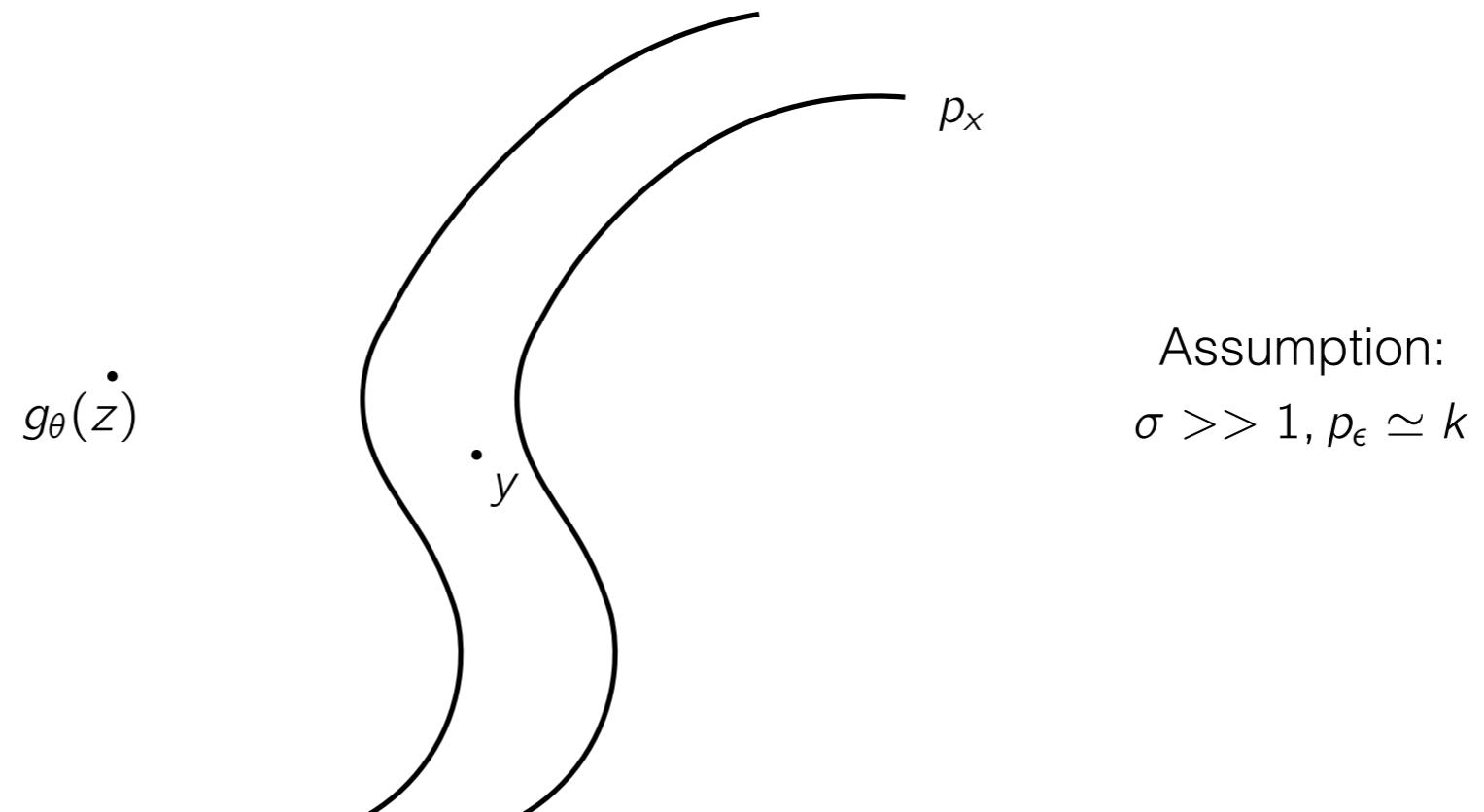
# Interpretation

$$\begin{aligned} E_{z \sim p_z} \left\{ \nabla_{\theta} \log(1 - D_{\phi}^*(g_{\theta}(z))) \right\} &= E_{z \sim p_z} \left\{ a(z) \int p_{\epsilon}(g_{\theta}(z) - y) \nabla_{\theta} \|g_{\theta}(z) - y\|^2 p_x(y) dy \right. \\ &\quad \left. - b(z) \int p_{\epsilon}(g_{\theta}(z) - y) \nabla_{\theta} \|g_{\theta}(z) - y\|^2 p_g(y) dy \right\} \end{aligned}$$



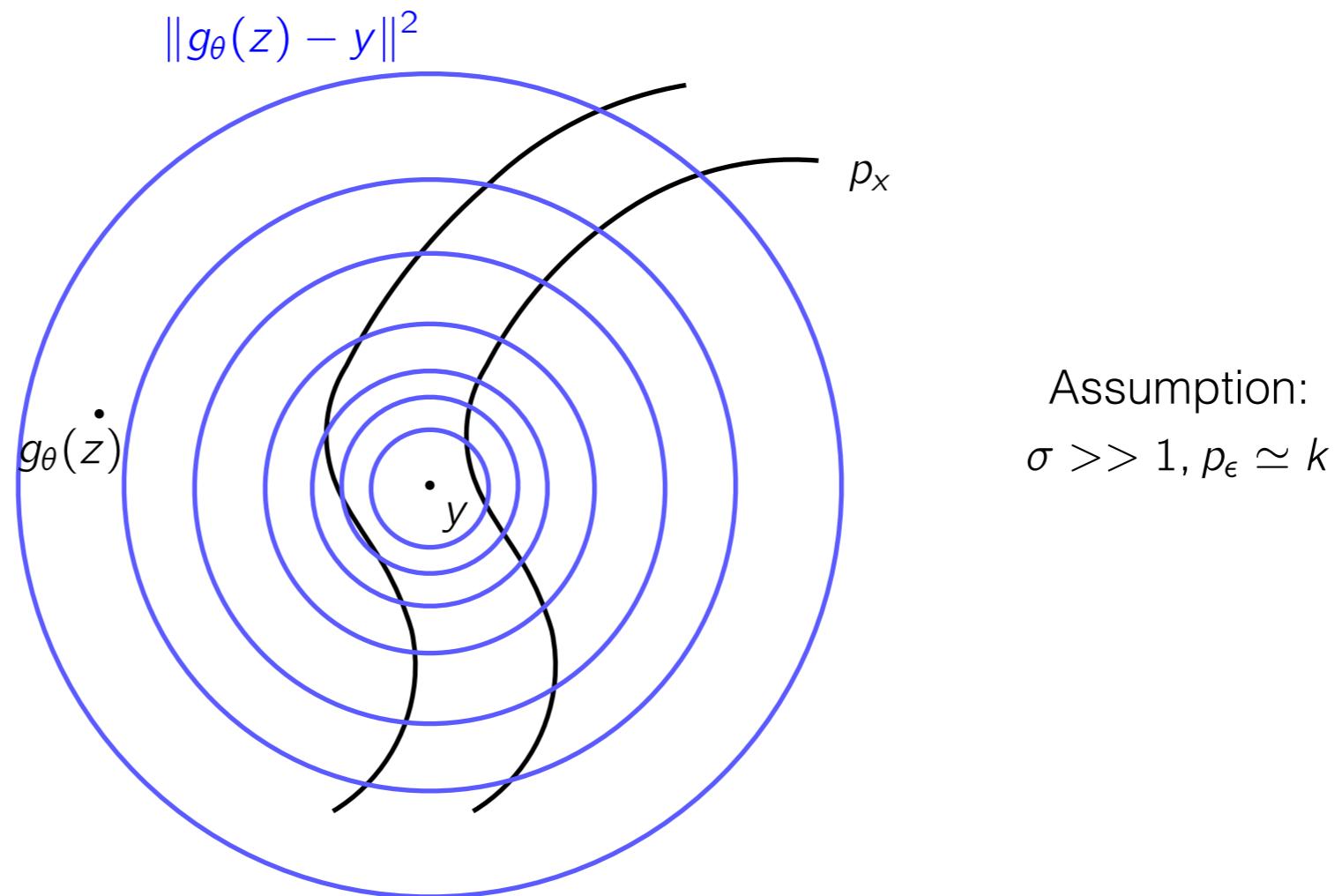
# Interpretation

$$E_{z \sim p_z} \left\{ \nabla_\theta \log(1 - D_\phi^*(g_\theta(z))) \right\} = E_{z \sim p_z} \left\{ a(z) \int p_\epsilon(g_\theta(z) - y) \nabla_\theta \|g_\theta(z) - y\|^2 p_x(y) dy \right. \\ \left. - b(z) \int p_\epsilon(g_\theta(z) - y) \nabla_\theta \|g_\theta(z) - y\|^2 p_g(y) dy \right\}$$



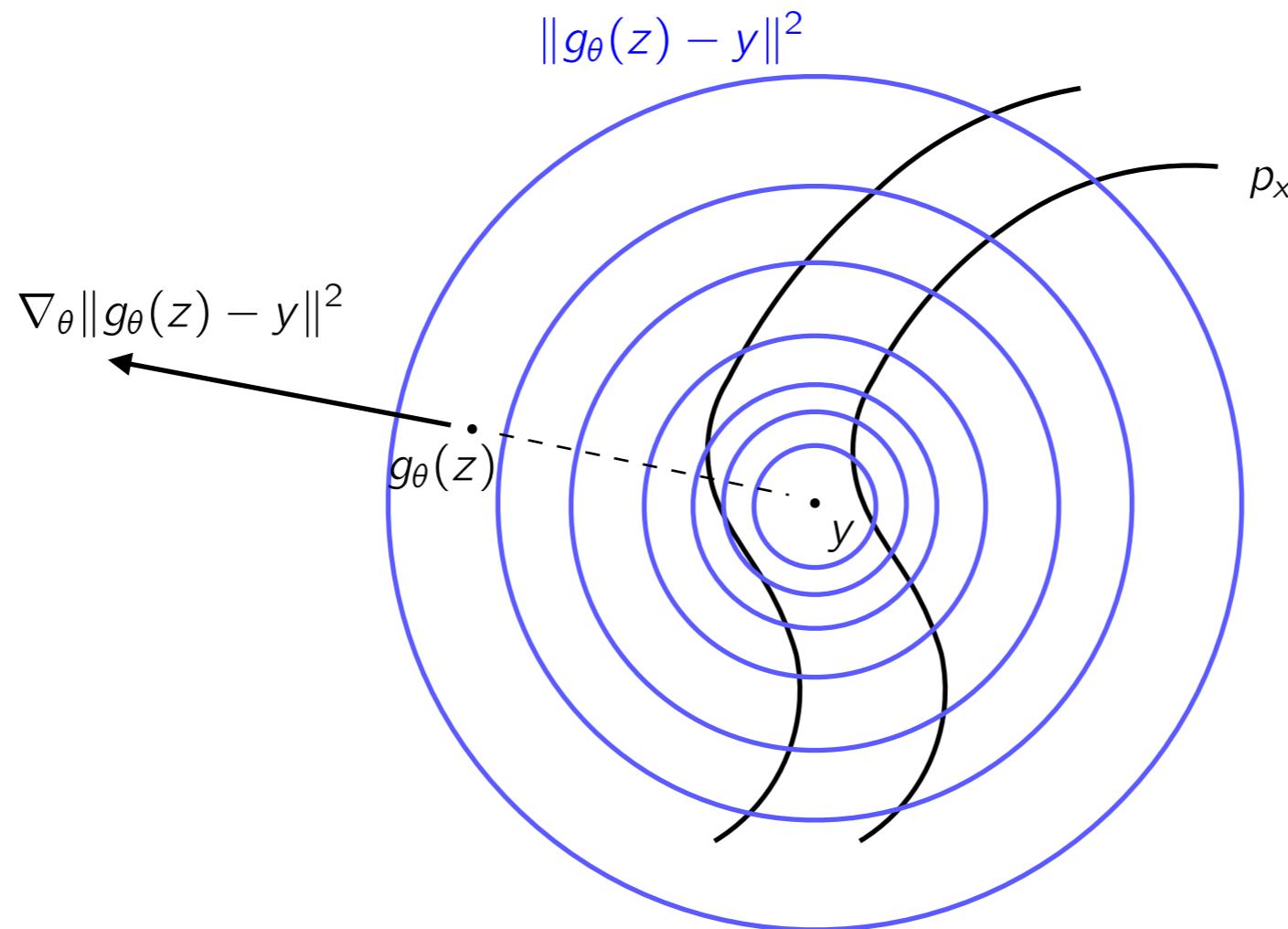
# Interpretation

$$\begin{aligned} E_{z \sim p_z} \left\{ \nabla_\theta \log(1 - D_\phi^*(g_\theta(z))) \right\} &= E_{z \sim p_z} \left\{ a(z) \int p_\epsilon(g_\theta(z) - y) \nabla_\theta \|g_\theta(z) - y\|^2 p_x(y) dy \right. \\ &\quad \left. - b(z) \int p_\epsilon(g_\theta(z) - y) \nabla_\theta \|g_\theta(z) - y\|^2 p_g(y) dy \right\} \end{aligned}$$



# Interpretation

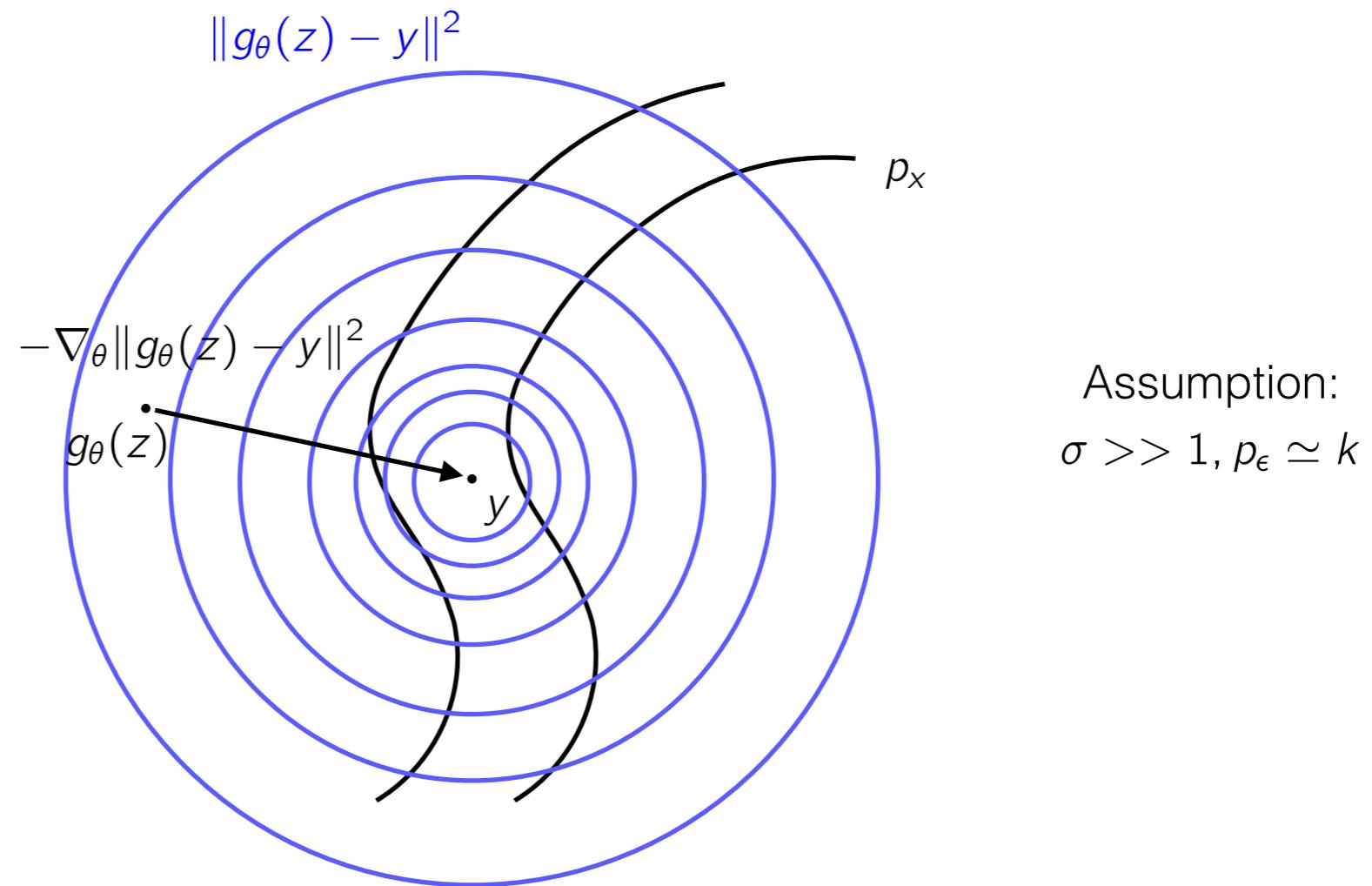
$$\begin{aligned} E_{z \sim p_z} \left\{ \nabla_{\theta} \log(1 - D_{\phi}^*(g_{\theta}(z))) \right\} &= E_{z \sim p_z} \left\{ a(z) \int p_{\epsilon}(g_{\theta}(z) - y) \nabla_{\theta} \|g_{\theta}(z) - y\|^2 p_x(y) dy \right. \\ &\quad \left. - b(z) \int p_{\epsilon}(g_{\theta}(z) - y) \nabla_{\theta} \|g_{\theta}(z) - y\|^2 p_g(y) dy \right\} \end{aligned}$$



Assumption:  
 $\sigma \gg 1, p_{\epsilon} \simeq k$

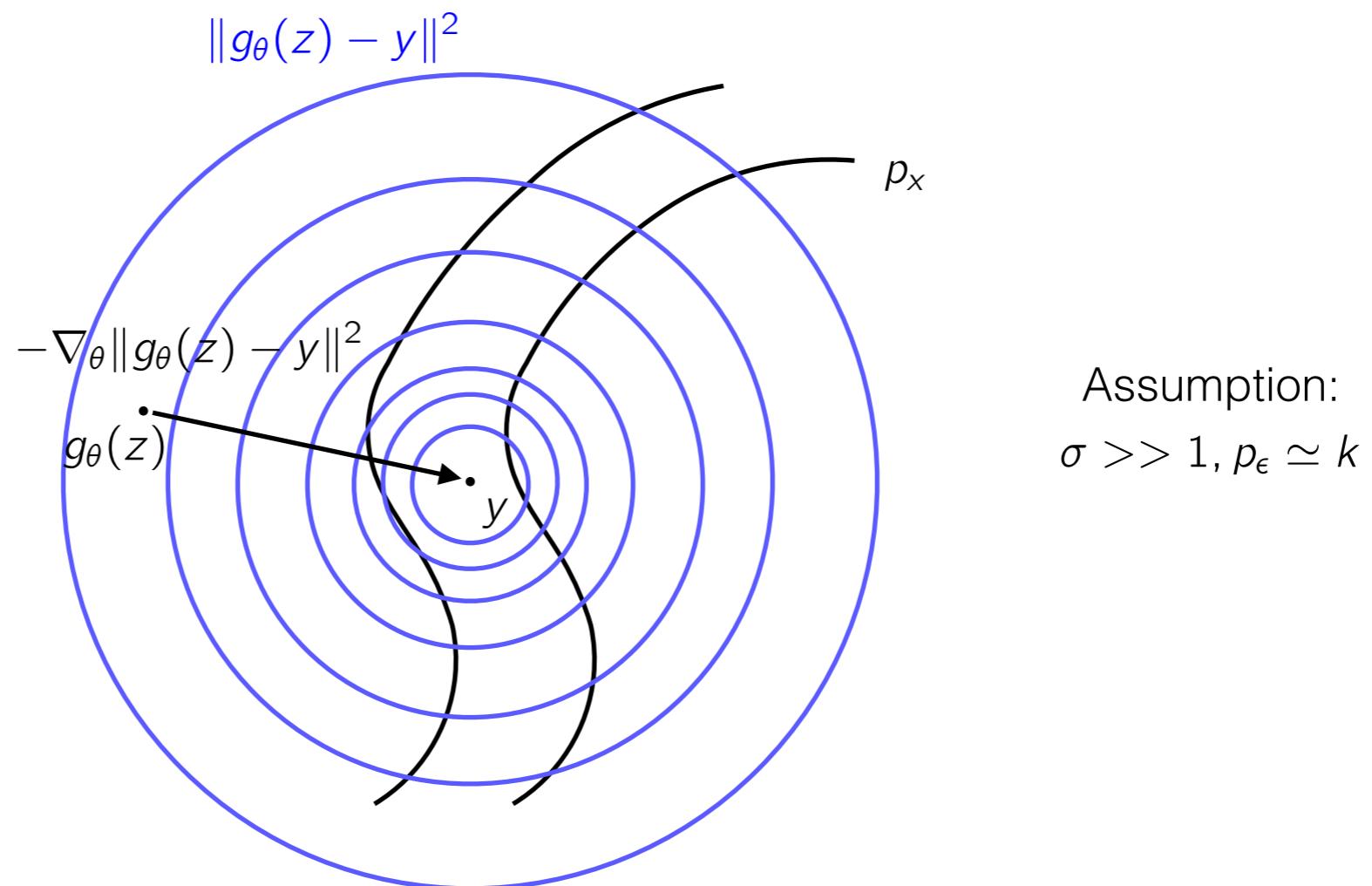
# Interpretation

$$E_{z \sim p_z} \left\{ \nabla_\theta \log(1 - D_\phi^*(g_\theta(z))) \right\} = E_{z \sim p_z} \left\{ a(z) \int p_\epsilon(g_\theta(z) - y) \nabla_\theta \|g_\theta(z) - y\|^2 p_x(y) dy \right. \\ \left. - b(z) \int p_\epsilon(g_\theta(z) - y) \nabla_\theta \|g_\theta(z) - y\|^2 p_g(y) dy \right\}$$



# Interpretation

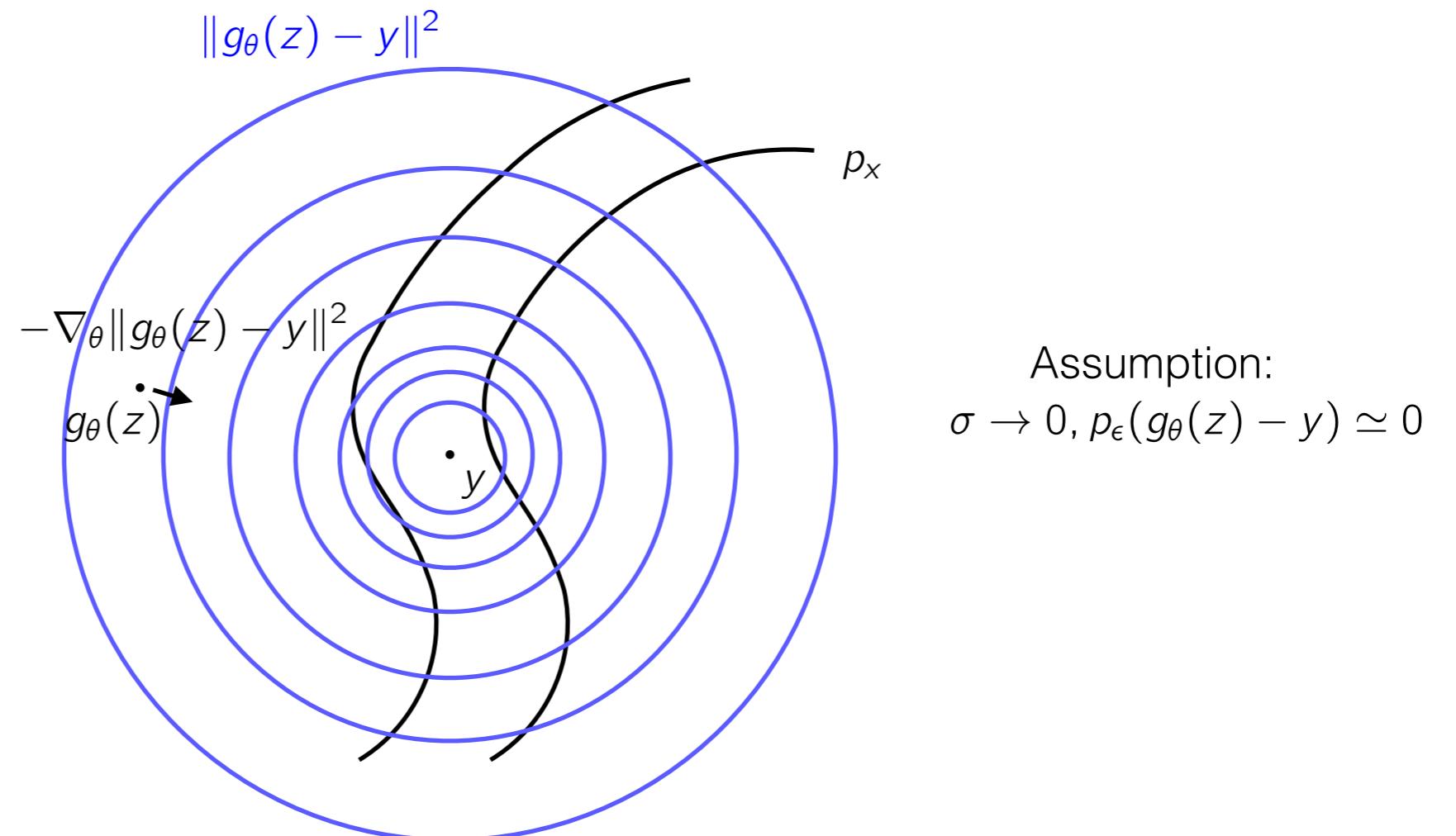
$$\begin{aligned} E_{z \sim p_z} \left\{ \nabla_{\theta} \log(1 - D_{\phi}^*(g_{\theta}(z))) \right\} &= E_{z \sim p_z} \left\{ a(z) \int p_{\epsilon}(g_{\theta}(z) - y) \nabla_{\theta} \|g_{\theta}(z) - y\|^2 p_x(y) dy \right. \\ &\quad \left. - b(z) \int p_{\epsilon}(g_{\theta}(z) - y) \nabla_{\theta} \|g_{\theta}(z) - y\|^2 p_g(y) dy \right\} \end{aligned}$$



The overall effect is to move generated points close to the data manifold (ATTRACTION TO HIGH DENSITY)

# Interpretation

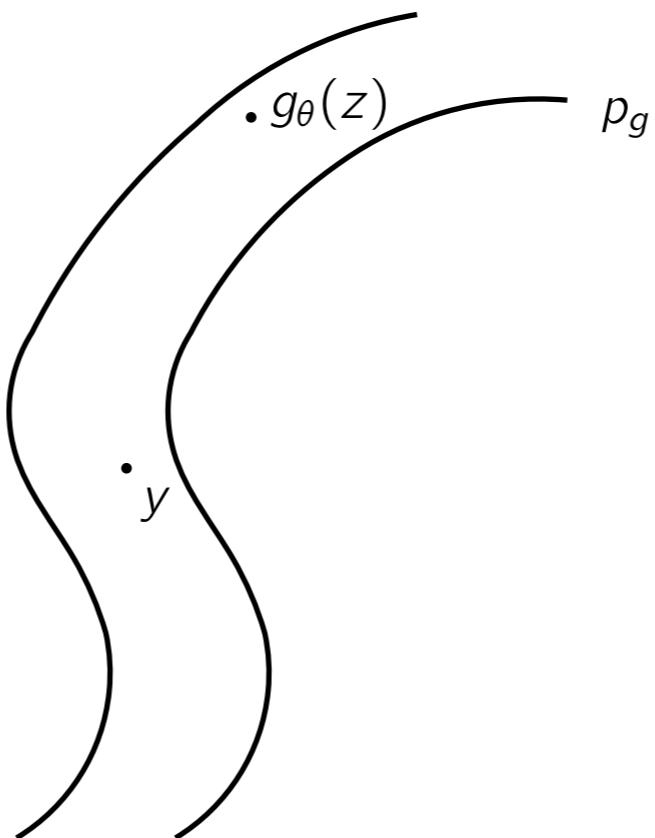
$$\begin{aligned} E_{z \sim p_z} \left\{ \nabla_{\theta} \log(1 - D_{\phi}^*(g_{\theta}(z))) \right\} &= E_{z \sim p_z} \left\{ a(z) \int p_{\epsilon}(g_{\theta}(z) - y) \nabla_{\theta} \|g_{\theta}(z) - y\|^2 p_x(y) dy \right. \\ &\quad \left. - b(z) \int p_{\epsilon}(g_{\theta}(z) - y) \nabla_{\theta} \|g_{\theta}(z) - y\|^2 p_g(y) dy \right\} \end{aligned}$$



VANISHING GRADIENTS

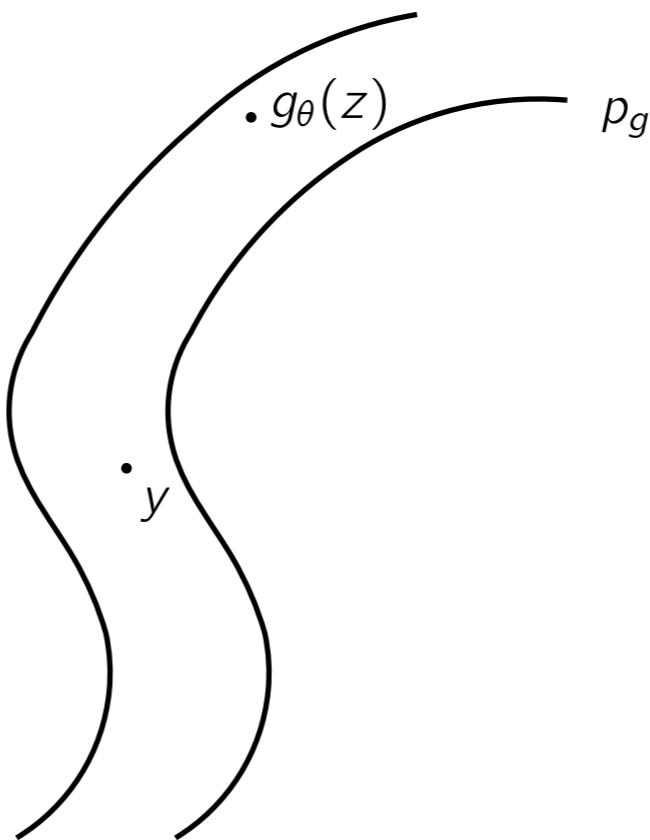
# Interpretation

$$E_{z \sim p_z} \left\{ \nabla_\theta \log(1 - D_\phi^*(g_\theta(z))) \right\} = E_{z \sim p_z} \left\{ a(z) \int p_\epsilon(g_\theta(z) - y) \nabla_\theta \|g_\theta(z) - y\|^2 p_x(y) dy \right. \\ \left. - b(z) \int p_\epsilon(g_\theta(z) - y) \nabla_\theta \|g_\theta(z) - y\|^2 p_g(y) dy \right\}$$



# Interpretation

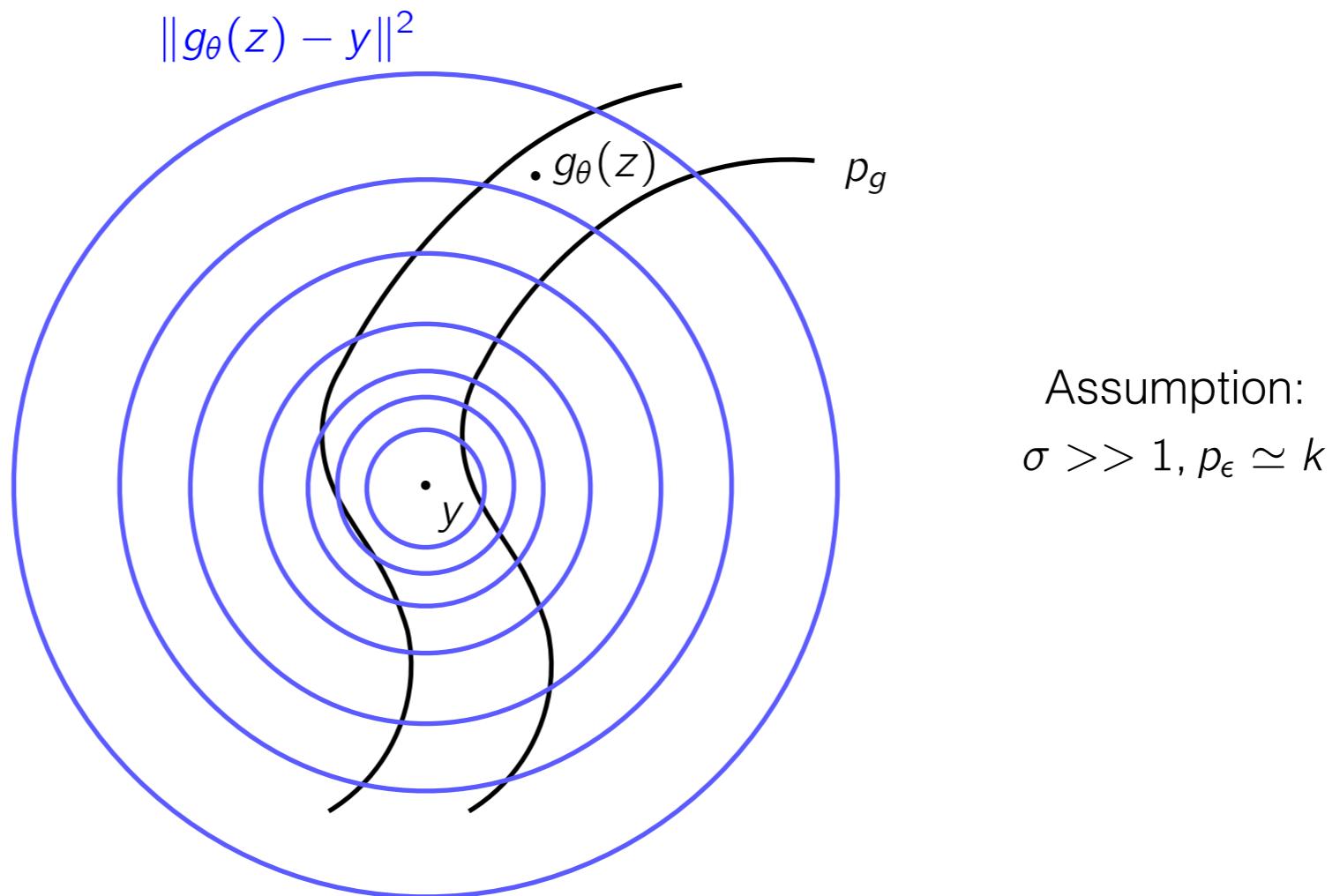
$$E_{z \sim p_z} \left\{ \nabla_\theta \log(1 - D_\phi^*(g_\theta(z))) \right\} = E_{z \sim p_z} \left\{ a(z) \int p_\epsilon(g_\theta(z) - y) \nabla_\theta \|g_\theta(z) - y\|^2 p_x(y) dy \right. \\ \left. - b(z) \int p_\epsilon(g_\theta(z) - y) \nabla_\theta \|g_\theta(z) - y\|^2 p_g(y) dy \right\}$$



Assumption:  
 $\sigma \gg 1, p_\epsilon \simeq k$

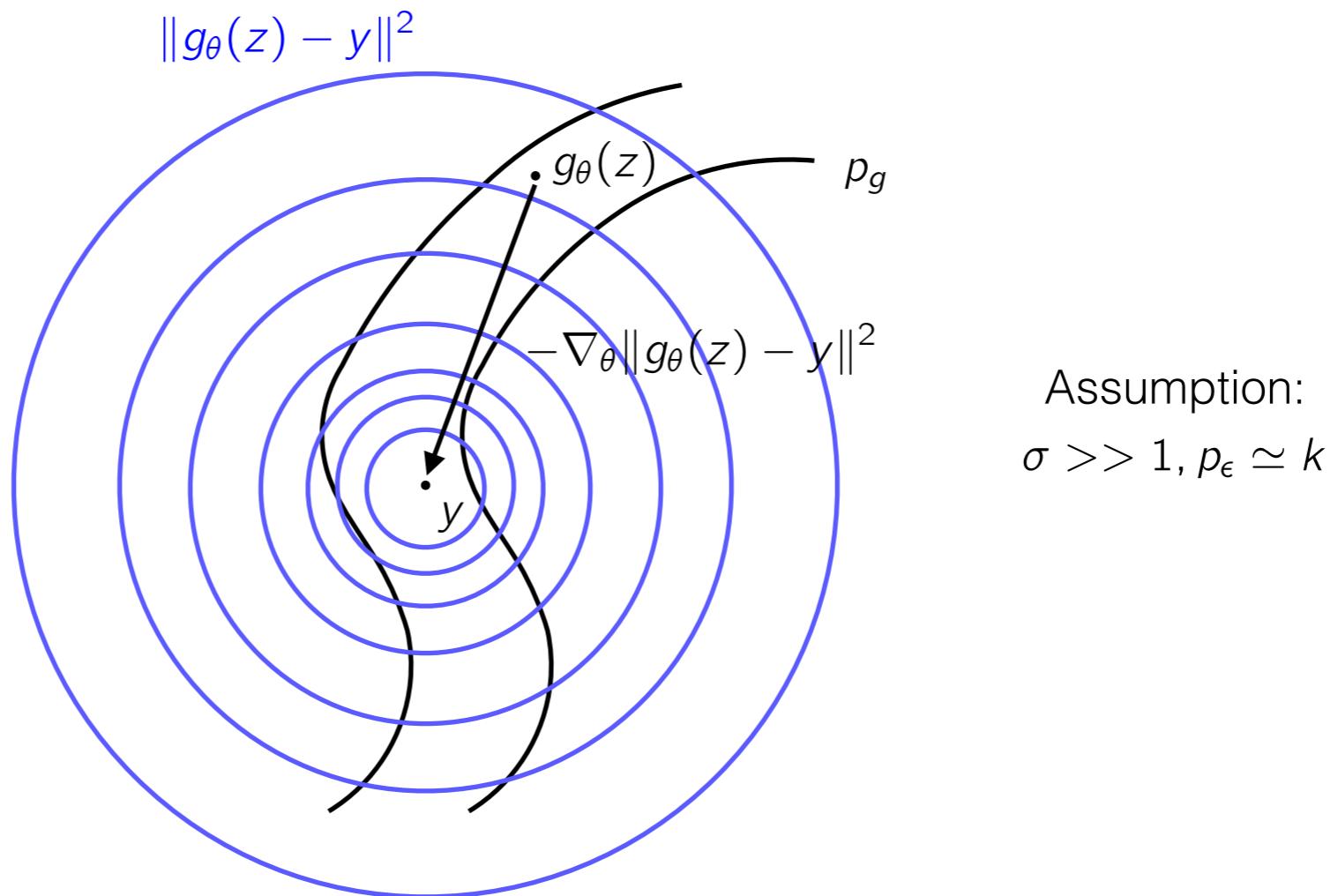
# Interpretation

$$\begin{aligned} E_{z \sim p_z} \left\{ \nabla_\theta \log(1 - D_\phi^*(g_\theta(z))) \right\} &= E_{z \sim p_z} \left\{ a(z) \int p_\epsilon(g_\theta(z) - y) \nabla_\theta \|g_\theta(z) - y\|^2 p_x(y) dy \right. \\ &\quad \left. - b(z) \int p_\epsilon(g_\theta(z) - y) \nabla_\theta \|g_\theta(z) - y\|^2 p_g(y) dy \right\} \end{aligned}$$



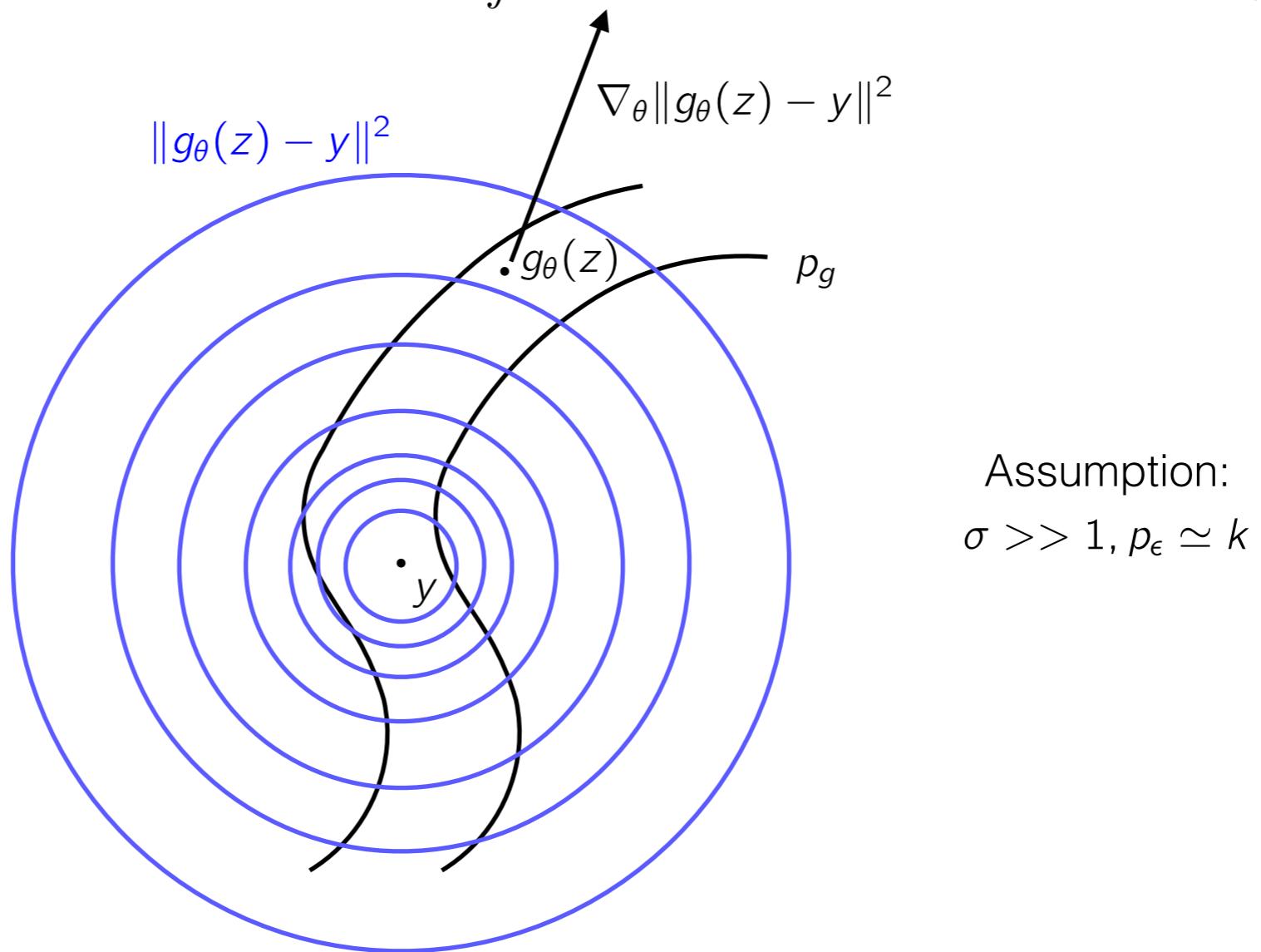
# Interpretation

$$\begin{aligned} E_{z \sim p_z} \left\{ \nabla_\theta \log(1 - D_\phi^*(g_\theta(z))) \right\} &= E_{z \sim p_z} \left\{ a(z) \int p_\epsilon(g_\theta(z) - y) \nabla_\theta \|g_\theta(z) - y\|^2 p_x(y) dy \right. \\ &\quad \left. - b(z) \int p_\epsilon(g_\theta(z) - y) \nabla_\theta \|g_\theta(z) - y\|^2 p_g(y) dy \right\} \end{aligned}$$



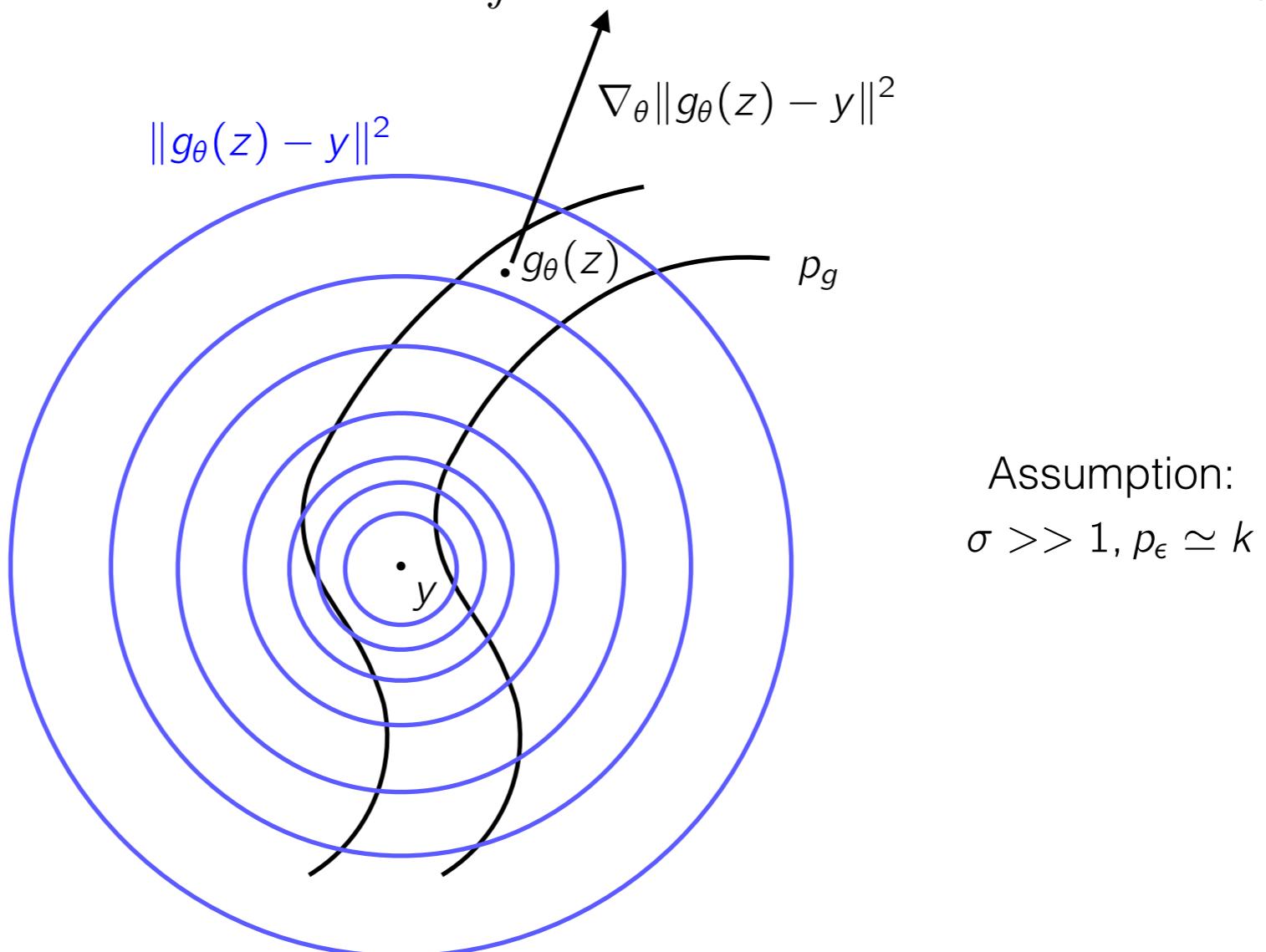
# Interpretation

$$E_{z \sim p_z} \left\{ \nabla_\theta \log(1 - D_\phi^*(g_\theta(z))) \right\} = E_{z \sim p_z} \left\{ a(z) \int p_\epsilon(g_\theta(z) - y) \nabla_\theta \|g_\theta(z) - y\|^2 p_x(y) dy \right. \\ \left. - b(z) \int p_\epsilon(g_\theta(z) - y) \nabla_\theta \|g_\theta(z) - y\|^2 p_g(y) dy \right\}$$



# Interpretation

$$E_{z \sim p_z} \left\{ \nabla_\theta \log(1 - D_\phi^*(g_\theta(z))) \right\} = E_{z \sim p_z} \left\{ a(z) \int p_\epsilon(g_\theta(z) - y) \nabla_\theta \|g_\theta(z) - y\|^2 p_x(y) dy \right. \\ \left. - b(z) \int p_\epsilon(g_\theta(z) - y) \nabla_\theta \|g_\theta(z) - y\|^2 p_g(y) dy \right\}$$

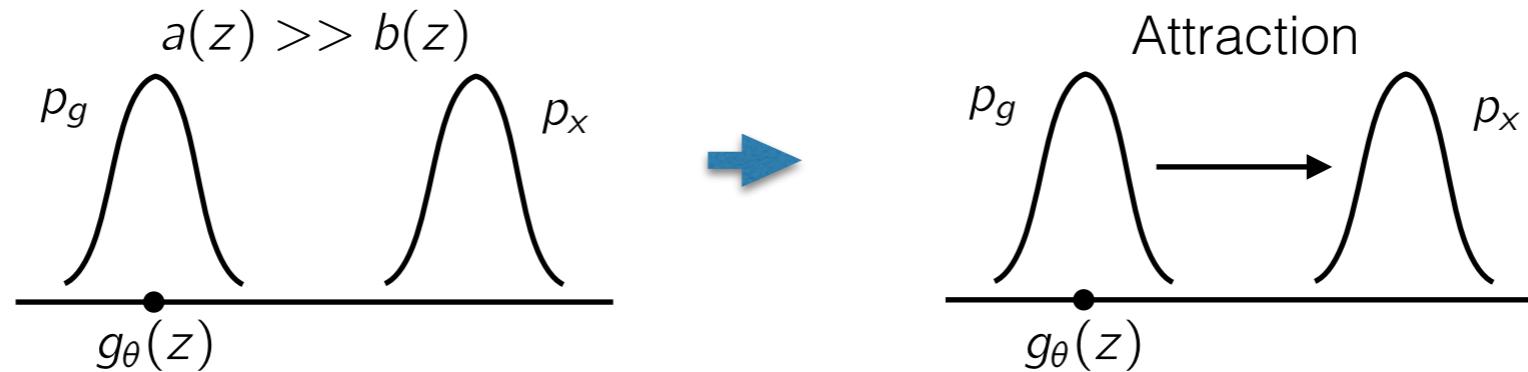


The overall effect is to stretch the generated manifold  
(STRETCHING HIGH DENSITY REGIONS)

# Interpretation

$$E_{z \sim p_z} \left\{ \nabla_{\theta} \log(1 - D_{\phi}^*(g_{\theta}(z))) \right\} = E_{z \sim p_z} \left\{ a(z) \int p_{\epsilon}(g_{\theta}(z) - y) \nabla_{\theta} \|g_{\theta}(z) - y\|^2 p_x(y) dy \right. \\ \left. - b(z) \int p_{\epsilon}(g_{\theta}(z) - y) \nabla_{\theta} \|g_{\theta}(z) - y\|^2 p_g(y) dy \right\}$$

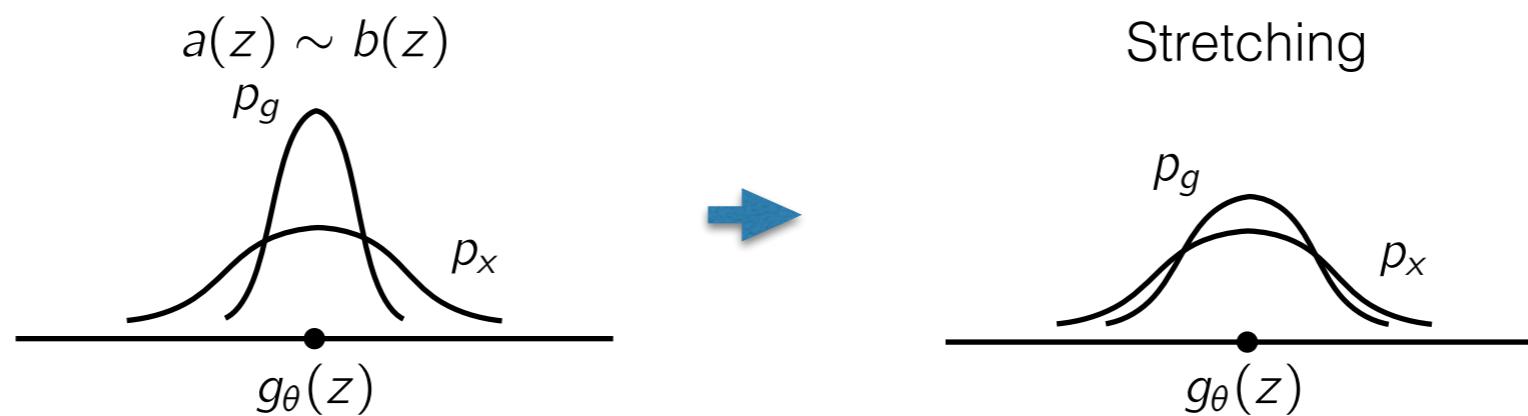
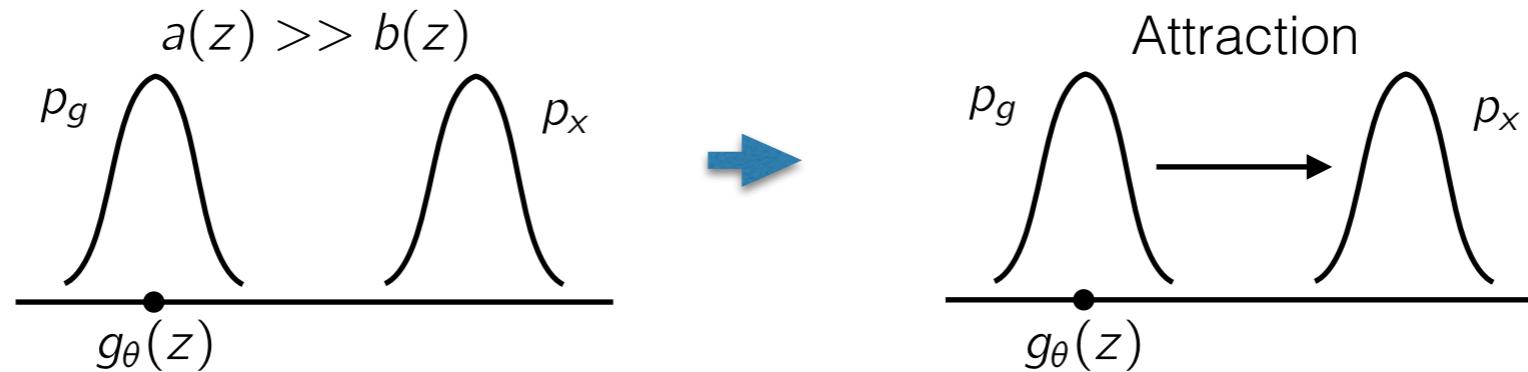
$$b(z) = a(z) \frac{p_{x+\epsilon}}{p_{g+\epsilon}}$$



# Interpretation

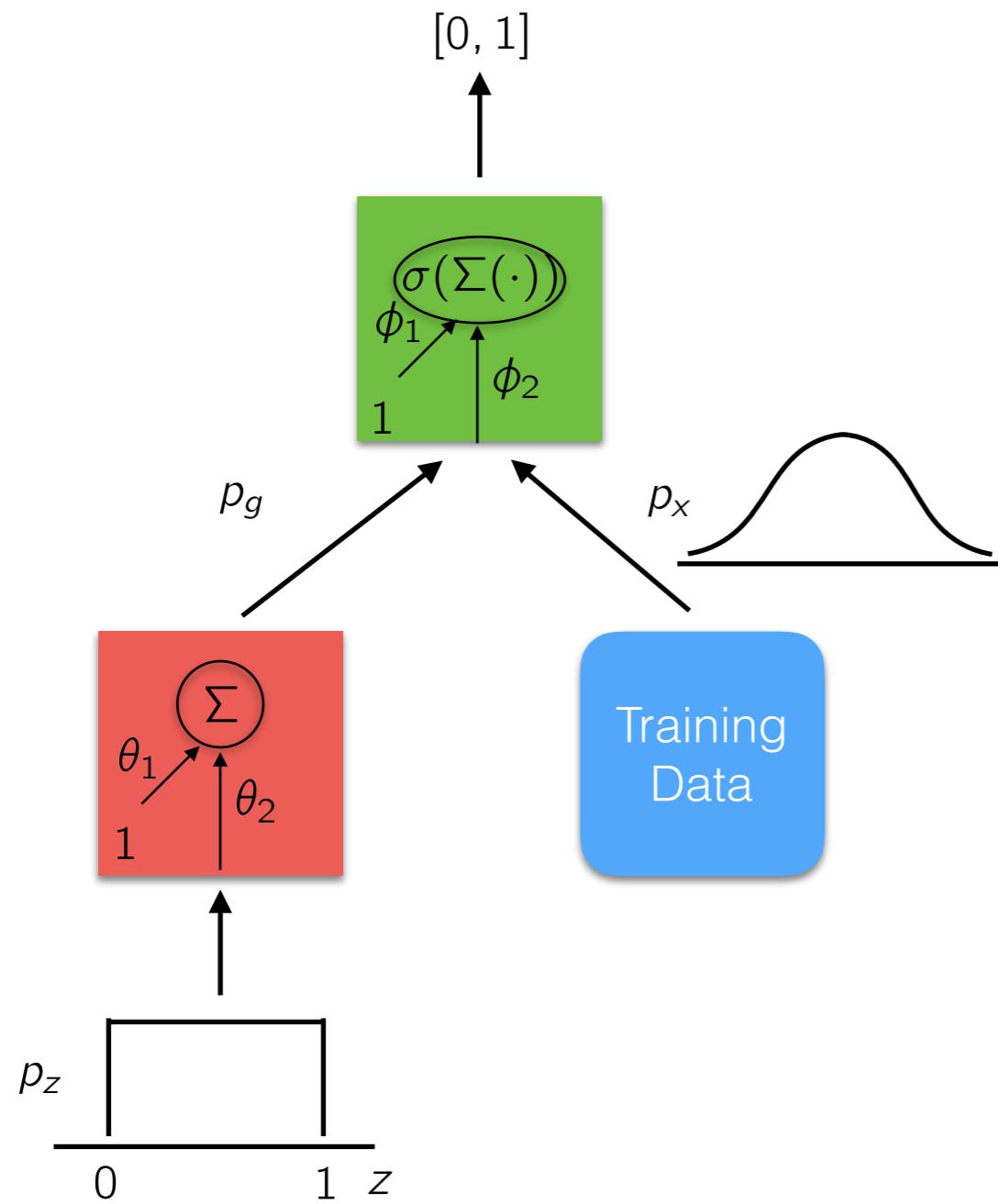
$$E_{z \sim p_z} \left\{ \nabla_{\theta} \log(1 - D_{\phi}^*(g_{\theta}(z))) \right\} = E_{z \sim p_z} \left\{ a(z) \int p_{\epsilon}(g_{\theta}(z) - y) \nabla_{\theta} \|g_{\theta}(z) - y\|^2 p_x(y) dy \right. \\ \left. - b(z) \int p_{\epsilon}(g_{\theta}(z) - y) \nabla_{\theta} \|g_{\theta}(z) - y\|^2 p_g(y) dy \right\}$$

$$b(z) = a(z) \frac{p_{x+\epsilon}}{p_{g+\epsilon}}$$



DEMO

# DEMO



**Normal:**

GAMES = 1000  
DISCRIMINATOR\_STEPS = 1  
GENERATOR\_STEPS = 1

**Strong discriminator:**

GAMES = 100  
DISCRIMINATOR\_STEPS = 500  
GENERATOR\_STEPS = 1

**Strong generator:**

GAMES = 10  
DISCRIMINATOR\_STEPS = 1  
GENERATOR\_STEPS = 500

# Open issues

# Open issues

1. Other solutions to the vanishing gradient problem?

# Open issues

1. Other solutions to the vanishing gradient problem?
2. Problem when the dimension of the support of the input distribution is lower than the dimension of the support of the data distribution



$$\dim(\text{supp}(p_z)) \geq \dim(\text{supp}(p_g))$$

# Open issues

POINT 1 -> example about learn how to study (motivating faster convergence)

1. Other solutions to the vanishing gradient problem?
2. Problem when the dimension of the support of the input distribution is lower than the dimension of the support of the data distribution

