

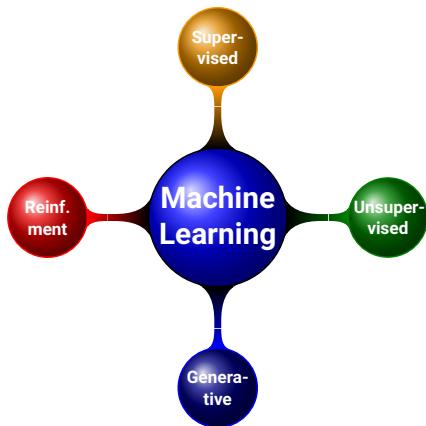
Statistical Theory for Generative Learning

Hanno Gottschalk (Math+, TU Berlin)

Summer School Applied Analysis, September 21 2026

The Task of Generative Learning

- **Supervised learning:** Learn a unknown conditional distribution $f(y|x)$ from pairs of data (Y_i, X_i) .
- **Unsupervised learning:** Learn a unknown distribution $f(x)$ from data X_j .
- **Generative learning:** Learn how to generate new samples from a unknown distribution $f(x)$ on the basis of data $X_j \sim f(x)$.



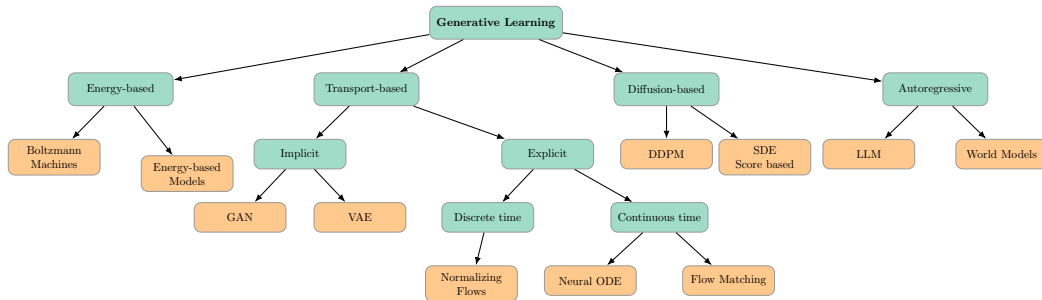
The Challenge of Generative Learning



From NVIDIA Style GAN, thispersondoesnotexist.com, Karras, Laine, Alia 2018

- **Challenge:** $X_j \in \mathbb{R}^d$ is high dimensional, $d \sim 10^6$.
- X_j could be images, a text messages, spoken language...
- Sampling from (log-) densities with MCMC is feasible, but need a density

Taxonomy of Generative Learning



Adversarial Learning Avoids Representing a Density

- Suppose $X_j \sim \mu$, μ probability measure on $[0, 1]^d$
- Transform noise $U_j \sim \lambda^{(d)}$ with **generator** $\phi : [0, 1]^d \rightarrow [0, 1]^d$

$$\phi_* \lambda^{(d)} = \lambda^{(d)} \circ \phi^{-1} \approx \mu$$

- $\phi_* \lambda^{(d)}(A) = \lambda^{(d)}(\phi^{-1}(A))$ is the image measure of the noise measure under ϕ .
- Learn ϕ in some hypothesis space $\mathcal{H}$ from the data X_j
- $\mathcal{H}$ might be high but finite dimensional – e.g. Neural Networks
- Or infinite dimensional, e.g. **Hölder space** $C^{k,\alpha}([0, 1]^d, \mathbb{R}^d)$.

Assumptions and Existence of a Generator in Hölder Spaces

- μ has a density wrt the Lebesgue measure on $[0, 1]^d$

$$d\mu(x) = f(x) d\lambda^{(d)}(x)$$

- $f(x)$ is a $C^{k,\alpha}([0, 1]^d, \mathbb{R})$ Hölder function
- $f(x) \geq \varepsilon > 0$ on $[0, 1]^d$.

Theorem (Realizability):

- (i) There exists a function $\phi_0 \in C^{k,\alpha}([0, 1]^d, \mathbb{R}^d)$ such that

$$\phi_{0*} \lambda^{(d)} = \mu.$$

- (ii) $\|\phi\|_{C^{k,\alpha}([0,1]^d, \mathbb{R}^d)}$ bounded in terms of $\|f\|_{C^{k,\alpha}([0,1]^d, \mathbb{R})}$ and ε .

Flow maps

- Want to realize ϕ as a flow endpoint w.r.t a vector field $v(x, t)$,

$$\dot{\phi}_t(x) = v(\phi(x), t), \quad \phi = \phi_1$$

- Does a suitable $v(x, t)$ exist?
- Transport equation

$$\partial_t f_t(x) = -\nabla \cdot (v(x, t)f_t(x)), \quad f_0(x) = f_\nu(x), \quad f_1(x) = f_\mu(x) \quad (!)$$

Beckmann Problem

- Model the influx $w(x) = v(x, t)f_t(x)$ time independent such that

$$-\nabla \cdot w(x) = f_\mu(x) - f_\nu(x), \quad x \in \Omega, \quad n \cdot w = 0, \quad x \in \partial\Omega. \quad (1)$$

- Then from the continuity equation for $v(x, t) = \frac{w(x)}{f_t(x)}$

$$\partial_t f_t(x) = -\nabla \cdot w(x) = f_\mu(x) - f_\nu(x) \Rightarrow f_t(x) = t f_\mu(x) - (1 - t) f_\nu(x)$$

- Beckmann's problem:

$$\frac{1}{2} \int_{\Omega} |w|^2 dx \rightarrow \min \quad \text{under constraints (1)}$$

Unrestricted Beckmann Problem

- Introduce Lagrangian multipliers $u \in C^1(\Omega, \mathbb{R})$, $u(x) = 0$ on $\partial\Omega$, $v \in C(\partial\Omega)$, $f = f_\mu - f_\nu$

$$\mathcal{L}(w, u, v) = \frac{1}{2} \int_{\Omega} |w|^2 dx + \int_{\Omega} (\nabla \cdot w + f) u dx + \int_{\partial\Omega} n \cdot (wv) dA$$

- Optimality conditions

$$\frac{\delta}{\delta v} \mathcal{L}(w, u, v) = 0 \Rightarrow n \cdot w = 0 \text{ on } \partial\Omega, \quad \frac{\delta}{\delta u} \mathcal{L}(w, u, v) = 0 \Rightarrow \nabla \cdot w + f = 0 \text{ on } \Omega$$

$$\frac{\delta}{\delta w} \mathcal{L}(w, u, v) = 0 \Rightarrow -\nabla u = w \text{ on } \Omega$$

Solution of the Optimality Conditions

- $u : \Omega \rightarrow \mathbb{R}$ fulfills Poisson equation with zero Neumann boundary conditions

$$\Delta u = f \text{ on } \Omega \quad \nabla u \cdot n = 0 \text{ on } \partial\Omega$$

- Weak formulation not coercive, but fulfills condition for Fredholm alternative with $g = 0$

$$\int_{\Omega} f \, dx - \int_{\partial\Omega} g \, dA = \int_{\Omega} f_{\mu} - f_{\nu} \, dx = 0.$$

- Classical solution exists (Nadri 2015).

Schauder Estimates

Theorem (Riedlinger, G. 2026)

- $u \in C^{k+1,\alpha}(\bar{\Omega}, \mathbb{R})$ (Schauder estimates, Agmon, Douglis, Nirenberg 59)
 - $\Rightarrow w \in C^{k+1,\alpha}(\bar{\Omega}, \mathbb{R}^d)$
 - $\Rightarrow v \in C^{k,\alpha}(\bar{\Omega} \times [0, 1], \mathbb{R}^d)$ (Asatryan et al, 2023)
 - $\Rightarrow \phi_t \in C^{k,(\alpha)}(\bar{\Omega} \times [0, 1], \bar{\Omega})$
 - Transport map is regular and can be well approximated by Neural Networks, (Yarotzky 2017 for ReLu NN, Belomnestny et al 2022 for ReQu NN).
-
- Extensions for parametric Beckmann problems are possible.
 - Of importance for conditional ('promptable') generative learning.

The Need for Adversarial Learning

- We want to estimate ϕ_0 with $\hat{\phi} \in \mathcal{H}$ based on data $X_1, \dots, X_n$ with Hypothesis space

$$\mathcal{H} = \{\phi : \phi, \phi^{-1} \in \mathcal{C}^{k,\alpha}([0, 1]^d, [0, 1]^d) : \|\phi\|_{\mathcal{C}^{k,\alpha}}, \|\phi^{-1}\|_{\mathcal{C}^{k,\alpha}} \leq K\}$$

- Standard procedure: compute **computationally infeasible!**

$$f_\phi(x) = |D\phi|^{-1} \circ \phi^{-1}(x) = |D\phi^{-1}(x)|$$

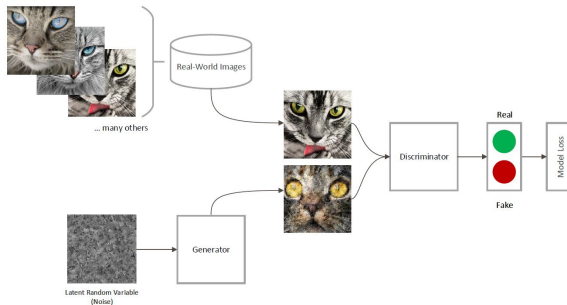
- Determine the neg log-Likelihood $\hat{L}_\phi = -\frac{1}{n} \sum_{j=0}^n \log f_\phi(X_j)$ and the ERM/MLE

$$\hat{\phi} = \arg \min_{\phi \in \mathcal{H}} \hat{L}_\phi$$

- Use the error decomposition and the ULLN with $L_\phi = \mathbb{E}[\hat{L}_\phi]$

$$\mathfrak{d}_{KL}(\mu || \hat{\phi}_* \lambda^{(d)}) \leq \underbrace{\inf_{\phi \in \mathcal{H}} \mathfrak{d}_{KL}(\mu || \phi_* \lambda^{(d)})}_{\text{red}} + \underbrace{\sup_{\phi \in \mathcal{H}} \hat{L}_\phi - \hat{L}_\phi}_{\text{red}} + 2 \sup_{\phi \in \mathcal{H}} |L_\phi - \hat{L}_\phi| \rightarrow 0 \text{ Unif. LLN}$$

Generative Adversarial Learning



- Min-Max two player game $\hat{\phi} \in \arg \min_{\phi \in \mathcal{H}} \sup_{D \in \mathcal{H}_D} \hat{L}(\phi, D, \{X_j\})$
- With $X_j \sim \mu, U_j \sim \lambda^{(d)}$, the empirical loss is defined as

$$\hat{L}(\phi, D, \{X_j\}) = \frac{1}{2n} \sum_{j=1}^n [\log(D(X_j)) + \log(1 - D(\phi(U_j)))]$$

The Kulback-Leibler-Divergence

Definition

- $\mu \ll \nu \Leftrightarrow \nu(A) = 0 \Rightarrow \mu(A) = 0$.
- $\mu \ll \nu \Rightarrow \mu = Z\nu, Z \geq 0, Z = L^1(\nu)$ Radon-Nikodyn derivative

$$Z = \frac{d\mu}{d\nu}.$$

- KL-divergence for $\mu \ll \nu$

$$\mathbb{d}_{KL}(\mu \parallel \nu) = \int f \left(\frac{d\mu}{d\nu} \right) d\nu, \quad f = t \log(t)$$

- Else $\mathbb{d}_{KL}(\mu \parallel \nu) = \infty$

Theorem:

$\mathbb{d}_{KL}(\mu \parallel \nu) \geq 0$ with equality iff $\mu = \nu$.

Proof: Apply Jensen's inequality and use that f is strictly convex.

Empirical and Theoretical Loss in Adv. Learning

Theorem (Goodfellow et al 2014):

(i) The expected loss

$$L(\phi, D, \{X_j\}) = \mathbb{E}_{X_j \sim \mu} [\hat{L}(\phi, D, \{X_j\})]$$

is maximized by $D_\phi = \frac{f(x)}{f(x) + f_\phi(x)}$, provided this is realizable in $\mathcal{H}_D$;

(ii)

$$\begin{aligned} L(\phi, D_\phi, \{X_j\}) &= \frac{1}{2} \left[\mathbb{d}_{KL} \left(\mu \parallel \frac{\phi_* \lambda^{(d)} + \mu}{2} \right) + \mathbb{d}_{KL} \left(\phi_* \lambda^{(d)} \parallel \frac{\phi_* \lambda^{(d)} + \mu}{2} \right) \right] \\ &\quad + \log(2) = \mathbb{d}_{JS} \left(\mu \parallel \phi_* \lambda^{(d)} \right) + \log(2). \end{aligned}$$

Empirical and Theoretical Loss in Adv. Learning

Theorem (Goodfellow et al 2014):

(i) The expected loss

$$L(\phi, D, \{X_j\}) = \mathbb{E}_{X_j \sim \mu} [\hat{L}(\phi, D, \{X_j\})]$$

is maximized by $D_\phi = \frac{f(x)}{f(x) + \hat{f}_\phi(x)}$, provided this is realizable in $\mathcal{H}_D$;

(ii)

$$\begin{aligned} L(\phi, D_\phi, \{X_j\}) &= \frac{1}{2} \left[\mathbb{d}_{KL} \left(\mu \left\| \frac{\phi_* \lambda^{(d)} + \mu}{2} \right\| \right) + \mathbb{d}_{KL} \left(\phi_* \lambda^{(d)} \left\| \frac{\phi_* \lambda^{(d)} + \mu}{2} \right\| \right) \right] \\ &\quad + \log(2) = \mathbb{d}_{JS} \left(\mu \parallel \phi_* \lambda^{(d)} \right) + \log(2). \end{aligned}$$

- A discriminator hypothesis space, s.t. D_ϕ is feasible, is given by

$$D_{\phi, \phi'} \in \mathcal{H}_D = \left\{ D \in C^{\alpha, k}([0, 1]^d, [0, 1]), \|D\|_{C^{k, \alpha}} \leq K' \right\} \quad \forall \phi, \phi' \in \mathcal{H}.$$

for K' suff. large.

proof

- (i) Write out the theoretical loss (under the assumption that densities f_μ and f_ϕ exist

$$\mathbb{E}_{X_j \sim \mu}[\hat{L}(\phi, D, \{X_j\})] = \int [\log(D)f_\mu + \log(1-D)f_\phi] dx$$

- Maximize the expressions in [...] for every x in $D(x)$ to achieve form of $D_\phi(x) = \frac{f_\mu(x)}{f_\mu(x) + f_\phi(x)}.$

■

- (ii) Insert a factor 2 in $\frac{f_\mu(x)}{f_\mu(x) + f_\phi(x)} \rightarrow \frac{2f_\mu(x)}{f_\mu(x) + f_\phi(x)}$ and correct with a $-\log(2)$.

Error Decomposition for Generative Adversarial Learning

- Asatryan, G., Lippert, Rottmann (2020), Biau, Cadre, Sangnier, Tanielian (2020)

$$\begin{aligned}\mathbb{d}_{\text{JS}}(\mu || \hat{\phi}_* \lambda^{(d)}) &= L(\hat{\phi}, D_{\hat{\phi}}) - \log(2) \\ &\leq \hat{L}(\hat{\phi}, D_{\hat{\phi}}) + \sup_{\phi, D} |L(\hat{\phi}, D_{\hat{\phi}}) - \hat{L}(\hat{\phi}, D_{\hat{\phi}})| - \log(2) \\ &\leq \hat{L}(\hat{\phi}, \hat{D}) + \sup_{\phi, D} |L(\hat{\phi}, D_{\hat{\phi}}) - \hat{L}(\hat{\phi}, D_{\hat{\phi}})| - \log(2) \\ &\leq \hat{L}(\phi_0, \hat{D}) + \sup_{\phi, D} |L(\hat{\phi}, D_{\hat{\phi}}) - \hat{L}(\hat{\phi}, D_{\hat{\phi}})| - \log(2) \\ &\leq L(\phi_0, \hat{D}) + 2 \sup_{\phi, D} |L(\hat{\phi}, D_{\hat{\phi}}) - \hat{L}(\hat{\phi}, D_{\hat{\phi}})| - \log(2) \\ &\leq L(\phi_0, D_{\phi_0}) + 2 \sup_{\phi, D} |L(\hat{\phi}, D_{\hat{\phi}}) - \hat{L}(\hat{\phi}, D_{\hat{\phi}})| - \log(2) \\ &= \cancel{\mathbb{d}_{\text{JS}}(\mu || \phi_{0*} \lambda^{(d)})} + 2 \sup_{\phi, D} |L(\hat{\phi}, D_{\hat{\phi}}) - \hat{L}(\hat{\phi}, D_{\hat{\phi}})| \rightarrow 0 \quad (\text{ULLN})\end{aligned}$$

Details on the Usage of the Uniform Law of Large Numbers

- Have to show that $\hat{L}(\phi, D, \{X_n\})$ $\mathcal{H}$ and $\mathcal{H}_D$ fulfill the requirements of the ULLN
- Problem $\mathcal{H}$ and $\mathcal{H}_D$ are not compact in the $C^{k,\alpha}$ -topology...
- ...but in the L^∞ -topology.
- As $\hat{L}(\phi, D, \{X_n\})$ is uniformly bounded (since $f(x) > \varepsilon$ by assumption) and continuous in the L^∞ -topology, we obtain:

Theorem (Asatyan, G. Lippert, Rottmann 2020)

Generative adversarial learning on (bounded subsets of) Hölder spaces is consistent, i.e. almost surely finds a correct generator ϕ_0 in the large sample limit.

The Empirical Loss as Subgaussian Process

- $\hat{L}_1(\phi, D, \{X_j\})$ is uniformly bounded and $\|\cdot\|_\infty$ -Lipshitz in $(\phi, D) \in \mathcal{H} \times \mathcal{H}_D$ with Lipshitz constant C_L
- By Hoeffding's inequality, $\hat{L}_1(\phi, D, \{X_j\})$ is a subgaussian process with variance proxy $\sigma^2 = C_L^2 \|\cdot\|_\infty^2 / 4$
- By Dudley's inequality

$$\begin{aligned}\mathbb{E}[\mathrm{d}_{\mathrm{JS}}(\mu || \hat{\phi}_* \lambda^{(d)})] &\leq 2\mathbb{E}[\sup_{\phi, D} |L(\hat{\phi}, D_{\hat{\phi}}) - \hat{L}(\hat{\phi}, D_{\hat{\phi}})|] \\ &\leq \frac{C}{\sqrt{n}} \int_0^\infty \sqrt{\log N(\mathcal{H} \times \mathcal{H}_D, \|\cdot\|_\infty, \varepsilon)} d\varepsilon.\end{aligned}$$

- I.e. the expected Jensen-Shannon distance between the learned mesure $\hat{\phi}_* \lambda^{(d)}$ and the true measure μ converges to 0

Covering Numbers and Chaining for Hölder Spaces

- It remains to show that the metric entropy integral is finite.

$$\int_0^\infty \sqrt{\log N(\mathcal{H} \times \mathcal{H}_D, \|\cdot\|_\infty, \varepsilon)} d\varepsilon < \infty$$

- For the covering numbers of Hölder spaces we know (van der Vaart/Wellner 1996)

$$\log N(B_{C^{k,\alpha}}(K), \|\cdot\|_\infty, \varepsilon) \leq C_1 \left(\frac{K}{\varepsilon}\right)^{\frac{d}{\alpha+k}}$$

- This the metric entropy integral converges, for high regularity $d/(k + \alpha) < 2$.

Further Results Based on Concentration Inequalities

Theorem (Asatryan, G., Lippert Rottmann 2020):

Let $K = \log^\beta n$ grow sufficiently slow and $\hat{\phi}, \hat{D}$ be estimated from $\mathcal{H}_K$ and $\mathcal{H}_{D,K}$. Then almost surely for all but finitely many n we have

$$\mathbb{d}_{JS}(\mu \| \hat{\phi}_* \lambda^{(d)}) \leq C_3 \frac{\log(n)^{4\beta(d+1)}}{\sqrt{n}}$$

- **Proof:** Combine Dudley's inequality with MC Diarmid's concentration inequality

$$\begin{aligned} Z(\phi, D) &= \sup_{\phi, D} |L(\hat{\phi}, D_{\hat{\phi}}) - \hat{L}(\hat{\phi}, D_{\hat{\phi}})| \\ &\quad - \mathbb{E}[\sup_{\phi, D} |L(\hat{\phi}, D_{\hat{\phi}}) - \hat{L}(\hat{\phi}, D_{\hat{\phi}})|] \end{aligned}$$

- $\Rightarrow$ For some $c > 0$ $P(\sup_{\phi, D} Z(\phi, D) > \delta) \leq e^{-n\delta/c}$

Better rates recently announced by Belomestny et al. 2021.

Turbulence in Fluid Dynamics

- Turbulent flow is **chaotic** by nature
- Computational Fluid Dynamics (CFD): Simulation of turbulences by numerically solving Navier-Stokes equations:

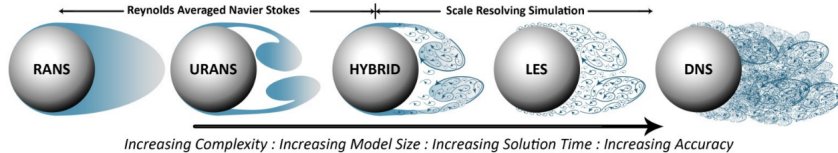
$$\frac{\partial \rho}{\partial t} + \nabla \cdot (\rho \mathbf{u}) = 0$$

$$\frac{\partial(\rho \mathbf{u})}{\partial t} + \nabla \cdot [\rho \mathbf{u} \otimes \mathbf{u}] = -\nabla p + \nabla \cdot \boldsymbol{\tau} + \rho \mathbf{f}$$

$$\frac{\partial(\rho \mathbf{e})}{\partial t} + \nabla \cdot ((\rho \mathbf{e} + p) \mathbf{u}) = \nabla \cdot (\boldsymbol{\tau} \cdot \mathbf{u}) + \rho \mathbf{f} \cdot \mathbf{u} + \nabla \cdot \dot{\mathbf{q}} + r$$

- But: Turbulent flow develops ever smaller and faster structures which are hard to resolve numerically

Simulation of Turbulence



Comparison of turbulence modeling approaches. Source: (J. Hart 2016)

- Modeling turbulences challenging in practice but highly technical relevant
- LES/DNS: Simulation of large or even all scales of turbulence but:
Enormous computational costs
- RANS/URANS: Struggle in simulation of fine details in vortex flow but:
Computational costs are sustainable $\Rightarrow$ mostly used in industry

Ergodicity: What Does “Chaotic” Actually Mean?

- Vortices always look different but in the long run their statistical properties are determined (**ergodicty**):

$$\lim_{T \rightarrow \infty} \frac{1}{T} \int_0^T f \circ \varphi_t(x_0) dt = \int_{\Omega} f(x) d\mu(x) \quad \forall x_0 \in \Omega. \quad (2)$$

⇒ The time average of a dynamical system equates with the ensemble average of its invariant measure.

- Probability measure μ on the flow configurations x encodes the statistical properties of the chaotic flow/dynamics $\varphi_t(x_0)$
- Goal: Sampling from the **unknown** measure $\mu \Rightarrow$ Can we **learn μ from data**?

Modeling Turbulence by Generative Adversarial Networks

Our contribution:

- GANs as another possibility to model turbulence
- Proof that GANs do converge for ergodic learning problems
- Generation of images that are as good as the output of the LES while requiring significantly less computation effort
- Numerical demonstration of generalization over changes in geometry
- Proof that physical quantities that characterize turbulence converge

Changing the Data Model from i.i.d. to Ergodic Data

- **Theorem (Drygala, Winhardt, di Mare, G. 2021)** Consider the Min-Max two player game for ergodic data

$$\hat{\phi} \in \arg \min_{\phi \in \mathcal{H}} \sup_{D \in \mathcal{H}_D} \hat{L}_T(\phi, D, x_0, \{U_j\})$$

- with $x_0 \in [0, 1]^d$, $U_j \sim \lambda^{(d)}$, the empirical loss **for ergodic flow** is defined as

$$\hat{L}_T(\phi, D, \{X_j\}) = \frac{1}{2T} \int_0^T \log(D(\varphi_t(x_0))) dt + \frac{1}{2T} \sum_{j=1}^{[T]} \log(1 - D(\phi(U_j))).$$

- Then the error decomposition of the previous theorem still holds and $\hat{\phi}_* \lambda^{(d)} \rightarrow \mu$ a.s. in JS-divergence.
- **Proof:** Use uniform ergodicity instead of ULLN ■

Lorenz Attractor

- Less complex **deterministic** ergodic system given by the system of ordinary differential equations

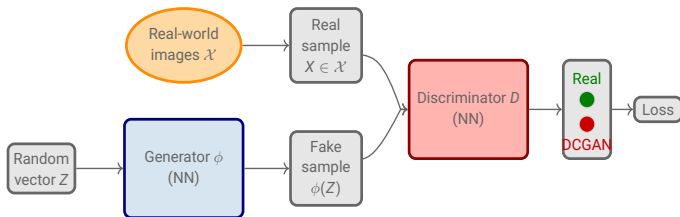
$$\frac{dx}{dt} = \sigma(y - x)$$

$$\frac{dy}{dt} = x(\rho - z) - y$$

$$\frac{dz}{dt} = xy - \beta z$$

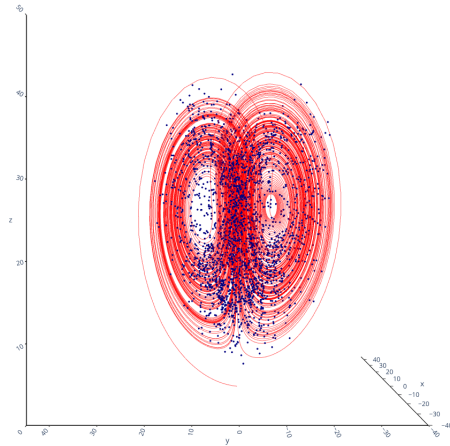
with $\sigma = 10$, $\beta = \frac{8}{3}$ and $\rho = 28$.

Lorenz Attractor - Training



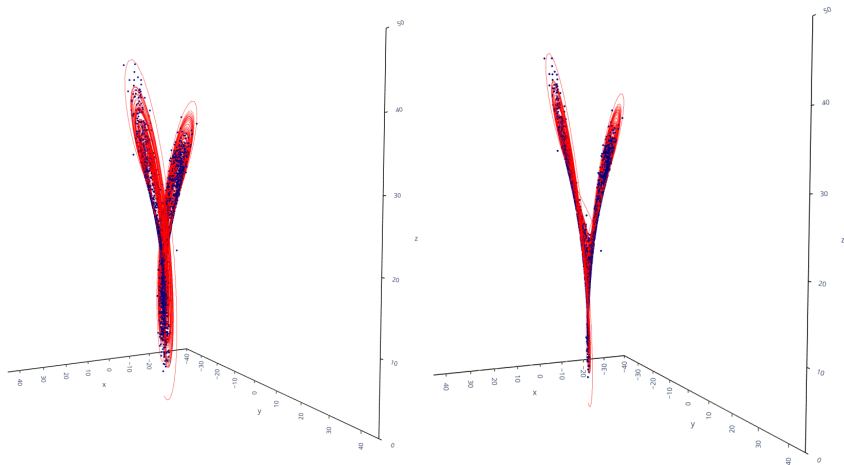
- Vanilla GAN for 200 000 epochs with batch size 20 000
- Trajectory started per epoch from randomly sampled initial point (x_0, y_0, z_0)
- Gaussian noise added to the network of discriminator and its real input data
- ϕ and D (deep) fully connected neural networks
- $U \in \mathbb{R}^{100}, U \sim U(0, 1)$

Lorenz Attractor - Results I



Trajectory consisting of 20 000 data points (red) and 3 000 synthesized data points (blue)

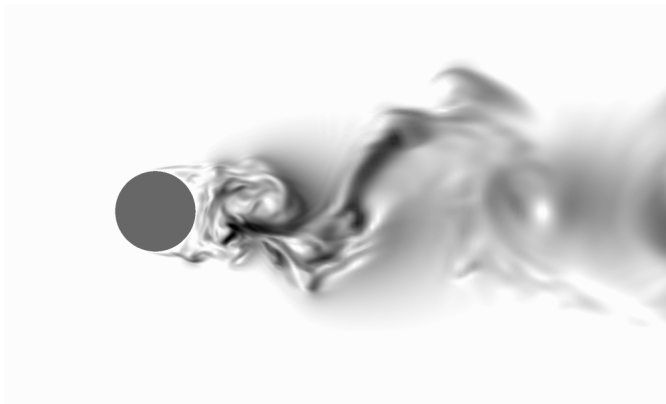
Lorenz Attractor - Results III



Rotated perspective from the trajectory of real data points (red) and the synthesized data points (blue).

Karman Vortex Street - Data

- Dataset: 5,000 images produced by LES
- Images: $w \times h = 1,000 \times 600$

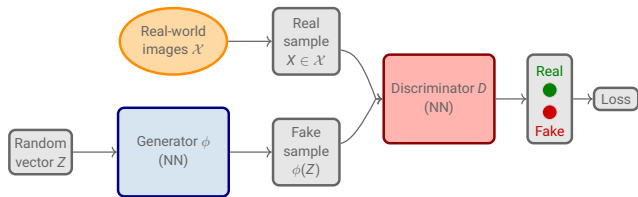


Example of the dataset.

Karman Vortex Street - Training and Inference I

Training

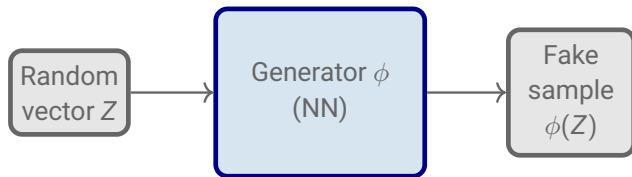
- GAN frameworks:
 - Vanilla GAN
 - Wasserstein GAN
 - Deep convolutional GAN
- Epochs: 200
- Batch size: 20
- Input:
 - 5,000 LES images
 - Image size: $k \times k$,
 $k \in \{64, 128, 256, \mathbf{512}\}$
 - Noise vector
 $Z \in \mathbb{R}^{100}, Z \sim N(0, 1)$



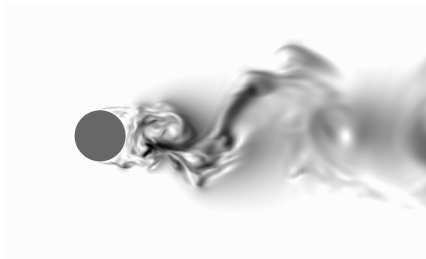
Karman Vortex Street - Training and Inference II

Inference

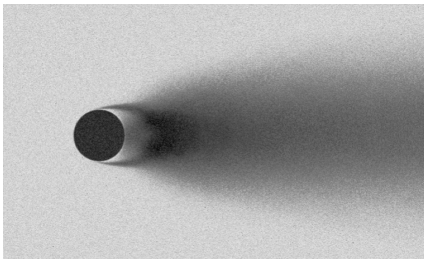
- Trained generator ϕ
- Input: Noise vector $Z \in \mathbb{R}^{100}, Z \sim N(0, 1)$



Karman Vortex Street - Vanilla GAN



LES image

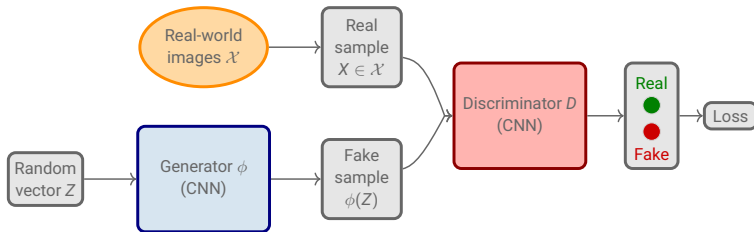


Fake image

Experiment settings:

- Architecture of ϕ and D : Fully connected NN with 5 layers
- Optimizer: Adam
- Learning rate: $2 \cdot 10^{-4}$

Deep Convolutional Generative Adversarial Network (DCGAN)



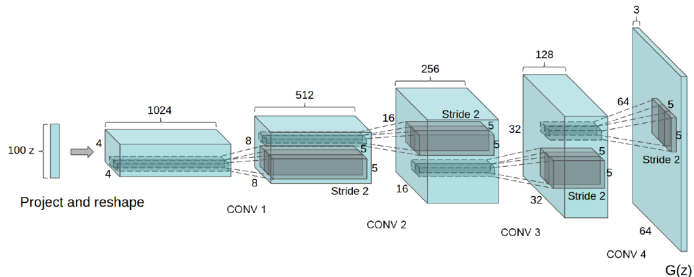
Architecture of DCGAN.

- Generator and discriminator are convolutional neural networks (CNNs)
- CNNs especially successful and applicable in field of image processing

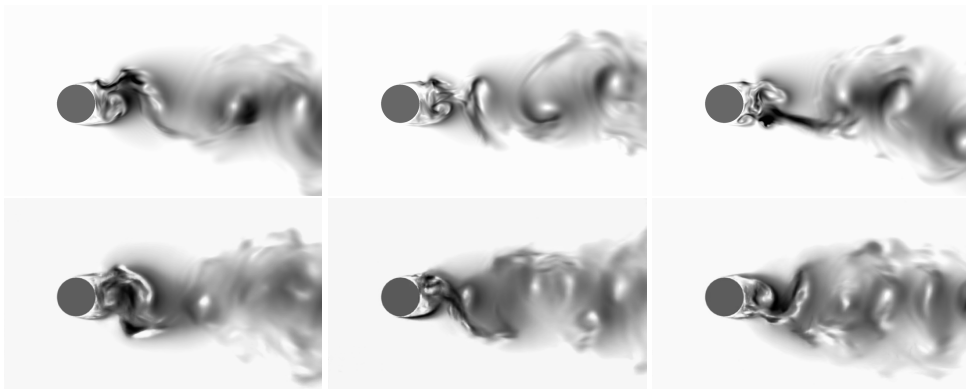
Deep Convolutional Generative Adversarial Network II

Guidelines to follow for **stable training** at **higher resolution** and **deeper architectures**:

- Stability: Apply batch normalization on the output layer of ϕ and the input layer of D
- Deeper architectures: Avoid fully-connected layers on top of convolutional features
- Higher resolution modeling: Leaky Rectified Linear Unit (ReLU) activation function for D
- ϕ and D learn own spatial up- or downsampling by replacing deterministic spatial pooling layers with (fractional-) strided atrous convolutions (Radford et al 2016)



Karman Vortex Street - Inference results after 2,000 Epochs



Karman Vortex Street - Inference results after 2,000 Epochs

LES



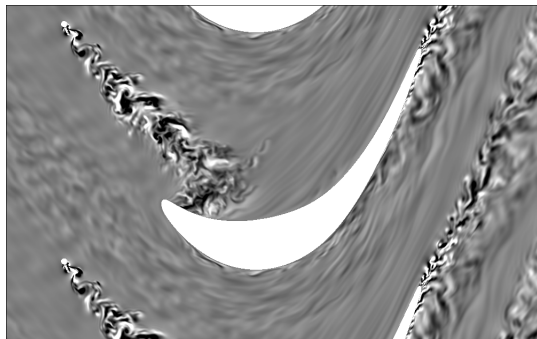
Fake



Comparison of LES (top) and fake images (bottom) produced by generator ϕ trained for 2 000 epochs.

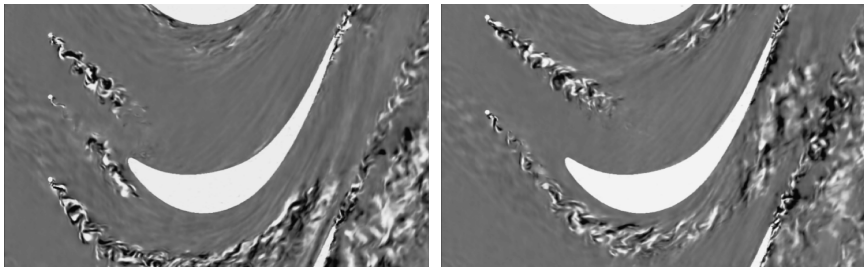
LPT Stator Under Periodic Wake Impact (LPT Stator) - Data

- Dataset: 2,250 images produced by LES
- Images: $w \times h = 1,000 \times 625$



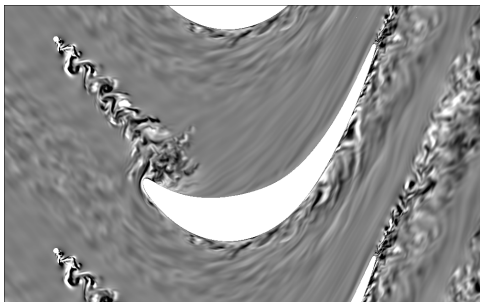
Example of the dataset.

LPT Stator - DCGAN: Inference Results after 2,000 Epochs I

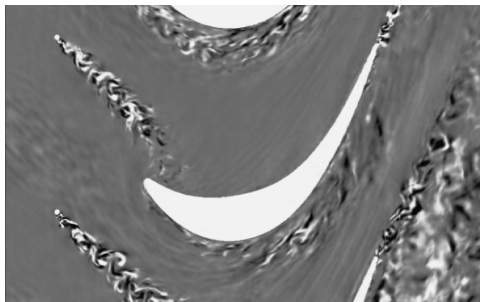


Examples of images generated by ϕ trained for 2 000 epochs with 2 250 images.

LPT Stator - DCGAN: Inference Results after 2,000 Epochs II

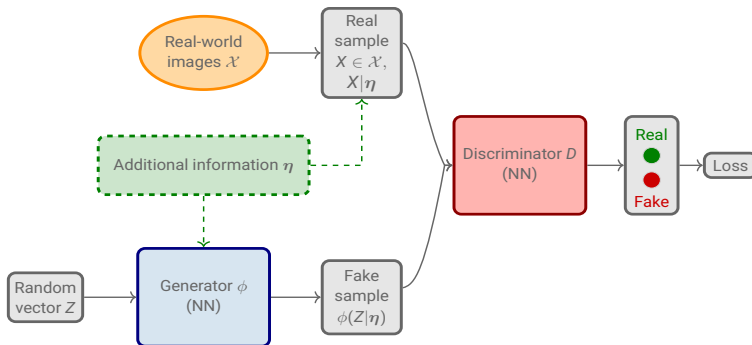


LES image



Fake image

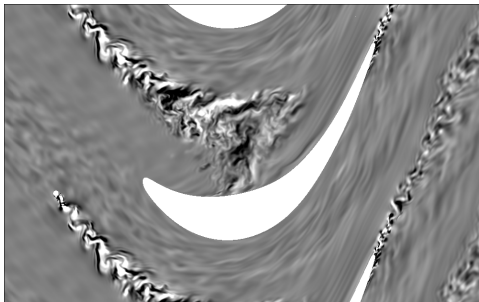
Conditional GAN (cGAN)



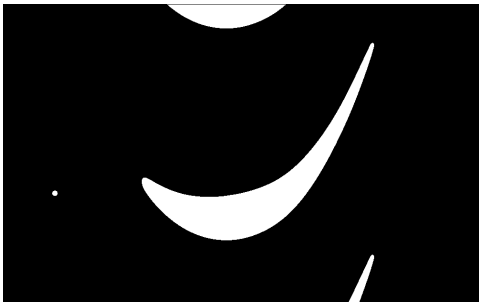
- Take control over the data production process by conditioning GAN framework
- Extension of loss function:

$$\mathcal{L}_{\text{cond}}(D, \phi) = \mathbb{E}_{\substack{X \sim \mu \\ \eta \sim \nu}} [\log(D(X|\eta))] + \mathbb{E}_{\substack{Z \sim \lambda \\ \eta \sim \nu}} [\log(1 - D(\phi(Z|\eta)))] \quad (3)$$

LPT Stator - Conditional Training



LES image



Binary segmentation mask

High-Resolution Image Synthesis with Conditional GANs

- Special form of cGAN
- Generation of **high-resolution photo-realistic** images by **conditioning** the input of the adversarial network on the corresponding **semantic label maps**
- Based on its former version pix2pix whose optimization problem is given as

$$\min_{\phi} \max_D \mathcal{L}_{cond}(D, \phi) \quad (4)$$

High-Resolution Image Synthesis via a Multi Scale Approach

- Introduction of **three innovations** for **improvement** of **photorealism** and **resolution** of the synthesized images:
 1. Coarse-to-fine generator: Decomposition of the generator into two sub-networks $\Rightarrow \phi = \{\phi_1, \phi_2\}$
 2. Multi-scale discriminators: Three discriminators D_1, D_2 and D_3 with same architecture but operating on different scales $\Rightarrow$ Modification of loss function:

$$\min_{\phi} \max_{D_1, D_2, D_3} \sum_{i=1}^3 \mathcal{L}_{cond}(\phi, D_i) \quad (5)$$

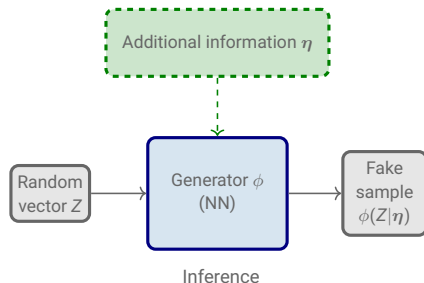
3. Feature matching loss $\mathcal{L}_{FM}$: Stabilize training

- Loss function in total:

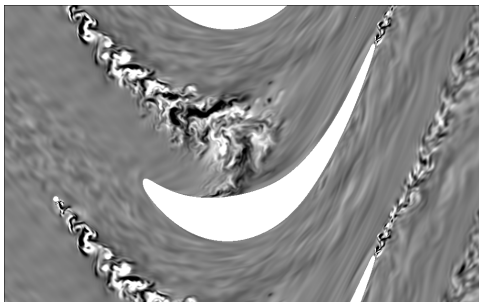
$$\min_{\phi} \left[\left(\max_{D_1, D_2, D_3} \sum_{i=1}^3 \mathcal{L}_{cond}(\phi, D_i) \right) + \gamma \sum_{i=1}^3 \mathcal{L}_{FM}(\phi, D_i) \right] \quad (6)$$

LPT-Stator - Experiment Settings

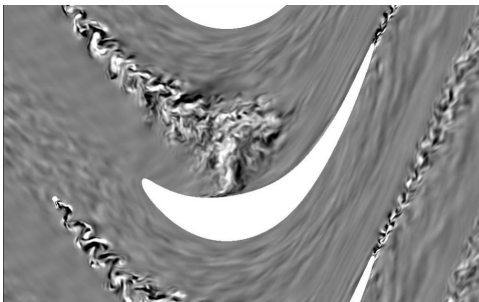
- Architecture: Adopted from original authors with small changes to avoid artifacts
- Epochs: 200
- Batch size: 10
- Input: 2,000 LES images, noise vector and masks of size $k \times k' = 992 \times 624$
- Optimizer: Adam
- Learning rate: $2 \cdot 10^{-4}$



LPT Stator - Inference Results I

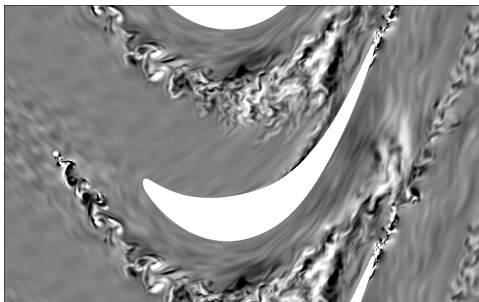


LES image

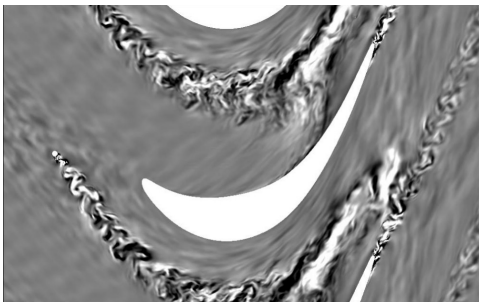


Fake image

LPT Stator - Results II

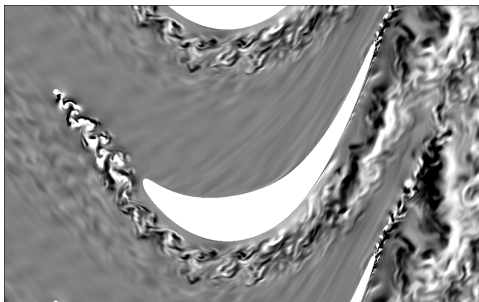


LES image

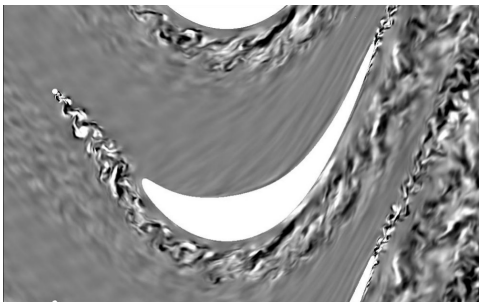


Fake image

LPT Stator - Results III

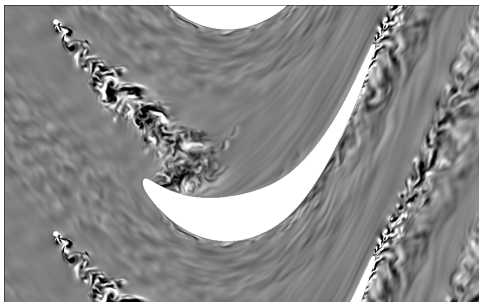


LES image

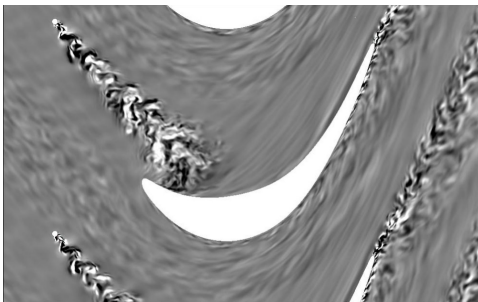


Fake image

LPT Stator - Results IV



LES image



Fake image

Computational Costs

► LES:

- Performed on 560 CPU cores of the CPU type Intel Xeon "Skylake" Gold 6132 @2.6 GHz
- Karman vortex street (5 000 images): 72 core weeks $\hat{=}$ 1 days
- LPT Stator (2250 images): 640 core weeks $\hat{=}$ 8 days

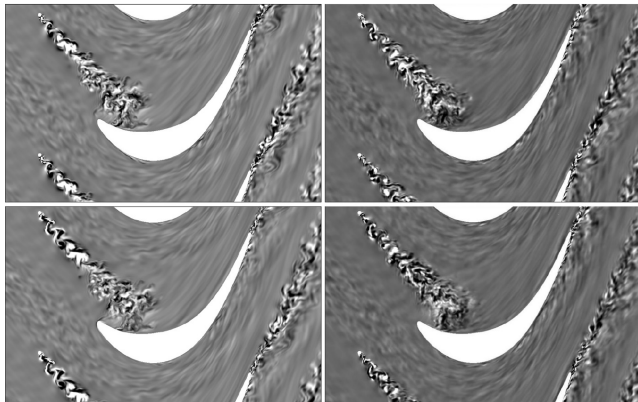
► GAN-Training:

- Performed on GPU of type Quadro RTX 8000 with 48 GB
- Karman vortex street (DCGAN, 2 000 epochs): 1.5 min/epoch
- LPT Stator (pix2pixHD, 200 epochs): 17 min/epoch

► GAN-Inference:

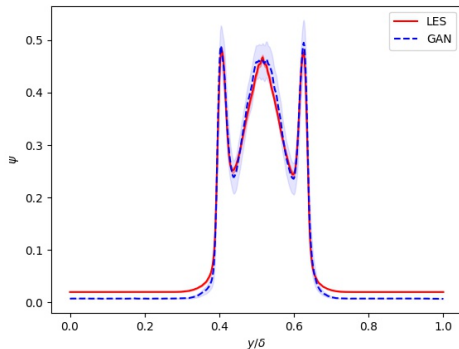
- Performed on GPU of type Quadro RTX 8000 with 48 GB
- Karman vortex street (DCGAN): 0.001 sec/image $\Rightarrow$ 5 seconds
- LPT Stator (pix2pixHD): 0.01 sec/image $\Rightarrow$ 22.5 seconds

Generalization for Unseen Geometry Configurations



- Repeat experiment with 5% of the images omitted around specific wake position
- During inference, position the wake at exactly the middle of that position
- results demonstrates generalization capability over unseen changes in geometry

Statistical Similarity vs Physics



- Measure strength of turbulence (variation of $c = |u|$ as a function of ξ)

$$\text{Var}[c(\xi)] = \lim_{T \rightarrow \infty} \frac{1}{T} \int_0^T |c(\xi, t) - \bar{c}(\xi)|^2 dt$$

Estimation of Physical Quantities

- All statistical physical quantities of the turbulent flow can be expressed through evaluation of functionals $\Psi(x)$ and averaging

$$\bar{\Psi} = \mathbb{E}_{X \sim \mu}[\Psi(X)]$$

- We can assume that $\Psi(x)$ is bounded on $[0, 1]^d$.
- We estimate $\bar{\Psi}$ by $\hat{\Psi}$

$$\hat{\Psi} = \mathbb{E}_{Z \sim \lambda^{(d)}}[\Psi(\hat{\phi}(Z))] = \lim_{N \rightarrow \infty} \frac{1}{N} \sum_{j=1}^N \Psi(\hat{\phi}(U_j)) \quad \text{a.s..}$$

■ $\Rightarrow$

$$|\bar{\Psi} - \hat{\Psi}| \leq C \sup_{\|\Psi'\| \leq 1} \left| \int \Psi' d\mu - \int \Psi' d\hat{\phi}_* \lambda^{(d)} \right| = C \|\mu - \hat{\phi}_* \lambda^{(d)}\|_{TV}.$$

Convergence of Physical Quantities

- **Proposition (Drygala, Winhard, di Mare, G. 2021):**

$$|\bar{\Psi} - \hat{\Psi}| \rightarrow 0 \quad \text{as } n \rightarrow \infty$$

- **Proof:** Enough to show: $\|\mu - \hat{\phi}_* \lambda^{(d)}\|_{TV} \rightarrow 0$.
- We obtain with $\hat{\mu} = \hat{\phi}_* \lambda^{(d)}$, triangular and Pinsker's inequality

$$\begin{aligned} \|\hat{\mu} - \mu\|_{TV} &\leq \left\| \hat{\phi}_* \lambda^{(d)} - \frac{\mu + \hat{\mu}}{2} \right\|_{TV} + \left\| \mu - \frac{\mu + \hat{\mu}}{2} \right\|_{TV} \\ &\leq \sqrt{2\mathfrak{d}_{KL} \left(\hat{\mu} \left\| \frac{\mu + \hat{\mu}}{2} \right\| \right)} + \sqrt{2\mathfrak{d}_{KL} \left(\mu \left\| \frac{\mu + \hat{\mu}}{2} \right\| \right)} \\ &\leq 2\sqrt{\mathfrak{d}_{KL} \left(\hat{\mu} \left\| \frac{\mu + \hat{\mu}}{2} \right\| \right) + \mathfrak{d}_{KL} \left(\mu \left\| \frac{\mu + \hat{\mu}}{2} \right\| \right)} = 2\sqrt{\mathfrak{d}_{JS}(\hat{\mu} \parallel \mu)} \rightarrow 0. \end{aligned}$$

Summary

- Theory of Generative Adversarial Learning can be understood in the framework of SLT
 - The (remaining) practical problems of GAN can be better understood as violations of assumptions
 - Synthetic data potentially can help tasks like automated driving (but beware the remaining domain gap)
 - Generative learning can be useful for applications beyond speech generation and computer vision
 - Further data models, like deterministic chaos, can be combined with generative adversarial learning.
1. H. Asatryan, G., M. Lippert, M. Rottmann, A Convenient Infinite Dimensional Framework for Generative Adversarial Learning, Electronic Journal Of Statistics 2023, arXiv:2011.12087
 2. H. Gottschalk, T. Riedlinger, Regularity of Solutions to Beckmann's Parametric Optimal Transport, arXiv:2603.19755
 3. C. Drygala, B. Winhard, F. di Mare, G., Generative Modeling of Turbulence, Physics of Fluids 2021 (editor's pick)
 4. C. Drygala, F. di Mare, G., Generaization capabilities of conditional GAN for turbulent flow, EUROGEN 2023.