Generative Models: Fundamentals and Applications

Lecture 9: Normalizing Flow

Shuigeng Zhou, Yuxi Mi
College of CSAI

December 1, 2025

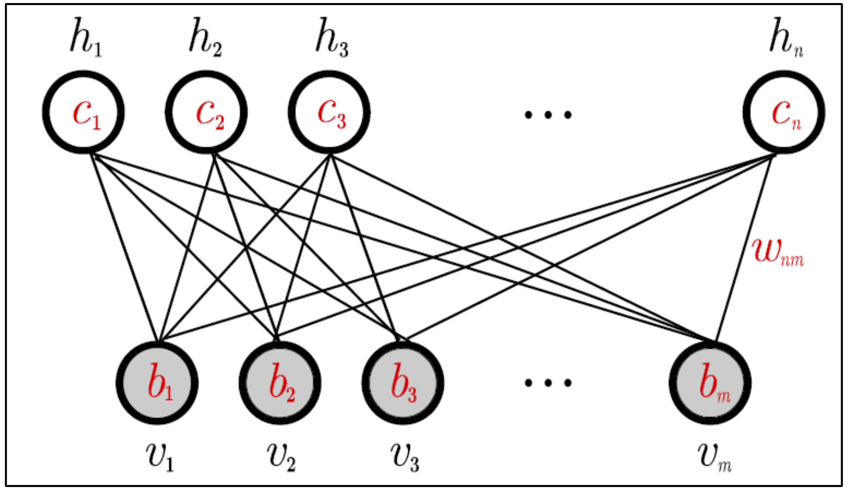
目录

- 背景知识
- 标准化流模型
- 最新进展

回顾

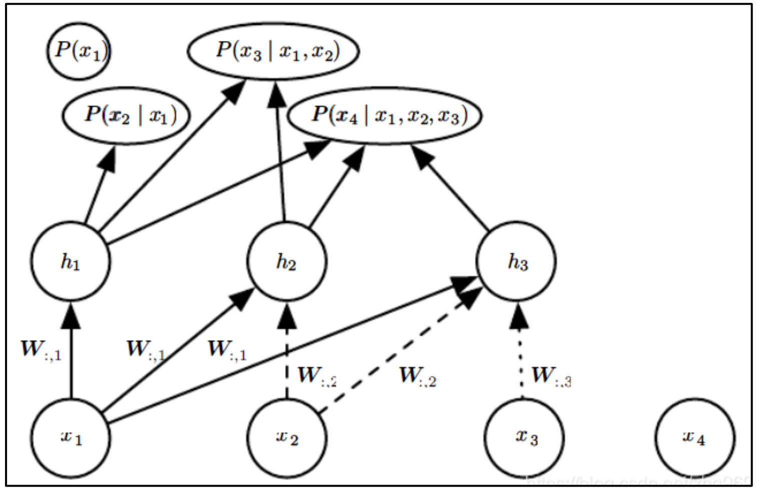
- 有可求解的极大似然函数
- 没有直接学习隐变量特征

能量模型



$$p_\theta(x) = \frac{1}{Z(\theta)} \exp(f_\theta(x))$$

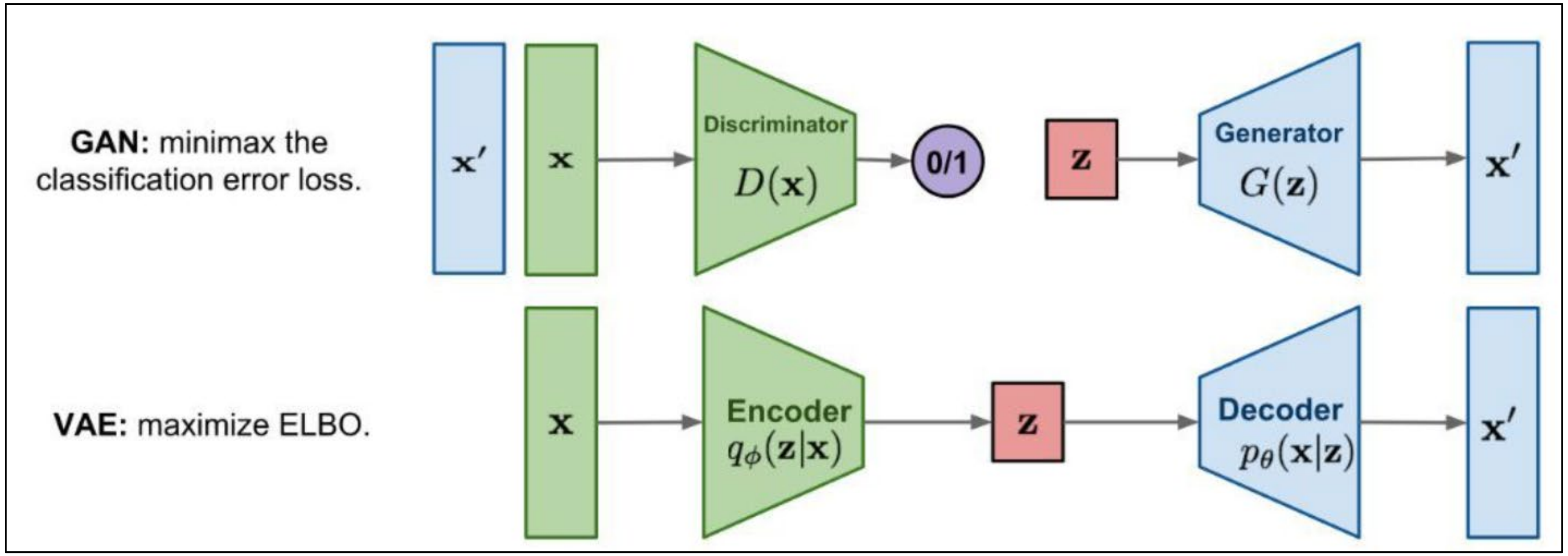
自回归模型



$$p(x|\theta) = \prod_{i=1}^{D} p_\theta(x_i|x_{<i})$$

回顾

- 没有可求解的极大似然函数
- 有隐变量特征表示

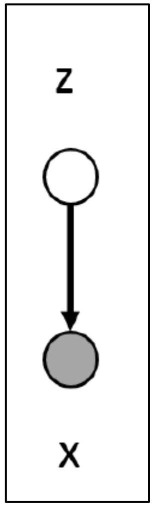


基于流的生成模型

- 特点
  - 有可求解的极大似然函数
  - 有隐变量特征表示
- 隐变量模型

$$\boxed{p(x) = \int_z p(z)p(x|z)\,dz}$$

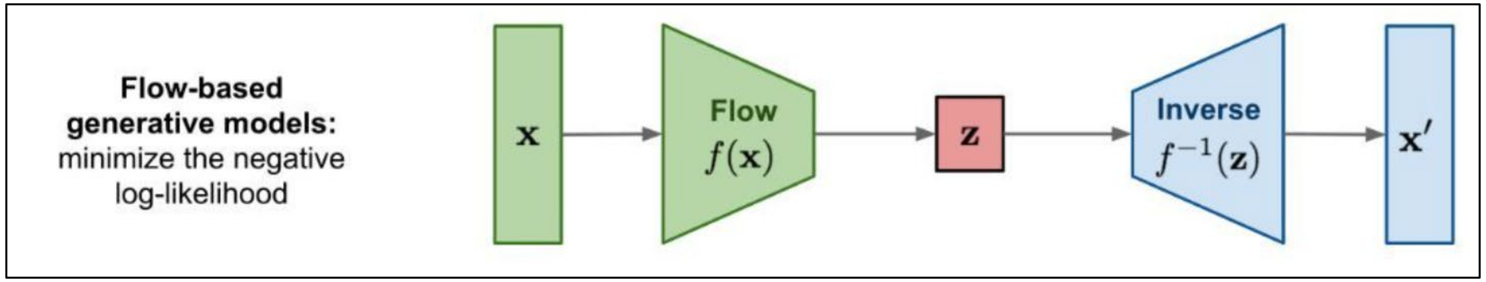
  - 简单分布 $p(z)$ 映射到复杂分布 $p(x)$

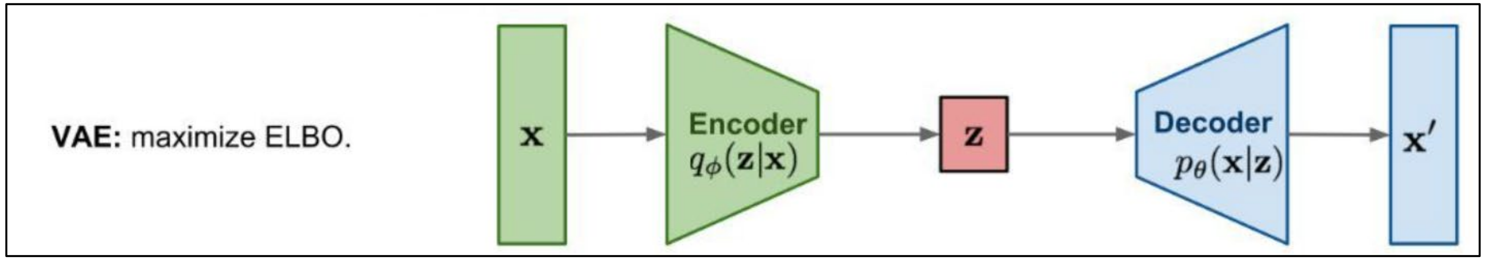


基于流的生成模型

- 模型结构：可逆的确定性变换

$$\boxed{f^{-1}(z) = f^{-1}\big(f(x)\big) = x}$$





变量变换（change of variable）

- 实例
  - 设 $Z \sim \text{Uniform}(0,1)$，$Z$ 的密度函数为 p_Z，求 $p_Z(0.5)$?
    - $Z = 0.5$ 时，$p_Z(0.5) = 1$
  - 令 $X = 2Z + 1$，X 的密度函数为 p_X，求 $p_X(2)$?
    - 直觉上：$X \sim \text{Uniform}(1,3)$，所以 $p_X(2) = 1/2$
    - 变量变换：

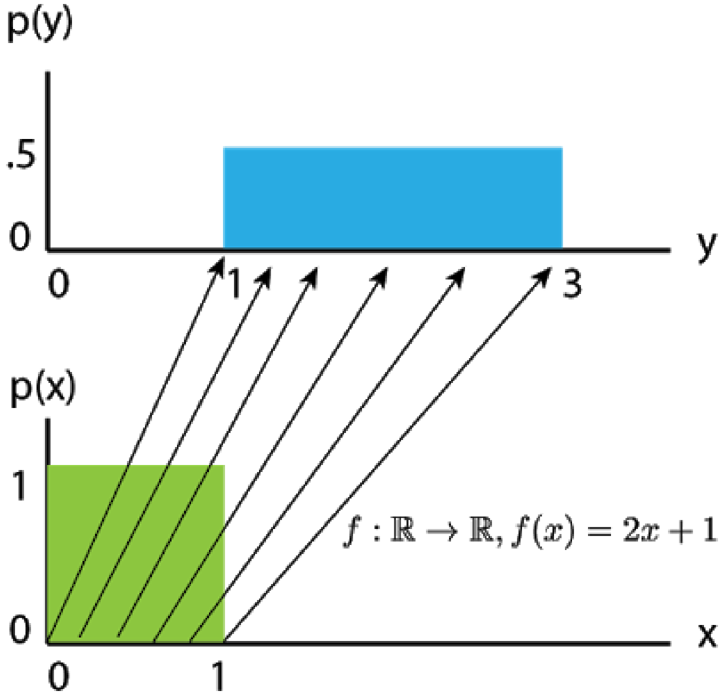
$$p_X(2) = \frac{dP_X(2)}{dX} = \frac{dP\,(X \le 2)}{dX} = \frac{dP\,(2Z+1 \le 2)}{dX} = \frac{dP\,(Z \le 0.5)}{dZ}\frac{dZ}{dX}$$

$$= \frac{dP_Z(0.5)}{dZ}\frac{dZ}{dX} = p_Z(0.5)\frac{dZ}{dX} = 1 \cdot \frac{1}{2} = \frac{1}{2}$$

变量变换（change of variable）

- 实例
  - 设 $X \sim \text{Uniform}(0,1)$，$X$ 的密度函数为 p_X
  - 令 $Y = 2X + 1$，Y 的密度函数 p_Y

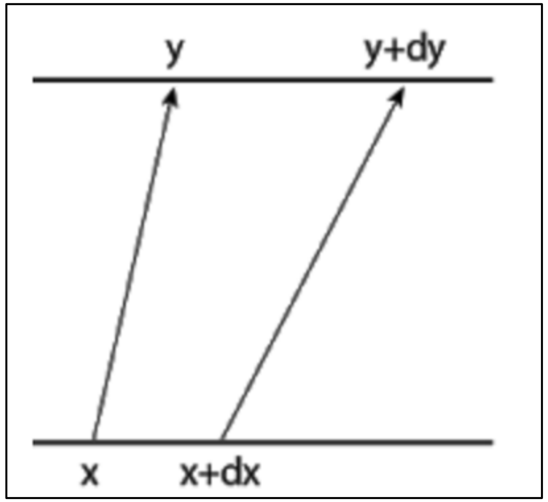
$$\boxed{\begin{aligned} p_Y(y) &= \frac{dP_Y(y)}{dY} = \frac{dP\,(2X+1\leq y)}{dX}\frac{dX}{dY} \\ &= p_X(x)\frac{dX}{dY} \end{aligned}}$$

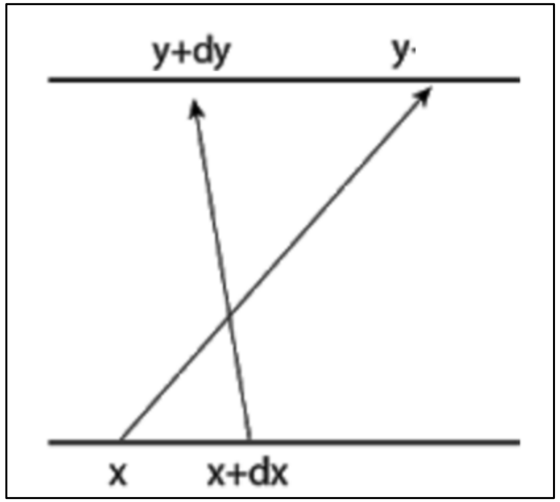


随机变量的变换

- 一般情形
  - 若 X 是连续型随机变量（使用变量变换公式）

$$\boxed{p_y(y) = p_x(x)\left|\frac{dx}{dy}\right|}$$





变量变换

- 一维变量
  - 给定单调可逆函数f: $x = f(z)$，$z = f^{-1}(x) \triangleq h(x)$

$$\boxed{p_X(x) = p_Z(h(x))|h'(x)|}$$

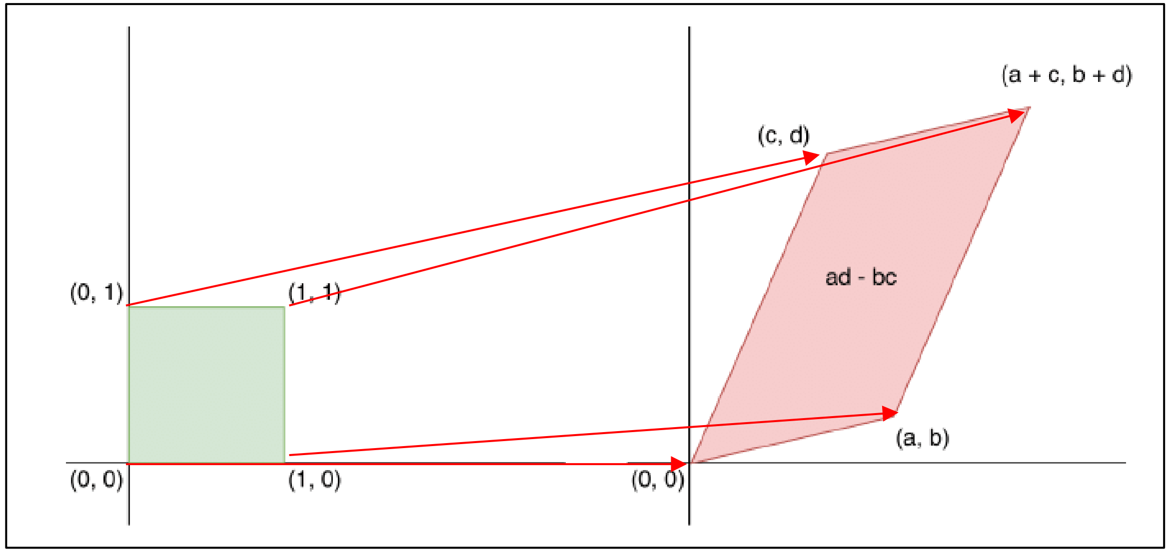
  - 回顾前例： 设 Z ~ Uniform(0,1), 令 $X = 2Z + 1$, $p_X(2)$?

$$\boxed{z = h(x) = \frac{1}{2}x - \frac{1}{2}}$$

$$\boxed{p_X(2) = p_Z(h(2))|h'(2)| = p_Z(0.5) \times \left|\frac{1}{2}\right| = 1 \times \frac{1}{2} = \frac{1}{2}}$$

变量变换

- 多维变量
  - 设 Z 在区间 $[0,1]^n$ 上均匀分布
  - 令 $X = AZ$，其中 A 是可逆方阵，记$W = A^{-1}$，求 X 的分布
    - 令$A = \begin{bmatrix} a & c \\ b & d \end{bmatrix}$
    - 原四边形的容量为 1
    - 新四边形的“容量”为 $|\det A| = |ad - bc|$



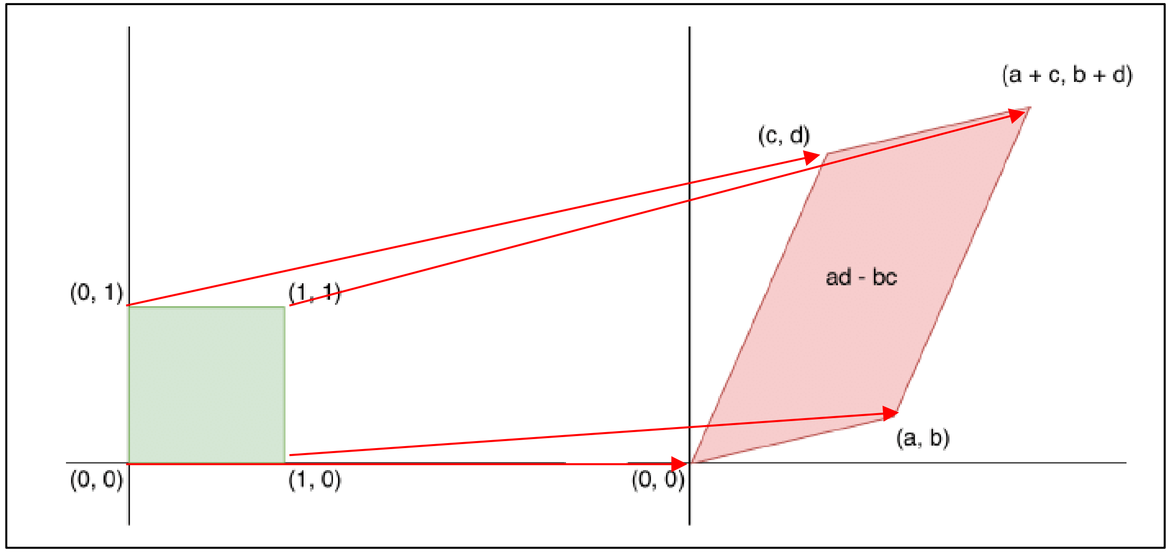
变量变换

- 设 Z 在区间 $[0,1]^n$ 上均匀分布
  - X 的“容量变化”（缩放因子）

$$\boxed{\frac{1}{|\det A|} = |\det W|}$$

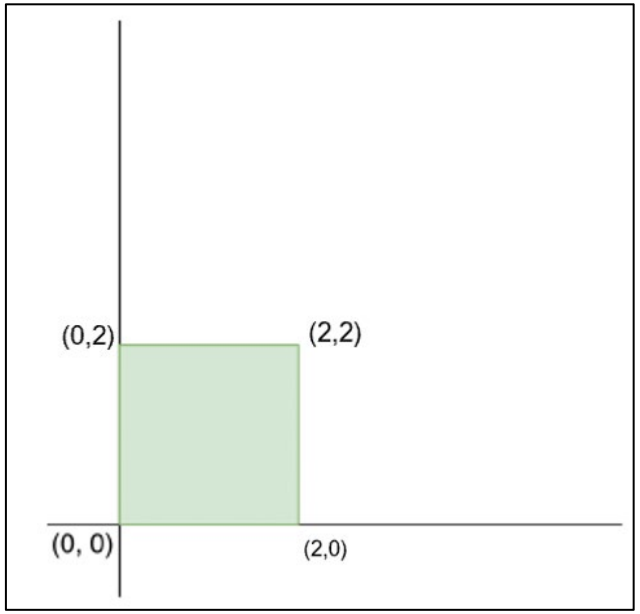
  - X 的分布

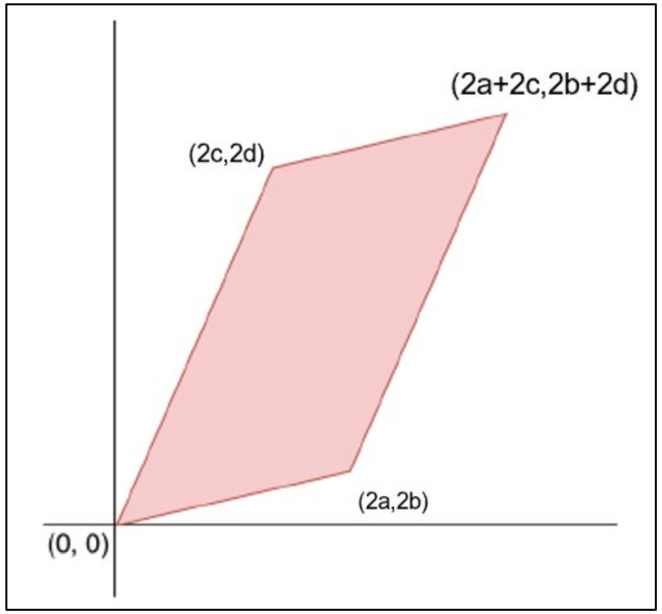
$$\boxed{\begin{aligned} p_X(\mathbf{x}) &= p_Z(W\mathbf{x})\,|\det(W)| \\ &= p_Z(W\mathbf{x})\,/\,|\det(A)| \end{aligned}}$$



变量变换

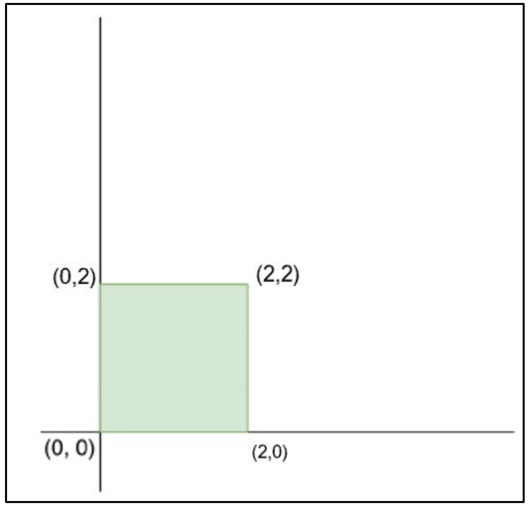
- 设 Z 在区间 $[0,2]^n$上均匀分布
  - 原四边形的容量：4
  - 新四边形的容量：$4(ad-bc)=4|\det A|$

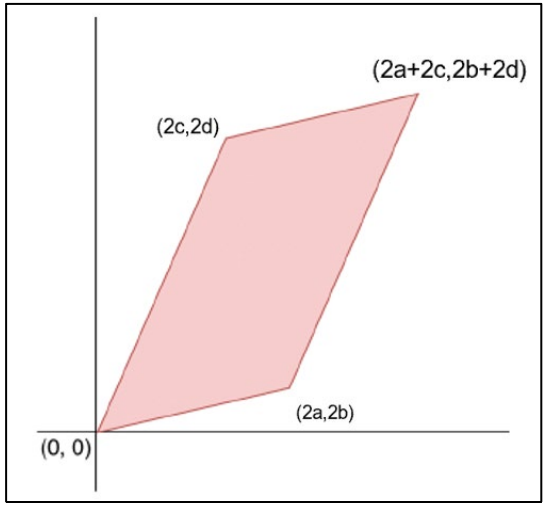




变量变换

- 设 Z 在区间 $[0,2]^n$ 上均匀分布
  - 容量变化仍为 $\boxed{\frac{1}{|\det A|} = |\det W|}$
  - 结论：给定线性变换 $X = AZ$，容量变化由 $|\det A|$ 唯一确定





变量变换

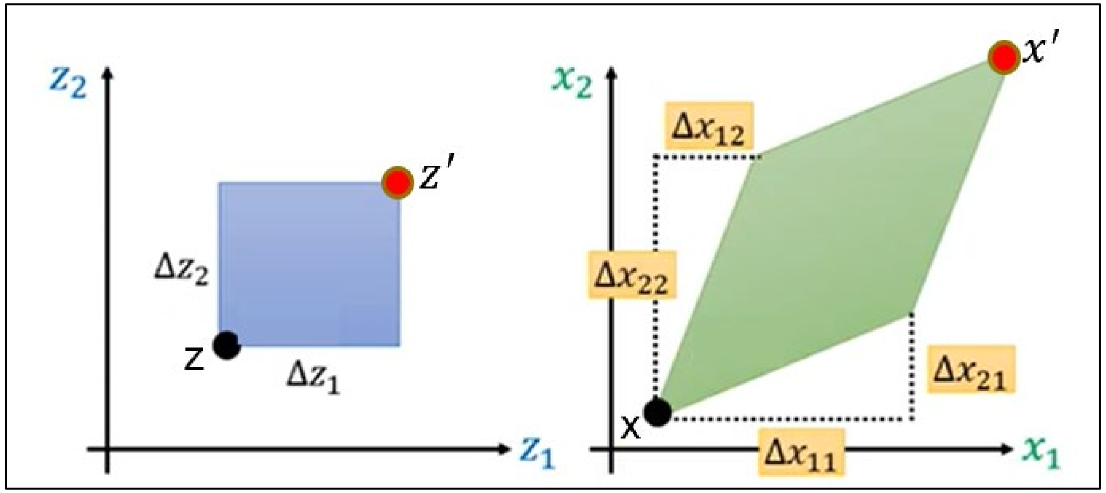
- 问题：给定非线性变换 $X = f(Z)$，容量变换由什么确定?
  - 函数 f 的雅可比行列式
    - 设$Z \in R^n, X \in R^m, X = f(Z)$
    - f 在某一 Z 处的偏导数 $\frac{\partial f}{\partial Z}$ 可写成雅可比矩阵：

$$\boxed{\frac{\partial f}{\partial Z} = \frac{\partial X}{\partial Z} = \begin{bmatrix} \frac{\partial X_1}{\partial Z_1} & \cdots & \frac{\partial X_1}{\partial Z_n} \\ \vdots & \ddots & \vdots \\ \frac{\partial X_m}{\partial Z_1} & \cdots & \frac{\partial X_m}{\partial Z_n} \end{bmatrix}}$$

    - 如果 $\frac{\partial f}{\partial Z}$ 方阵（$m = n$），称其行列式为雅可比行列式

变量变换

- 实例：$X, Z \in R^2$
  - 记 $z = (z_1, z_2)^T$， $x = f(z) = (x_1, x_2)^T$
  - 考虑充分小邻域下的变换：
    - $z' = (z_1 + \Delta z_1, z_2 + \Delta z_2)^T$
    - $x' = f(z')$
  - 充分小邻域下：
    - 非线性变换可用线性变换近似
    - 此时缩放比例由雅可比行列式决定



变量变换

- 雅克比矩阵

$$J_f = \begin{bmatrix} \partial x_1/\partial z_1 & \partial x_1/\partial z_2 \\ \partial x_2/\partial z_1 & \partial x_2/\partial z_2 \end{bmatrix}$$

- 雅克比逆矩阵

$$J_{f^{-1}} = \begin{bmatrix} \partial z_1/\partial x_1 & \partial z_1/\partial x_2 \\ \partial z_2/\partial x_1 & \partial z_2/\partial x_2 \end{bmatrix}$$

  - 如果x $= f(z) = (z_1 + z_2, 2z_1)^T$，则$J_f, J_{f^{-1}}$？

$$\Rightarrow \quad J_f = \begin{bmatrix} 1 & 1 \\ 2 & 0 \end{bmatrix}$$

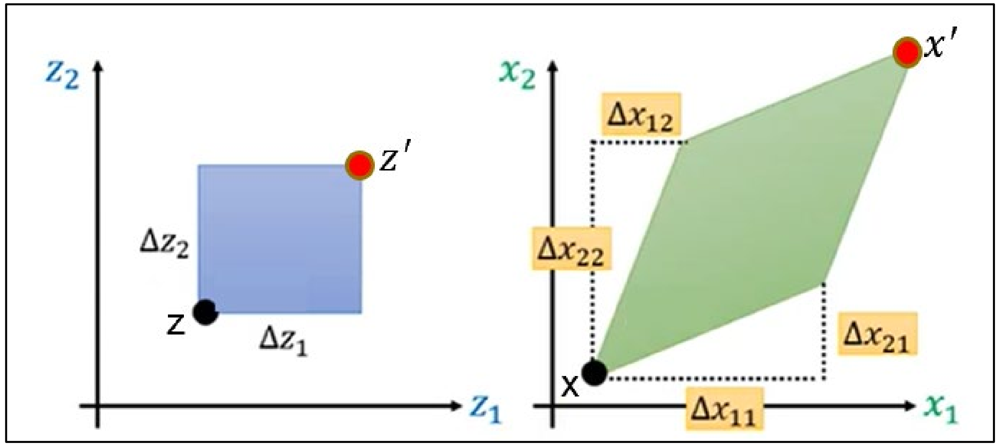
$$z = \begin{bmatrix} z_1 \\ z_2 \end{bmatrix} = f^{-1}\left(\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}\right) = \begin{bmatrix} \frac{x_2}{2} \\ x_1 - \frac{x_2}{2} \end{bmatrix} \Rightarrow J_{f^{-1}} = \begin{bmatrix} 0 & \frac{1}{2} \\ 1 & -\frac{1}{2} \end{bmatrix}$$

变量变换

- 容量变换

$$p(x')\left|det\begin{bmatrix}\Delta x_{11} & \Delta x_{21}\\ \Delta x_{12} & \Delta x_{22}\end{bmatrix}\right| = p(z')\Delta z_1 \Delta z_2$$

$$\begin{aligned}
p(z') &= p(x')\left|\frac{1}{\Delta z_1 \Delta z_2} det\begin{bmatrix}\Delta x_{11} & \Delta x_{21}\\ \Delta x_{12} & \Delta x_{22}\end{bmatrix}\right| \\
&= p(x')\left|det\begin{bmatrix}\Delta x_{11}/\Delta z_1 & \Delta x_{21}/\Delta z_1\\ \Delta x_{12}/\Delta z_2 & \Delta x_{22}/\Delta z_2\end{bmatrix}\right| \\
&= p(x')\left|det\begin{bmatrix}\partial x_1/\partial z_1 & \partial x_2/\partial z_1\\ \partial x_1/\partial z_2 & \partial x_2/\partial z_2\end{bmatrix}\right| \\
&= p(x')\left|det\begin{bmatrix}\partial x_1/\partial z_1 & \partial x_1/\partial z_2\\ \partial x_2/\partial z_1 & \partial x_2/\partial z_2\end{bmatrix}\right|
\end{aligned}$$



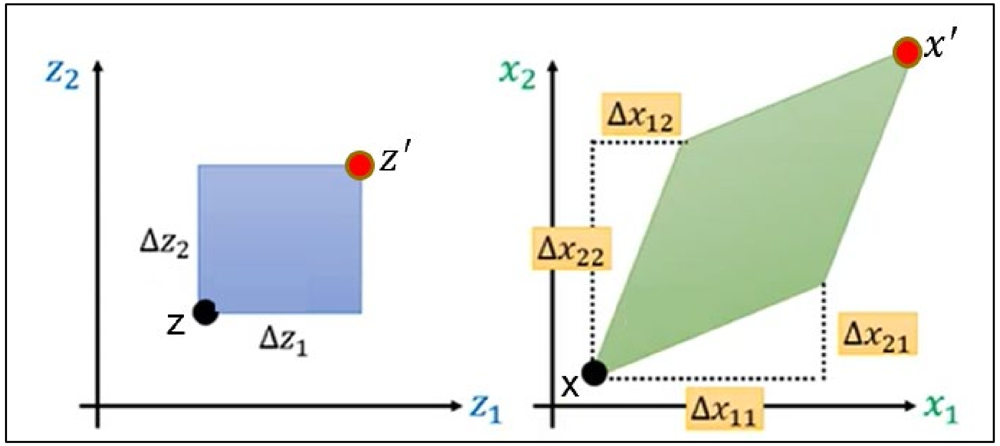
变量变换

- 容量变换

$$p(z') = p(x')\left|det(J_f)\right|$$

$$\Rightarrow \quad p(x') = p(z')\left|\frac{1}{det(J_f)}\right|$$

$$\Rightarrow \quad p(x') = p(z')\left|det(J_{f^{-1}})\right|$$



变量变换

- 变量变换定理
  - 给定Z和X之间的映射关系$f: R^n \longmapsto R^n$可逆，满足$X = f(Z)$以及$Z = f^{-1}(X)$，则有

$$p_X(\mathbf{x}) = p_Z\left(\mathbf{f}^{-1}(\mathbf{x})\right)\left|\det\left(\frac{\partial \mathbf{f}^{-1}(\mathbf{x})}{\partial \mathbf{x}}\right)\right|$$

- 注意
  - X和Z都是连续随机变量，且维度相等
  - Fact：对任意可逆矩阵A，$\det(A^{-1}) = \det(A)^{-1}$

$$p_X(\mathbf{x}) = p_Z(\mathbf{z})\left|\det\left(\frac{\partial \mathbf{f}(\mathbf{z})}{\partial \mathbf{z}}\right)\right|^{-1}$$

目录

- 背景知识
- 标准化流模型
- 最新进展

标准化流模型

- 概率图
  - 推断模型

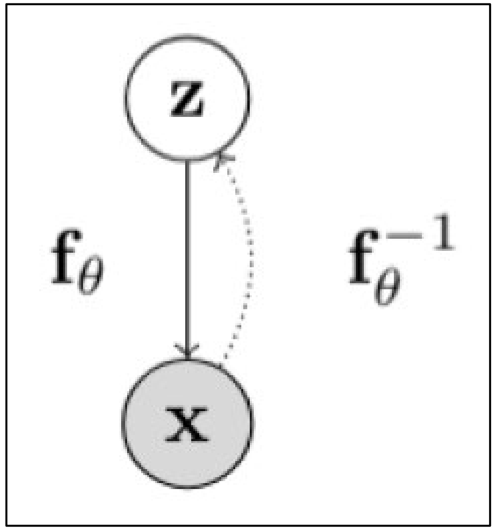
$$\boxed{Z = \mathbf{f}_\theta^{-1}(X)}$$

  - 生成模型

$$\boxed{X = \mathbf{f}_\theta(Z)}$$

    - 注意：X和Z的维度相等
    - 似然函数

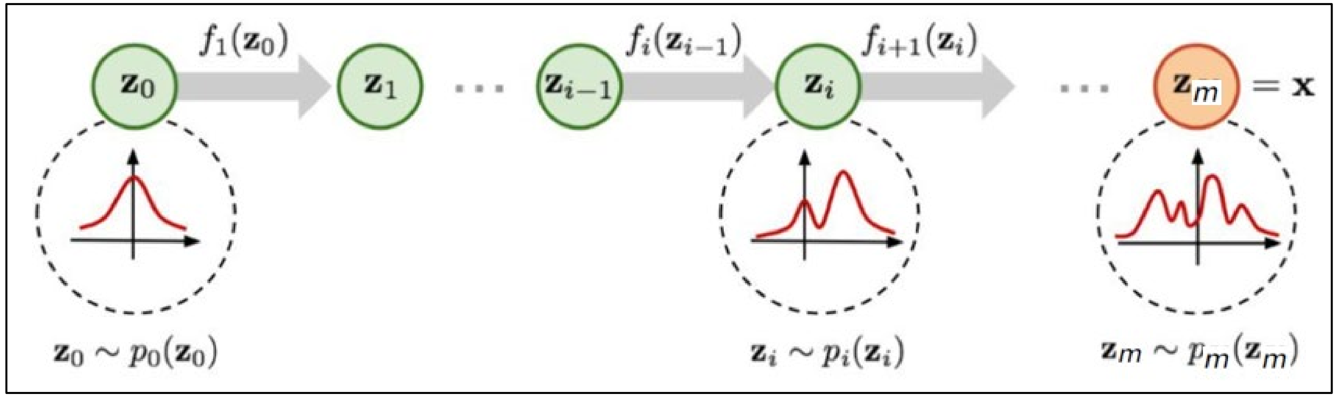
$$\boxed{p_X(\mathbf{x};\theta) = p_Z\left(\mathbf{f}_\theta^{-1}(\mathbf{x})\right)\left|\det\left(\frac{\partial \mathbf{f}_\theta^{-1}(\mathbf{x})}{\partial \mathbf{x}}\right)\right|}$$



标准化流模型

- 标准化（normalizing）
  - 通过变量变换以后的变量积分为1，满足概率分布函数的定义
- 流（flow）

$$\boxed{\mathbf{z}_m := \mathbf{f}_\theta^m \circ \cdots \circ \mathbf{f}_\theta^1(\mathbf{z}_0) = \mathbf{f}_\theta^m(\mathbf{f}_\theta^{m-1}(\cdots(\mathbf{f}_\theta^1(\mathbf{z}_0)))) \triangleq \mathbf{f}_\theta(\mathbf{z}_0)}$$

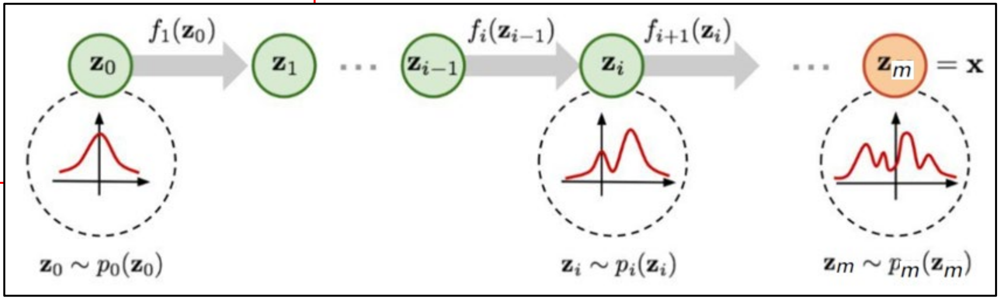


标准化流模型

- 似然函数

$$
\begin{aligned}
p_X(x;\theta) &= p_Z(z_{m-1})\left|det\left(\frac{\partial\left(f_\theta^m\right)^{-1}(z_m)}{\partial z_m}\right)\right| \\
&= p_Z(z_{m-2})\left|det\left(\frac{\partial\left(f_\theta^{m-1}\right)^{-1}(z_{m-1})}{\partial z_{m-1}}\right)\right|\left|det\left(\frac{\partial\left(f_\theta^m\right)^{-1}(z_m)}{\partial z_m}\right)\right| \\
&\vdots \\
&= p_Z(z_0)\prod_{i=1}^m\left|det\left(\frac{\partial\left(f_\theta^i\right)^{-1}(z_i)}{\partial z_i}\right)\right| \\
&= p_Z\left(f_\theta^{-1}(x)\right)\left|det\left(\frac{\partial f_\theta^{-1}(x)}{\partial x}\right)\right|
\end{aligned}
$$

$$
\log p_X(x;\theta) = \log p_Z(z_0) + \sum_{i=1}^m \log\left|\det J_{(f_\theta^i)^{-1}}(z_i)\right|
$$



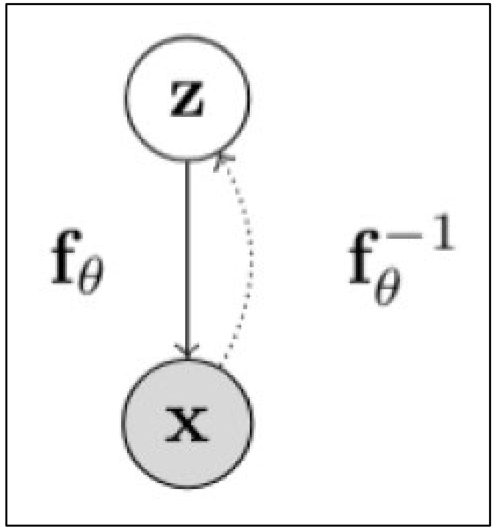
标准化流模型

- 极大似然估计

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

- 精确似然
  - 逆变换 $f^{-1}: x \mapsto z$
  - 变量变换公式
- 采样生成新数据
  - 变换$f: z \mapsto x$

$$\boxed{\mathbf{z} \sim p_Z(\mathbf{z}) \quad \mathbf{x} = \mathbf{f}_\theta(\mathbf{z})}$$



标准化流模型

- 先验分布 $p_Z(z)$
  - 易于采样和计算似然函数，如各向同性高斯
- 可逆变换 f
  - 容易计算
    - 似然函数计算逆变换$f^{-1}: x \mapsto z$
    - 采样生成数据$f: z \mapsto x$
- 雅克比行列式计算
  - 时间复杂度为$O(D^3)$
  - 选择特殊结构的雅克比矩阵，如下三角矩阵

标准化流模型

- 平面流（Planar flow）

数学描述：A 可逆矩阵，u, v 列向量。则有：

$$\det(\mathrm{A}+\mathrm{u}\mathrm{v}^{\mathrm{T}})=\det(\mathrm{A})(1+\mathrm{v}^{\mathrm{T}}\mathrm{A}^{-1}\mathrm{u})$$

  - 选择可逆变换

$$\boxed{\mathbf{x}=\mathbf{f}_\theta(\mathbf{z})=\mathbf{z}+\mathbf{u}h(\mathbf{w}^T\mathbf{z}+b)}$$

  - 雅克比行列式
    - 记 $\psi(\mathbf{z})=h'(\mathbf{w}^\top\mathbf{z}+b)\mathbf{w}$

$$\boxed{\begin{aligned}\left|\det\frac{\partial\mathbf{f}_\theta(\mathbf{z})}{\partial\mathbf{z}}\right| &= \left|\det(I+h'(\mathbf{w}^T\mathbf{z}+b)\mathbf{u}\mathbf{w}^T)\right| \\ &= \left|1+h'(\mathbf{w}^T\mathbf{z}+b)\mathbf{u}^T\mathbf{w}\right| \\ &= \left|1+\mathbf{u}^\top\psi(\mathbf{z})\right|\end{aligned}}$$

标准化流模型

- 平面流（Planar flow）
  - 流的长度为K

$$\boxed{\mathbf{z}_K = f_K \circ f_{K-1} \circ \ldots \circ f_1(\mathbf{z})}$$

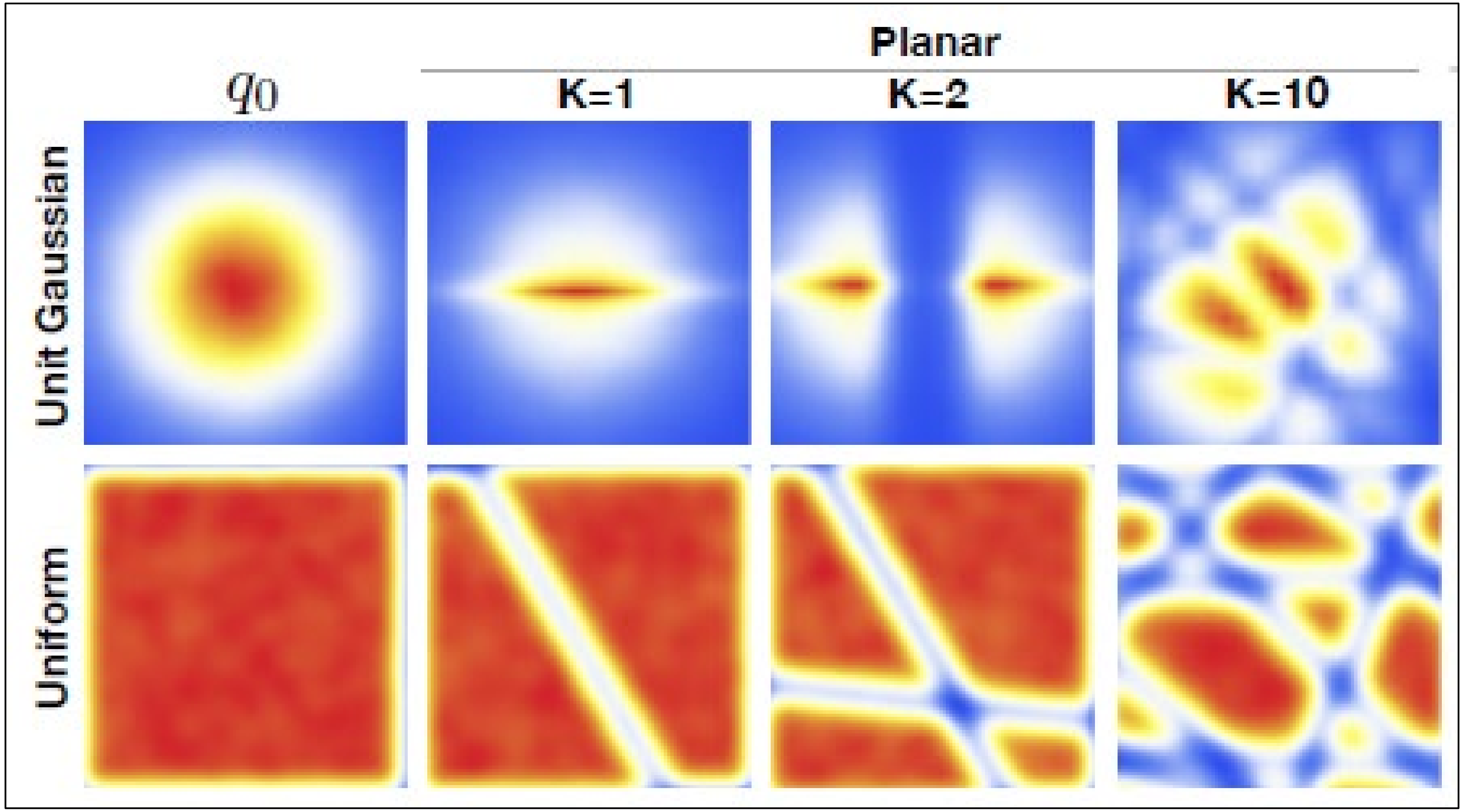
  - 似然函数

$$\boxed{p_X(x;\theta) = q_K(z_K) = q_0(z) \prod_{i=1}^{K} \left| det\left(\frac{\partial \left(f_\theta^i\right)^{-1}(z_i)}{\partial z_i} \right) \right|}$$

  - 对数似然函数

$$\boxed{\ln q_K(\mathbf{z}_K) = \ln q_0(\mathbf{z}) - \sum_{k=1}^{K} \ln \left|1 + \mathbf{u}_k^\top \psi_k(\mathbf{z}_k)\right|}$$

标准化流模型



目录

- 背景知识
- 标准化流模型
- 最新进展
  - Nonlinear Independent Components Estimation (NICE)
  - Real-valued Non-volume Preserving (Real NVP)
  - Glow
  - Masked Autoregressive Flow (MAF)
  - Inverse Autoregressive Flow (IAF)

NICE

- 雅克比行列式计算
  - 时间复杂度为$O(D^3)$
  - 选择特殊结构的雅克比矩阵，如下三角矩阵
- NICE：将f的雅克比矩阵变换为三角矩阵
  - 加性耦合层（additive coupling layer）
    - 将输入的一部分维度保持不变，而另一部分维度进行平移变换
  - 尺度变换层（rescaling layer）

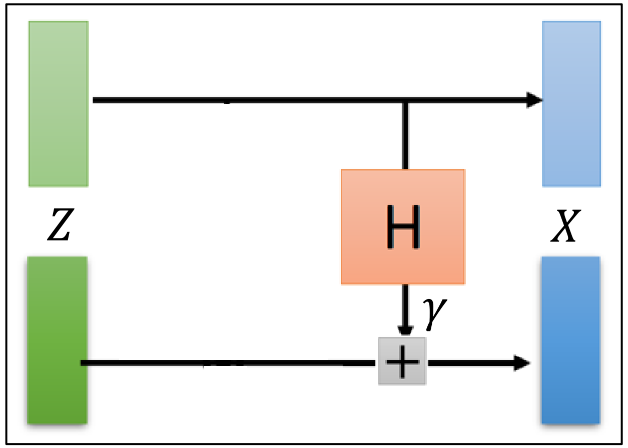
NICE

- 将隐变量Z分成两个不相交部分$z_{1:d}$和$z_{d+1:D}$
- 变换$f: Z \mapsto X$

  $\boxed{x_{i\le d} = z_i}$ $\boxed{x_{i>d} = z_i + \gamma_i}$

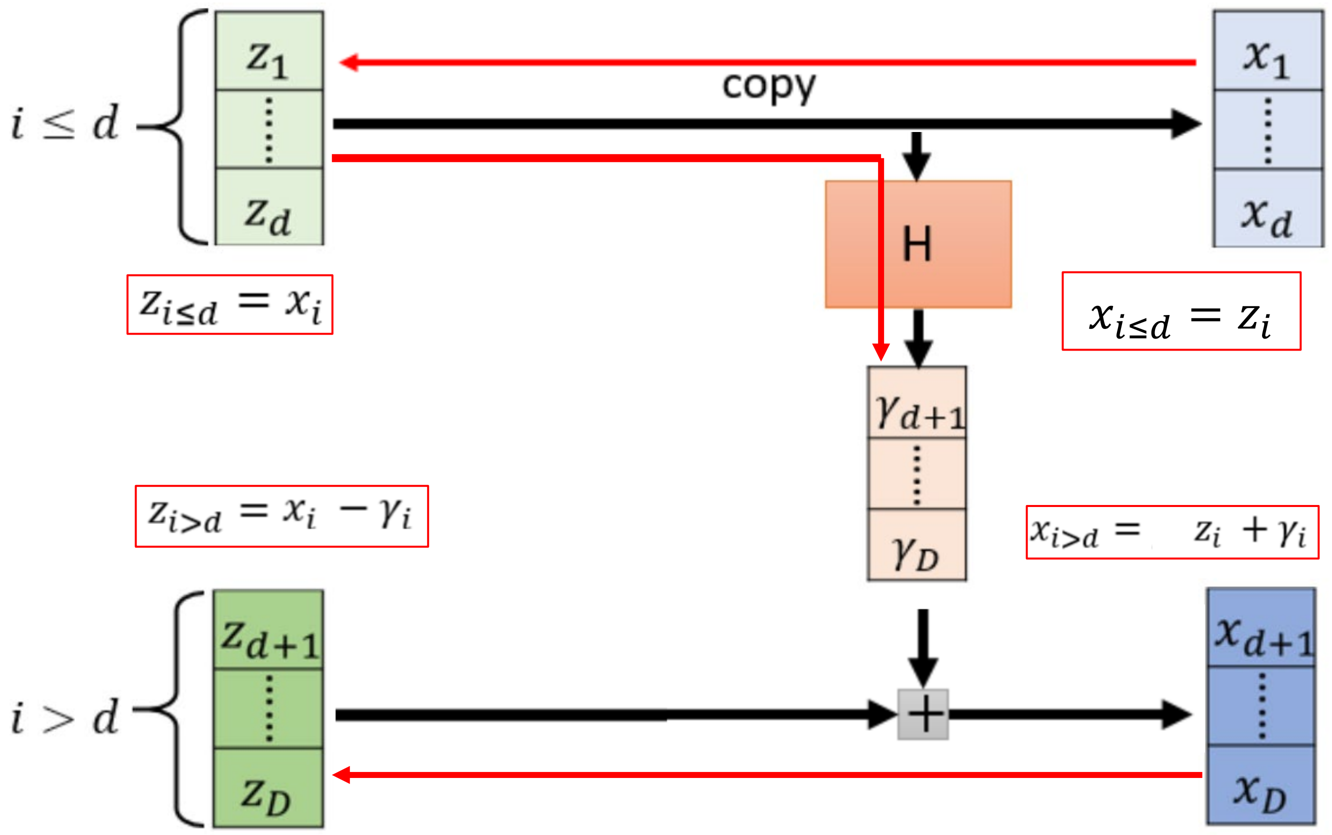
- 逆变换$f^{-1}: X \mapsto Z$

  $\boxed{z_{i\le d} = x_i}$ $\boxed{z_{i>d} = x_i - \gamma_i}$



- H可以是任意复杂函数，不需要可逆

NICE


$i \leq d$
z_1
z_d
copy
x_1
x_d
H
$z_{i \leq d} = x_i$
$x_{i \leq d} = z_i$
γ_{d+1}
γ_D
$z_{i>d} = x_i - \gamma_i$
$x_{i>d} = z_i + \gamma_i$
$i > d$
z_{d+1}
z_D
x_{d+1}
x_D

NICE

- 映射$x = f(z)$

$$\boxed{x_{i\le d} = z_i}$$

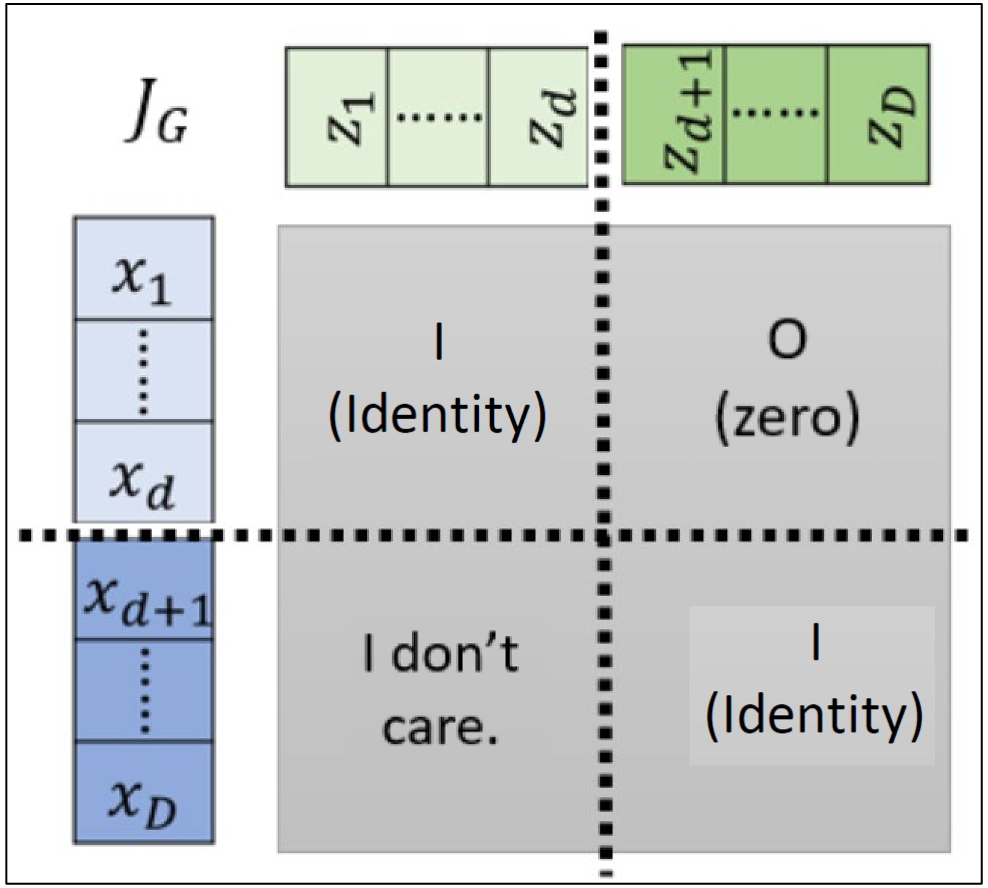
$$\boxed{x_{i>d} = z_i + \gamma_i}$$

  - 雅克比矩阵

$$\boxed{J = \frac{\partial \mathbf{x}}{\partial \mathbf{z}} = \begin{pmatrix} I_d & 0 \\ \frac{\partial \mathbf{x}_{d+1:n}}{\partial \mathbf{z}_{1:d}} & I_{n-d} \end{pmatrix}}$$

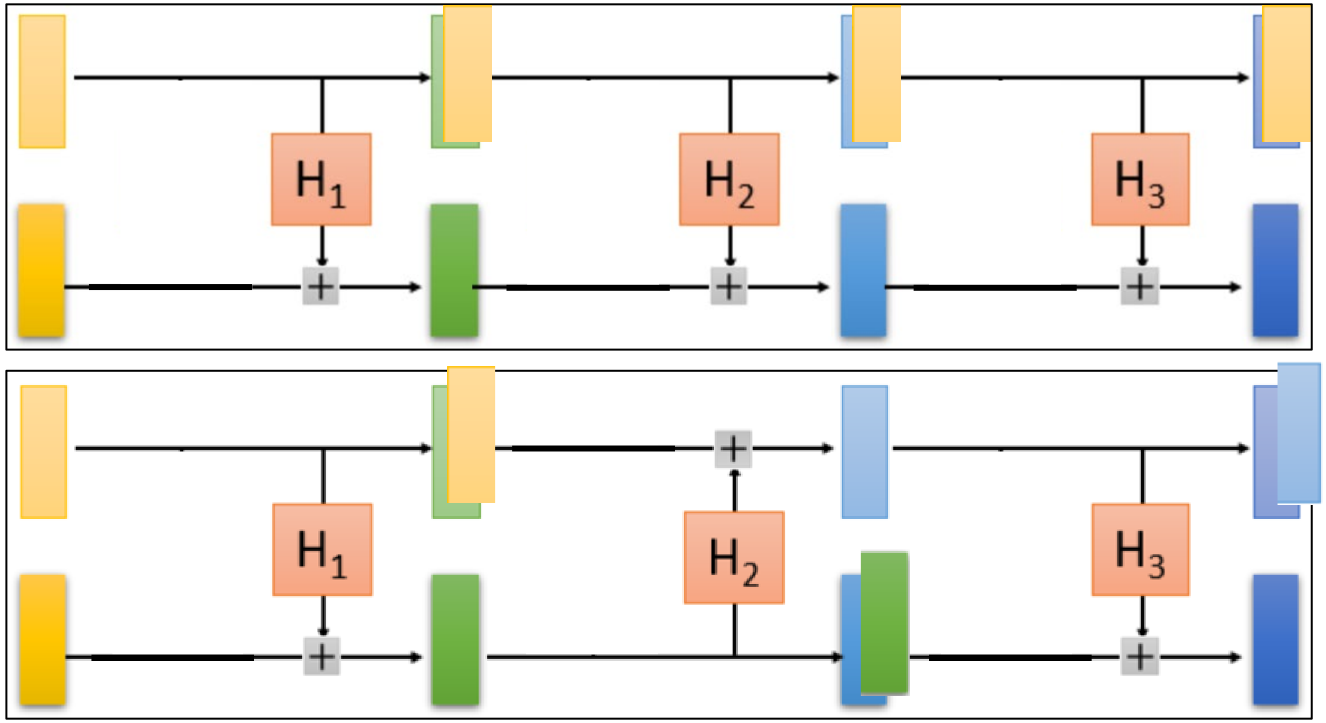
  - 行列式

$$\boxed{\det(J_G) = 1}$$



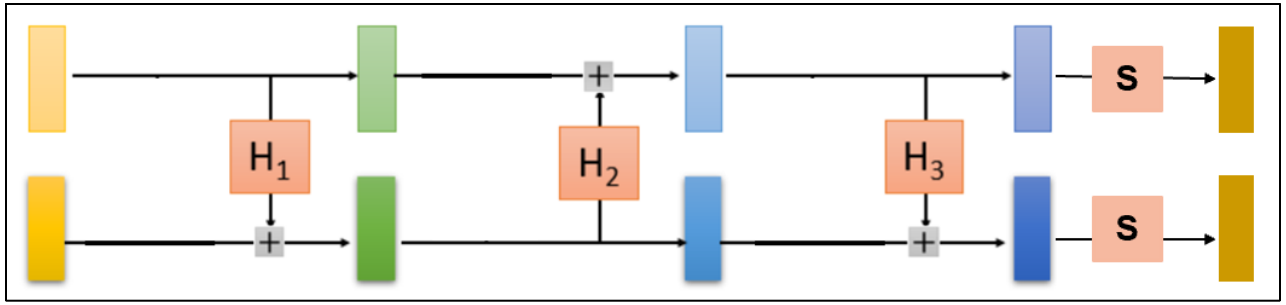
NICE

- 叠加：Flow的序列长度为K
  - 相邻两层交替做恒等变换



NICE

- 尺度变换层
  - 存在比较严重的维度浪费问题
  - 对最后一层每个维度的特征做尺度变换



  - 变换$f: Z \mapsto X$ $$\boxed{x_i = s_i z_i}$$
  - 逆变换$f^{-1}: X \mapsto Z$ $$\boxed{z_i = \frac{x_i}{s_i}}$$

NICE

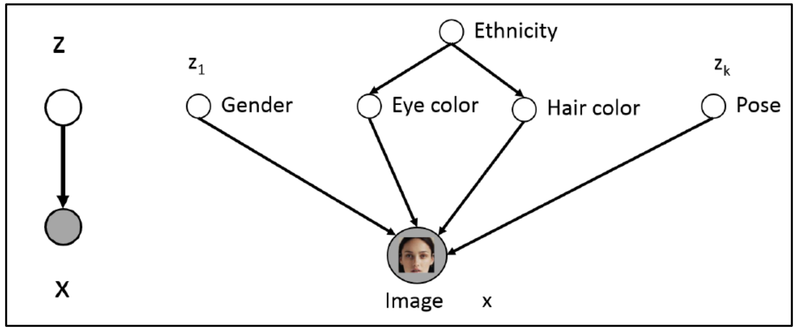
- 尺度变换层
  - S为对角矩阵
  - 雅克比行列式

$$\boxed{\det(J_G) = \prod_{d=1}^{D} S_d}$$

  - 识别该维度的重要程度
    - 等价于将先验分布的方差（标准差）也作为训练参数，如果方差足够小，那么该特征就恒为均值 0，故认为该维度所表示的流形坍缩为一个点，从而总体流形的维度减 1，暗含了降维的可能

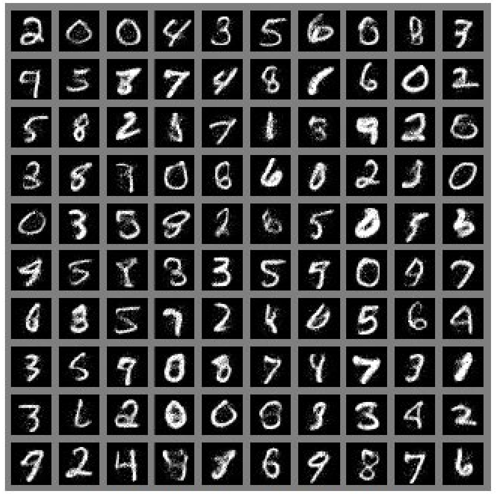
NICE

- 先验分布 $p_Z(z)$： 高斯模型
  - 对原始数据x进行编码时，输出的编码特征 $z = f^{-1}(x)$ 的各个维度是解耦
  - 由于 z 的每个维度的独立性，理论上我们控制改变单个维度时，就可以看出生成图像是如何随着该维度的改变而改变，从而发现该维度的含义



NICE

- 生成数据集



(a) Model trained on MNIST

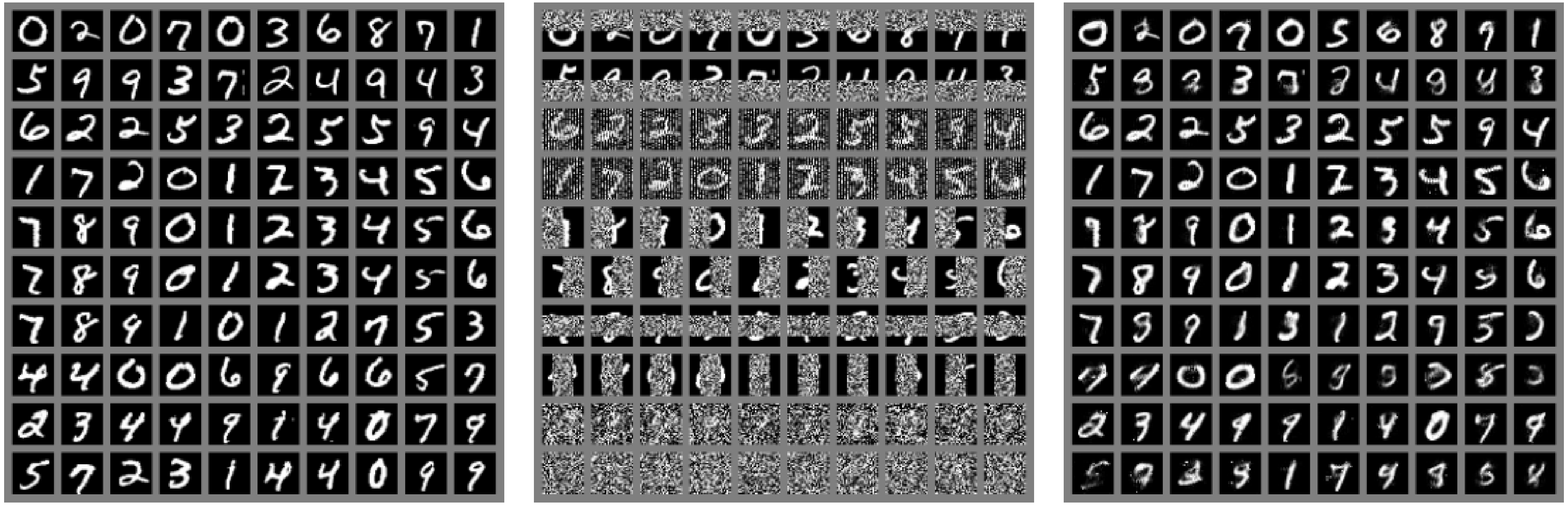
(b) Model trained on TFD

(c) Model trained on SVHN

(d) Model trained on CIFAR-10

NICE

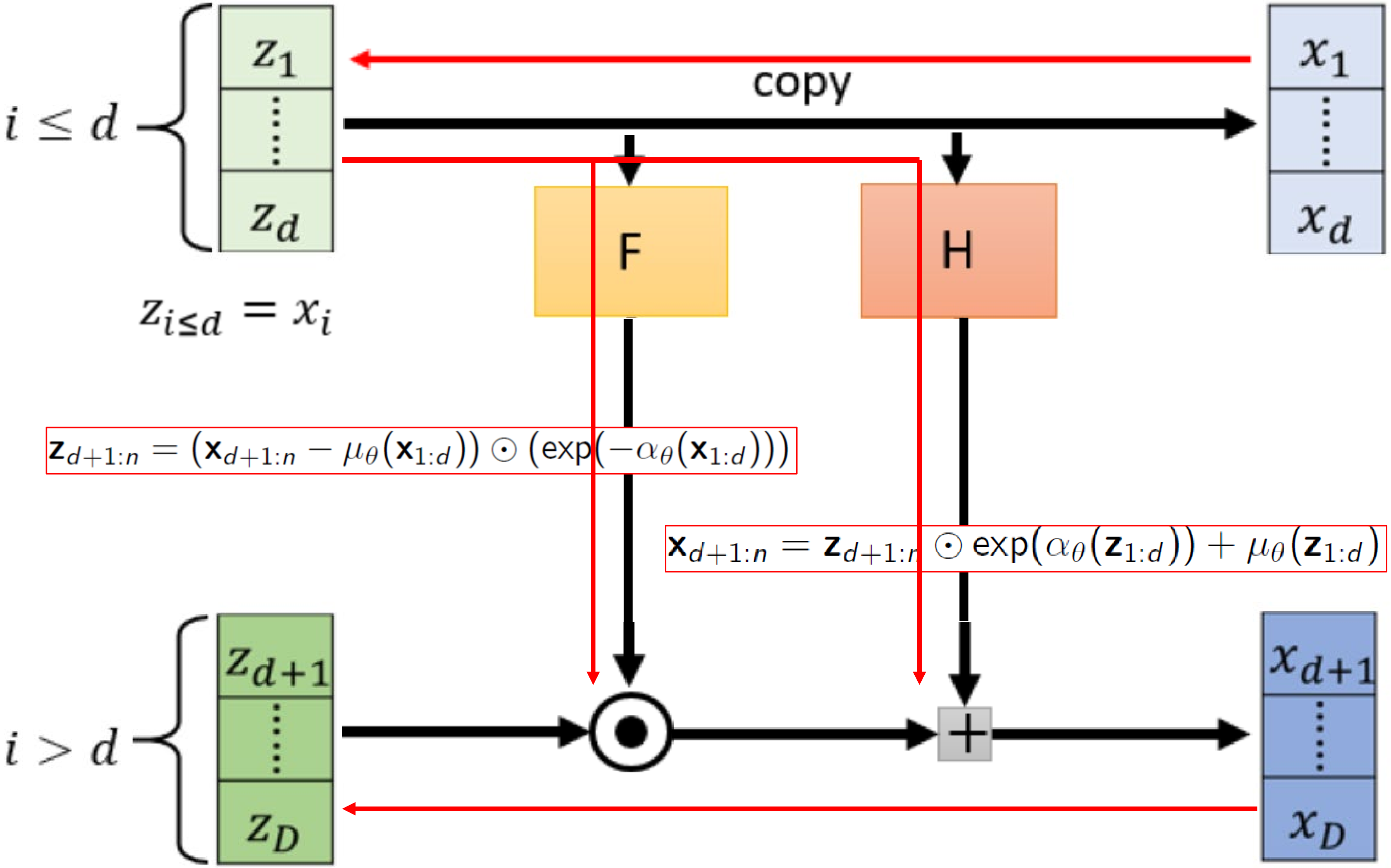
- 图像补全



(a) MNIST test examples (b) Initial state (c) MAP inference of the state

Real NVP (non-volume preserving)

- 将f的雅克比矩阵变换为三角矩阵
  - 仿射耦合层（affine coupling layer）
    - 将输入的一部分维度保持不变，而另一部分维度进行仿射变换
  - 不需要尺度变换层
    - 当仿射层的平移量等于0时，仿射变换可以看成是一种尺度变换



$i \leq d$
z_1
z_d
copy
x_1
x_d
F
H
$z_{i \le d} = x_i$
$\mathbf{z}_{d+1:n} = (\mathbf{x}_{d+1:n} - \mu_\theta(\mathbf{x}_{1:d})) \odot (\exp(-\alpha_\theta(\mathbf{x}_{1:d})))$
$\mathbf{x}_{d+1:n} = \mathbf{z}_{d+1:n} \odot \exp(\alpha_\theta(\mathbf{z}_{1:d})) + \mu_\theta(\mathbf{z}_{1:d})$
$i > d$
z_{d+1}
z_D
x_{d+1}
x_D

Real NVP

- 将隐变量Z分成两个不相交部分$z_{1:d}$和$z_{d+1:D}$
- 变换$f: Z \mapsto X$

$$\boxed{x_{i\le d} = z_i}$$

$$\boxed{\mathbf{x}_{d+1:n} = \mathbf{z}_{d+1:n} \odot \exp(\alpha_\theta(\mathbf{z}_{1:d})) + \mu_\theta(\mathbf{z}_{1:d})}$$

- 逆变换$f^{-1}: X \mapsto Z$

$$\boxed{z_{i\le d} = x_i}$$

$$\boxed{\mathbf{z}_{d+1:n} = (\mathbf{x}_{d+1:n} - \mu_\theta(\mathbf{x}_{1:d})) \odot (\exp(-\alpha_\theta(\mathbf{x}_{1:d})))}$$

- F和H可以是任意复杂函数，不需要可逆

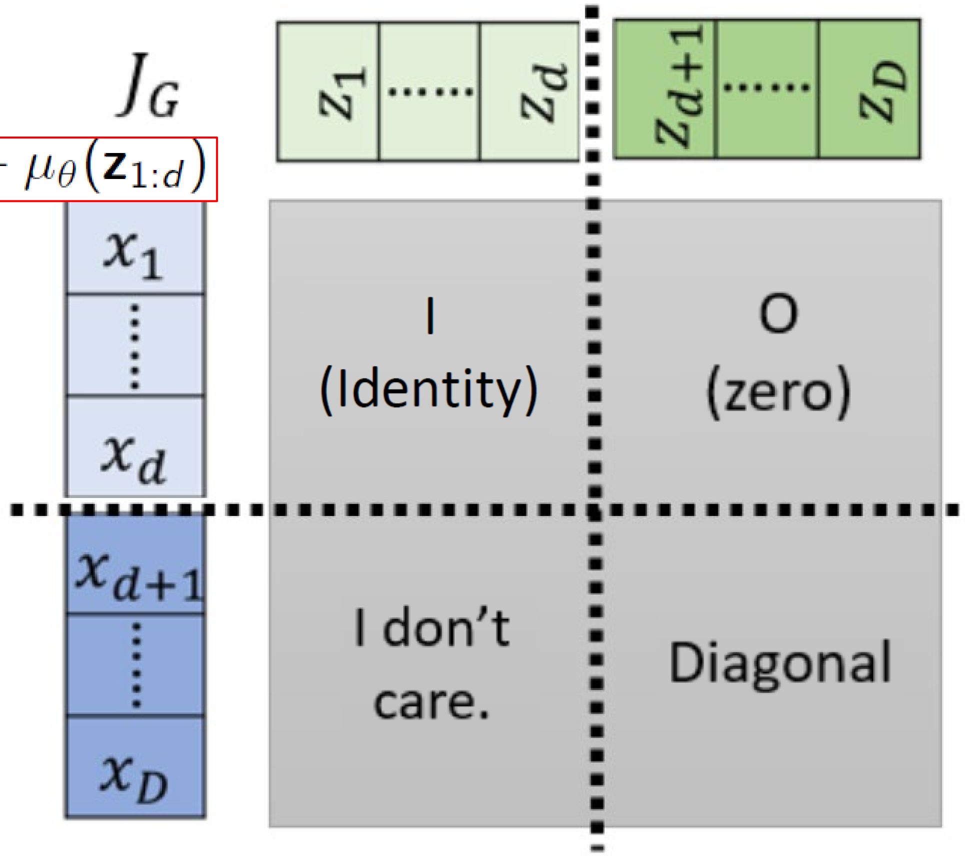
Real NVP

- 雅克比矩阵

$$\boxed{x_{i\le d} = z_i}$$

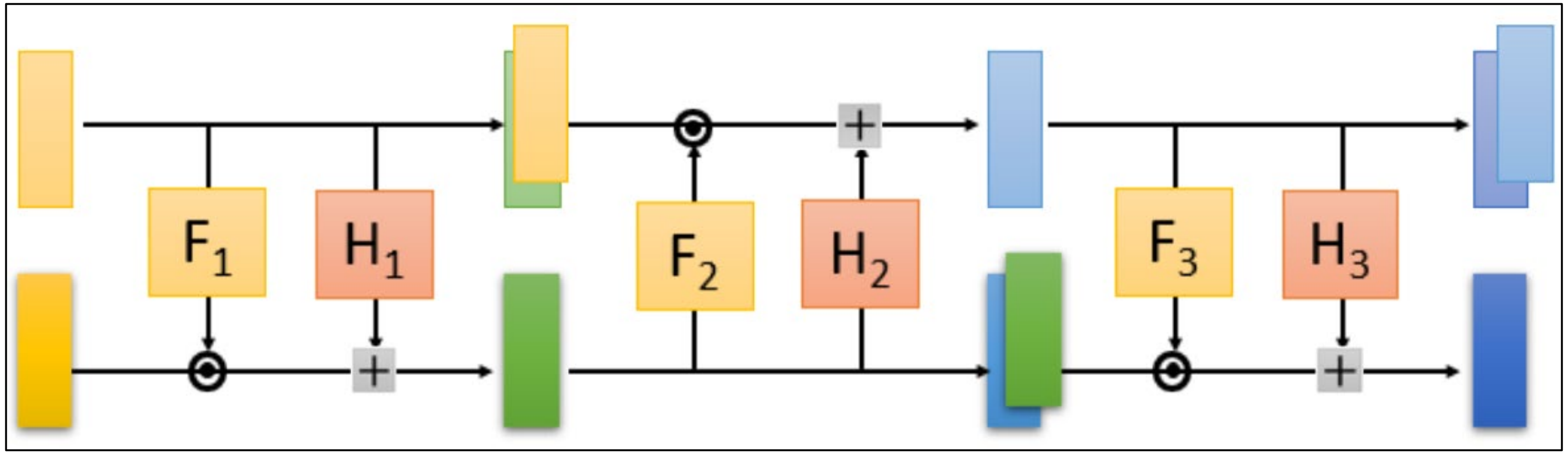
$$\boxed{\mathbf{x}_{d+1:n} = \mathbf{z}_{d+1:n} \odot \exp(\alpha_\theta(\mathbf{z}_{1:d})) + \mu_\theta(\mathbf{z}_{1:d})}$$

$$det(J_G) = \frac{\partial x_{d+1}}{\partial z_{d+1}}\frac{\partial x_{d+2}}{\partial z_{d+2}}\cdots\frac{\partial x_D}{\partial z_D} = \prod_{i=d+1}^{n}\exp(\alpha_\theta(\mathbf{z}_{1:d})_i) = \exp\left(\sum_{i=d+1}^{n}\alpha_\theta(\mathbf{z}_{1:d})_i\right)$$



Real NVP

- K个仿射耦合层叠加
  - 相邻两层交替做恒等变换



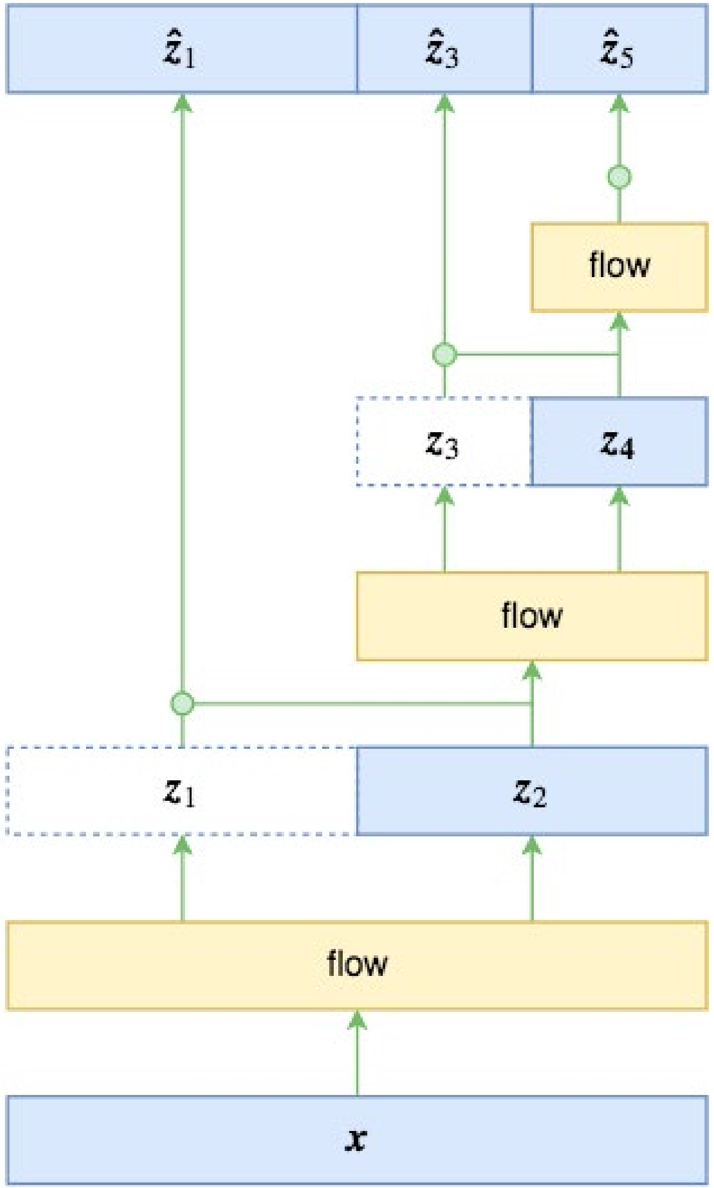
  - 雅克比行列式

$$\frac{\partial (f_b \circ f_a)}{\partial x_a^T}(x_a) = \frac{\partial f_a}{\partial x_a^T}(x_a) \cdot \frac{\partial f_b}{\partial x_b^T}\left(x_b = f_a(x_a)\right)$$

$$\det(A \cdot B) = \det(A)\det(B).$$

Real NVP

- 多尺度结构
  - “flow 运算”指的是多个仿射耦合层的复合后，输出跟输入的大小一样，这时候将输入对半分开两半



Real NVP

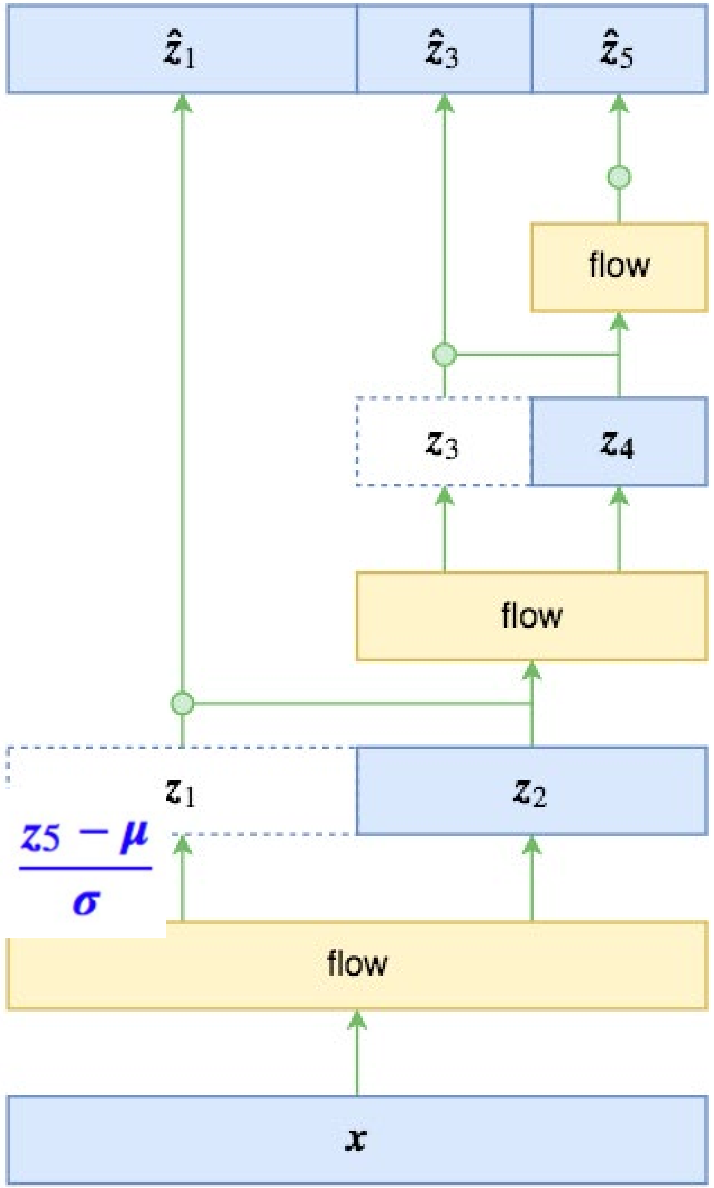
- 多尺度结构
  - 抛弃了 p(z) 是标准正态分布的直接假设，而采用了一个组合式的条件分布

$$p(z_1, z_3, z_5) = p(z_1 | z_2)p(z_3 | z_4)p(z_5)$$

  - $\hat{z}_1, \hat{z}_3, \hat{z}_5$服从标准正态分布

$$\hat{z}_1 = \frac{z_1 - \mu(z_2)}{\sigma(z_2)}, \quad \hat{z}_3 = \frac{z_3 - \mu(z_4)}{\sigma(z_4)}, \quad \hat{z}_5 = \frac{z_5 - \mu}{\sigma}$$

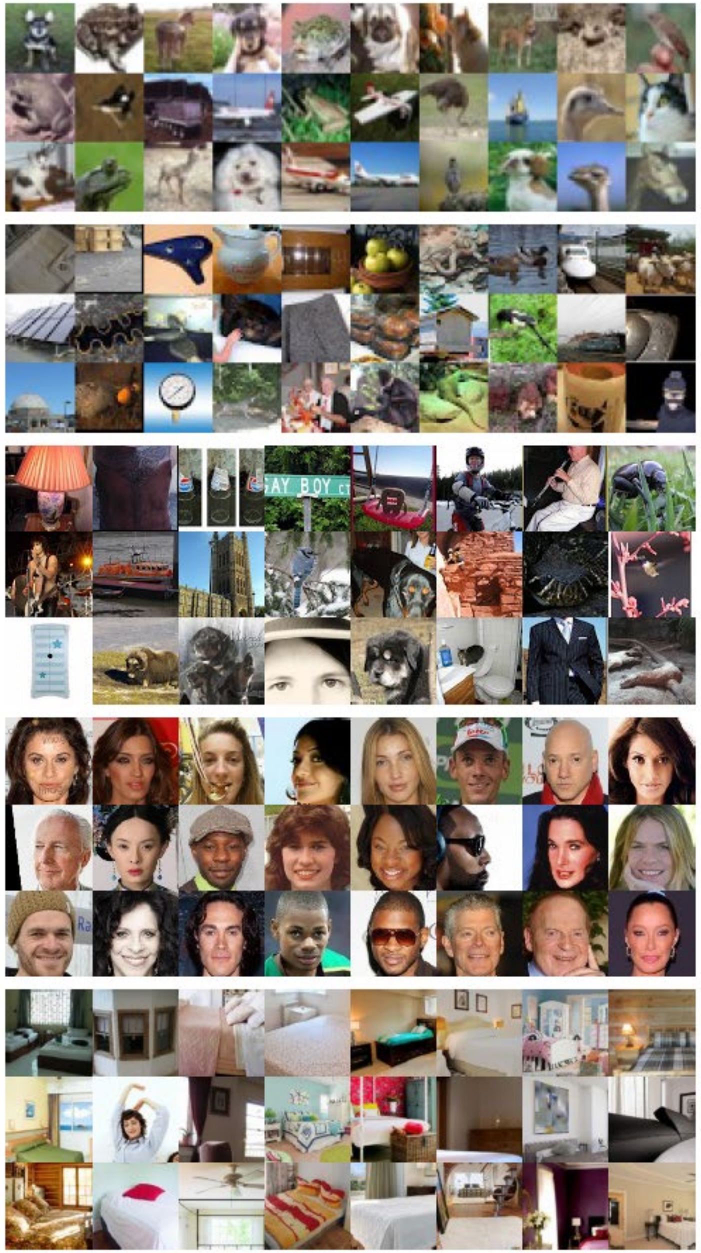
    - 通过控制每个条件分布的方差来抑制维度浪费问题

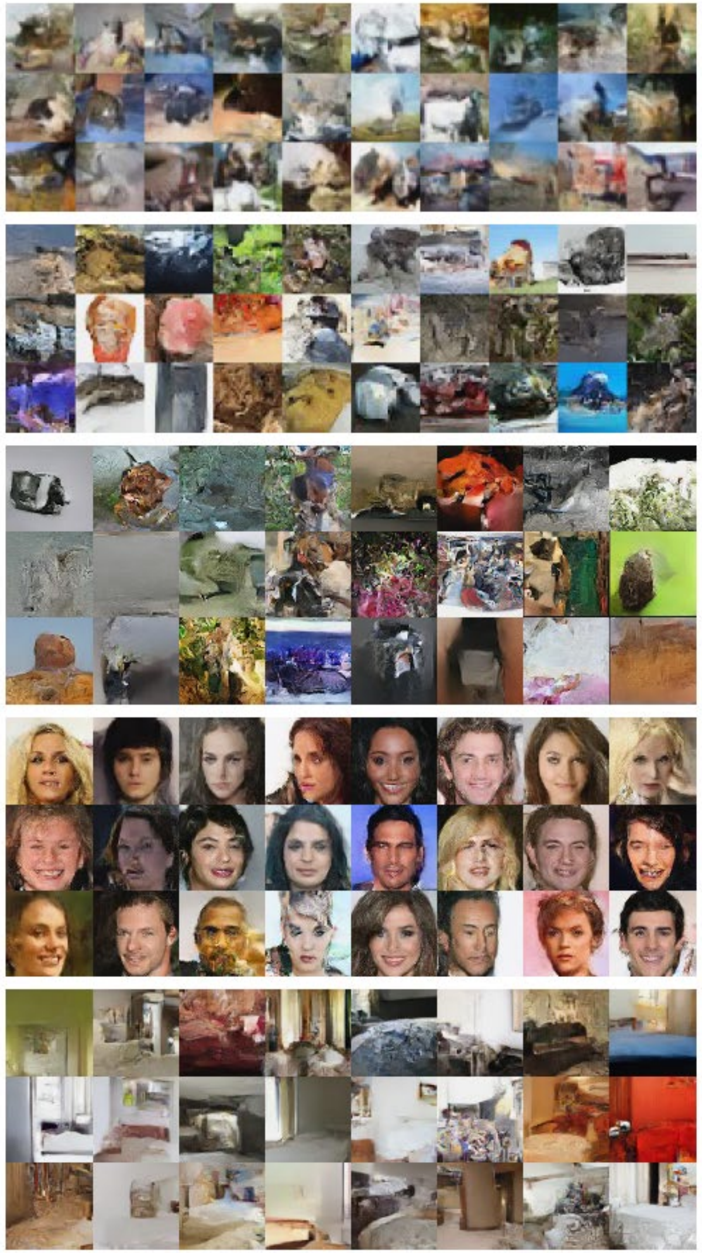


Real NVP

左图：
真实数据集

右图：
生成数据集





Real NVP

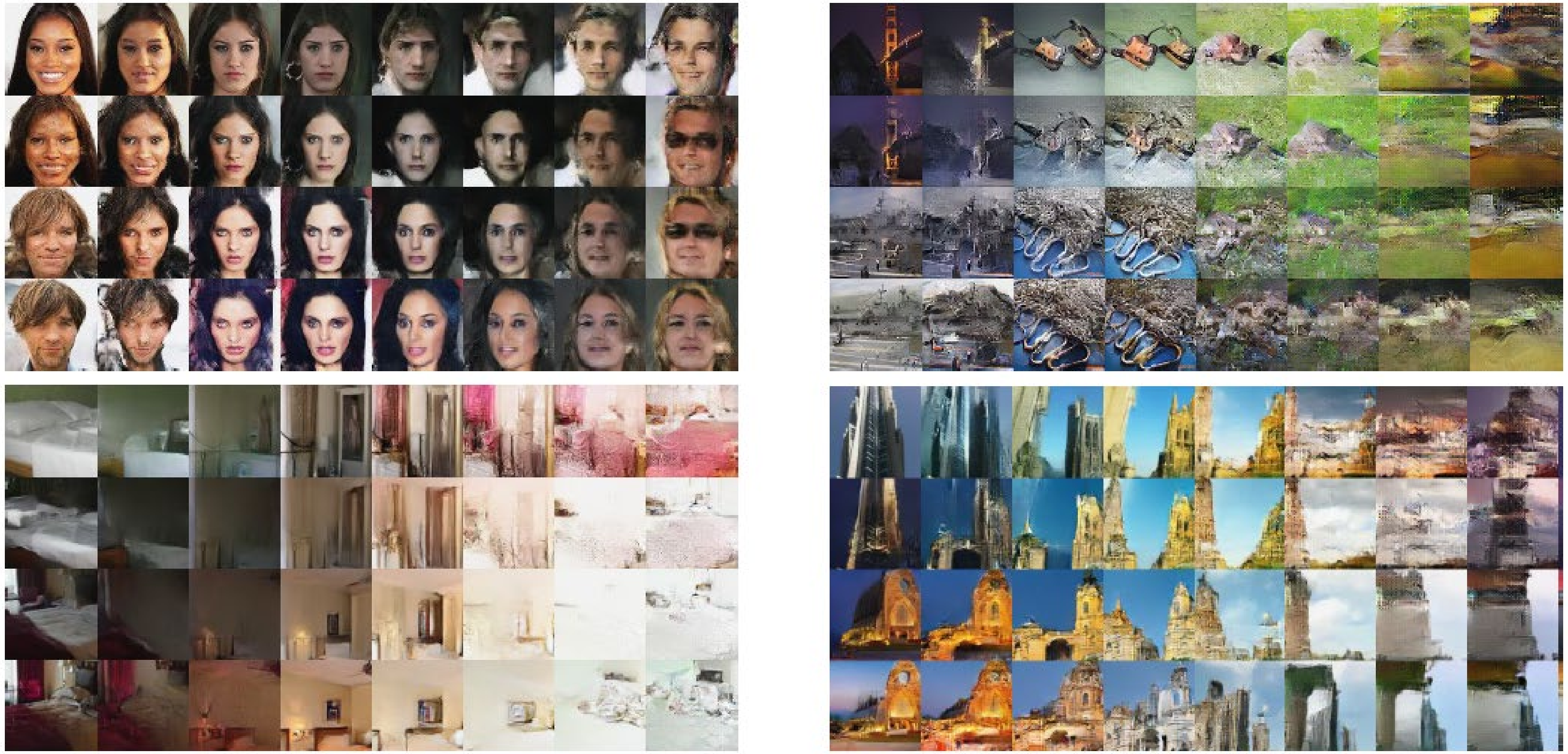


Figure 6: Manifold generated from four examples in the dataset. Clockwise from top left: CelebA, Imagenet (64 × 64), LSUN (tower), LSUN (bedroom).

$$\boxed{z = \cos(\phi)\left(\cos(\phi')z_{(1)} + \sin(\phi')z_{(2)}\right) + \sin(\phi)\left(\cos(\phi')z_{(3)} + \sin(\phi')z_{(4)}\right)}$$

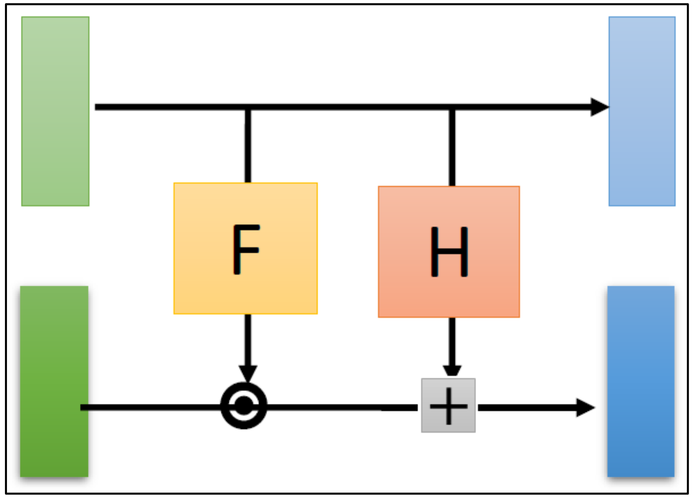
NICE vs. Real NVP

- 将隐变量Z分成两个不相交部分$z_{1:d}$和$z_{d+1:D}$
- 变换$f: Z \mapsto X$

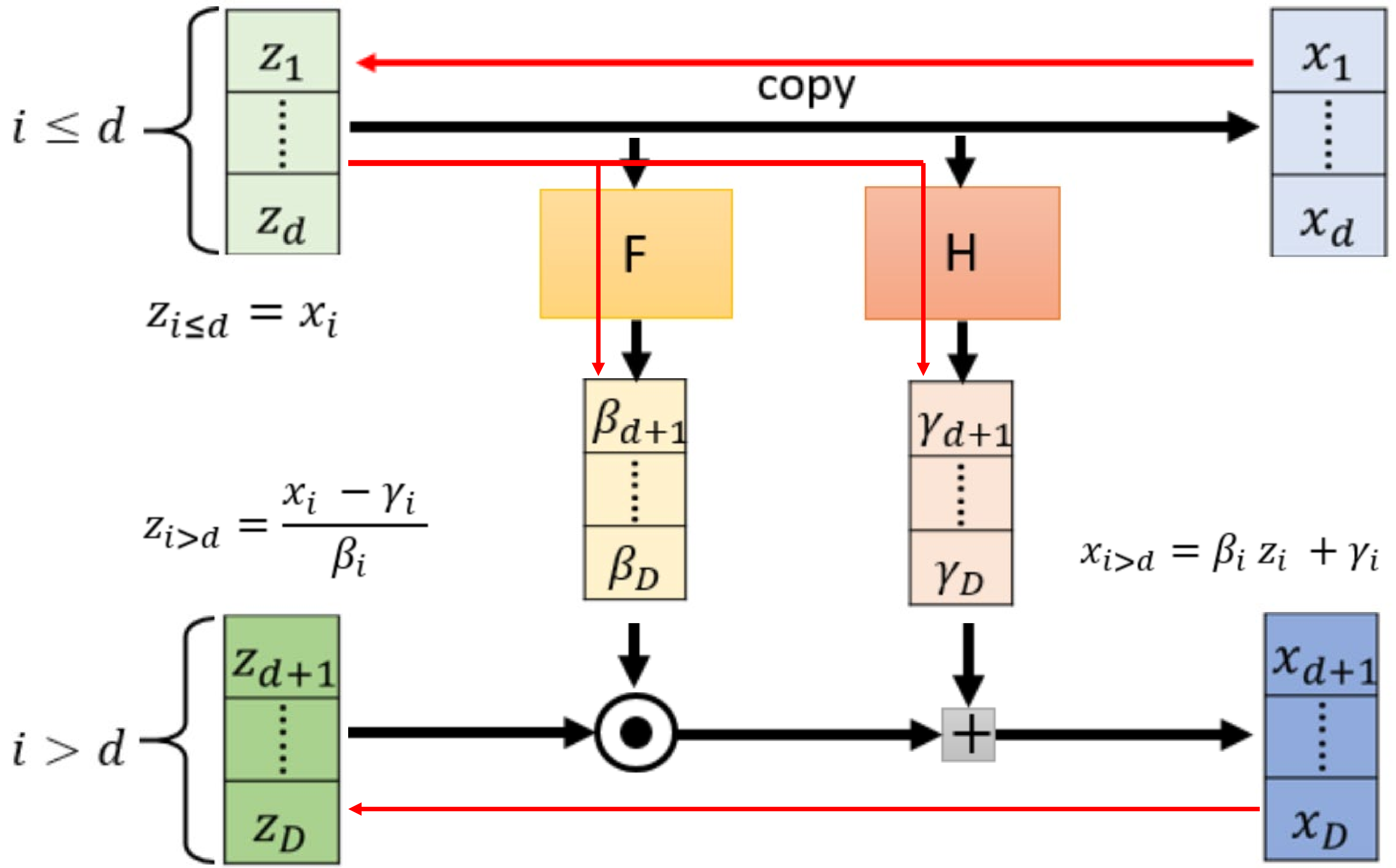
  $\boxed{x_{i\le d} = z_i}$ $\boxed{x_{i>d} = \beta_i z_i + \gamma_i}$

- 逆变换$f^{-1}: X \mapsto Z$

  $\boxed{z_{i\le d} = x_i}$ $\boxed{z_{i>d} = \frac{x_i - \gamma_i}{\beta_i}}$



- F和H可以是任意复杂函数，不需要可逆


$i \le d$
z_1
z_d
copy
x_1
x_d
F
H
$z_{i \le d} = x_i$
β_{d+1}
β_D
γ_{d+1}
γ_D
$z_{i>d} = \frac{x_i - \gamma_i}{\beta_i}$
$x_{i>d} = \beta_i z_i + \gamma_i$
$i > d$
z_{d+1}
z_D
x_{d+1}
x_D

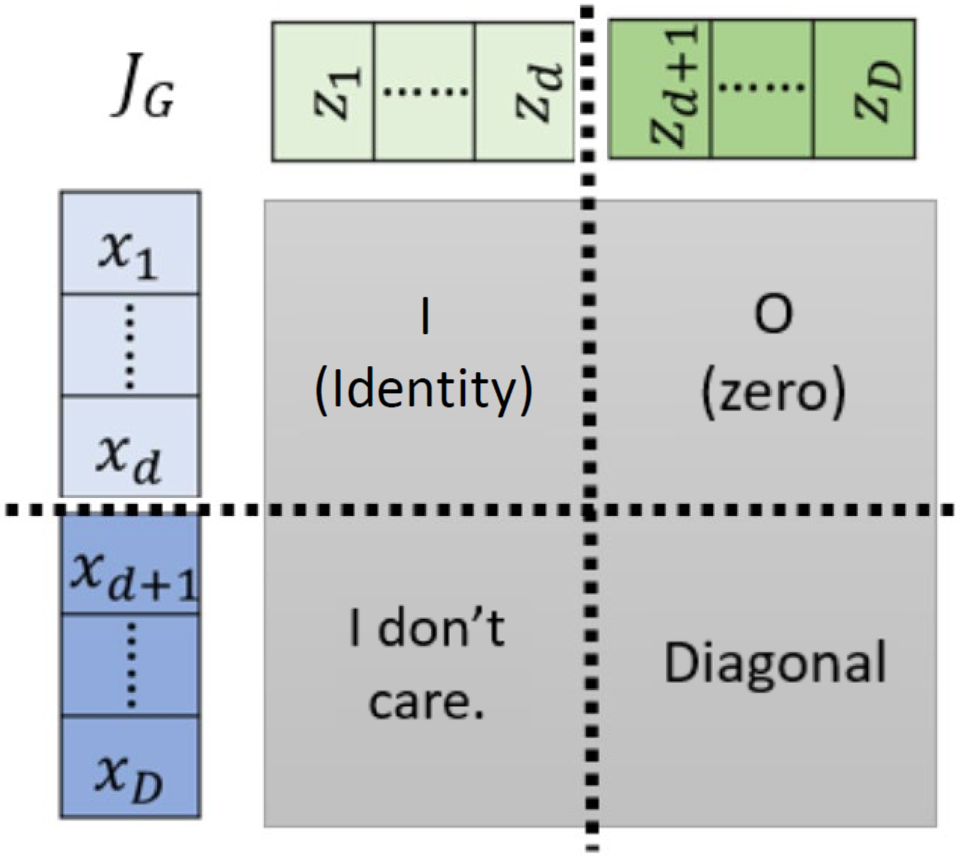
NICE vs. Real NVP

- 雅克比矩阵

$$\boxed{x_{i\le d} = z_i}$$

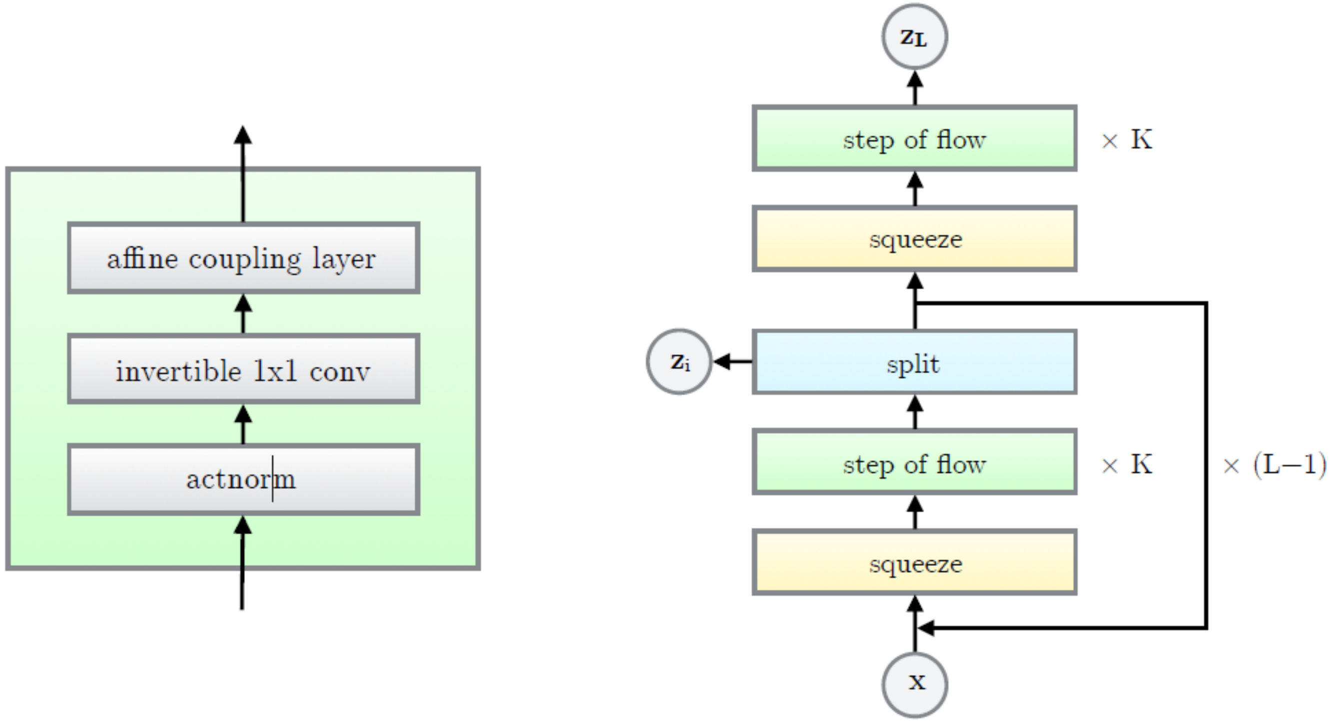
$$x_{i>d} = \beta_i\, z_i + \gamma_i$$

$$\boxed{\begin{aligned} &det(J_G) \\ &= \frac{\partial x_{d+1}}{\partial z_{d+1}} \frac{\partial x_{d+2}}{\partial z_{d+2}} \cdots \frac{\partial x_D}{\partial z_D} \\ &= \beta_{d+1}\beta_{d+2} \cdots \beta_D \end{aligned}}$$



Glow

- 采用了可逆的1 x 1卷积核打乱输入数据，简化了网络模型



(a) One step of our flow.
(b) Multi-scale architecture (Dinh et al., 2016).

Glow

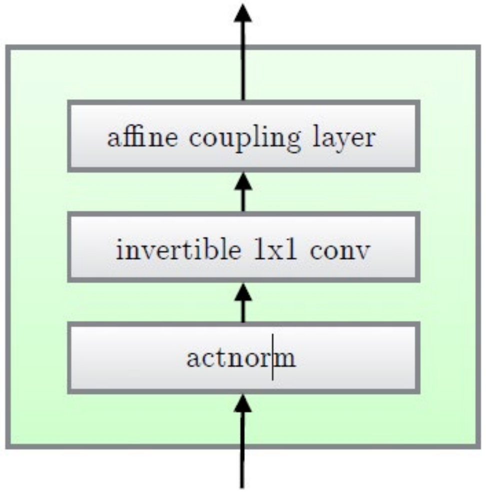
- 可逆的1x1卷积核
  - 采用矩阵乘法实现对各个维度数据的打乱，使信息混合得更加充分
    - 置换矩阵

$$\begin{pmatrix}2\\1\\4\\3\end{pmatrix}=\begin{pmatrix}0&1&0&0\\1&0&0&0\\0&0&0&1\\0&0&1&0\end{pmatrix}\begin{pmatrix}1\\2\\3\\4\end{pmatrix}$$

  - 可逆矩阵：正交矩阵
  - 对数行列式

$$\boxed{h = xW}$$

$$\boxed{\log\left|\det\left(\frac{d\,\mathrm{conv2D}(\mathbf{h};\mathbf{W})}{d\,\mathbf{h}}\right)\right| = h\cdot w\cdot\log|\det(\mathbf{W})|}$$

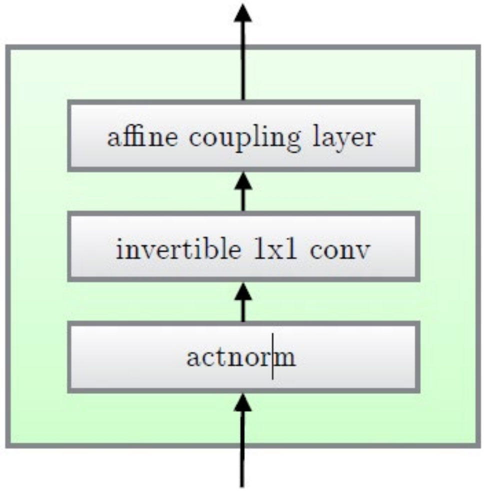


Glow

- 可逆的1x1卷积核
  - LU分解，使得计算更加方便

$$\boxed{\mathbf{W} = \mathbf{P}\mathbf{L}(\mathbf{U} + \mathrm{diag}(\mathbf{s}))}$$

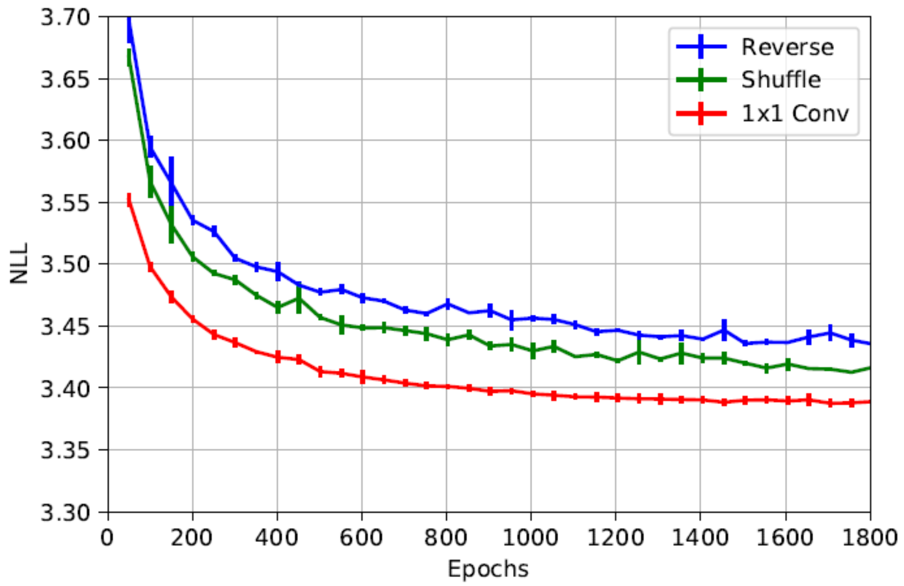
$$\Longrightarrow \boxed{\log|\det(\mathbf{W})| = \mathrm{sum}(\log|\mathbf{s}|)}$$



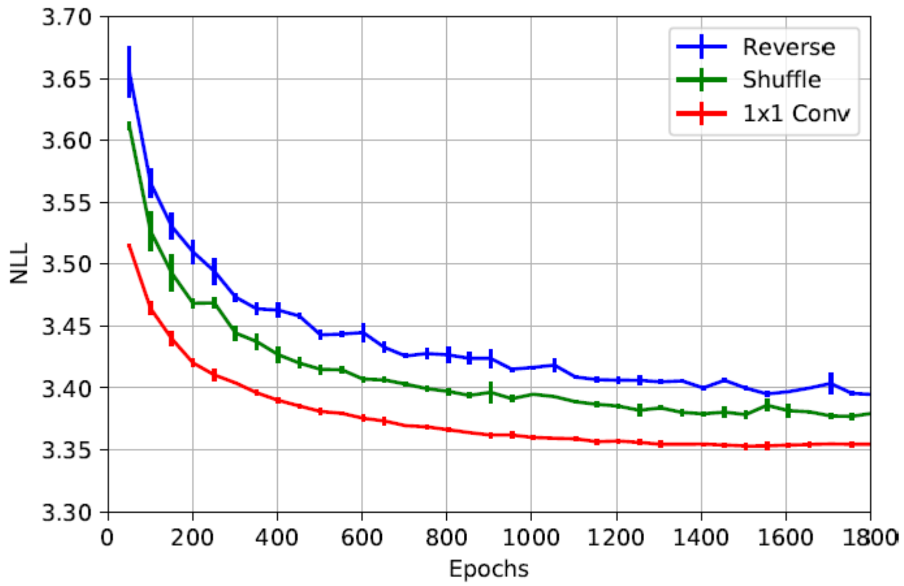
  - 技巧
    - 先随机生成一个正交矩阵，然后做 LU 分解，得到 P,L,U
    - 固定 P，也固定 U 的对角线的正负号，然后约束 L 为对角线全 1 的下三角阵，U 为上三角阵，优化训练 L,U 的其余参数

Glow

- 可逆的1x1卷积核可提高似然值



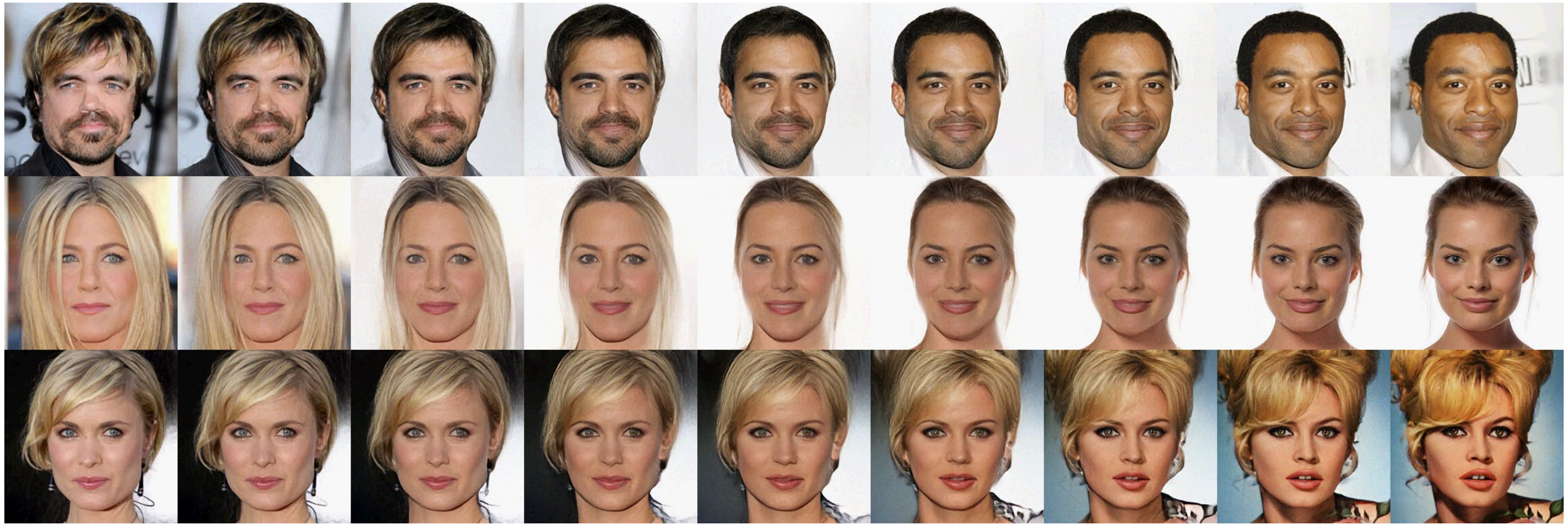
(a) Additive coupling.



(b) Affine coupling.

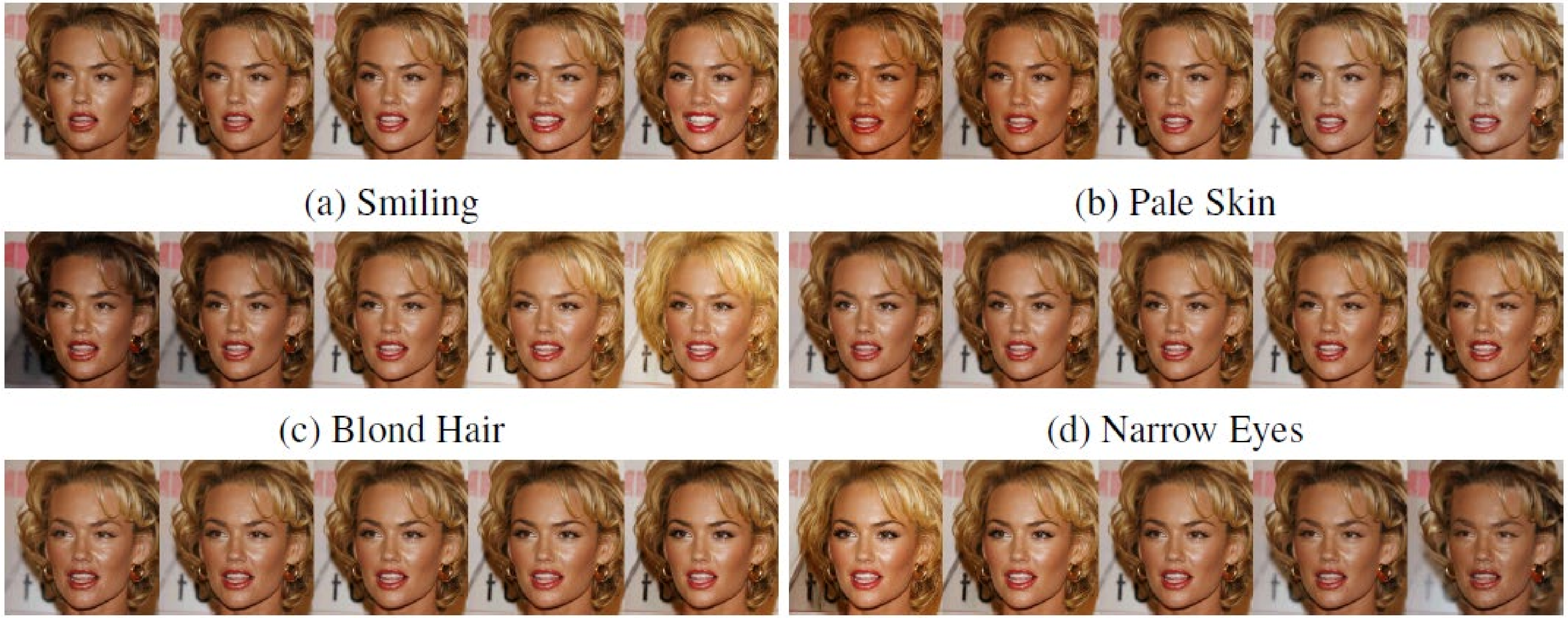
Glow

- 隐空间做线性插值



Glow

- 控制图片生成



(a) Smiling (b) Pale Skin

(c) Blond Hair (d) Narrow Eyes

(e) Young (f) Male

Glow

- 模型深度

L=4 L=6

自回归模型

- 高斯自回归模型

$$\boxed{p(\mathbf{x}) = \prod_{i=1}^{n} p(x_i|\mathbf{x}_{<i})}$$

$$\boxed{p(x_i \mid \mathbf{x}_{<i}) = \mathcal{N}(\mu_i(x_1, \cdots, x_{i-1}), \exp(\alpha_i(x_1, \cdots, x_{i-1}))^2}$$

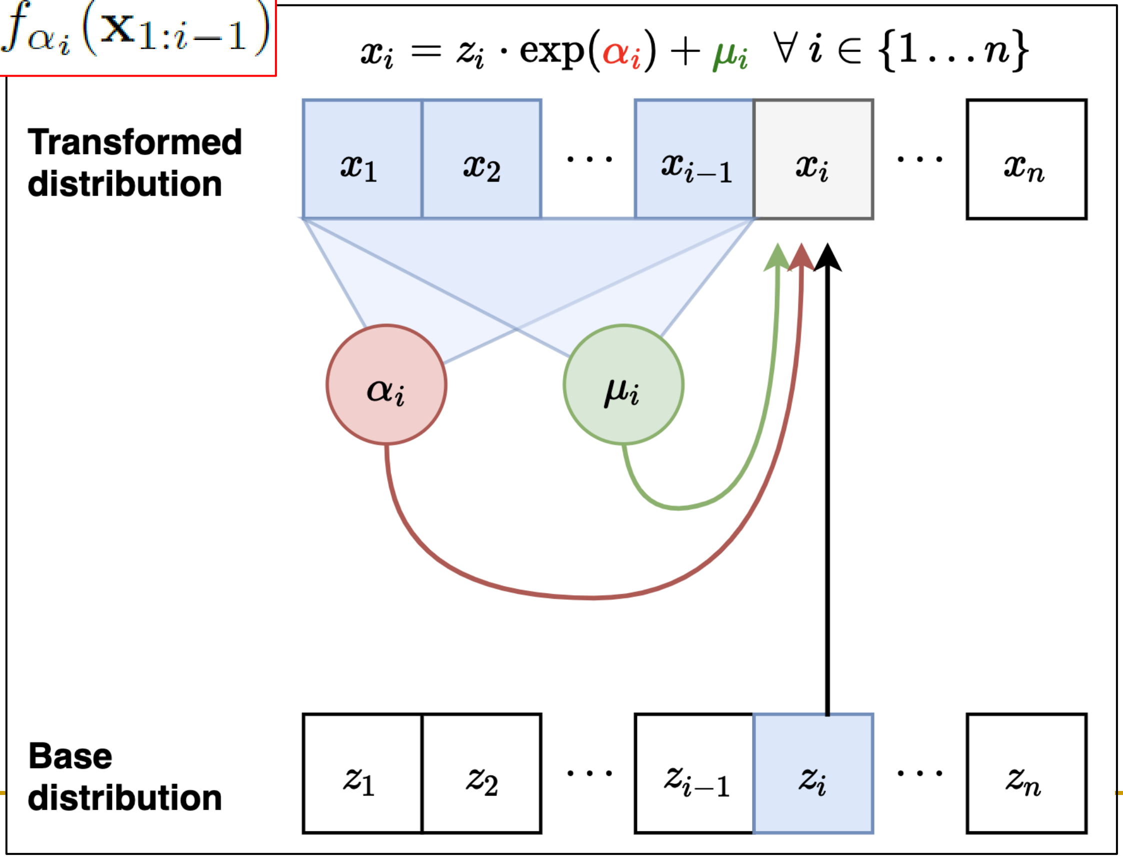
- 生成数据过程
  - 采样$z_i \sim \mathcal{N}(0,1), i = 1, \dots, n$
  - 令$x_1 = \exp(\alpha_1) z_1 + \mu_1$，计算$\mu_2(x_1), \alpha_2(x_1)$
  - 令$x_2 = \exp(\alpha_2) z_2 + \mu_2$，计算$\mu_3(x_1, x_2), \alpha_3(x_1, x_2)$
  - 令$x_3 = \exp(\alpha_3) z_3 + \mu_3 \dots \dots$

掩码自回归流（MAF）

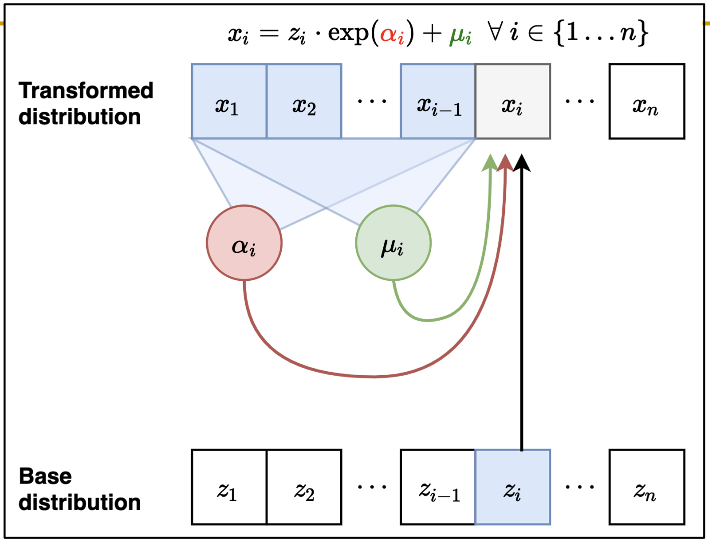
- 自回归网络

$$\mu_i = f_{\mu_i}(\mathbf{x}_{1:i-1}) \quad \text{and} \quad \alpha_i = f_{\alpha_i}(\mathbf{x}_{1:i-1})$$

  - 均值μ_i
  - 方差$(\exp \alpha_i)^2$



掩码自回归流（MAF）



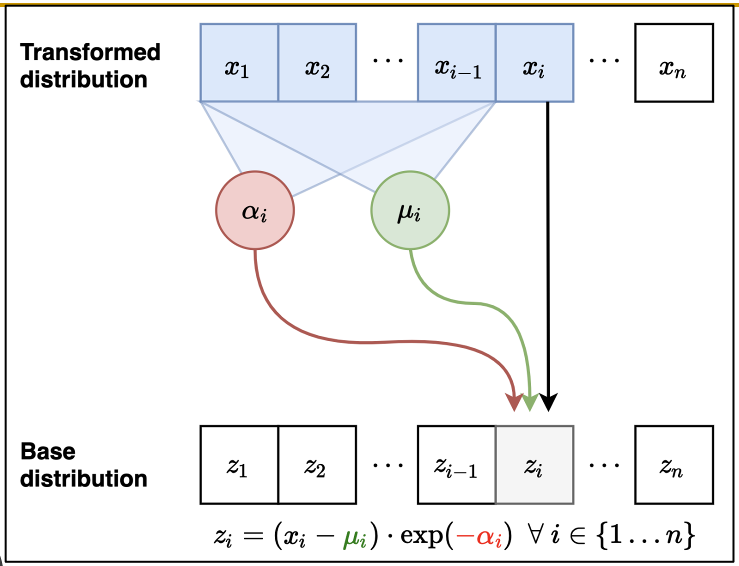
- 变换$f: Z \mapsto X$

$$\boxed{x_i = \exp(\alpha_i)\ z_i + \mu_i}$$

  - 令$x_1 = \exp(\alpha_1)\, z_1 + \mu_1$，计算$\mu_2(x_1), \alpha_2(x_1)$
  - 令$x_2 = \exp(\alpha_2)\, z_2 + \mu_2$，计算$\mu_3(x_1, x_2), \alpha_3(x_1, x_2)$
  - 令$x_3 = \exp(\alpha_3)\, z_3 + \mu_3 \ldots\ldots$
- 雅克比矩阵

$$\boxed{J_G = \begin{bmatrix} \exp(\alpha_1) & 0 & \cdots & 0 \\ \frac{\partial x_2}{\partial z_1} & \exp(\alpha_2) & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial x_n}{\partial z_1} & \frac{\partial x_n}{\partial z_2} & \cdots & \exp(\alpha_n) \end{bmatrix}}$$

掩码自回归流（MAF）



- 逆变换$f^{-1}: X \mapsto Z$

$$\boxed{z_i = \exp(-\alpha_i)\,(x_i - \mu_i)}$$

  - 并行计算μ_i, α_i（使用MADE）
  - 令$z_1 = \exp(-\alpha_1)\,(x_1 - \mu_1)$
  - 令$z_2 = \exp(-\alpha_2)\,(x_2 - \mu_2)$
  - 令$z_3 = \exp(-\alpha_3)\,(x_3 - \mu_3)$
- 雅克比逆矩阵

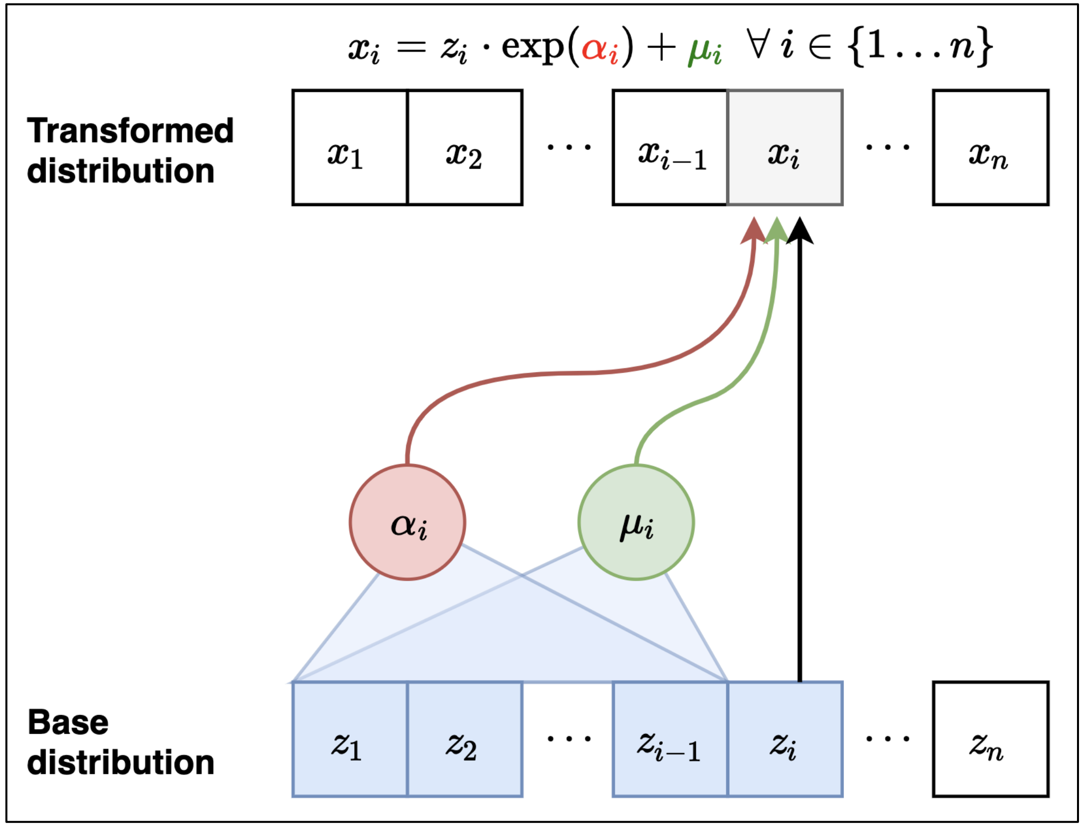
$$\boxed{J_{G^{-1}} = \begin{bmatrix} \exp(-\alpha_1) & 0 & \cdots & 0 \\ \frac{\partial z_2}{\partial x_1} & \exp(-\alpha_2) & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial z_n}{\partial x_1} & \frac{\partial z_n}{\partial x_2} & \cdots & \exp(-\alpha_n) \end{bmatrix}}$$

逆自回归流（IAF）

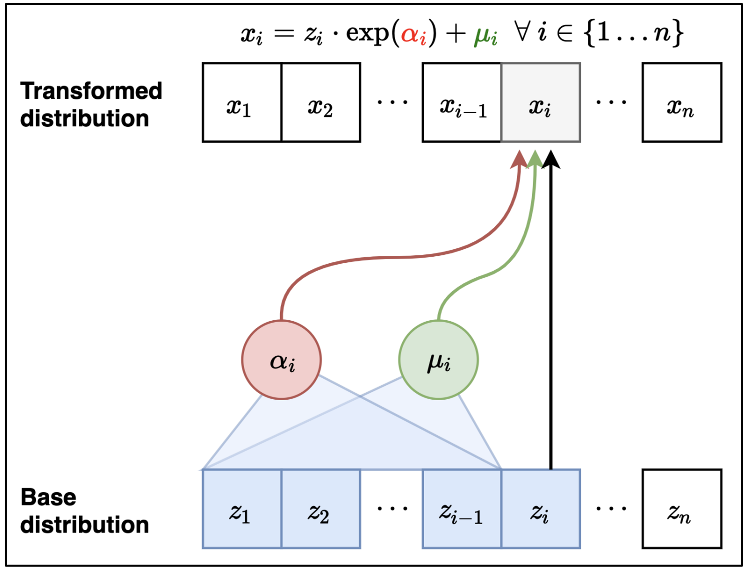
- 自回归网络

$$\mu_i = f_{\mu_i}(z_{1:i-1})$$
$$\alpha_i = f_{\alpha_i}(z_{1:i-1})$$

  - 均值μ_i
  - 方差$(\exp \alpha_i)^2$



逆自回归流（IAF）



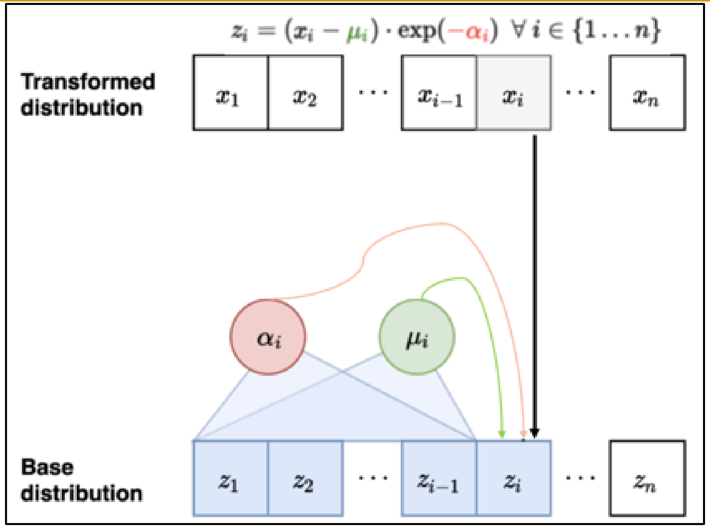
- 变换$f: Z \mapsto X$

$$\boxed{x_i = \exp(\alpha_i)\ z_i + \mu_i}$$

  - 采样$z_i \sim \mathcal{N}(0,1), i = 1, \dots, n$
  - 并行计算μ_i, α_i（使用MADE）
  - 令$x_1 = \exp(\alpha_1)\, z_1 + \mu_1$
  - 令$x_2 = \exp(\alpha_2)\, z_2 + \mu_2 \dots\dots$
- 雅克比矩阵

$$\boxed{J_G = \begin{bmatrix} \exp(\alpha_1) & 0 & \cdots & 0 \\ \frac{\partial x_2}{\partial z_1} & \exp(\alpha_2) & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial x_n}{\partial z_1} & \frac{\partial x_n}{\partial z_2} & \cdots & \exp(\alpha_n) \end{bmatrix}}$$

逆自回归流（IAF）



- 逆变换$f^{-1}: X \mapsto Z$

$$\boxed{z_i = \exp(-\alpha_i)\,(x_i - \mu_i)}$$

  - 令$z_1 = \exp(-\alpha_1)\,(x_1 - \mu_1)$，计算$\mu_2(z_1), \alpha_2(z_1)$
  - 令$z_2 = \exp(-\alpha_2)\,(x_2 - \mu_2)$, 计算$\mu_3(z_1, z_2), \alpha_3(z_1, z_2)$
  - 令$z_3 = \exp(-\alpha_3)\,(x_3 - \mu_3), \dots \dots$

- 雅克比逆矩阵

$$\boxed{J_{G^{-1}} = \begin{bmatrix} \exp(-\alpha_1) & 0 & \cdots & 0 \\ \frac{\partial z_2}{\partial x_1} & \exp(-\alpha_2) & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial z_n}{\partial x_1} & \frac{\partial z_n}{\partial x_2} & \cdots & \exp(-\alpha_n) \end{bmatrix}}$$

MAF vs. IAF

- 交换MAF编码器的变量z和x，就是IAF的生成器

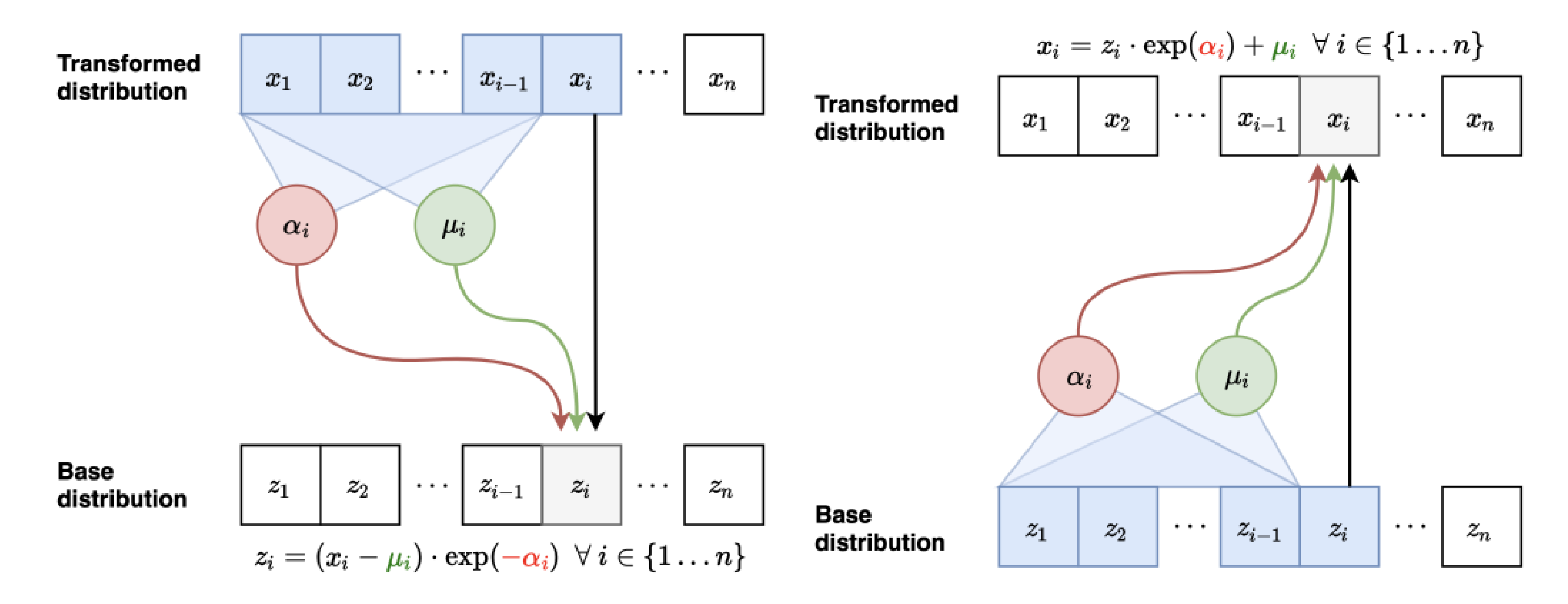


Figure: Inverse pass of MAF (**left**) vs. Forward pass of IAF (**right**)

MAF vs. IAF

- 交换MAF生成器的变量z和x，就是IAF的编码器

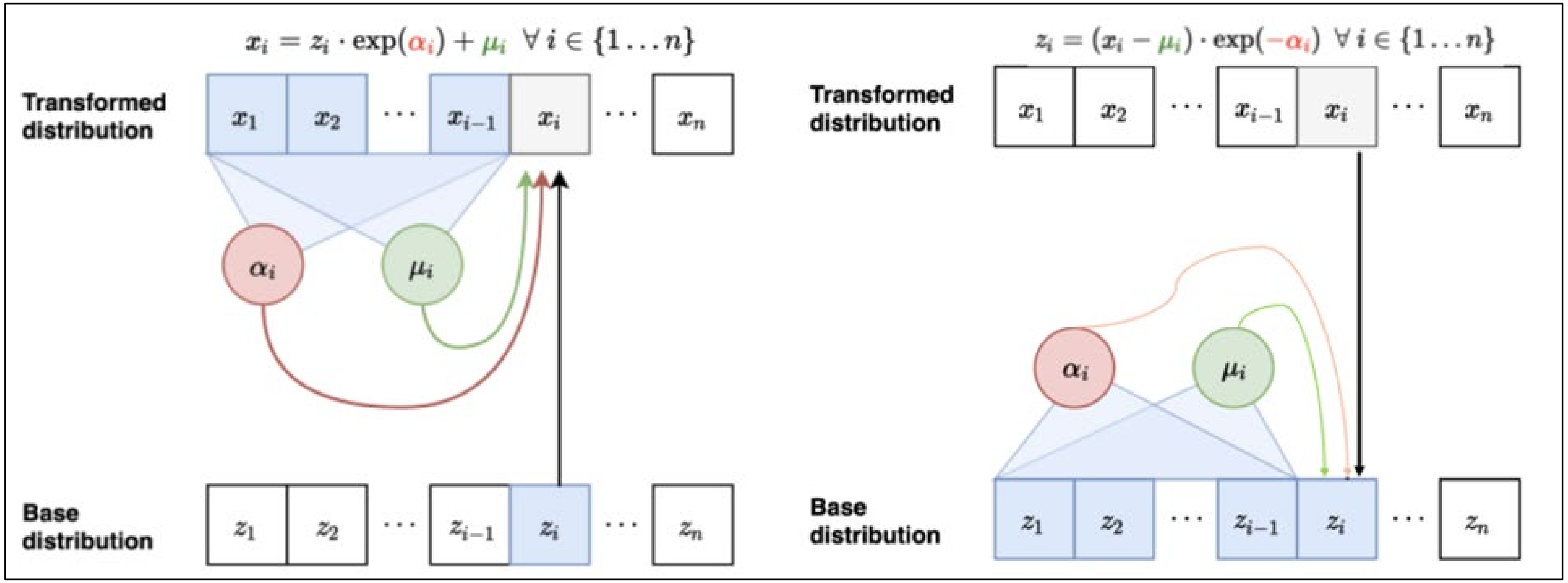


Figure: Forward pass of MAF (left) vs. Inverse pass of IAF (right)

MAF vs. IAF

- 计算代价
  - MAF计算似然函数快，采样慢
  - IAF计算似然函数慢，采样快
- MAF更适合极大似然函数估计
- IAF更适合实时生成数据

Thanks!

Questions?