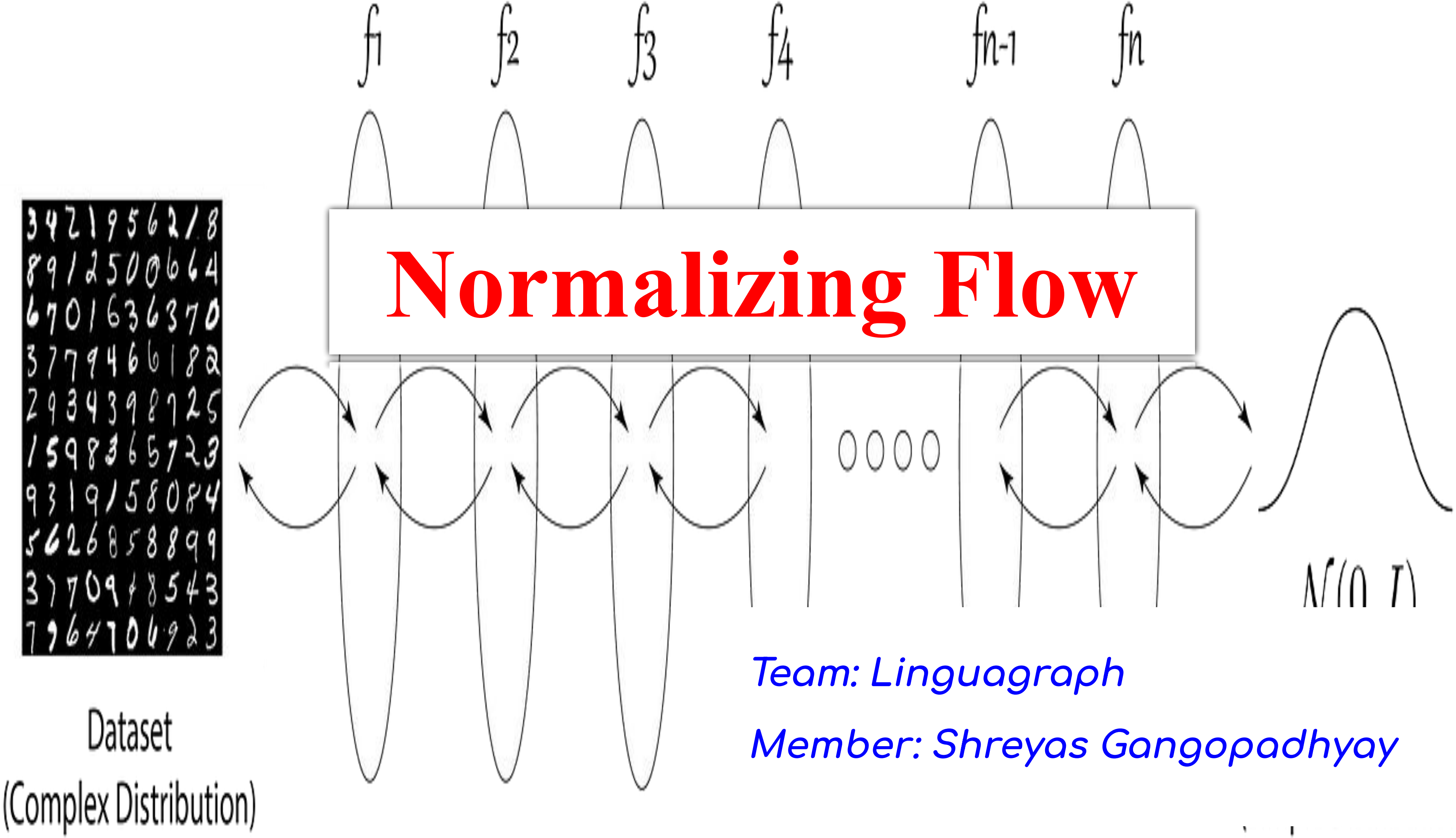
Normalizing Flow



Team: Linguagraph

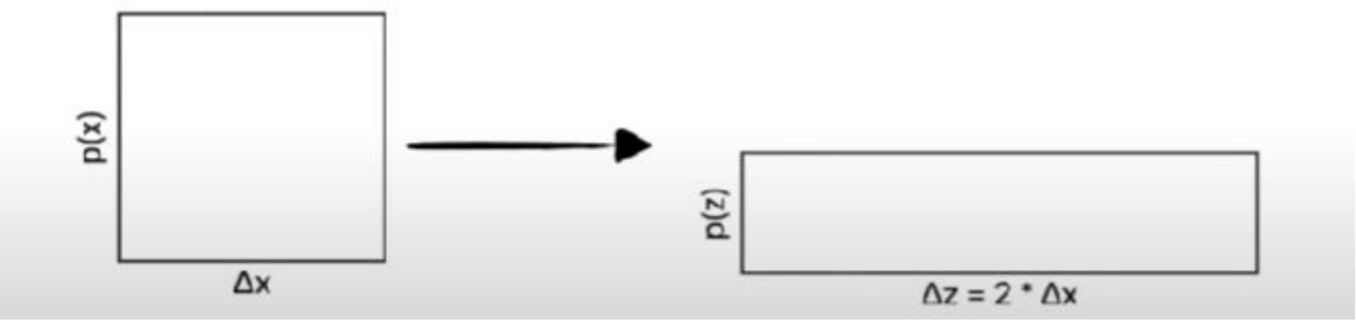
Member: Shreyas Gangopadhyay

Mapping The Random Variable

$$|p(x)\Delta x| = |P(z)\Delta z|$$

$$p(x) = P(z)|d(z)/d(x)|$$

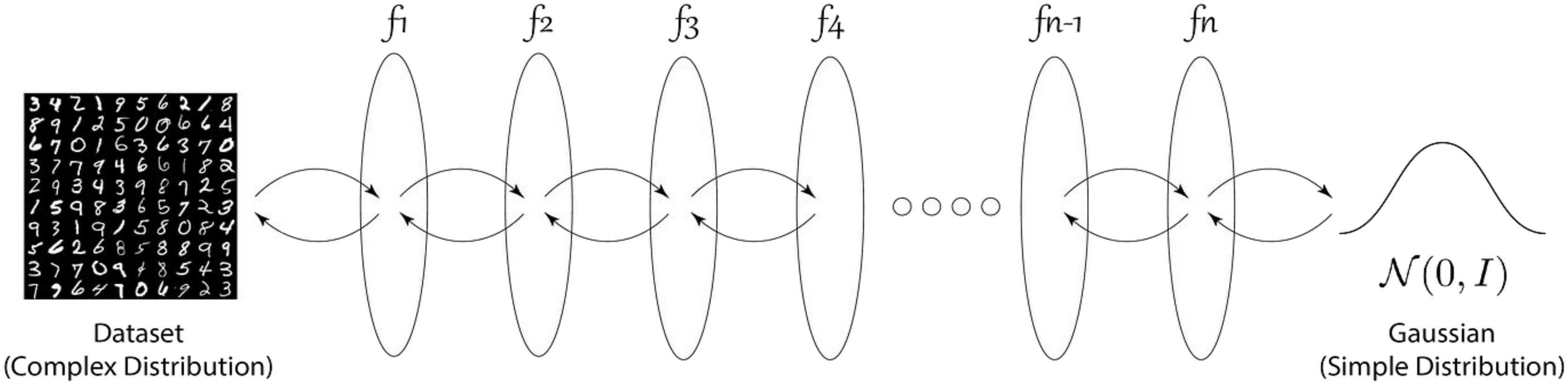
$$\log(p(x)) = \log(P(z)) + \log(|d(z)/d(x)|)$$



What is Normalizing Flow?

Deep Learning based method to map a one distribution to another one.

The random variable is mapped from one space to another using a series of bijective functions.



How Does It Work?

$$z \sim P(z) = \mathcal{N}(z;0,1)$$

$$x = f(z) = f_1 \ldots\ldots f_n\ (z)$$

$$f: Z \longrightarrow X$$, f is invertible.

Can we say : $P_\Theta(x) = P_\Theta(f^{-1}(x))$ **?**

The answer is **No.**

Change Of Variables Formula:

$$P_\Theta(x) = P_\Theta(f^{-1}(x)) \left|det(\partial(f^{-1}(x))/\partial(x))\right|$$

$$\Rightarrow P_\Theta(x) = P_\Theta(z) \left|det(\partial(z)/\partial(x))\right|$$

Jacobian Determinant

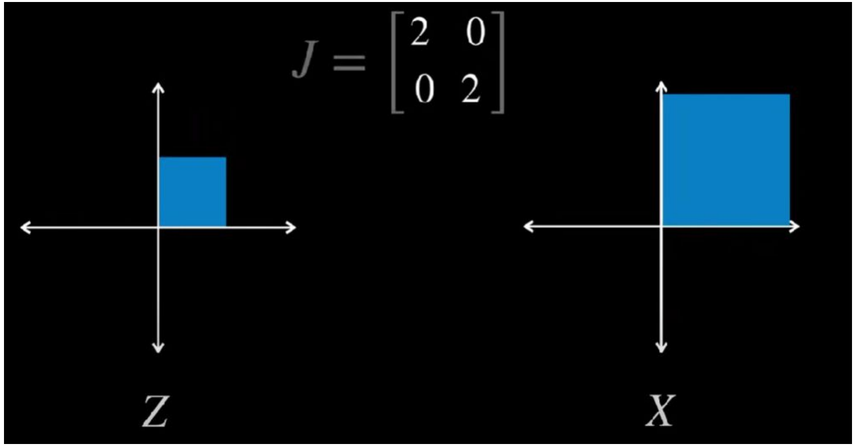
What does the determinant of Jacobian Signify?

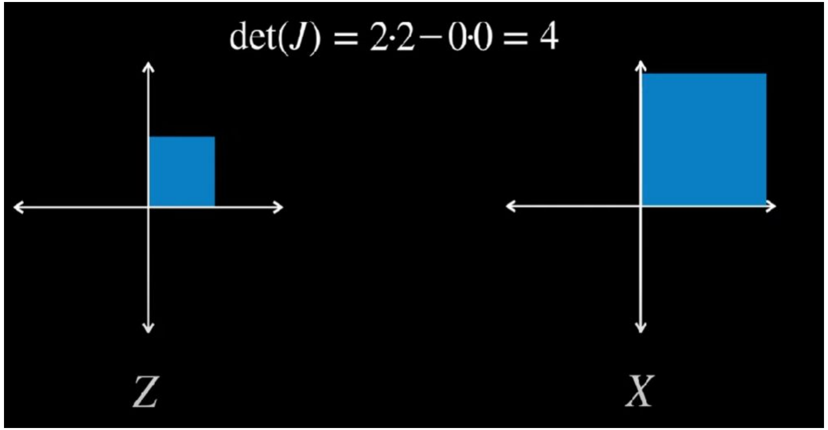
1. The previous equation shows the probability of x is equal to the probability of z multiplied by a scalar constant.

2. A probability distribution function must always integrate to 1.

3. The Jacobian determinant signifies how much the transformation contracts or expands a space.

4. This factor ensures that the new density function of x also satisfies the requirement of x.

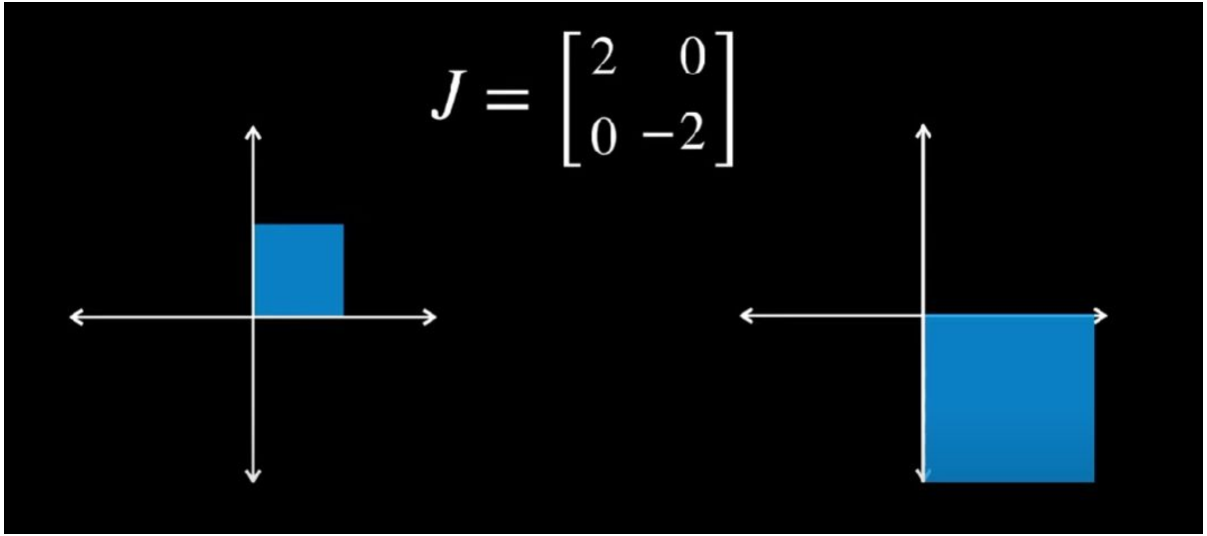




Note : $|\det(\partial(z)/\partial(x))| = |1/\det(\partial(x)/\partial(z))|$

Another Note:

The orientation of the space is of no consequence because we are taking into account the magnitude of the determinant.



Log-Likelihood

$$P_{\Theta}(x) = P_{\Theta}(z)\ \Pi|\det(\partial(f_i^{-1})/\partial(x))|$$

Taking Logarithm on both sides:

$$\log(P_{\Theta}(x)) = \log(P_{\Theta}(z)) + \Sigma \log(|\det(\partial(f_i^{-1})/\partial(x))|)$$

The final objective is to maximise **$\log(P_{\Theta}(x))$**.

Applications

Normalizing flows can be used to generate new data based on the training data.

For generative models we typically assume:

z to be the latent variables

(We assume the latent variables to be sampled from a normal distribution)

x to be the observed variables.

$$\text{maximise}\left(\log\left(P_{\Theta}(z)\right) + \Sigma \log\left(\left|\det\left(\partial(f_i^{-1})/\partial(x)\right)\right|\right)\right)$$

Latent variables sampled from a normal distribution

Model we want to train

Types of Normalizing Flows

1. Planar Flows : Expands or contracts the distribution along a specific direction.
2. Radial Flows : Modifies the distribution along a certain point.
3. Residual Flows : These use mapping functions like $g(x) = x + F(x)$

Case Study : VITS(TTS Model)

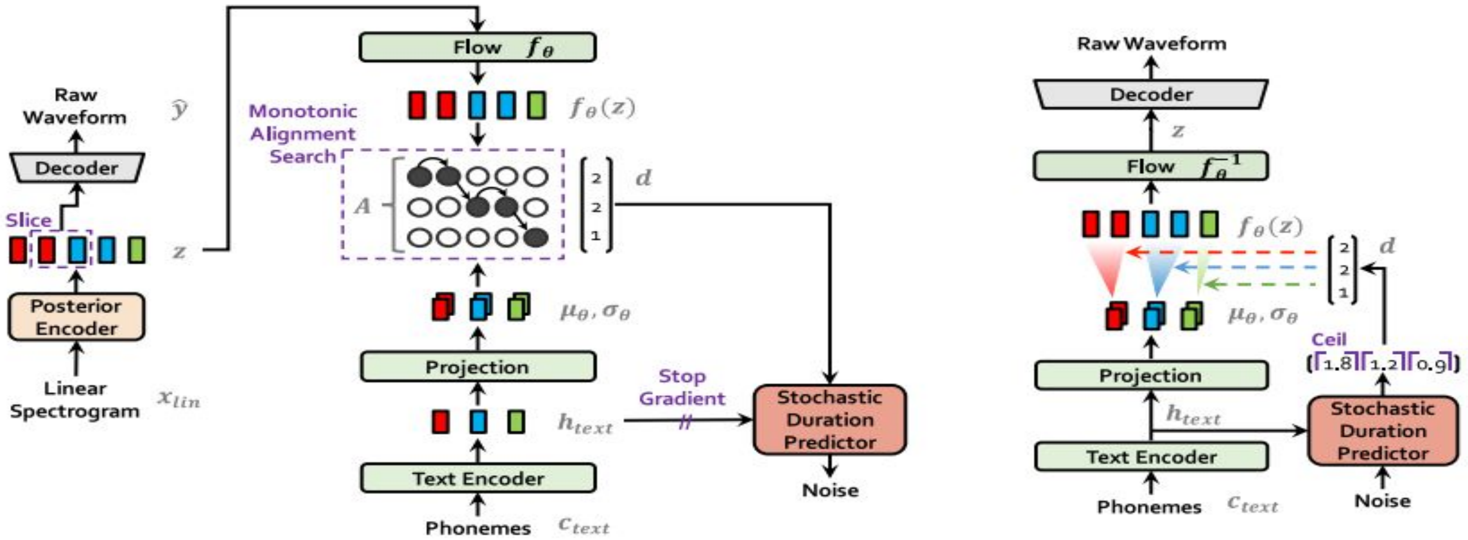
VITS stand for *Conditional Variational AutoEncoder With Adversarial Training.*

Normalizing Flow finds an application over there, in mapping the simple prior distribution into a more complex distribution to improve expresiveness.

$$p_\theta(z|c) = N(f_\theta(z); \mu_\theta(c), \sigma_\theta(c)) \left| \det \frac{\partial f_\theta(z)}{\partial z} \right|,$$

$$c = [c_{text}, A]$$

Architecture of VITS



(a) Training procedure

(b) Inference procedure

Thank You