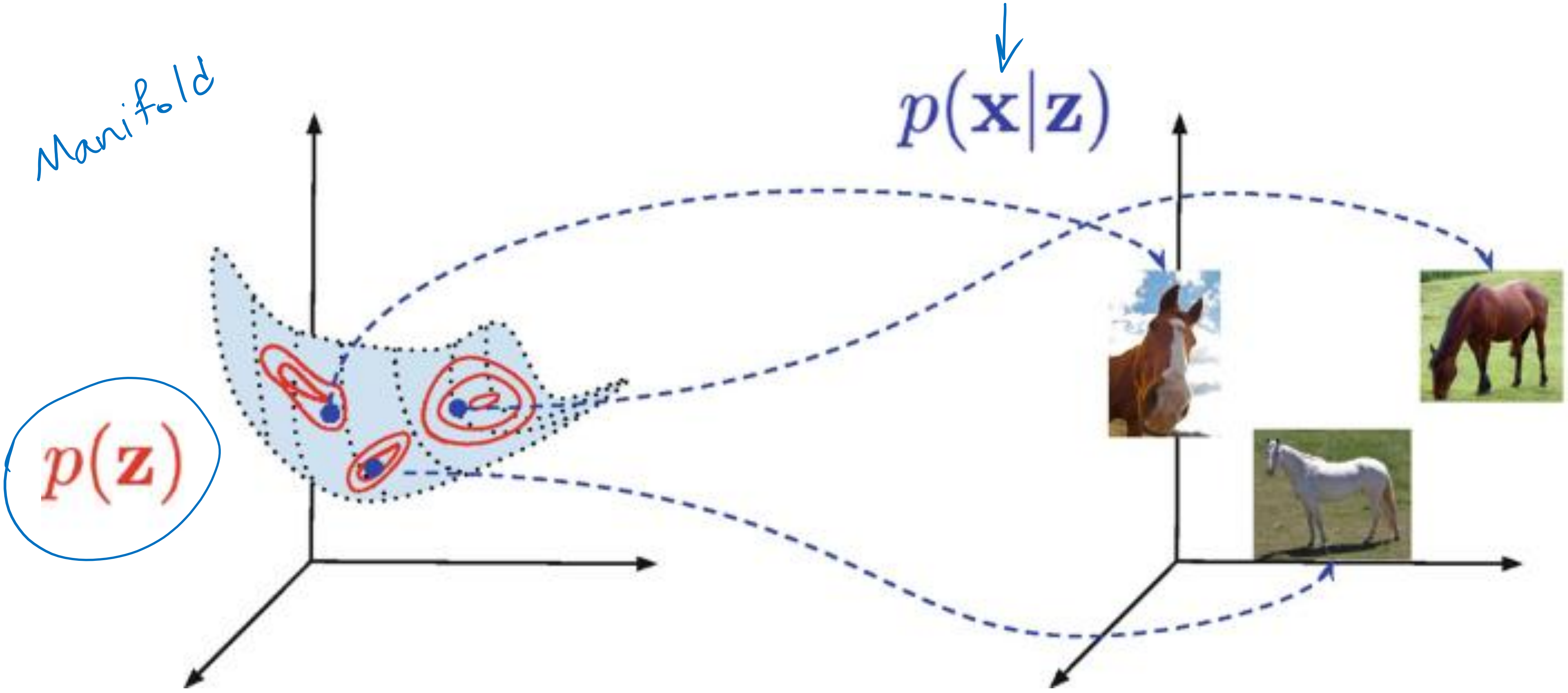
Normalizing Flows

Mostafa Tavassolipour

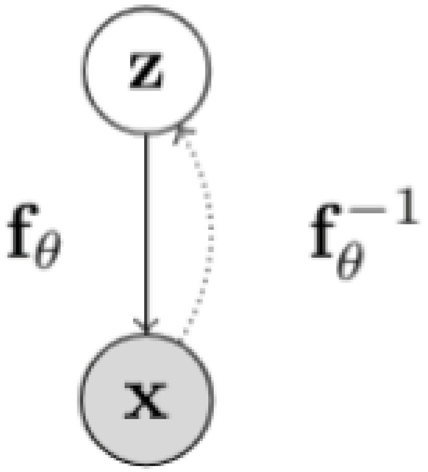
Fall 2024

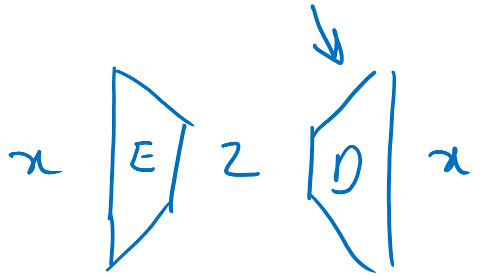
Latent Variable Models

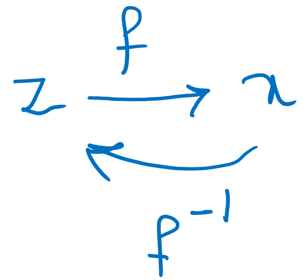


Normalizing Flow

- In a **normalizing flow** model, the mapping between Z and X, given by $f_\theta : \mathbb{R}^n \to \mathbb{R}^n$, is deterministic and invertible such that $X = f_\theta(Z)$ and $Z = f_\theta^{-1}(X)$



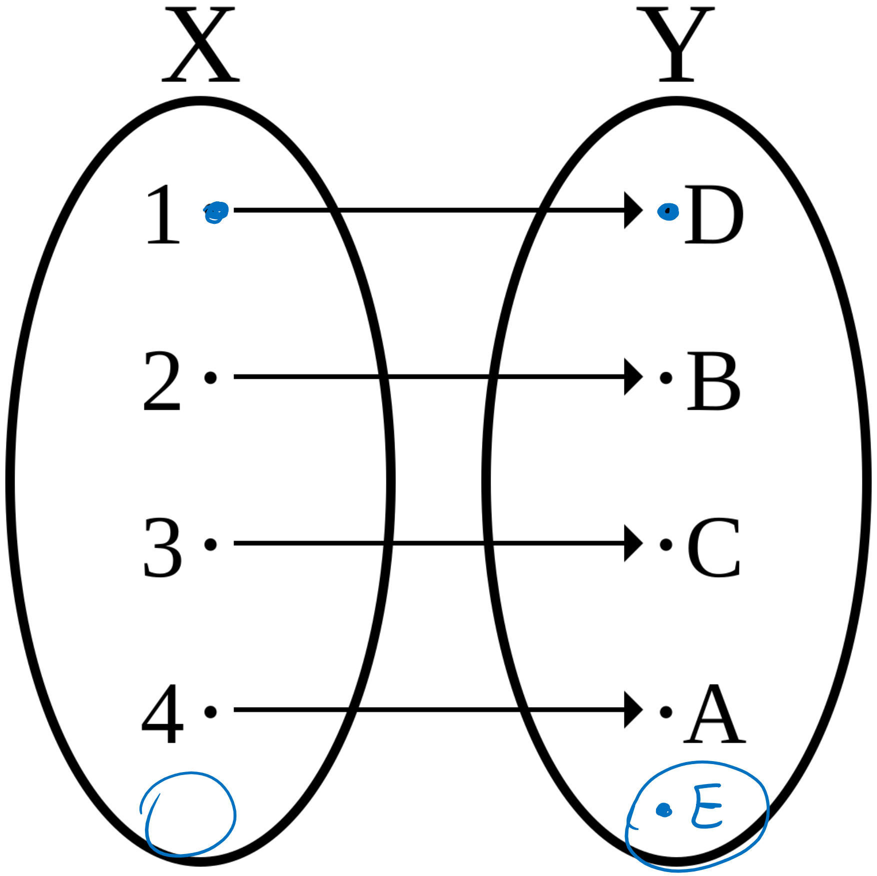


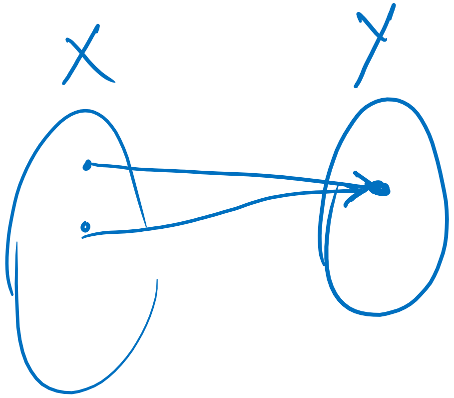


- Note: $\boldsymbol{x}, \boldsymbol{z}$ need to be continuous and have the same dimension.

Bijective Transformation

$Z \xrightarrow{f} X$

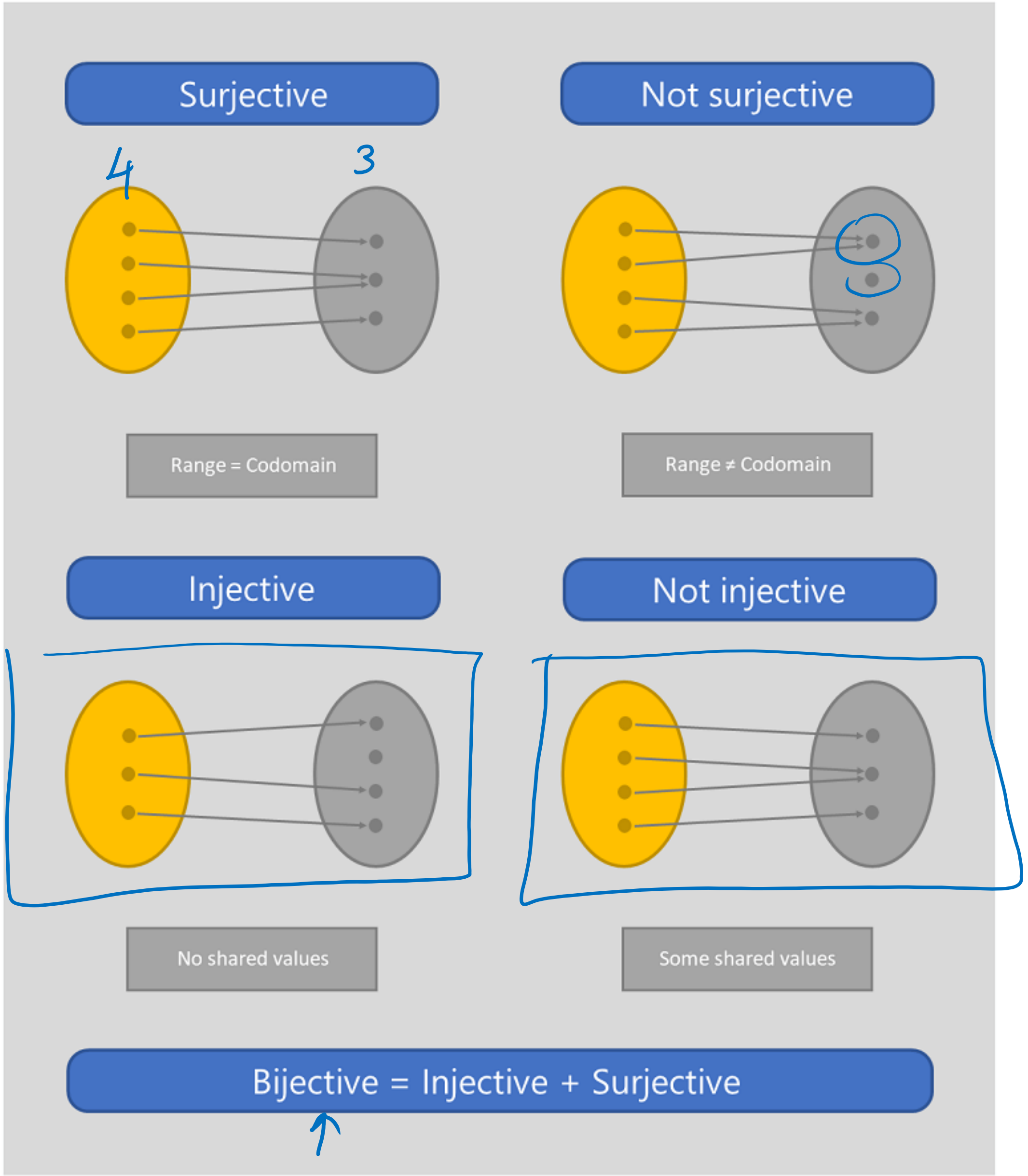




Non-bijective Transformation

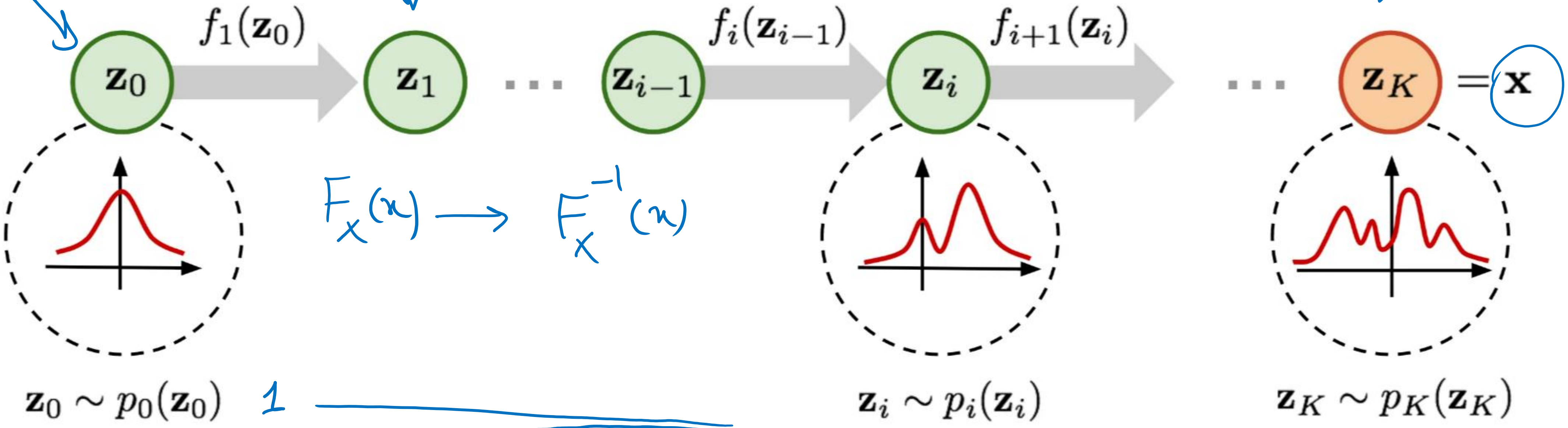
$z \in \mathbb{R}^5$

$x \in \mathbb{R}^{10}$



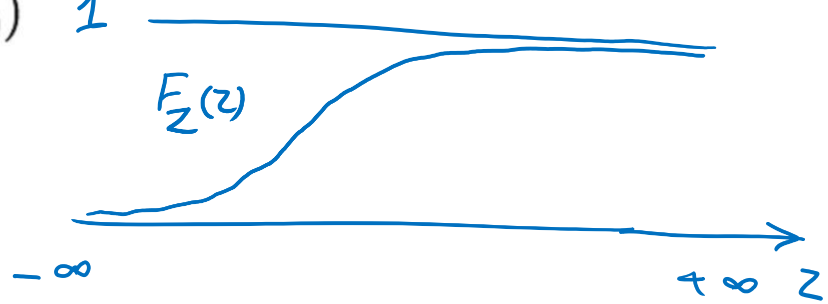
Normalizing Flows

$$x = \underbrace{f_K \circ f_{K-1} \circ \cdots \circ f_1}_{f}(z_0) \Rightarrow \boxed{x = f(z)} \rightarrow \boxed{z = f^{-1}(x)}$$



$F_X(x) \longrightarrow F_X^{-1}(x)$

$p_Z(z) \Rightarrow F_Z(z)$



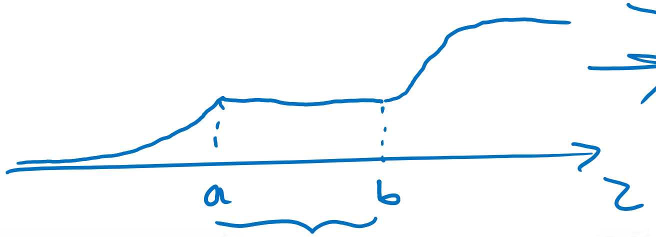
$y = F_Z(z)$ $\quad$ $y \sim U(0,1)$

$W = F_X^{-1}(y)$

$$\boxed{X = F_X^{-1}(F_Z(z))}$$

Example

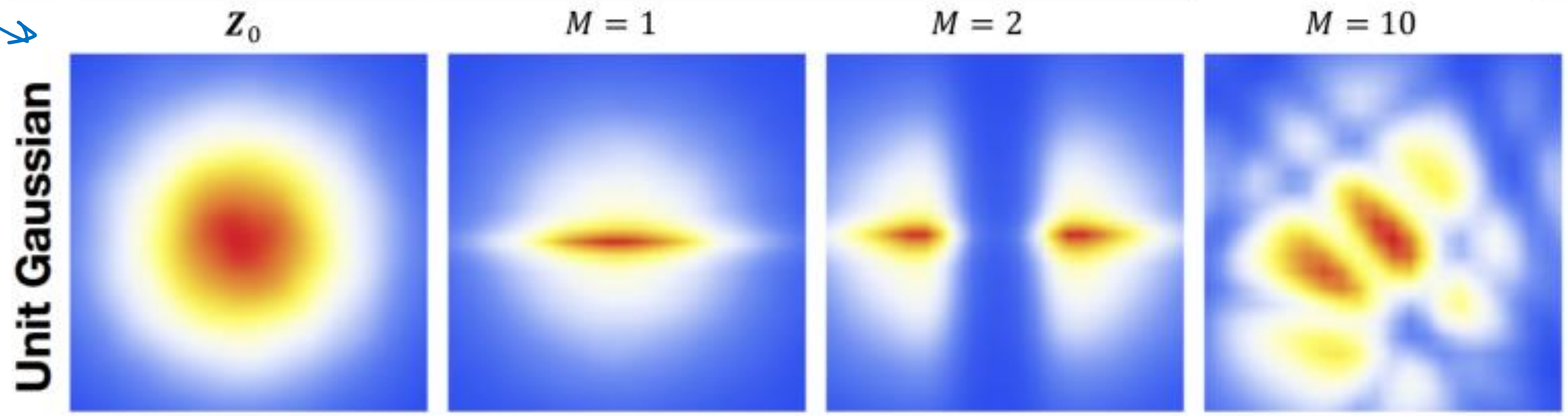
$Pr\{a \le z \le b\} = 0$



$\boxed{\begin{matrix} P(z) > 0 \\ P(X) > 0 \end{matrix}}$

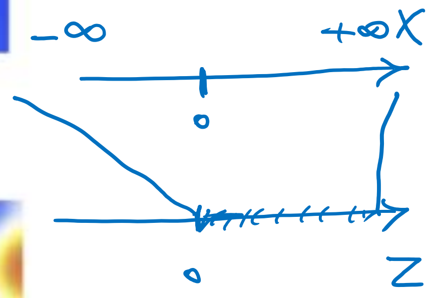
- Base distribution: Gaussian

$p(z) > 0$



$X \sim N(0, I)$

$z = X^2$

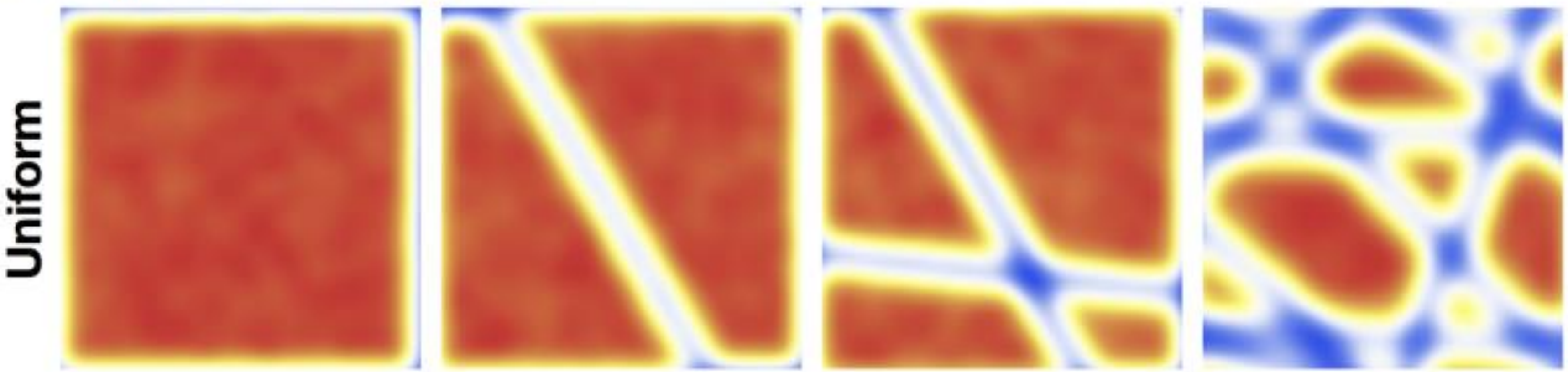


- Base distribution: Uniform

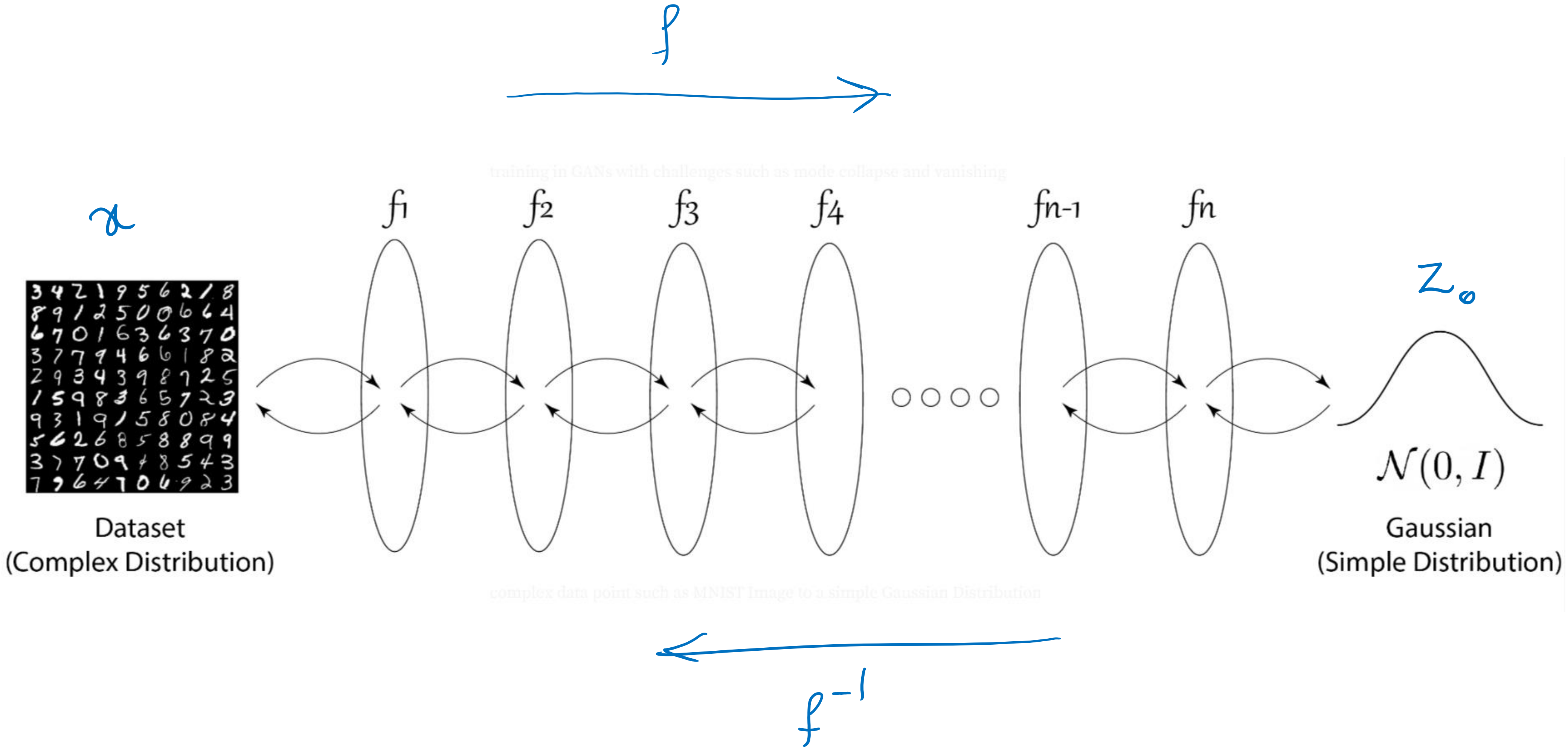
$z \to X$

$\boxed{U \to X}$

$\boxed{P(X) > 0}$



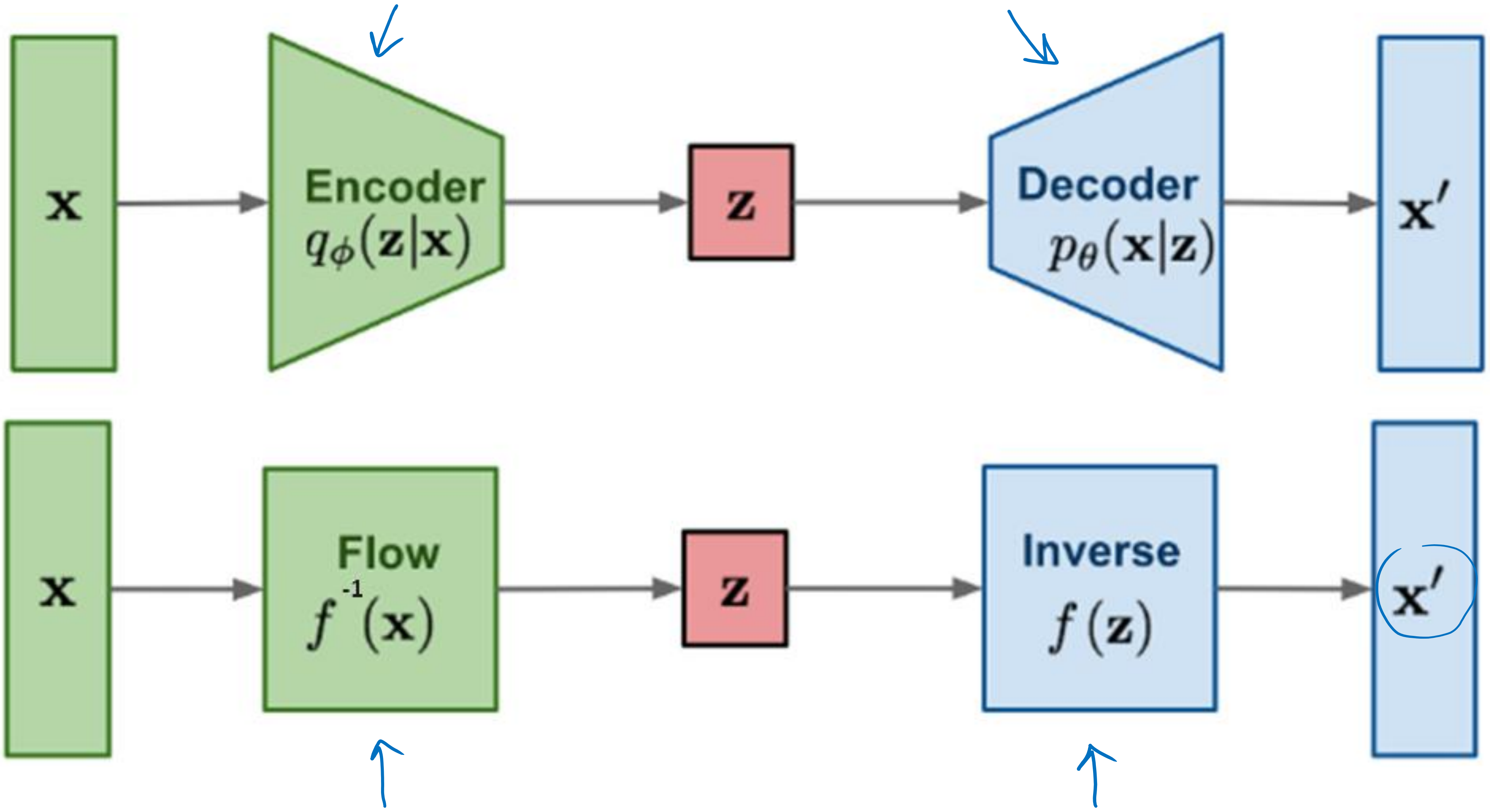
Example: MNIST Dataset



Normalizing Flows vs. VAE

VAE

NF :



x

Marginal Likelihood: VAE vs NF

Decoder

$z \rightarrow \boxed{f_\theta} \rightarrow x$

$x \rightarrow \boxed{f_\theta^{-1}} \rightarrow z$

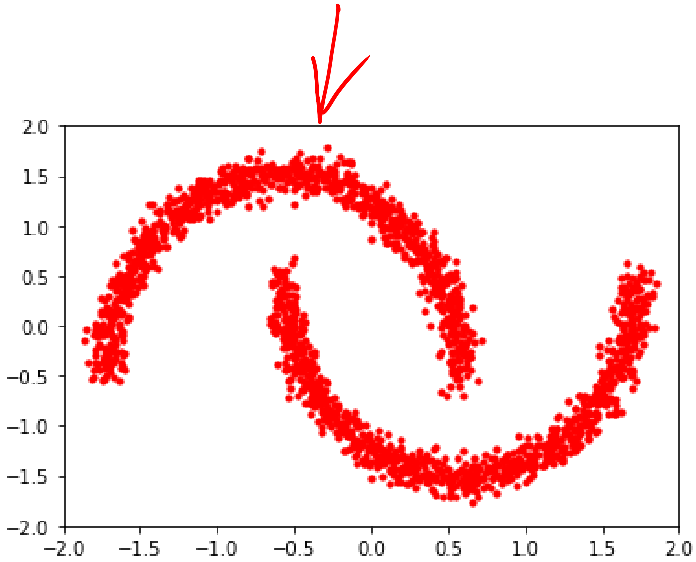
$D = \{x_1, x_2, \dots, x_n\}$

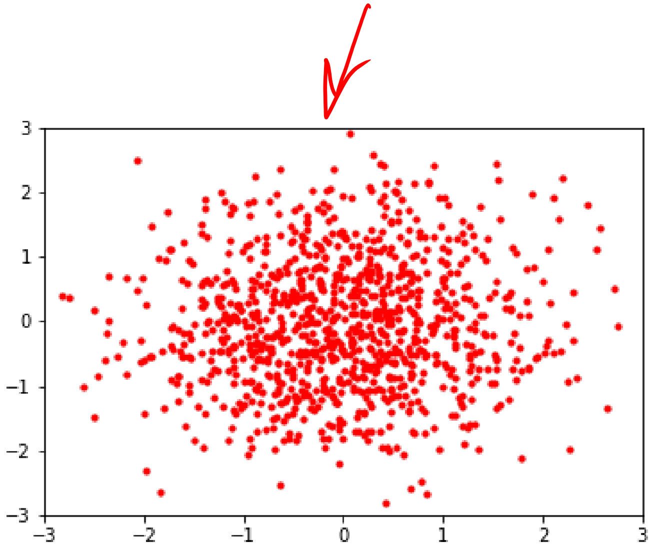
$$\hat{\theta}_{ML} = \arg\max_\theta \log P(D) = \arg\max_\theta \log P(x_1, \dots, x_n)$$

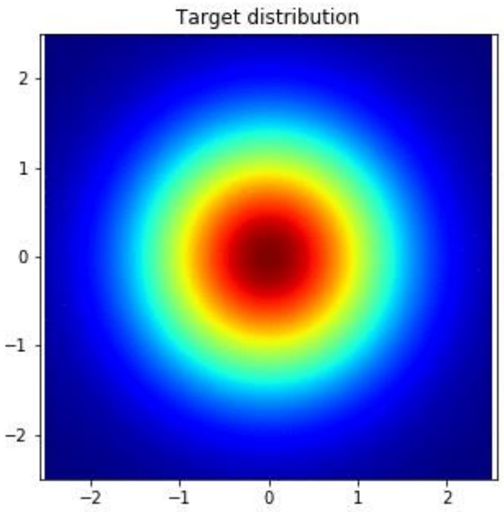
$$= \arg\max_\theta \boxed{\sum_{i=1}^{n} \underline{\log P(x_i)} \quad \checkmark}$$

$\log P(x) \geq ELBO$

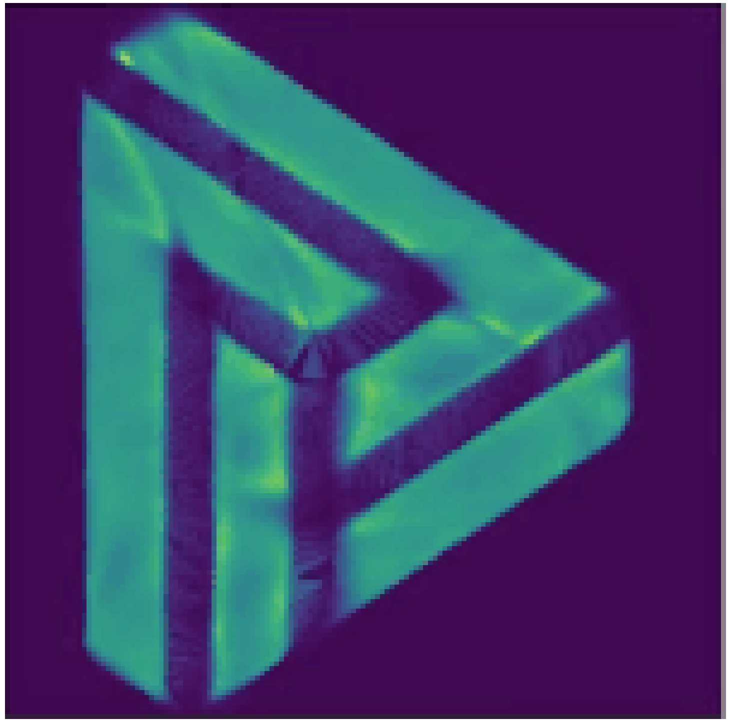
From Normal to Complex Distribution







Normalizing Flow: another Example



Samples from Normalizing Flow

Samples from the Glow model (Source)

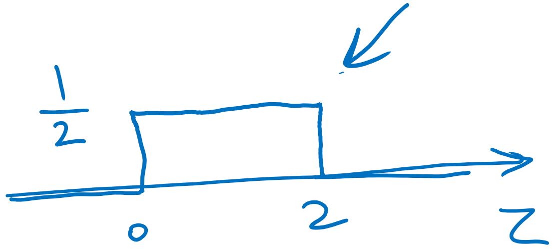
کیفیت ① X

سرعت ② ✓

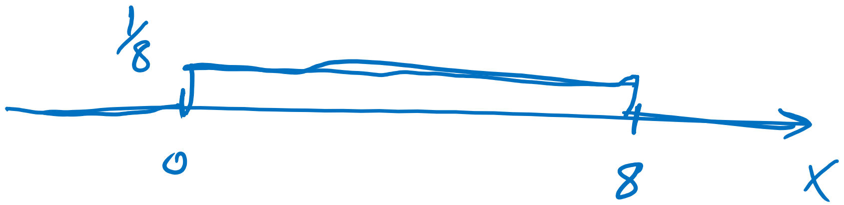
تنوع ③ X

Change of Variable Formula

- Let $Z \sim \mathcal{U}[0,2]$

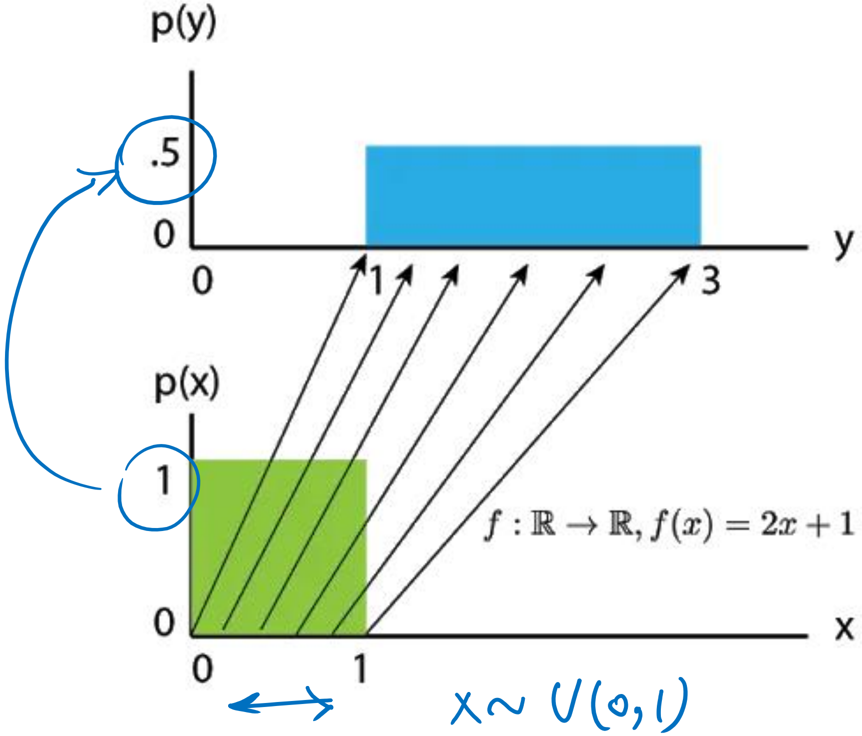


- $\boxed{P_Z(z) = \frac{1}{2}}$
- Let $X = 4Z$ $\Rightarrow$ $0 \leq X \leq 8$



- $P_X(4) = ?$

Change of Variable Formula



$$P_Y(y) = \frac{1}{2} P_X(x) = \frac{1}{2} P_X\left(\frac{y-1}{2}\right)$$

$$\boxed{Y = 2X + 1}$$

$$\boxed{X = \frac{y-1}{2}}$$

Change of variable formula (1-D case)

- If $\boxed{X = f(Z)}$ and $f(.)$ is monotone with inverse $Z = \underbrace{f^{-1}(X) = h(X)}$, then:

$$\boxed{P_X(x) = P_Z\big(h(x)\big)\,|h'(x)|}$$

Change of formula: Example

$$Z \sim \mathcal{U}[0,2]$$

$$X = f(Z) = \exp(Z) = e^{Z}$$

- What is $P_X(x) = ?$

$$P_X(x) = \underbrace{P_Z(f^{-1}(x))} \left| f^{-1\prime}(x) \right|$$

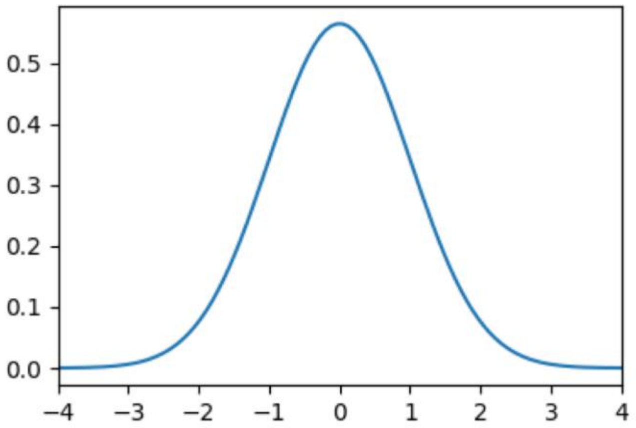
$$Z = f^{-1}(x) = \underbrace{\ln x}$$

$$= \frac{1}{2} \times \left| \frac{1}{x} \right| = \frac{1}{2x}$$

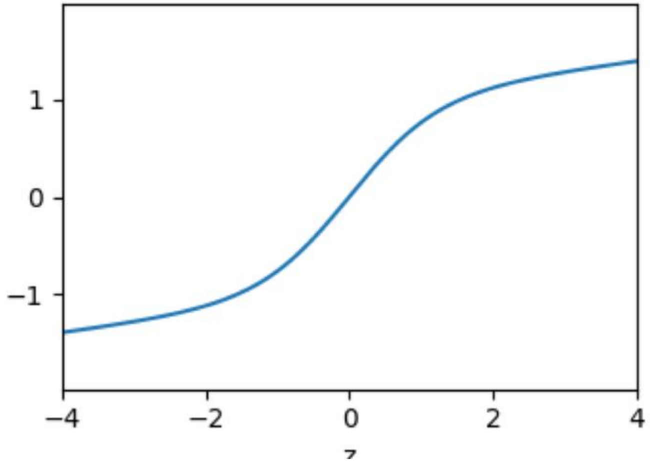
$$1 \le x \le e^2$$

Change of variable: 1-D case

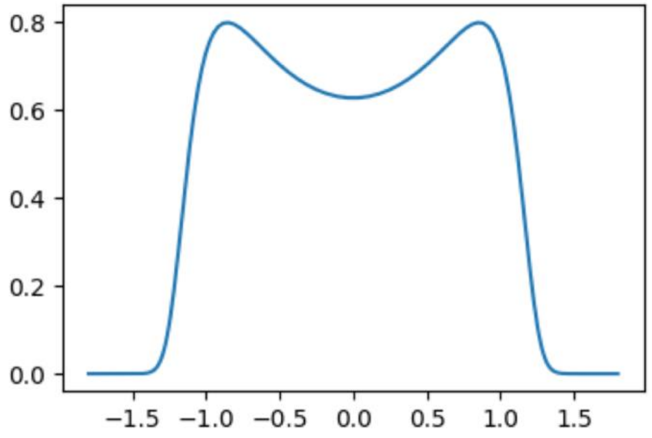
$p_Z(\mathbf{z}) = \mathcal{N}(0,1)$



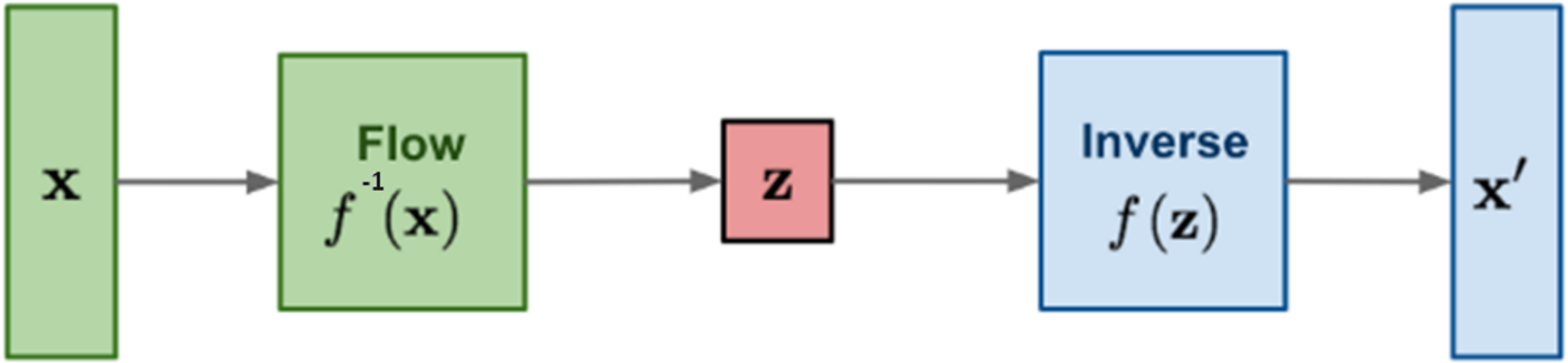
$\mathbf{x} = g(\mathbf{z})$



$p_X(\mathbf{x})$

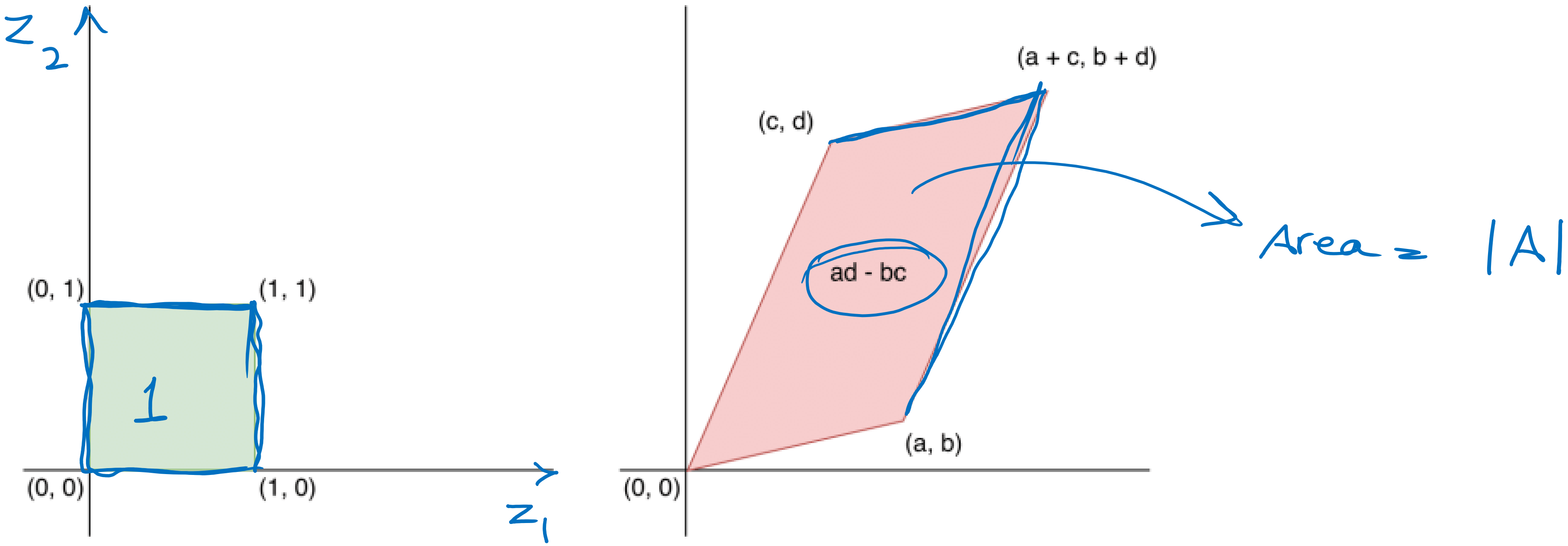


Change of Variable Formula: 1-Dimensional



$$P_X(x) = P_Z\left(f^{-1}(x)\right)\left|\frac{\partial f^{-1}(x)}{\partial x}\right|$$

Change of variable: n-Dimensional



تبدیل خطی

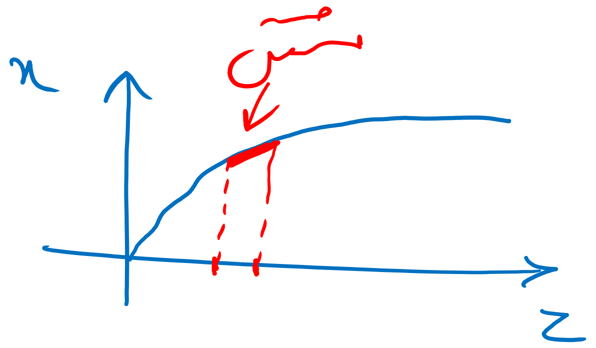
$$\boxed{\boldsymbol{x} = A\boldsymbol{z}}$$

$$\boxed{z = A^{-1} x}$$

$$A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$$

$$\frac{ad - bc}{1}$$

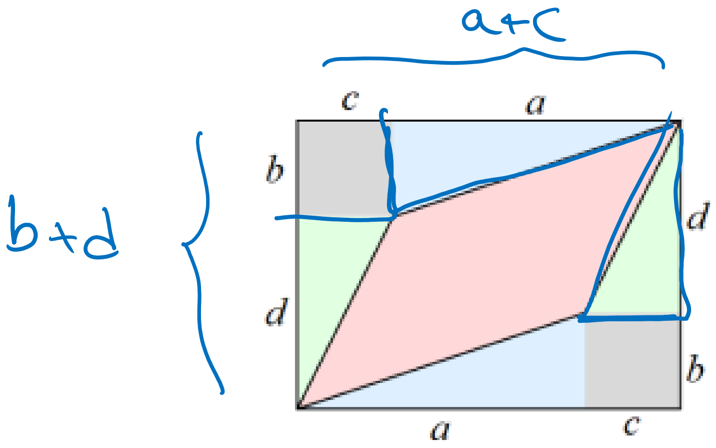
Determinants and Volume



$$\det(A) = \det\begin{pmatrix} a & c \\ b & d \end{pmatrix} = ad - bc$$

$f: \mathbb{R}^n \to \mathbb{R}^n$

$J_{n\times n}$



$(a+c)(b+d) - ab - 2bc - cd = ad - bc$

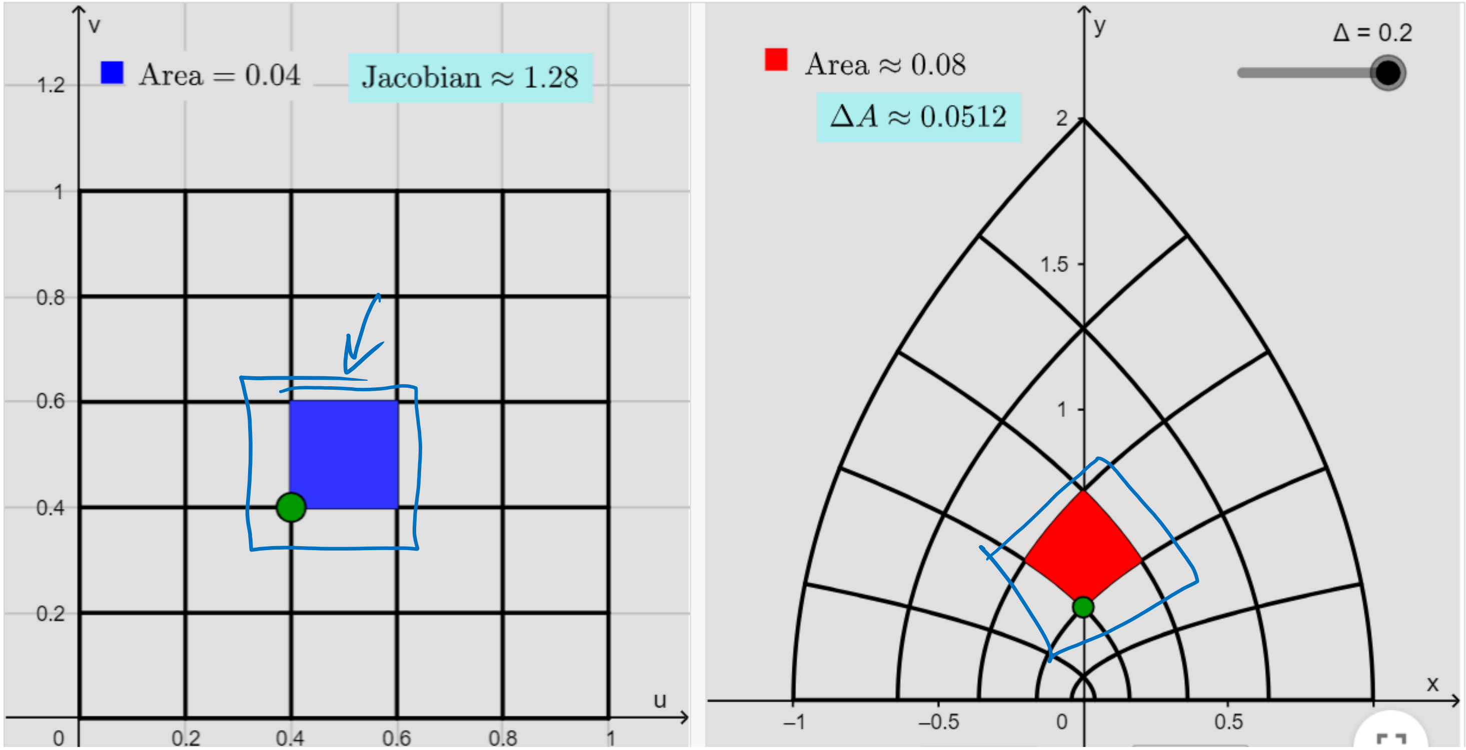
$X = AZ$

$$P_X(x) = \frac{1}{|\det(A)|} P_Z(A^{-1}x)$$

$X = f(z) \longrightarrow X \simeq Az$

Jacobian Determinant

$$\Delta A = |J| \times Area\ of\ square$$



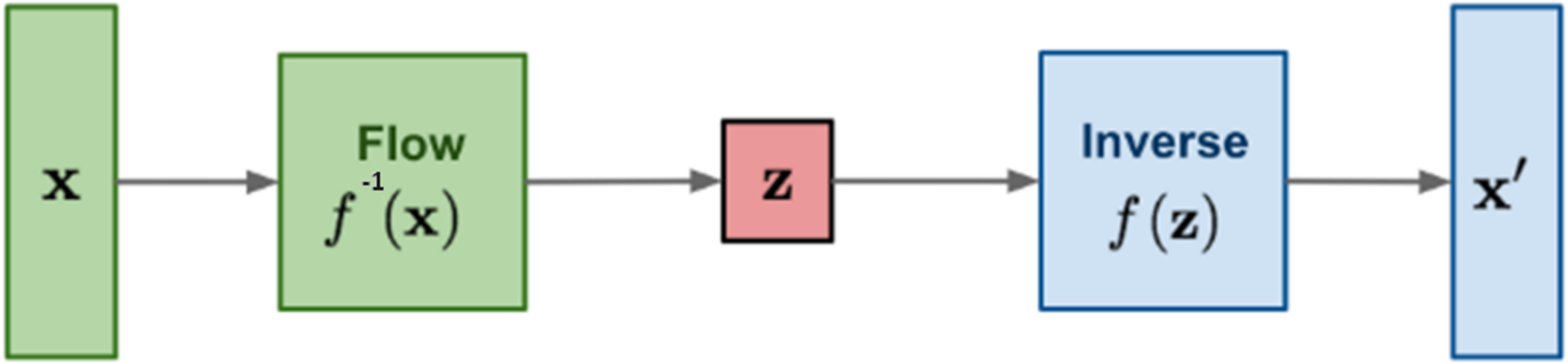
https://www.geogebra.org/m/qM777NYH

Jacobian Matrix

$$\boldsymbol{x} = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} \xrightarrow{f} \boldsymbol{y} = \begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix}$$

$$\frac{\partial(y_1, \ldots, y_n)}{\partial(x_1, \ldots, x_n)} = \begin{bmatrix} \boxed{\dfrac{\partial y_1}{\partial x_1} \quad \cdots \quad \dfrac{\partial y_1}{\partial x_n}} \\ \vdots \qquad \ddots \qquad \vdots \\ \boxed{\dfrac{\partial y_n}{\partial x_1} \quad \cdots \quad \dfrac{\partial y_n}{\partial x_n}} \end{bmatrix}_{n \times n}$$

Change of Variable Formula: n-Dimensional



$$\boxed{P_X(x) = \underbrace{P_Z\left(f_\theta^{-1}(x)\right)}\ \left|\det\left(\underbrace{\frac{\partial f_\theta^{-1}(x)}{\partial x}}_{\text{Jacobian}}\right)\right|}$$

Learning and Inference

$p_x(x) = p_z(f^{-1}(x))\,|J|$

- Learning via **Maximum Likelihood:**

max

$$\max_{\theta} \log p_X(\mathcal{D};\theta) = \sum_{x\in\mathcal{D}} \log p_Z\left(f_\theta^{-1}(x)\right) + \log\left|\det\left(\frac{\partial f_\theta^{-1}(x)}{\partial x}\right)\right|$$

$\mathcal{N}(0, I)$

- **Sampling:**

$$z \sim p_Z(z) \quad x = f_\theta(z)$$

- **Latent Representation:**

$$z = f_\theta^{-1}(x)$$

Calculation of the Determinant

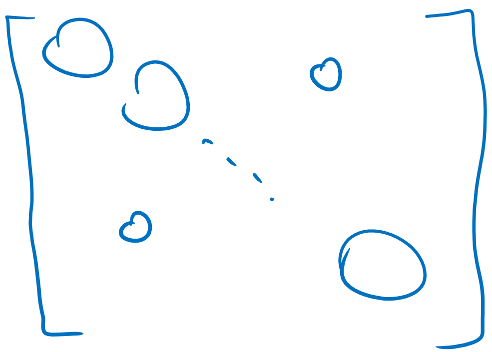
$n = 10000$

- Computing the determinant for an $n \times n$ matrix is $\boldsymbol{O(n^3)}$: prohibitively expensive within a learning loop!

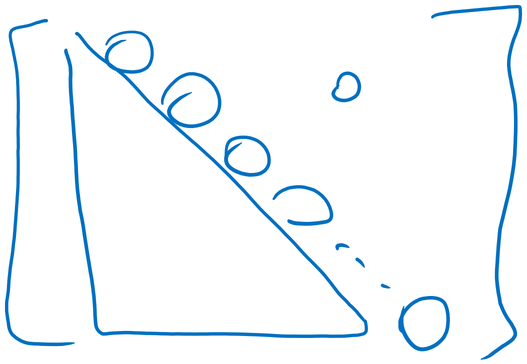
$O(10^{12})$

- **Key idea**: Choose transformations so that the resulting Jacobian matrix has **special structure**. For example, the determinant of a **triangular** matrix is the product of the diagonal entries, i.e., an $\boldsymbol{O(n)}$ operation.

$J =$



$J =$



$x = Az$

$|A|$