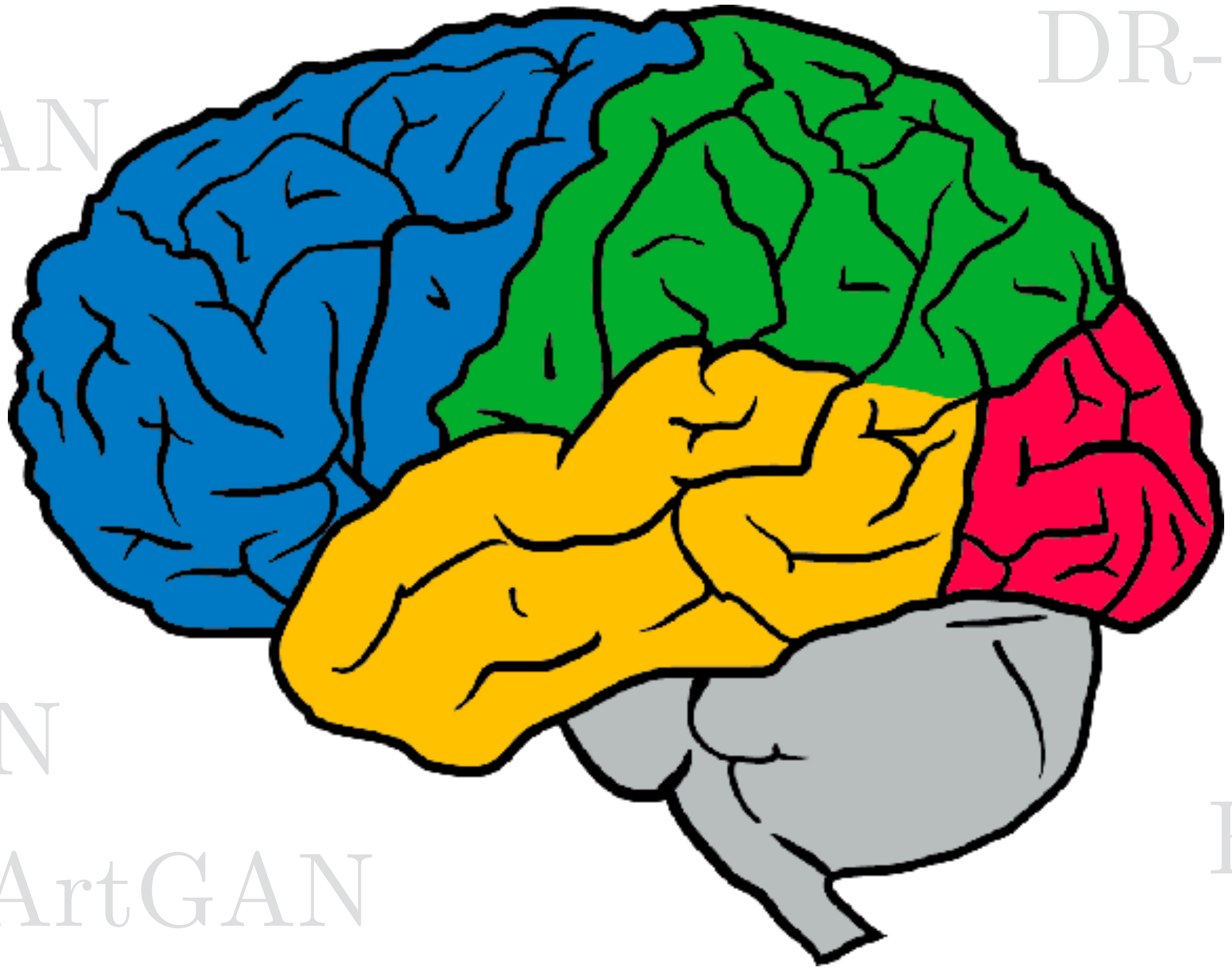
Bridging Theory and Practice of GANs

Ian Goodfellow, Staff Research Scientist, Google Brain

NIPS 2017 Workshop: Deep Learning: Bridging Theory and Practice

Long Beach, 2017-12-09



MedGAN ID-CGAN Progressive GAN LR-GAN CGAN IcGAN b-GAN LS-GAN AffGAN LAPGAN DiscoGAN MPM-GAN AdaGAN LSGAN InfoGAN CatGAN SN-GAN AMGAN iGAN IAN CoGAN McGAN MIX+GAN MGAN FF-GAN GoGAN B-GAN C-VAE-GAN C-RNN-GAN DR-GAN DCGAN CCGAN AC-GAN DRAGAN MAGAN 3D-GAN GMAN BiGAN GAWWN DualGAN CycleGAN alpha-GAN Bayesian GAN GP-GAN AnoGAN WGAN-GP EBGAN DTN Context-RNN-GAN MAD-GAN ALI BEGAN AL-CGAN MARTA-GAN f-GAN ArtGAN MalGAN

Generative Modeling

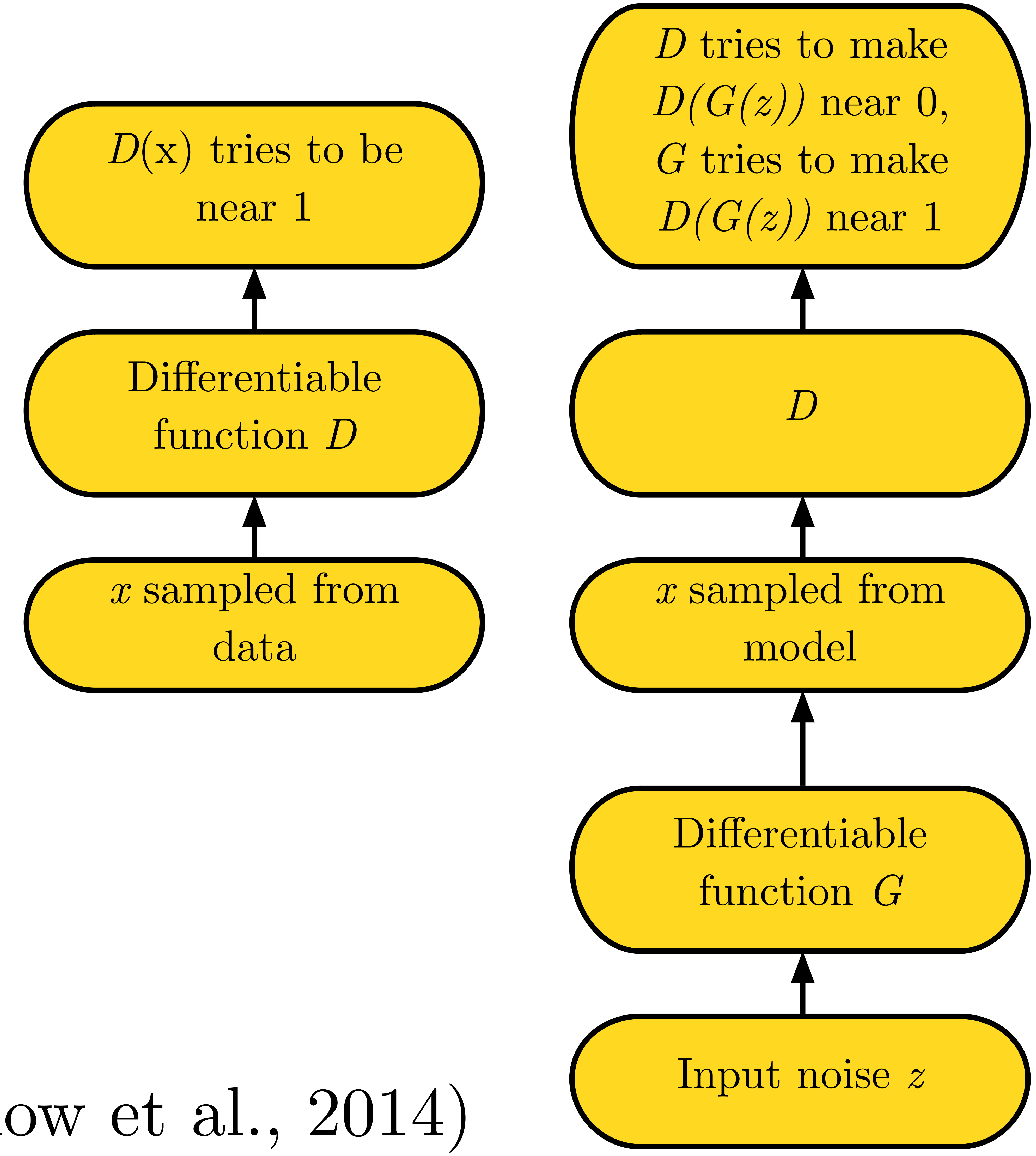
- Density estimation

- Sample generation

Training examples Model samples

(Goodfellow 2017)

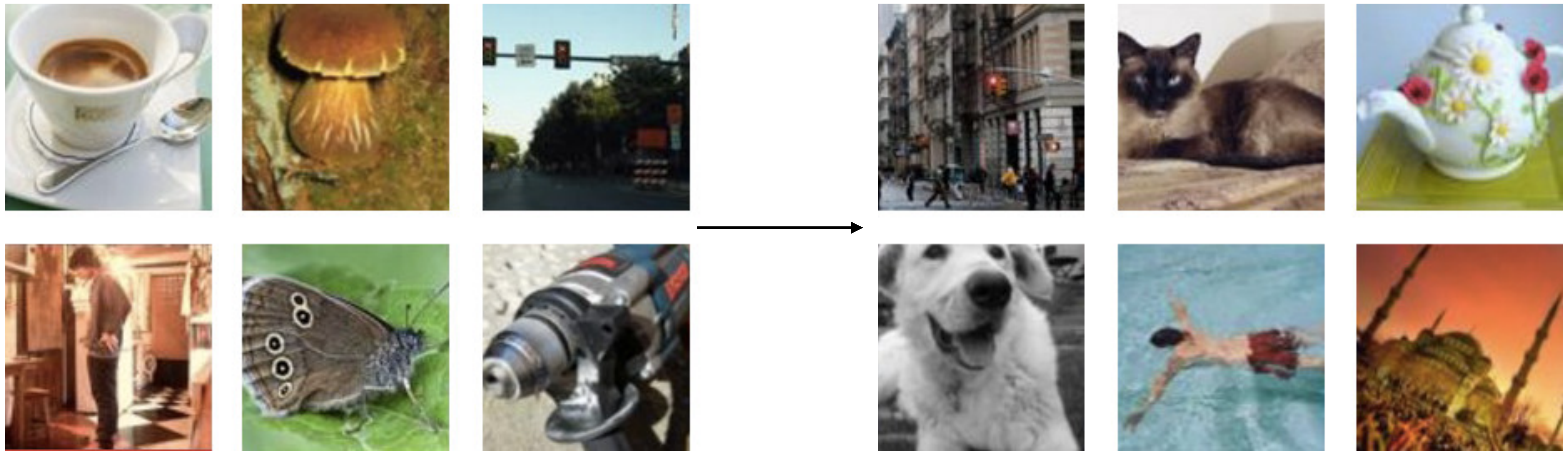
Adversarial Nets Framework



(Goodfellow et al., 2014)

(Goodfellow 2017)

How long until GANs can do this?



Training examples Model samples

Progressive GANs

(Karras et al., 2017)

Spectrally Normalized GANs

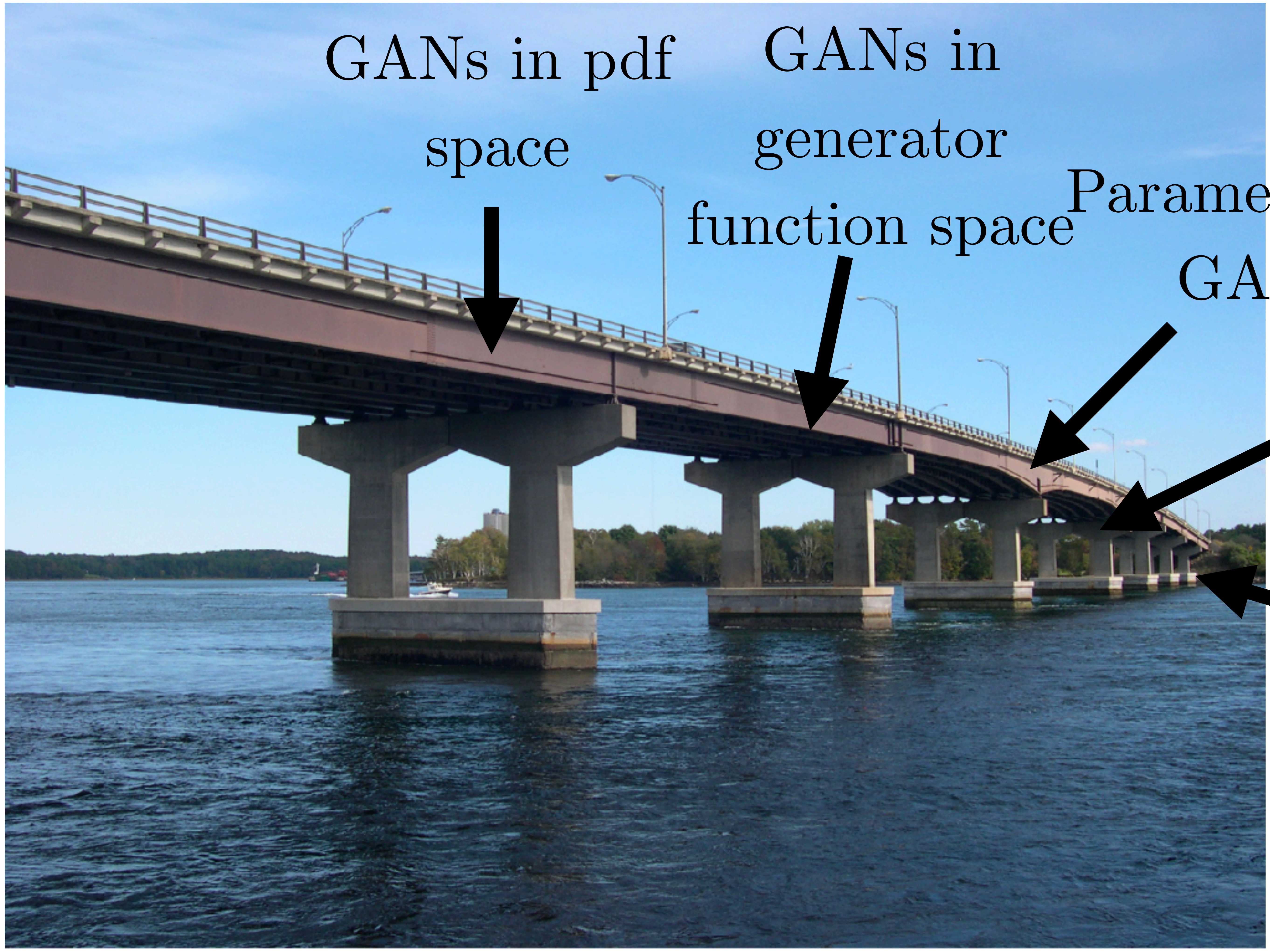
Welsh Springer Spaniel

Palace

Pizza

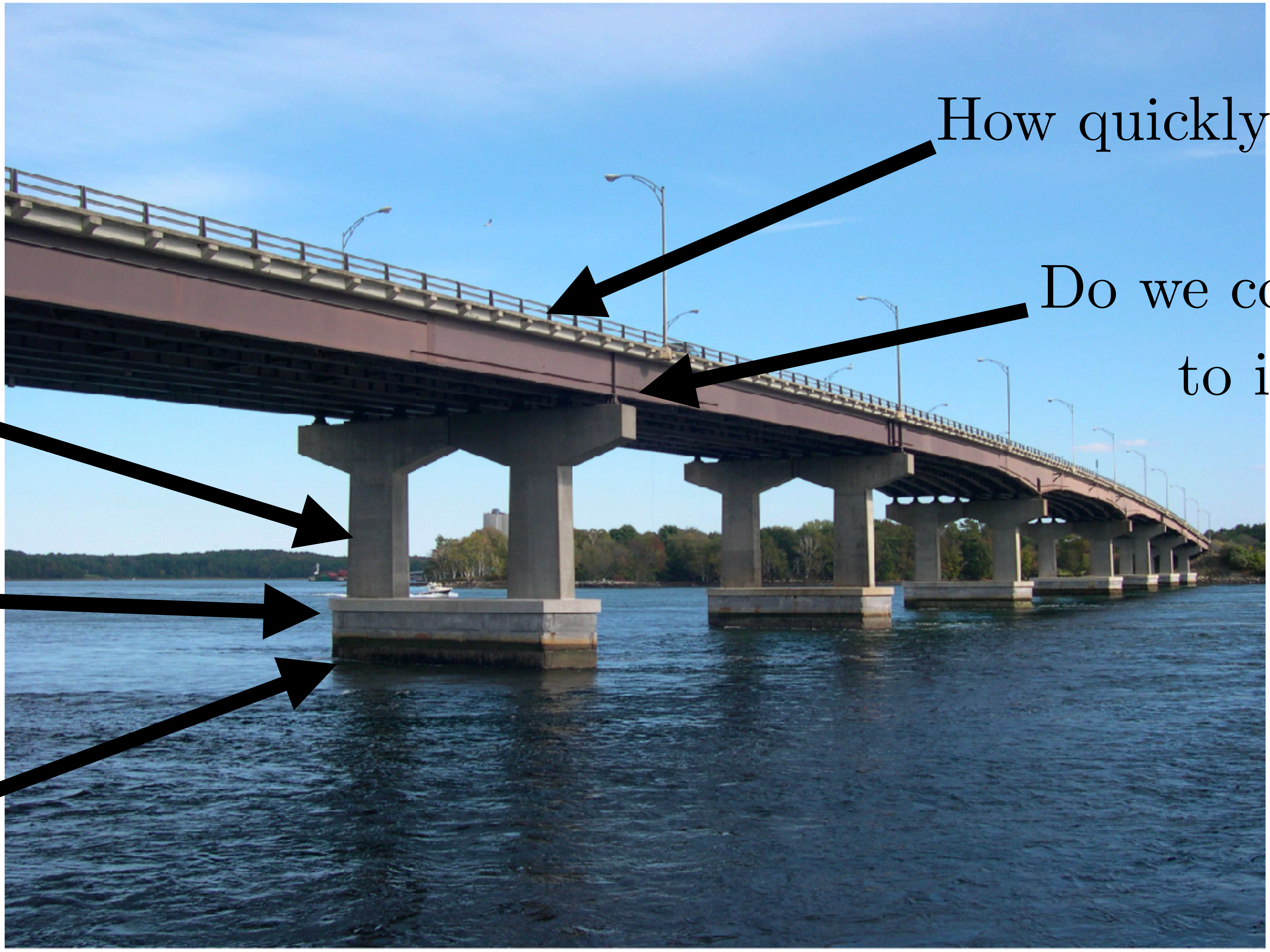
(Miyato et al., 2017)

Building a bridge from simple to complex theoretical models



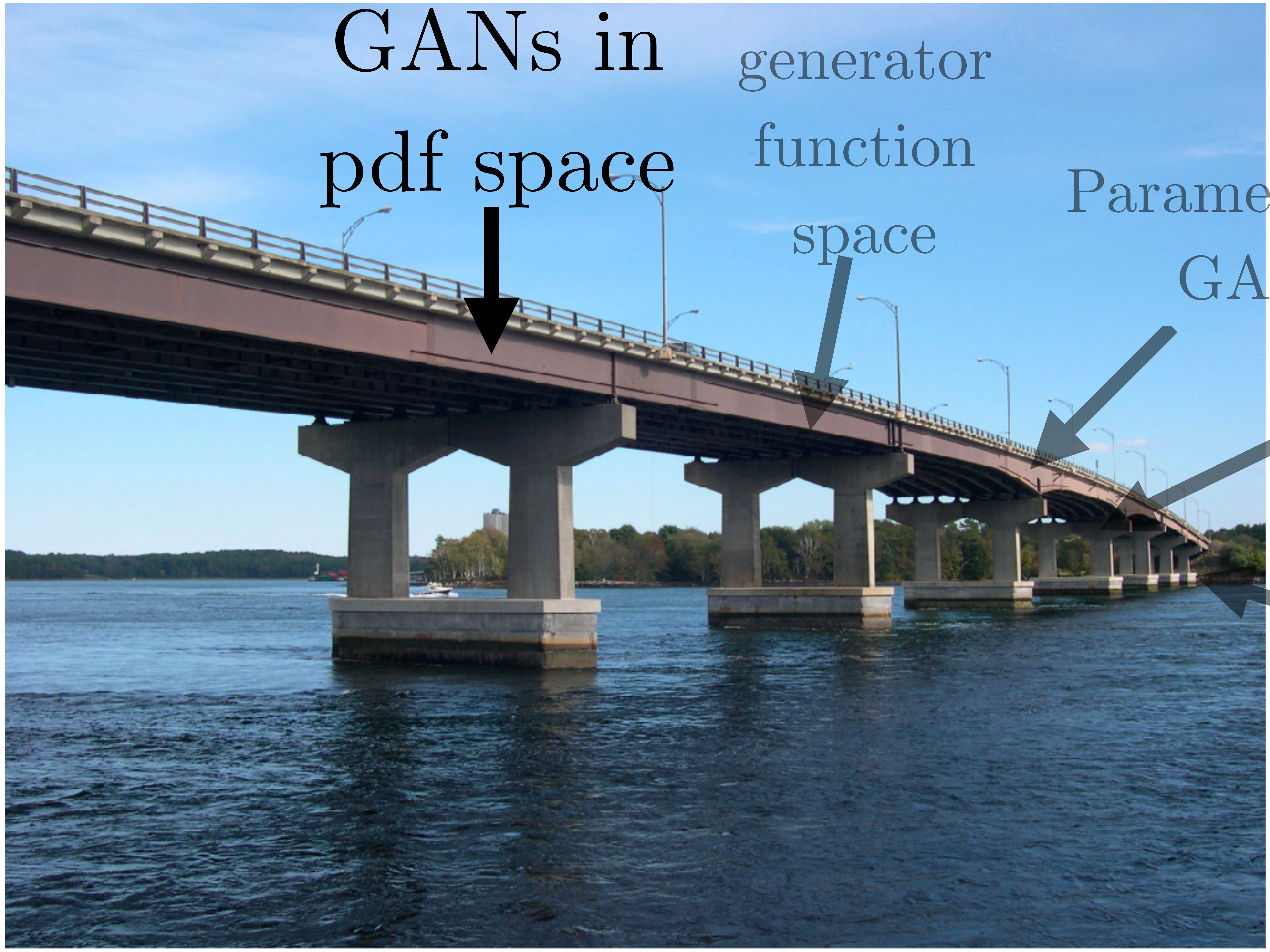
(Goodfellow 2017)

Building a bridge from intuition to theory



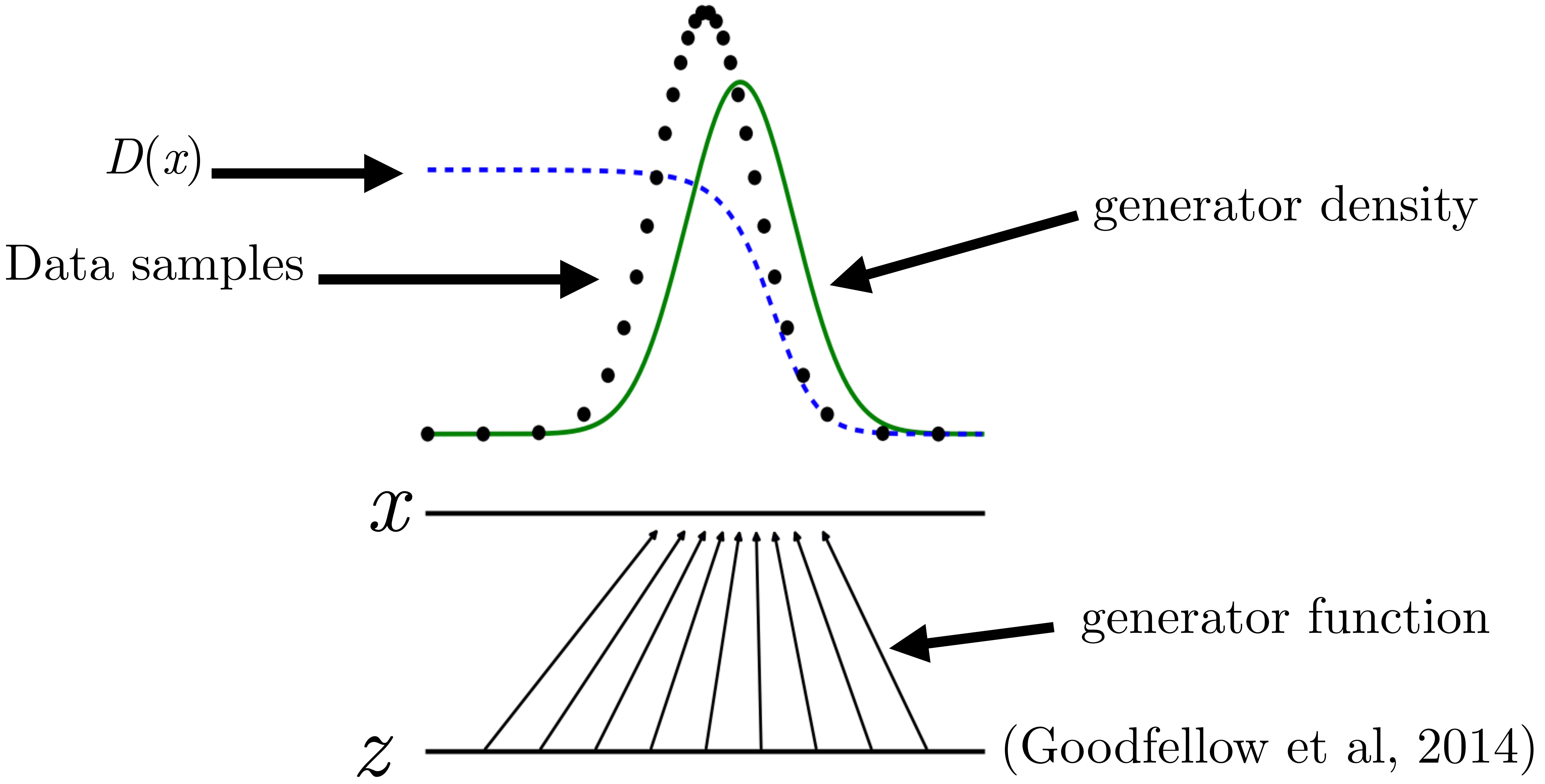
(Goodfellow 2017)

Building the bridge



(Goodfellow 2017)

Optimizing over densities



(Goodfellow 2017)

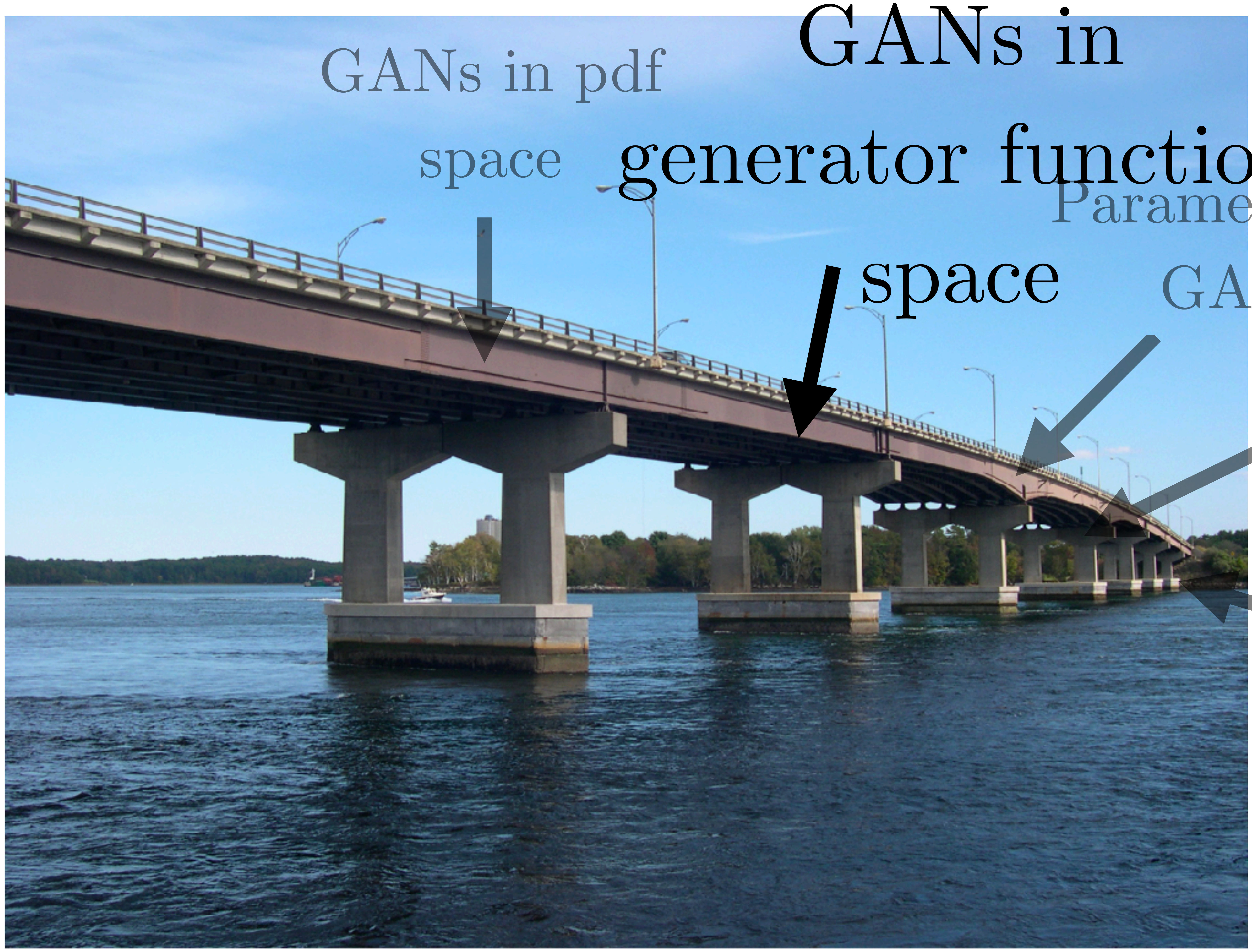
Tips and Tricks

- A good strategy to simplify a model for theoretical purposes is to work in *function space*.
  - Binary or linear models are often too different from neural net models to provide useful theory.
- Use *convex analysis* in this function space.

Results

- Goodfellow et al 2014:
  - Nash equilibrium exists
  - Nash equilibrium corresponds to recovering data-generating distribution
  - Nested optimization converges
- Kodali et al 2017: simultaneous SGD converges

Building a bridge from simple to complex theoretical models

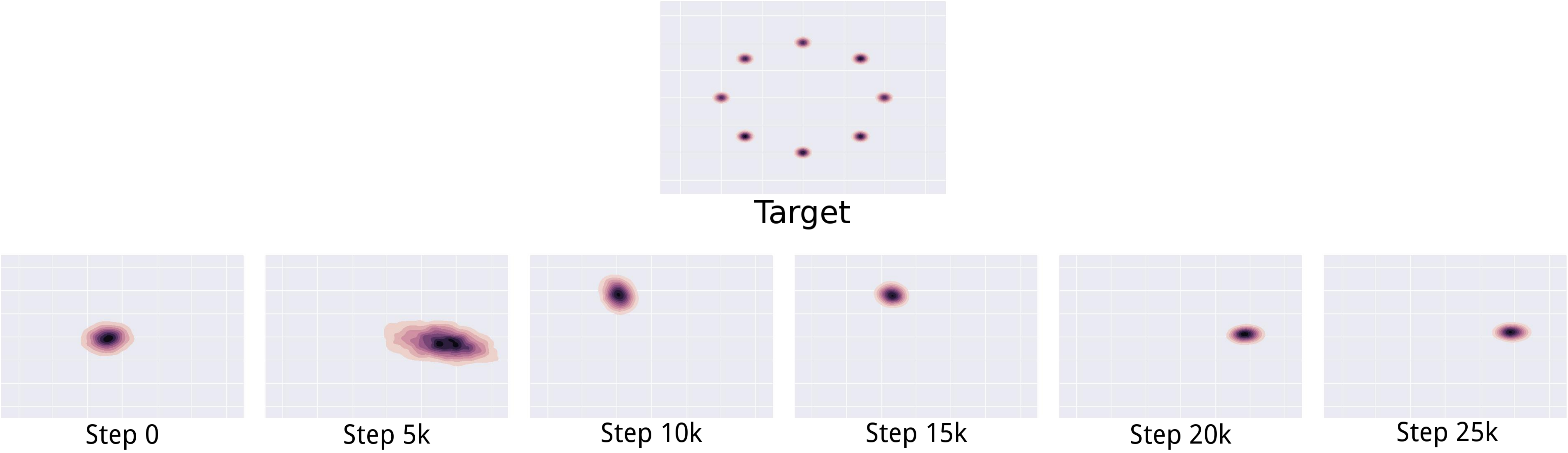


(Goodfellow 2017)

Non-Equilibrium Mode Collapse

$$\min_G \max_D V(G, D) \neq \max_D \min_G V(G, D)$$

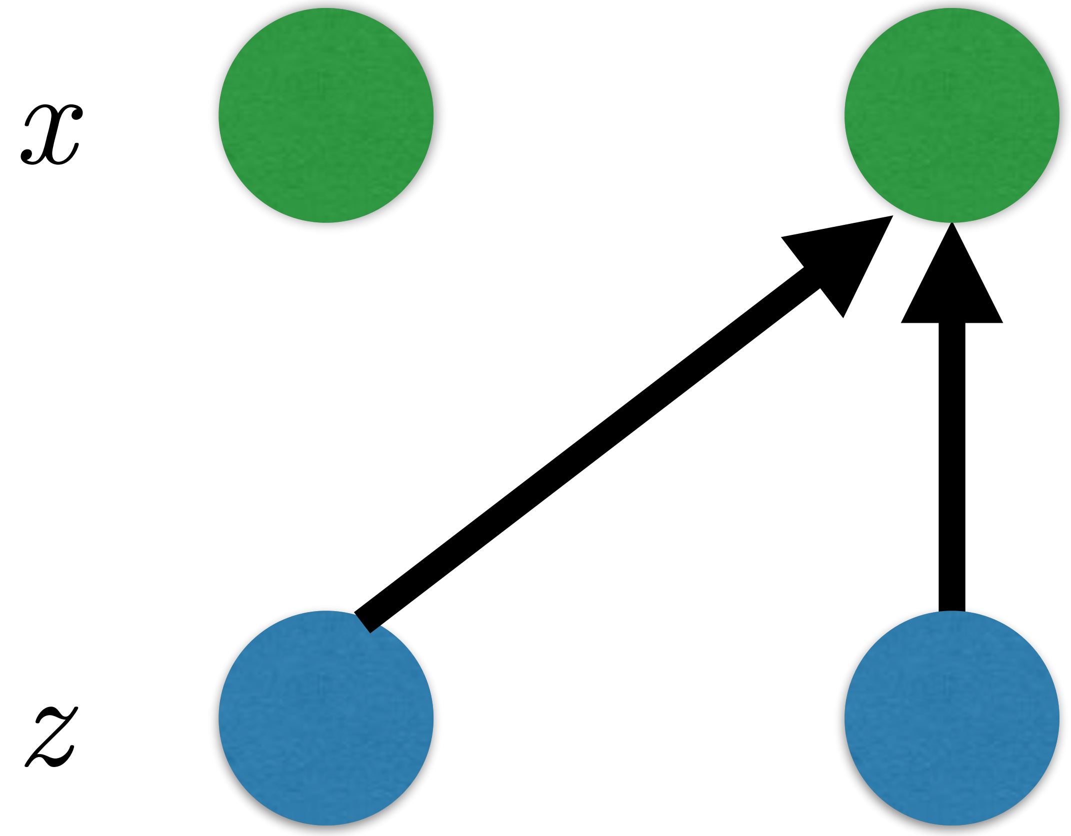
- D in inner loop: convergence to correct distribution
- G in inner loop: place all mass on most likely point



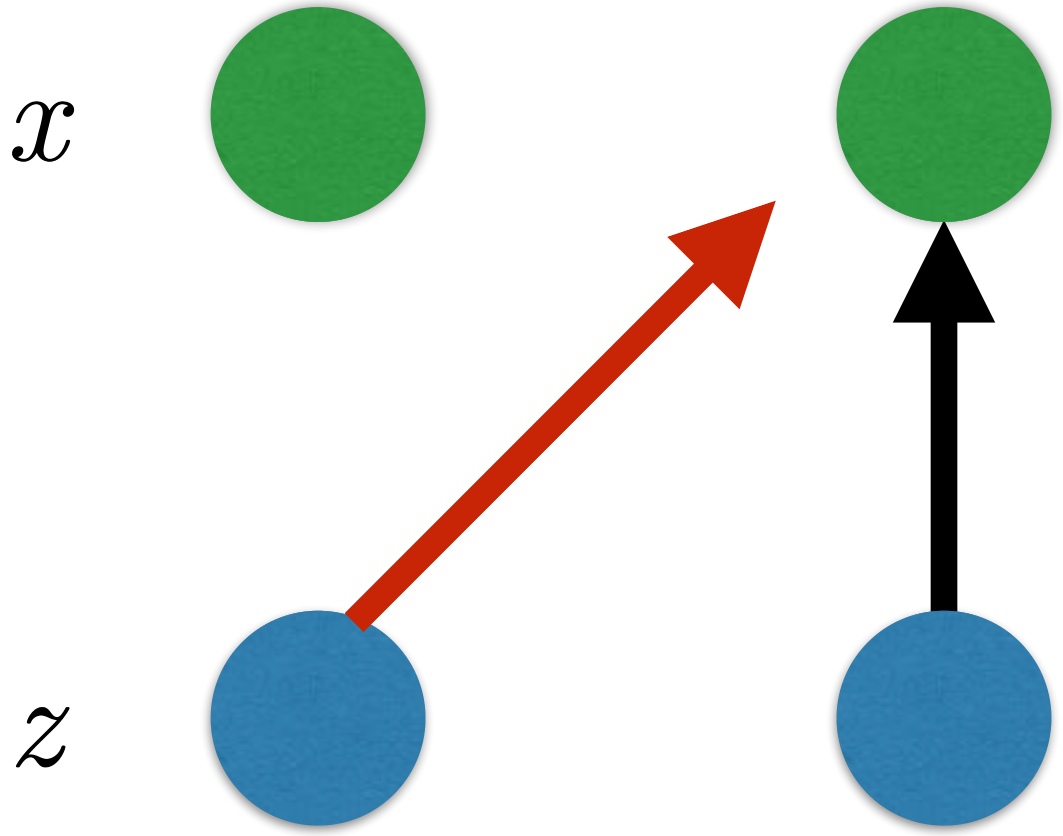
(Metz et al 2016)

Equilibrium mode collapse

Mode collapse

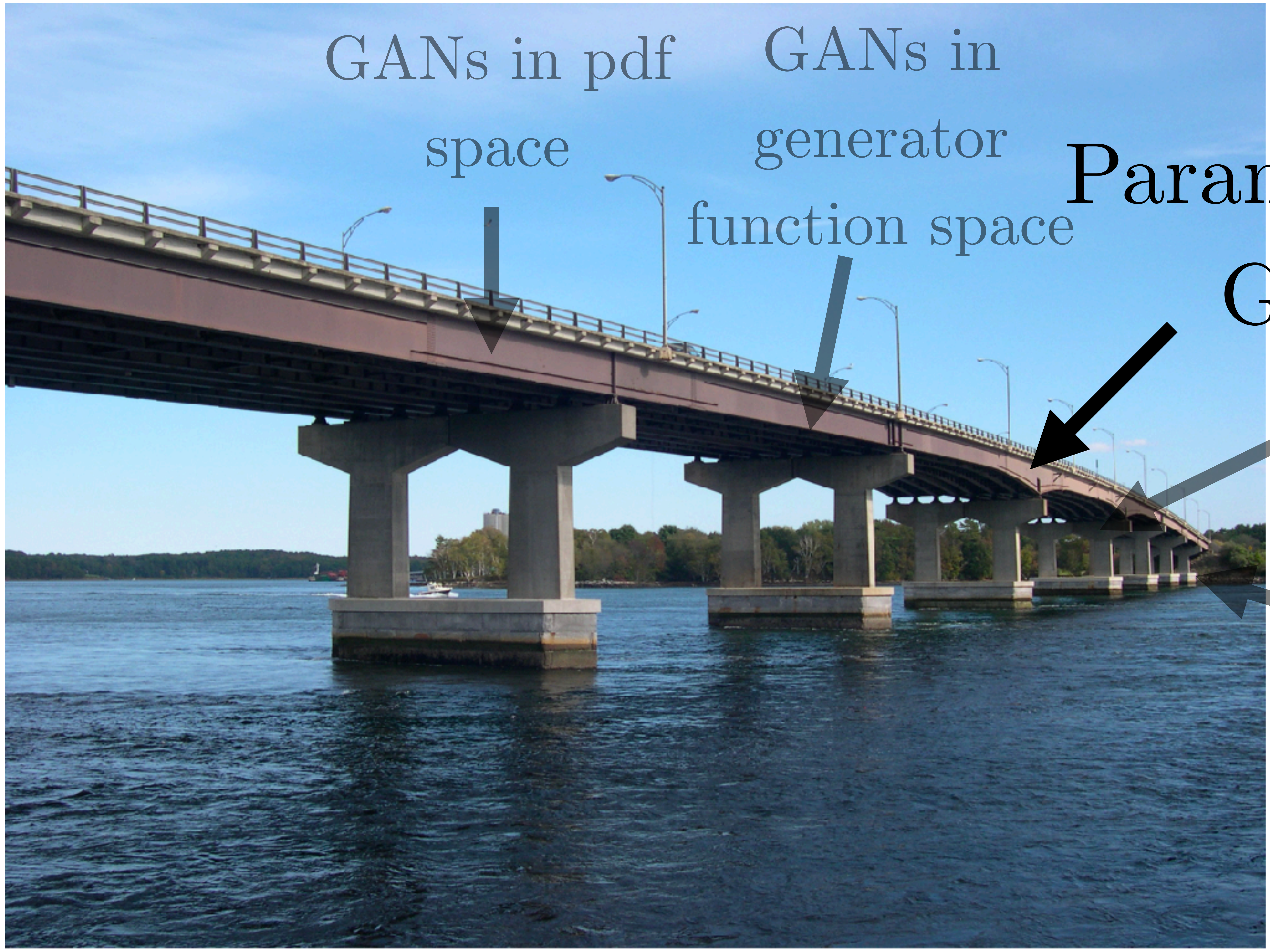


Neighbors in generator function space are worse



(Appendix A1 of Unterthiner et al, 2017)

Building a bridge from simple to complex theoretical models



(Goodfellow 2017)

Simple Non-convergence Example

- For scalar x and y, consider the value function:

$$V(x, y) = xy$$

- Does this game have an equilibrium? Where is it?

- Consider the learning dynamics of simultaneous gradient descent with infinitesimal learning rate (continuous time). Solve for the trajectory followed by these dynamics.

$$\frac{\partial x}{\partial t} = -\frac{\partial}{\partial x} V(x(t), y(t))$$

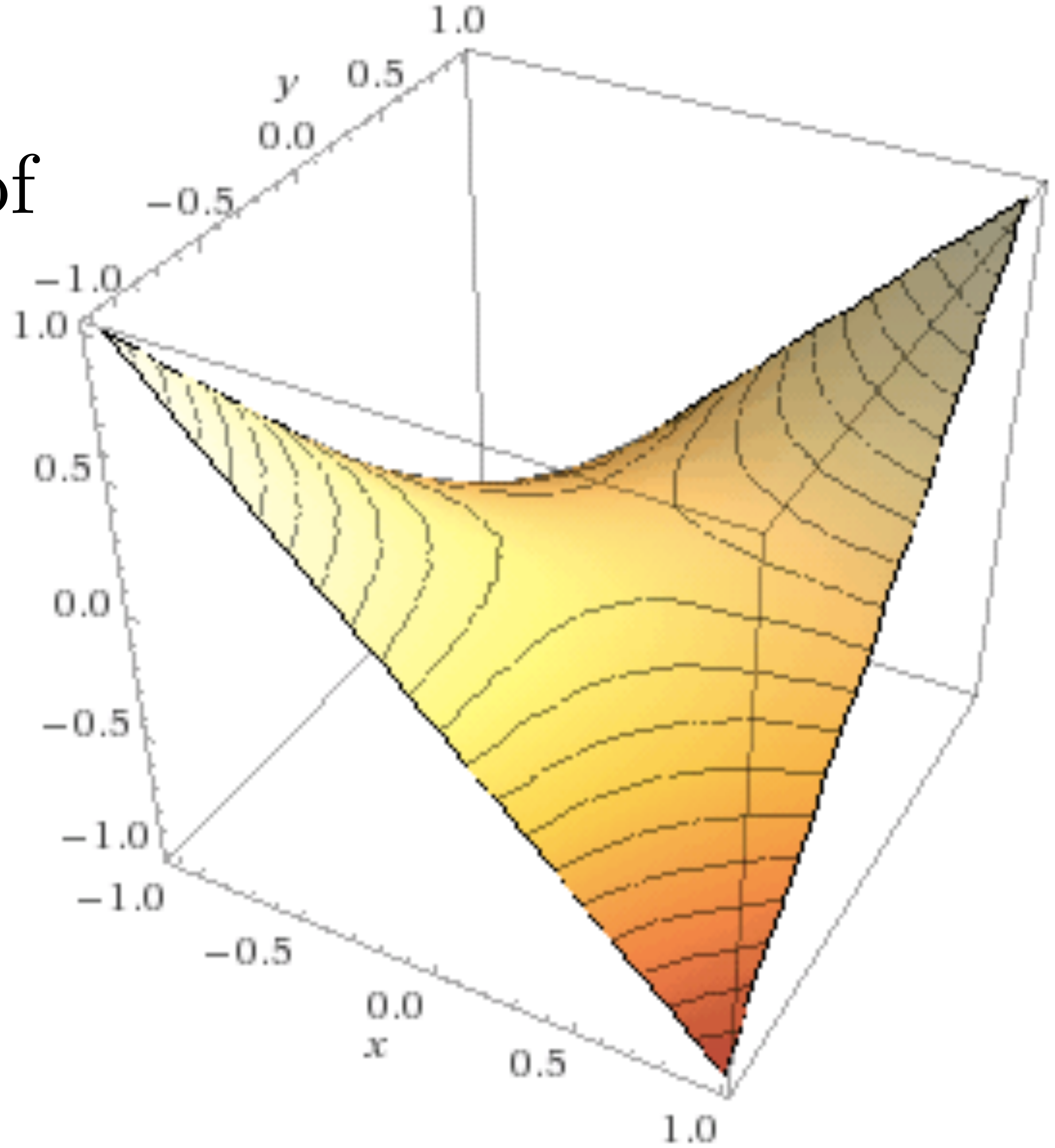
$$\frac{\partial y}{\partial t} = \frac{\partial}{\partial y} V(x(t), y(t))$$

(Goodfellow 2017)

Solution

This is the canonical example of a saddle point.

There is an equilibrium, at $x = 0$, $y = 0$.



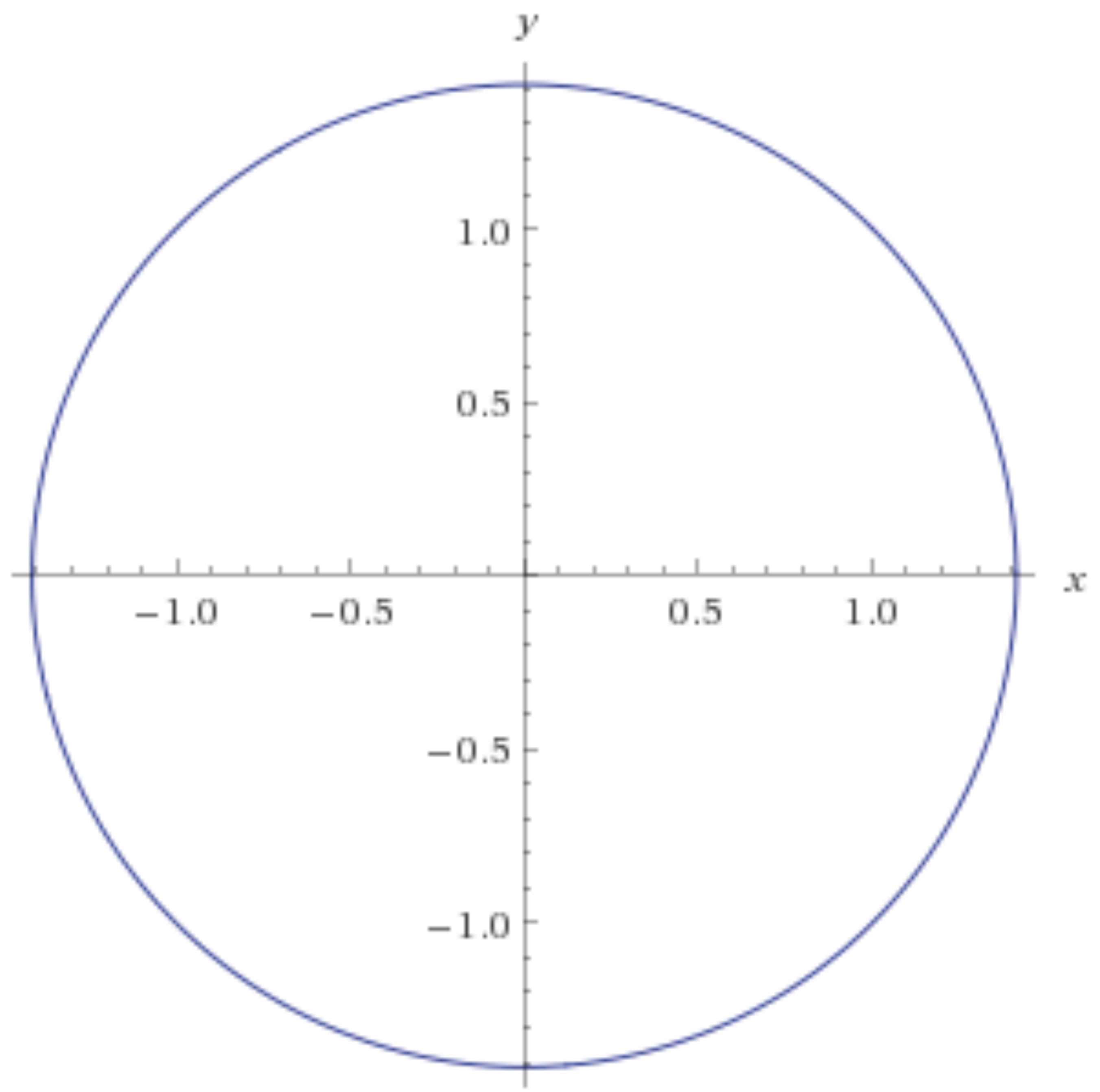
(Goodfellow 2017)

Solution

- The dynamics are a circular orbit:

$$x(t) = x(0)\cos(t) - y(0)\sin(t)$$
$$y(t) = x(0)\sin(t) + y(0)\cos(t)$$

Discrete time gradient descent can spiral outward for large step sizes



(Goodfellow 2017)

Tips and Tricks

- Use *nonlinear dynamical systems* theory to study behavior of optimization algorithms
- Demonstrated and advocated especially by Nagarajan and Kolter 2017

Results

- The good equilibrium is a stable fixed point (Nagarajan and Kolter, 2017)
- Two-timescale updates converge (Heusel et al, 2017)
  - Their recommendation: use a different learning rate for G and D
  - My recommendation: decay your learning rate for G
- Convergence is very inefficient (Mescheder et al, 2017)

Intuition for the Jacobian

How firmly does player 1 want to stay in place?

How much can player 1 dislodge player 2?

	$\boldsymbol{g}^{(1)}$	$\boldsymbol{g}^{(2)}$
$\boldsymbol{\theta}^{(1)}$	$\boldsymbol{H}^{(1)}$	$\nabla_{\boldsymbol{\theta}^{(1)}} g^{(2)}$
$\boldsymbol{\theta}^{(2)}$	$\nabla_{\boldsymbol{\theta}^{(2)}} g^{(1)}$	$\boldsymbol{H}^{(2)}$

How much can player 2 dislodge player 1?

How firmly does player 2 want to stay in place?

(Goodfellow 2017)

What happens for GANs?

		D	G
		$\boldsymbol{g}^{(1)}$	$\boldsymbol{g}^{(2)}$
D	$\boldsymbol{\theta}^{(1)}$	$\boldsymbol{H}^{(1)}$	$\nabla_{\boldsymbol{\theta}^{(1)}} g^{(2)}$
G	$\boldsymbol{\theta}^{(2)}$	$\nabla_{\boldsymbol{\theta}^{(2)}} g^{(1)}$	$\boldsymbol{H}^{(2)}$

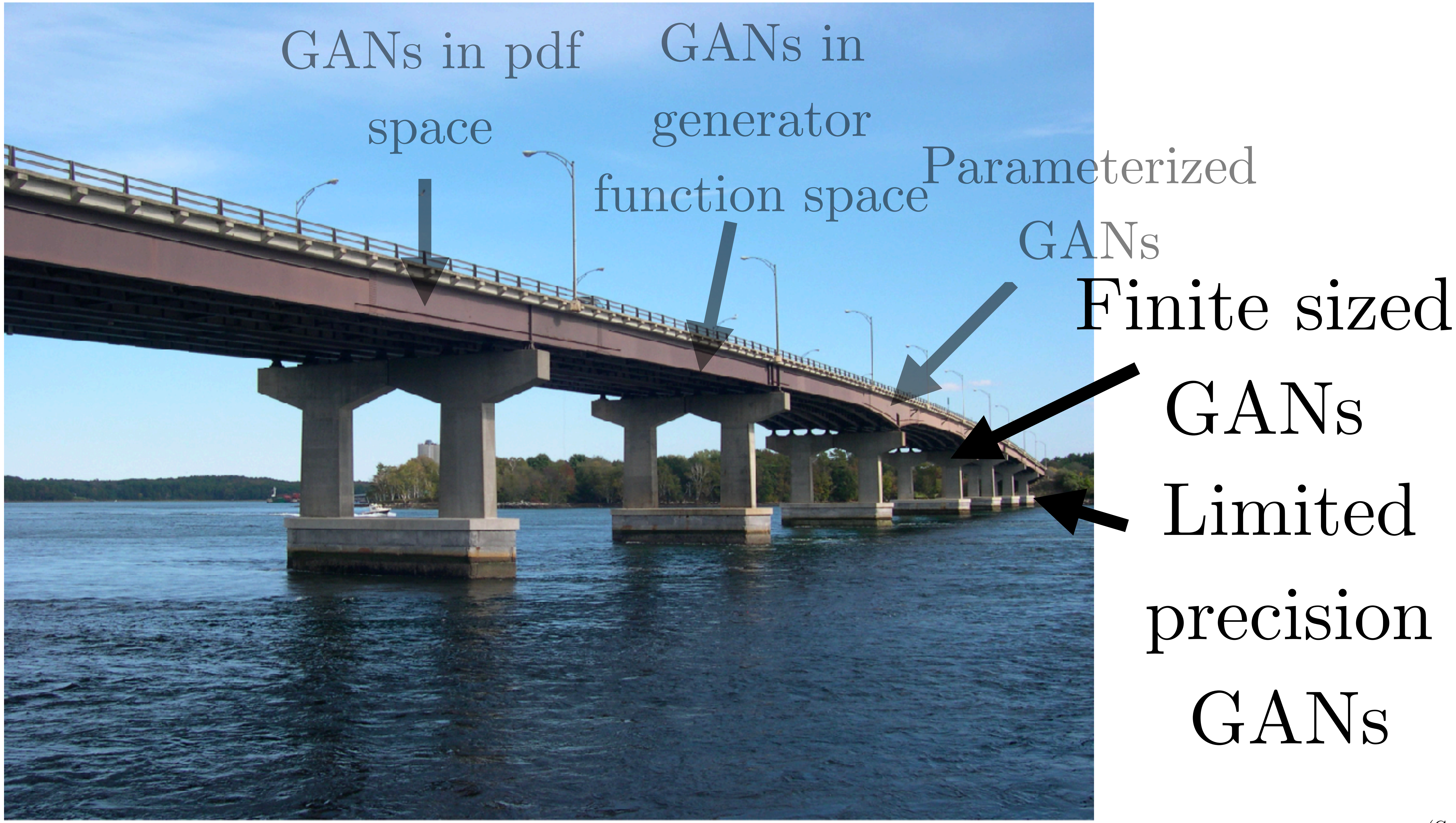
All zeros!

The optimal discriminator is constant.

Locally, the generator does not have any "retaining force"

(Goodfellow 2017)

Building a bridge from simple to complex theoretical models



(Goodfellow 2017)

Does a Nash equilibrium exist, in the right place?

- PDF space: yes
- Generator function space: yes, but there can also be bad equilibria
- What about for neural nets with a finite number of finite-precision parameters?
  - Arora et al, 2017: yes... for mixtures
    - Infinite mixture
    - Approximate an infinite mixture with a finite mixture

Open Challenges

- Design an algorithm that avoids bad equilibria in generator function space OR reparameterize the problem so that it does not have bad equilibria

- Design an algorithm that converges *rapidly* to the equilibrium

- Study the *global* convergence properties of the existing algorithms

(Goodfellow 2017)