香港科技大學
THE HONG KONG UNIVERSITY OF SCIENCE AND TECHNOLOGY

Learning Deep Generative models via Variational Gradient Flow (VGrow)

Prof. Can Yang

Email: macyang@ust.hk

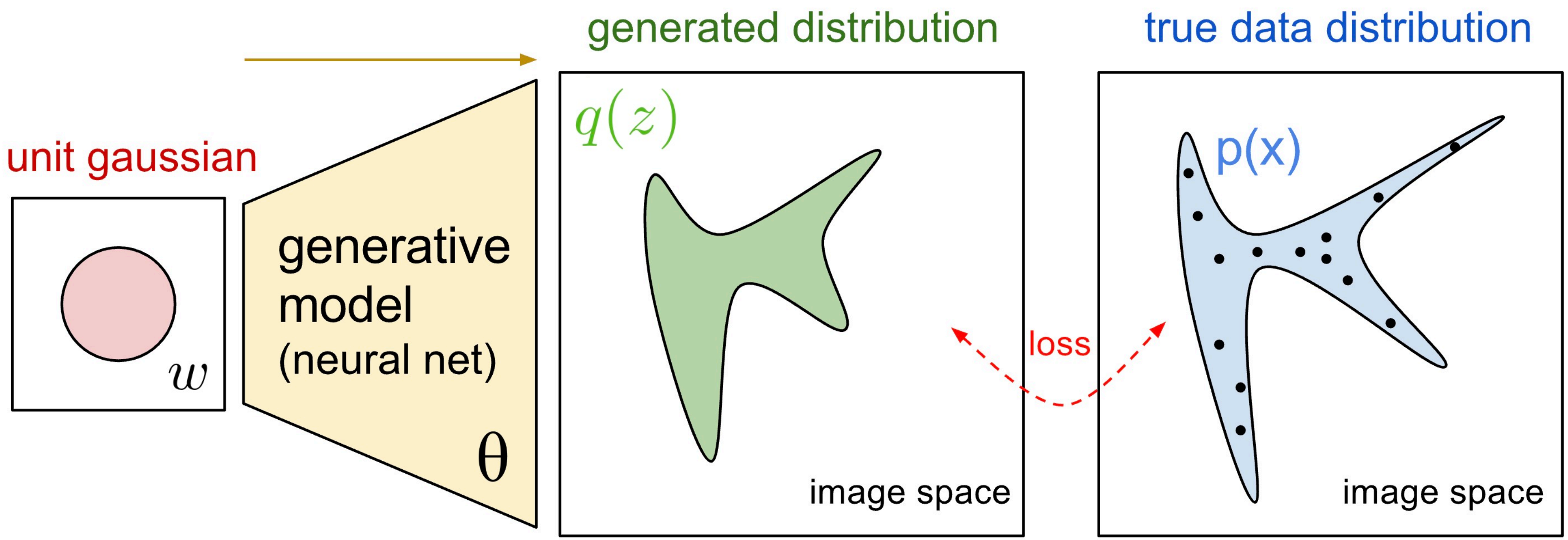
Data Science,

Department of Mathematics

School of Science

Joint work with Yuan Gao, Yuling Jiao, Gefei Wang,Yao Wang, Yang Wang, Shunkang Zhang.

Generative model



$$\min_{T_\theta} \mathbb{D}(q(z)|p(x))$$

Figure Credit: OpenAI

Intuition

- Consider a batch of particles $\{z_i\}, i = 1, \ldots, n$ with distribution $q(z)$
- Update these particles $\{z_i\}$ by a small amount (preserve continuity),

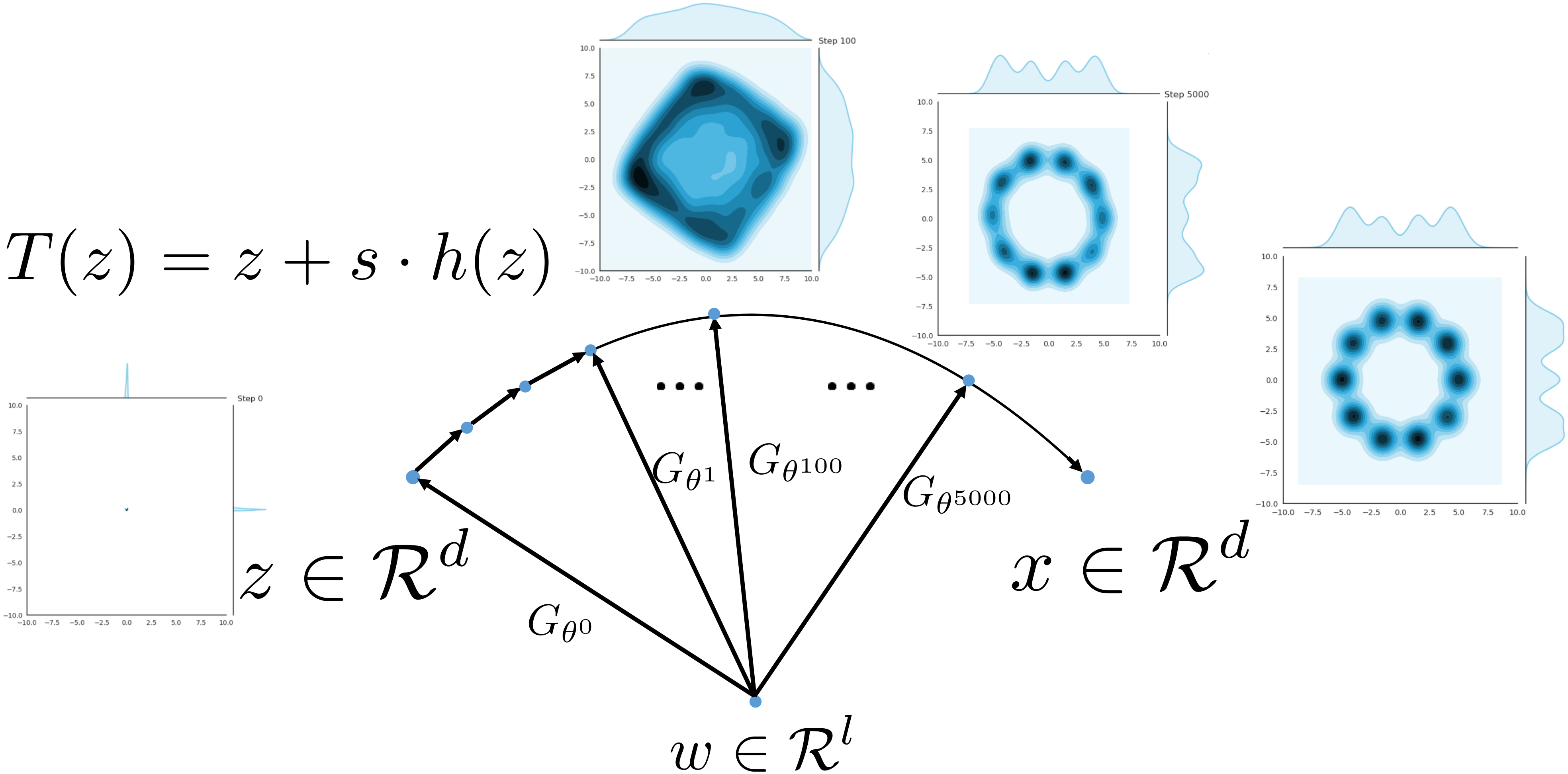
$$T(z) = z + s \cdot h(z)$$

such that the distribution of $\{T(z_i)\}$, denoted as $\tilde{q}(z)$, is closer to $p(x)$, the distribution of $\{x_i\}$

$$\mathbb{D}(\tilde{q}(z)|p(x)) \leq \mathbb{D}(q(z)|p(x))$$

Variational Gradient Flow (VGrow)

$$T(z) = z + s \cdot h(z)$$



Find h( )

- Consider f-divergence

$$\mathbb{D}_f(q(x)|p(x)) = \int p(x) f(\frac{q(x)}{p(x)}) dx$$

where $f : \mathbb{R}^1 \to \mathbb{R}^1$ is convex and $f(1) = 0$.

- KL, JS, Jeffery and log-D divergences are the special cases of f-divergence.
- By calculating functional gradient (MATH part), we have

$$h(x) = -f''(r(x)) \nabla r(x)$$

where $r(x) = \frac{q(x)}{p(x)}$

- Recall that $log\frac{p(x)}{q(x)} = d(x)$, $r(x)$ can be estimated as $\hat{r}(x) = \exp(-\hat{d}(x))$

Generated Portrait (based on Wiki Art)

Connection with Differential Equation

$$\min \frac{1}{2n}\|y - X\beta\|_2^2$$

Gradient Method

$$\beta^{(k)} = \beta^{(k-1)} + \epsilon \cdot \frac{X^T}{n}(y - X\beta^{(k-1})$$

From Gradient Method to ODE

$$\frac{\beta^{(k)} - \beta^{(k-1)}}{\epsilon} = \frac{X^T}{n}(y - X\beta^{(k-1)}) \qquad \epsilon \to 0 \qquad \frac{d\beta(t)}{dt} = \frac{X^T}{n}(y - X\beta(t))$$

$$\beta(t) = (X^T X)^+ \left(I - \exp\left(-t\frac{X^T X}{n}\right)\right) X^T y$$

Similarly

$$z^{(k)} = z^{(k-1)} + s \cdot h(z^{(k-1)}) \qquad \frac{dz(t)}{dt} = h(z(t))$$

Summary

- Proposed a general framework to learn deep generative models via Variational Gradient Flow (VGrow) on probability spaces.
- Proved: The evolving distribution of $\{z_i\}$ that asymptotically converges to the target distribution p(x) is governed by a vector field, which is the negative gradient of the first variation of the f-divergence between q(z) and p(x). (Based *Vlasov-Fokker-Planck* equation)
- Established connections of VGrow with other popular methods, such as VAE, GAN and flow-based methods (Stein Variational Gradient).
- We also evaluated several commonly used divergences, including Kullback-Leibler, Jensen-Shannon, Jeffrey divergences as well as our newly discovered “logD” divergence which serves as the objective function of the logD-trick GAN.