04 正规化流与可逆密度变换

从生成样本到精确密度

学习目标

- 解释可逆分布变换，以及概率质量与局部体积怎样决定密度。
- 说明耦合层怎样求逆、计算行列式，并复合为完整模型。
- 用精确似然解释参数学习，区分连续密度、离散概率与生成结果。

核心路线保留质量守恒、体积直观、耦合机制和训练；深入节点展开积分证明、矩阵偏导、数值演算与参数导数。

1 可逆模型与密度接口

怎样让同一模型既能生成样本，又能计算观测密度？

一个映射连接两个连续空间

给定一个真实观测，怎样计算模型在该处的密度，并用它学习参数?

正规化流选择可计算似然作为学习信号；VAE与GAN也能学习生成模型，采用的学习信号不同。

取易采样、易求密度的基分布 $p_Z = \mathcal{N}(\mathbf{0}, \boldsymbol{I})$，用同维光滑双射连接两个空间： $\boldsymbol{x} = g_\theta(\boldsymbol{z}), \boldsymbol{z} = g_\theta^{-1}(\boldsymbol{x}), \boldsymbol{z}, \boldsymbol{x} \in \mathbb{R}^d$.

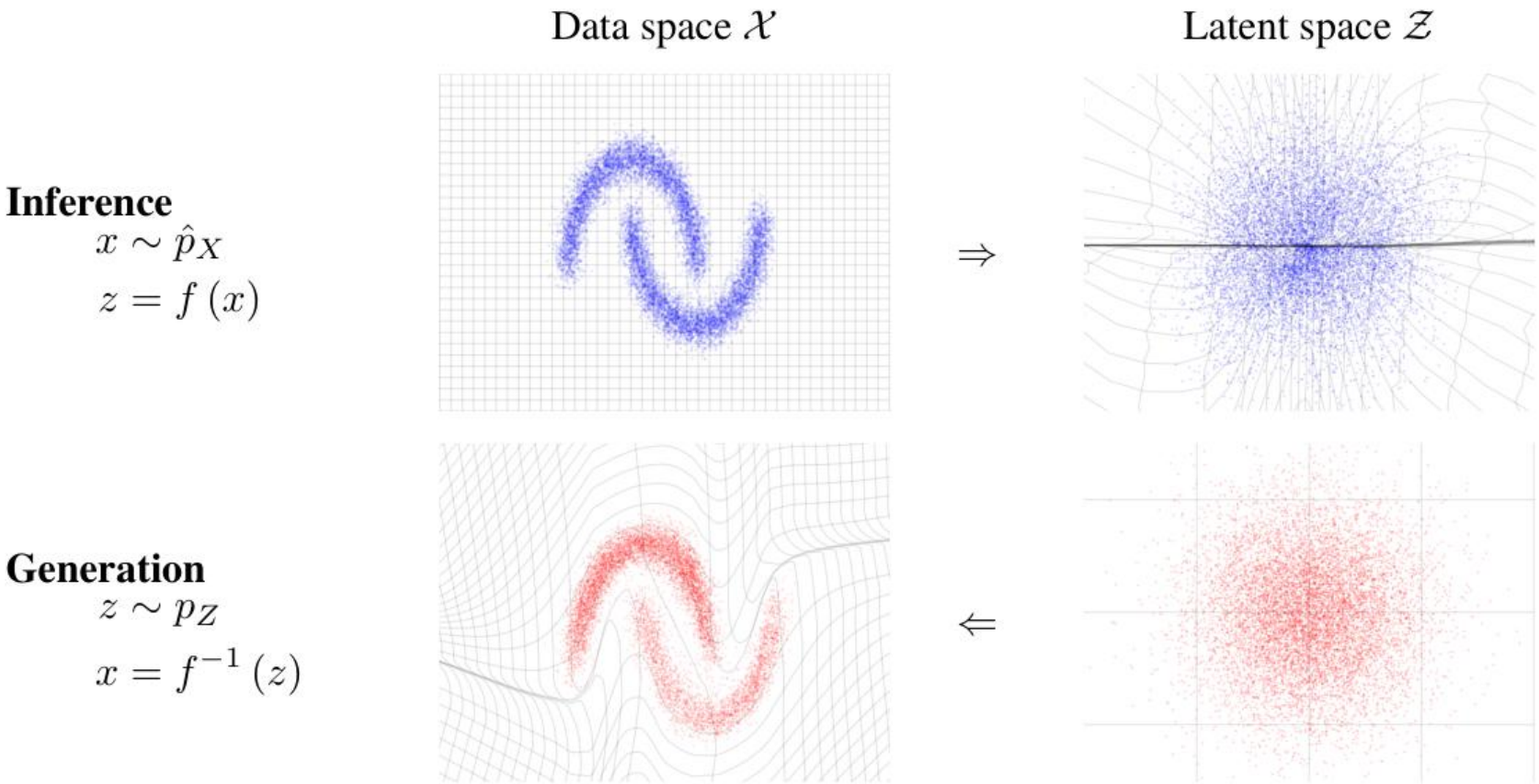


两方向共享参数 θ；不是两个独立网络

点的对应已经确定，下一步还要计算邻域的体积变化。

论文中的分布与网格变换

先读蓝色推断，再反读红色生成，观察分布与网格形变。

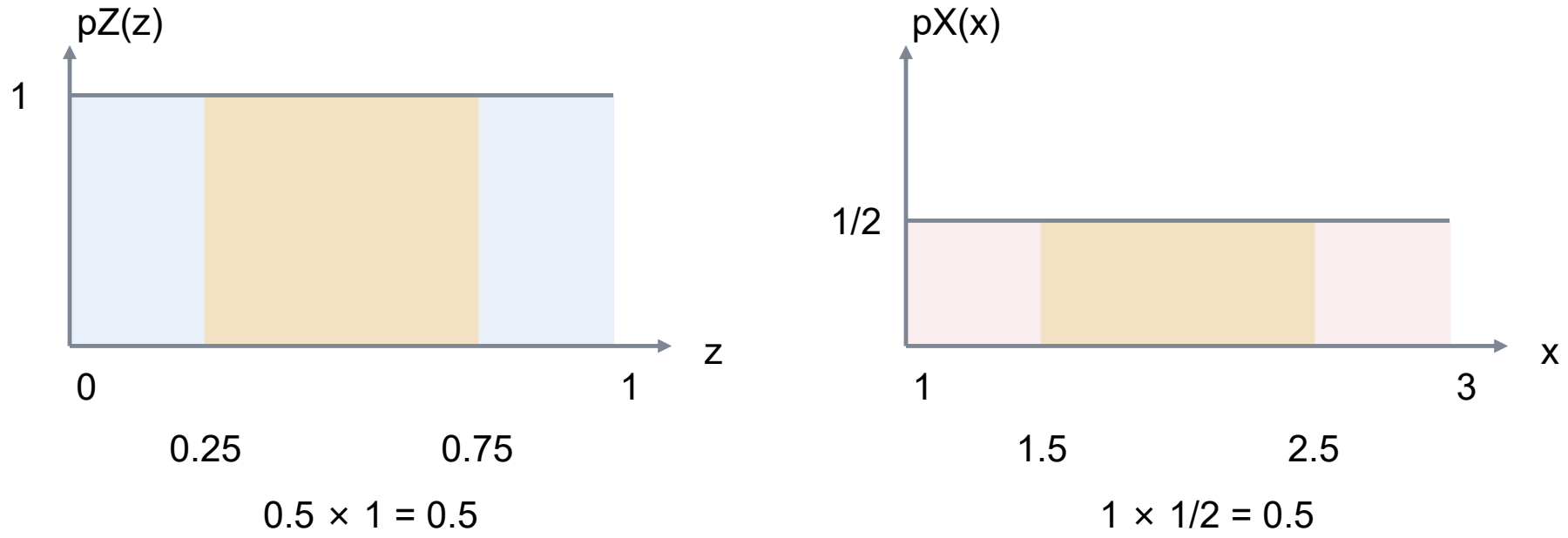


原文 $f: x \to z$，本讲 $g = f^{-1}$。来源：Real NVP v3，Fig.1，第2页；论文示例。

样本变换之后还需要密度修正

区间伸长后，同一概率质量需要更低的密度。

教学构造：$Z \sim \mathrm{Uniform}(0,1)$，$X = 2Z + 1$。



阴影面积均为概率0.5；两空间各用自己的横轴刻度

$P(0.25 \le Z \le 0.75) = 0.5 = P(1.5 \le X \le 2.5).$

宽度扩大两倍，密度降低一半；总面积仍为1。多维时，长度因子推广为局部体积因子。

密度沿生成方向补偿体积

一维长度因子推广为多维局部体积因子：体积扩大，单位体积的密度下降。先给出双向计算接口，随后解释它为什么成立。

对$g_\theta: \mathbb{R}^d \to \mathbb{R}^d$，**雅可比矩阵**的元素为 $[\boldsymbol{J}_g(\boldsymbol{z})]_{ij} = \frac{\partial g_i}{\partial z_j}$. 它描述小位移如何线性近似地变化。$|det\boldsymbol{J}_g(\boldsymbol{z})|$是局部体积放大倍数；绝对值去除方向符号。

设同维映射为可微且逆可微的双射、雅可比非奇异，基密度已归一化，则在对应点$\boldsymbol{z} = g_\theta^{-1}(\boldsymbol{x})$：

$p_\theta(\boldsymbol{x}) = p_Z(\boldsymbol{z})|det\boldsymbol{J}_g(\boldsymbol{z})|^{-1}, \log p_\theta(\boldsymbol{x}) = \log p_Z(\boldsymbol{z}) - \log|det\boldsymbol{J}_g(\boldsymbol{z})| = \log p_Z(g_\theta^{-1}(\boldsymbol{x})) + \log|det\boldsymbol{J}_{g^{-1}}(\boldsymbol{x})|.$

- 沿生成方向，体积放大时密度缩小，雅可比在$\boldsymbol{z}$处求值
- 沿逆方向使用逆雅可比，在对应的$\boldsymbol{x}$处求值

耦合结构提供可计算的逆与行列式

为什么不能直接使用普通神经网络作为流？ 一般网络不保证可逆；即使可逆，求逆与稠密雅可比行列式仍可能昂贵。耦合层把这两个计算要求落实到结构。

一种可执行结构是**仿射耦合层**：把输入分为两块，保留第一块，用它产生第二块的缩放和平移。$\boldsymbol{y}_a = \boldsymbol{u}_a, \boldsymbol{y}_b = \boldsymbol{u}_b \odot e^{\boldsymbol{s}_\theta(\boldsymbol{u}_a)} + \boldsymbol{t}_\theta(\boldsymbol{u}_a)$. ⊙为逐元素乘法，指数也逐元素计算；$\boldsymbol{s}_\theta, \boldsymbol{t}_\theta$输出与更新块同维的向量。

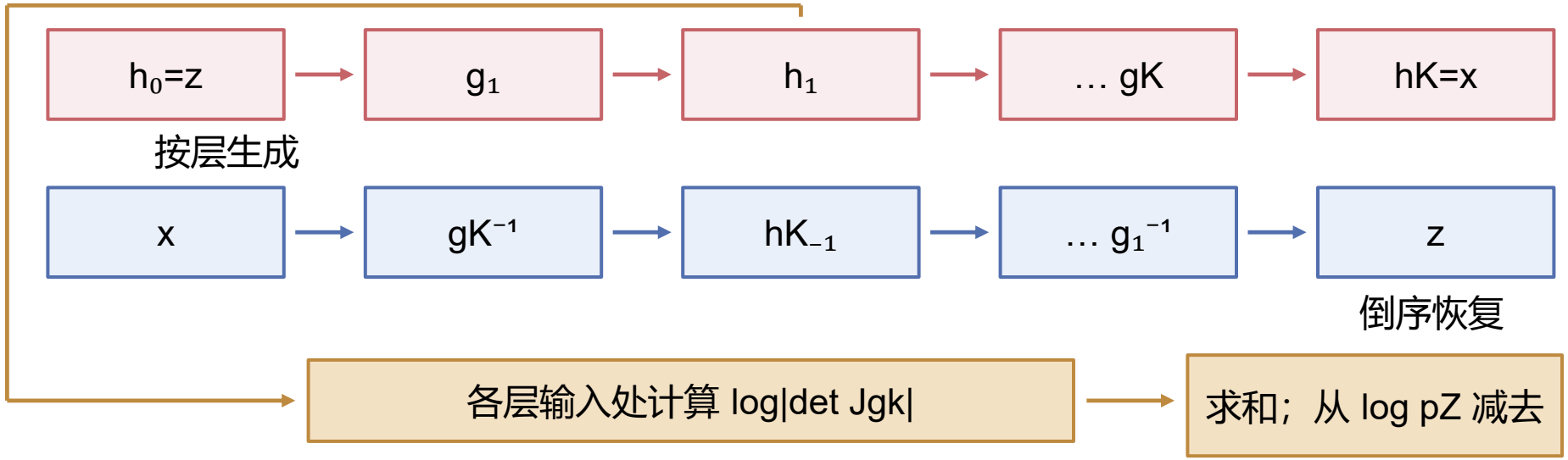
逆向按以下步骤执行：

1. 读取不变块：$\boldsymbol{u}_a = \boldsymbol{y}_a$。
2. 计算$\boldsymbol{s}_\theta(\boldsymbol{y}_a), \boldsymbol{t}_\theta(\boldsymbol{y}_a)$。
3. 恢复更新块：$\boldsymbol{u}_b = [\boldsymbol{y}_b - \boldsymbol{t}_\theta(\boldsymbol{y}_a)] \odot e^{-\boldsymbol{s}_\theta(\boldsymbol{y}_a)}$。

其生成方向logdet为$\sum_j s_{\theta,j}(\boldsymbol{u}_a)$。非零指数尺度使简单仿射步骤可逆，求逆无需反解两个条件网络。随后用结构图解释保留条件与三角雅可比；条件网络的执行成本仍须计入。

层复合的密度修正逐层累加

若$\boldsymbol{h}_0 = \boldsymbol{z}$，$\boldsymbol{h}_k = g_k(\boldsymbol{h}_{k-1})$，$\boldsymbol{h}_K = \boldsymbol{x}$，则 $g_\theta = g_K \circ \cdots \circ g_1, g_\theta^{-1} = g_1^{-1} \circ \cdots \circ g_K^{-1}$. 实际求逆先执行最右端的$g_K^{-1}$，再依次恢复更早状态。对数密度为 $\log p_\theta(\boldsymbol{x}) = \log p_Z(\boldsymbol{z}) - \sum_{k=1}^{K} \log|\, det \boldsymbol{J}_{g_k}(\boldsymbol{h}_{k-1})|$.



交替划分更新块，可使前一层保留的分量在后续层改变。每层仍需单独满足可逆和计算条件。

中间状态是变换后的向量，通常没有直接的像素含义；逐层推进也不等于一张图片逐渐显现。

2 变量替换与耦合变换

解释密度与耦合的依据，深入演算按基础选讲。

一维区间的概率保持不变

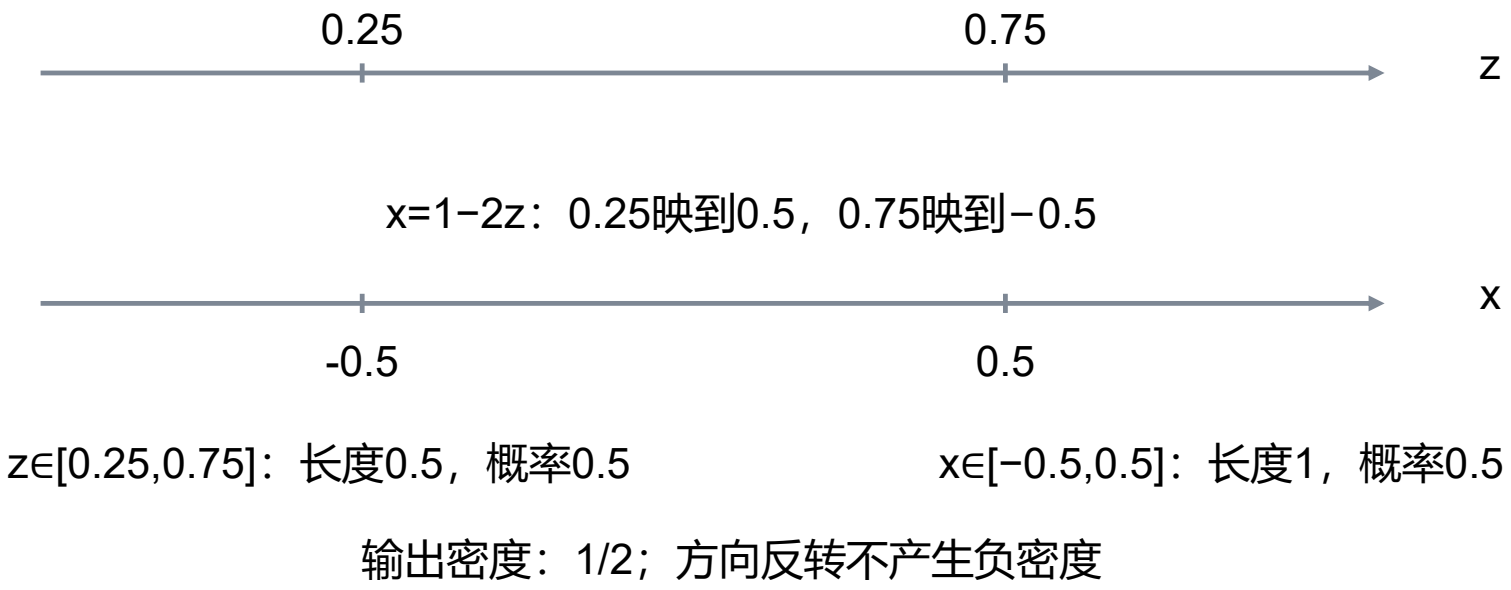
先对应事件，再写积分，最后读出局部密度。 设$X = g(Z)$，g严格递增、可微且导数非零，$a < b$，则$P(a \le Z \le b) = P(g(a) \le X \le g(b))$. 用密度积分表示，并在右侧换元$x = g(z)$：$\int_a^b p_Z(z)dz = \int_{g(a)}^{g(b)} p_X(x)dx = \int_a^b p_X(g(z))g'(z)dz$. 小区间任取，故在密度定义的几乎处处意义下：$p_Z(z) = p_X(g(z))g'(z), p_X(g(z)) = p_Z(z)/g'(z)$.

在上一例$g(z) = 2z + 1$中，导数2就是区间伸长倍数，均匀密度随之除以2。

方向反转解释导数绝对值

若g严格递减，区间端点顺序反转，但长度和概率仍非负。对应的局部长度因子为$|g'(z)|$，所以两种单调方向统一为 $p_X(g(z)) = \frac{p_Z(z)}{|g'(z)|}$.

对$Z \sim \mathrm{Uniform}(0,1)$，取$X = 1 - 2Z$，$g'(z) = -2$。输出区间为$[-1,1]$，密度是$1/2$；若直接除以$-2$会产生负密度。



导数符号描述坐标方向是否反转，绝对值描述非负的长度变化。

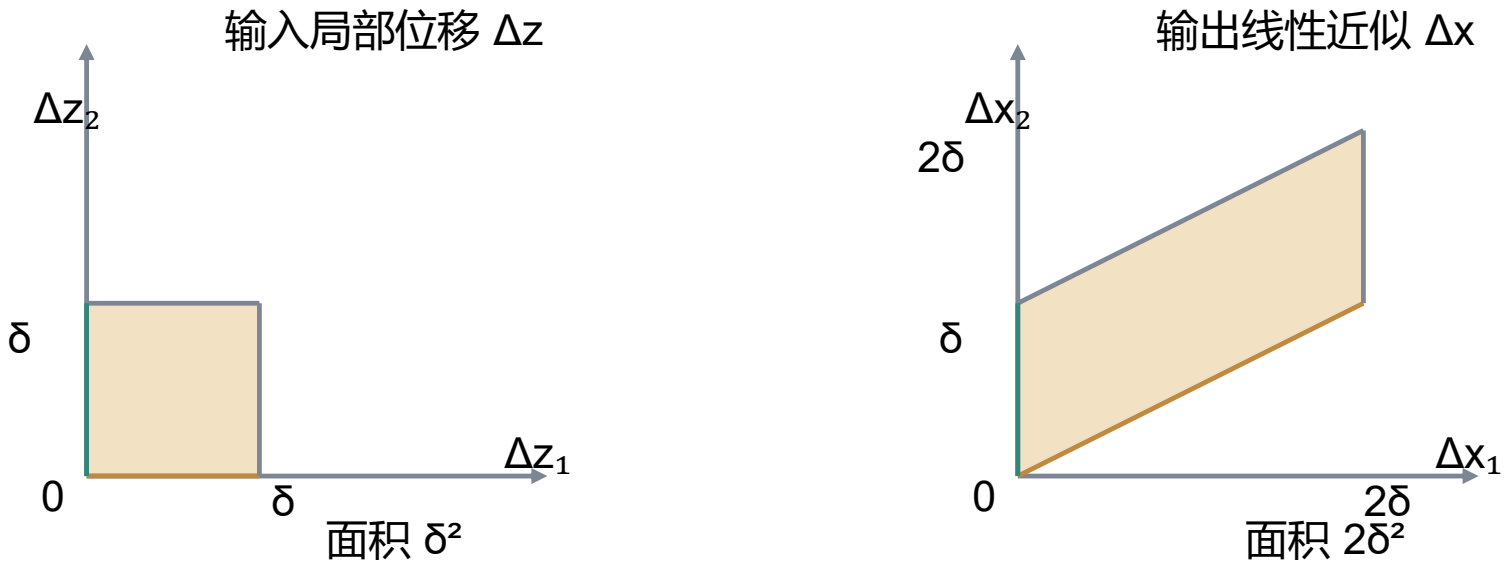
雅可比把小区域的位移与体积联系起来

多维变换在$\boldsymbol{z}$附近可线性近似： $g(\boldsymbol{z}+\Delta\boldsymbol{z})=g(\boldsymbol{z})+\boldsymbol{J}_g(\boldsymbol{z})\Delta\boldsymbol{z}+o(\|\Delta\boldsymbol{z}\|)$. 雅可比每一列描述沿一个输入坐标方向的小位移如何变化。线性变换$\boldsymbol{A}$把单位小方块变为平行多面体，其体积倍数为$|det\boldsymbol{A}|$。

二维局部示例取$\boldsymbol{A}=\begin{pmatrix}2 & 0\\ 1 & 1\end{pmatrix}$，$|det\boldsymbol{A}|=2$。

局部面积随矩阵变换

教学构造取$\boldsymbol{A} = \begin{pmatrix} 2 & 0 \\ 1 & 1 \end{pmatrix}$，把边长$\delta > 0$的正方形变为面积加倍的平行四边形。金色与青色分别对应两条基边的像。



A 的两列：(2,1) 与 (0,1)

对非线性g，小区域体积在极限中按$\left|det\boldsymbol{J}_g\right|$变化。

有限大小区域一般不能用同一个局部雅可比作精确替代。

多维密度与逆方向的对数符号

同维光滑双射使小区域的概率质量相等，局部体积因子给出 $p_Z(\boldsymbol{z})d\boldsymbol{z} = p_X(g(\boldsymbol{z}))|det\boldsymbol{J}_g(\boldsymbol{z})|d\boldsymbol{z}$. 这是变量替换公式的局部解释；完整多维积分定理的分析学证明不在本讲范围。

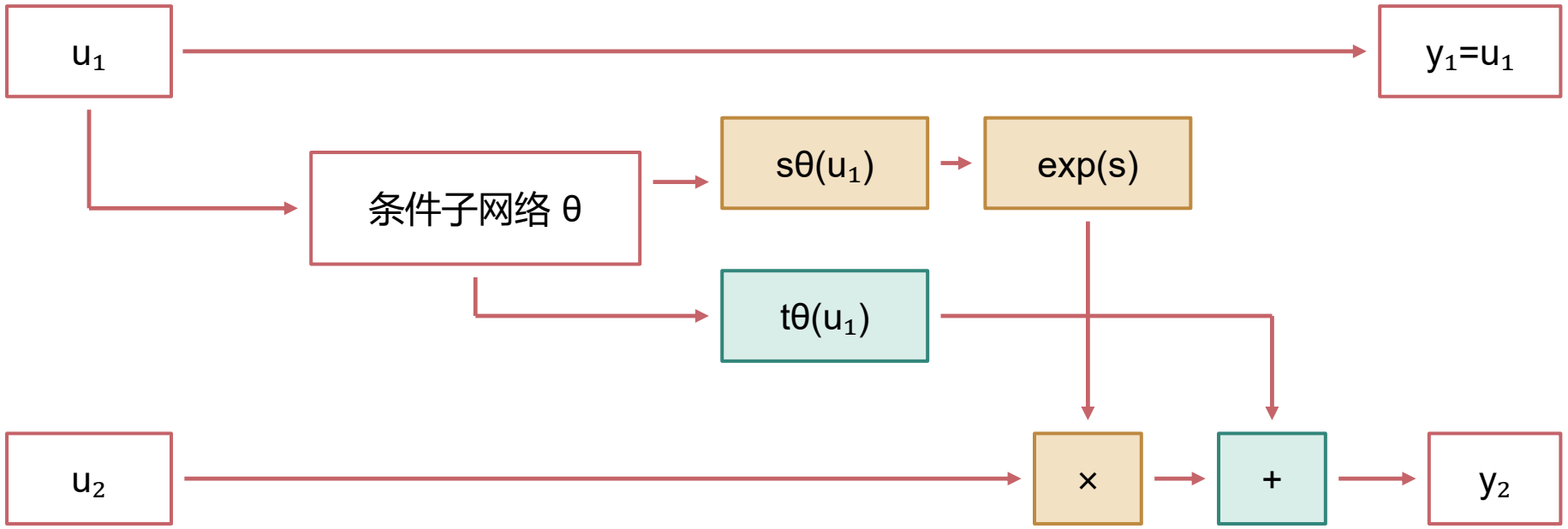
由$g^{-1}(g(\boldsymbol{z})) = \boldsymbol{z}$及链式法则： $\boldsymbol{J}_{g^{-1}}(\boldsymbol{x})\boldsymbol{J}_g(\boldsymbol{z}) = \boldsymbol{I}, \boldsymbol{x} = g(\boldsymbol{z}), |det\boldsymbol{J}_{g^{-1}}(\boldsymbol{x})| = |det\boldsymbol{J}_g(\boldsymbol{z})|^{-1}$. 因此生成方向修正为除法，逆方向修正为乘法；取对数后分别是减logdet与加logdet。

这些关系要求非奇异雅可比及相应连续密度；一般多对一或变维映射不能直接套用。公式成立后，还需要让逆映射与行列式在计算上可行，下一页据此选择耦合结构。

保持一部分输入作为变换条件

怎样保留足够信息，让复杂条件网络仍能组成易求逆的层？ 保留第一块作条件，第二块只执行非零缩放与平移： $y_1 = u_1, y_2 = u_2 \exp(s_\theta(u_1)) + t_\theta(u_1)$. 金色 s_θ 给对数尺度；青色 t_θ 给平移。

固定 u_1 时， $e^{s_\theta(u_1)} > 0$ 改变长度， $t_\theta(u_1)$ 移动位置。



不变分量保存了计算条件函数所需的信息。对子网络自身的可逆性没有要求；它们输出的尺度和平移参与的是可直接反解的仿射运算。

原论文计算图如何保留可恢复的条件

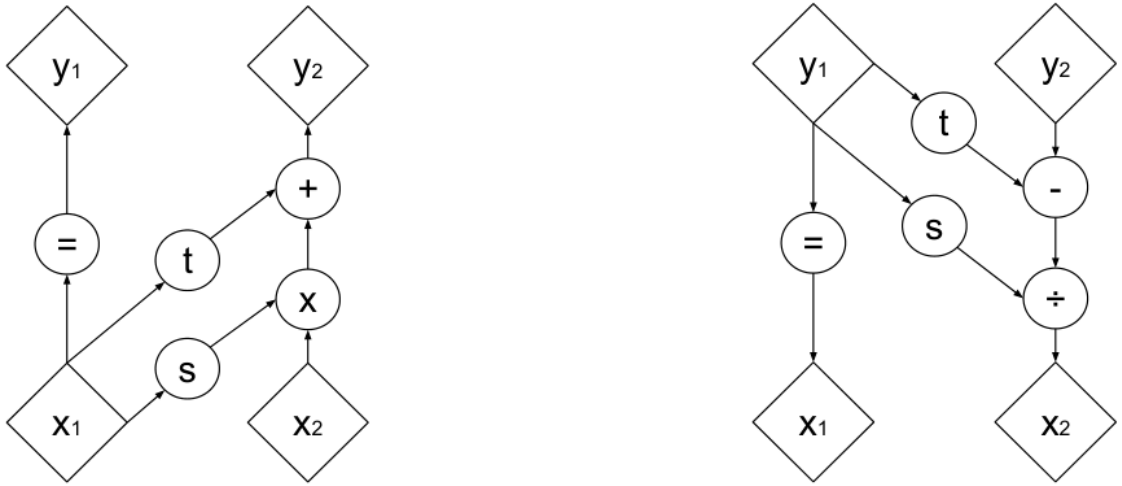
- 图(a)从下向上计算，左支保持不变，并提供尺度与平移
- 图(b)从上向下恢复，先读不变支，再减平移、除尺度

原图输入 (x_1, x_2) 对应本课 (u_1, u_2)，输出仍为 (y_1, y_2)。乘除尺度对应 $e^{s\theta}$ 及其倒数。

原文整体 $f: x \to z$ 对应本课 g^{-1}。(a)的“forward”只指局部耦合正向，不代表整体生成方向。

两个方向使用同一条件值重算条件函数，求逆无需反解条件网络。下一页给出代数过程。

原图来自Dinh等Real NVP v3（2017）Fig.2，PDF第4页；未展开条件网络内部层级。

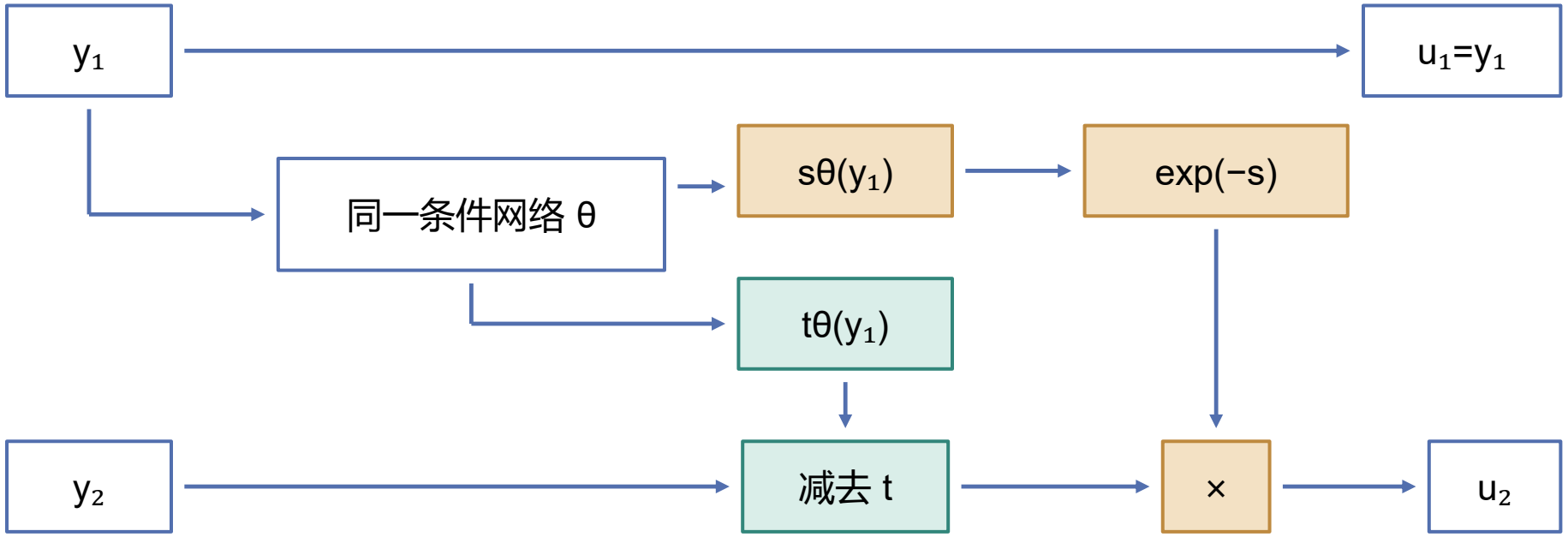


(a) Forward propagation　　(b) Inverse propagation

先恢复条件，再反解简单运算

给定(y_1, y_2)，先由$y_1 = u_1$得到原条件。第二式移项并除以非零尺度：$y_2 - t_\theta(y_1) = u_2 e^{s_\theta(y_1)} \Longrightarrow u_2 = [y_2 - t_\theta(y_1)]e^{-s_\theta(y_1)}$.

因此逆运算为 $u_1 = y_1, u_2 = [y_2 - t_\theta(y_1)]e^{-s_\theta(y_1)}$.



先恢复不变块，再重算条件，最后减平移、除尺度。极端尺度仍可能在浮点计算中溢出或下溢。

三角雅可比使行列式可直接计算

记左下偏导为 $c(\boldsymbol{u})$，先看完整三角矩阵。

$$\boldsymbol{J}_g(\boldsymbol{u}) = \begin{pmatrix} 1 & 0 \\ c(\boldsymbol{u}) & e^{s_\theta(u_1)} \end{pmatrix}.$$

灰色零项：不依赖u_2

金色对角项：固定u_1的尺度

青色左下项：乘法与链式法则

$$c(\boldsymbol{u}) = u_2 e^{s_\theta(u_1)} s'_\theta(u_1) + t'_\theta(u_1).$$

$$det\boldsymbol{J}_g = e^{s_\theta(u_1)}, \log|det\boldsymbol{J}_g| = s_\theta(u_1).$$

分块推广：

$$\begin{pmatrix} \boldsymbol{I} & \boldsymbol{0} \\ \partial \boldsymbol{y}_b / \partial \boldsymbol{u}_a & diag(e^{\boldsymbol{s}_\theta(\boldsymbol{u}_a)}) \end{pmatrix}$$

$$\log|det\boldsymbol{J}_g| = \sum_j s_{\theta,j}(\boldsymbol{u}_a).$$

三角行列式只取对角项，不必显式构造左下块；训练仍需对 s_θ, t_θ 及逆路径求导。

同一二维例核对正向、逆向与密度

设定：$s(u_1) = au_1, a = \ln 2, t(u_1) = u_1^2$。基分布为二维标准高斯，输入 $\boldsymbol{u} = (1, -1)^{\mathsf{T}}$。

01 正向计算

$$y_1 = 1, y_2 = (-1)e^{\ln 2} + 1 = -1.$$

02 逆向恢复

$$u_1 = 1, u_2 = (-1 - 1)e^{-\ln 2} = -1.$$

恢复原输入 $(1, -1)^{\mathsf{T}}$。

03 雅可比与密度修正

$$\boldsymbol{J}_g(1, -1) = \begin{pmatrix} 1 & 0 \\ 2 - 2\ln 2 & 2 \end{pmatrix}, det\boldsymbol{J}_g = 2.$$

$$\log p_Z(1, -1) = -\ln(2\pi) - 1 \approx -2.837877.$$

$$\log p_Y(1, -1) = -2.837877 - \ln 2 \approx -3.531024.$$

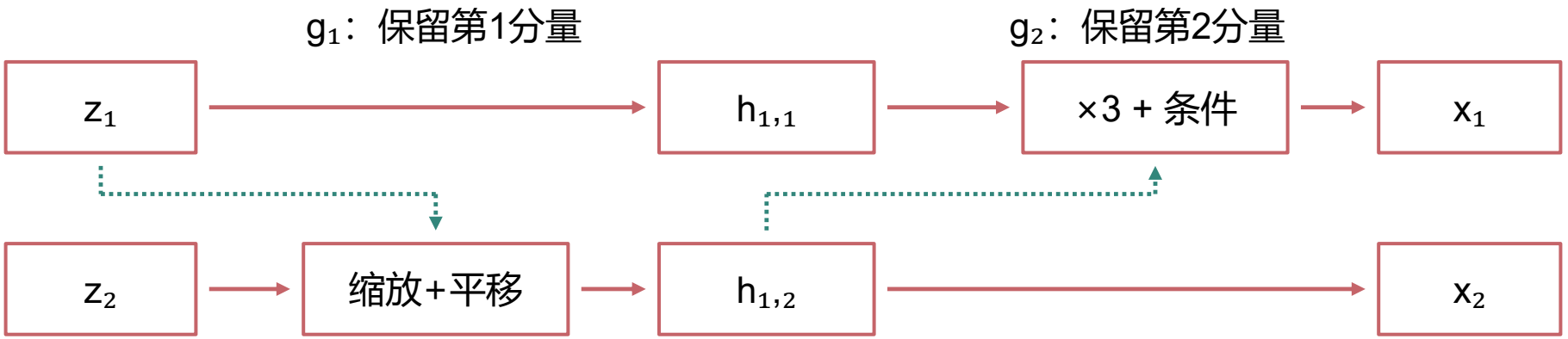
$$p_Y(1, -1) \approx 0.029275.$$

坐标恰好相同，局部面积却扩大2倍，密度仍须除以2。

交替耦合的结构图

前层保留的分量在后层更新，逆向倒序恢复。

第二层取尺度3，以条件分量作平移；后页代入数值。



逆向：$x \rightarrow g_2^{-1} \rightarrow h_1 \rightarrow g_1^{-1} \rightarrow z$

交替更新分量并倒序求逆

单层保留一块分量，以便恢复条件。后续层可换一个划分，使先前不变的分量也被更新。

沿用前例为第一层$\boldsymbol{h}_1 = g_1(\boldsymbol{z})$，再取第二层 $x_2 = h_{1,2}, x_1 = 3h_{1,1} + h_{1,2}$. 第二层保留第二分量、更新第一分量；这是对数尺度ln3、平移为条件分量的耦合。输入$\boldsymbol{z} = (1,-1)^{\mathsf{T}}$先得到$\boldsymbol{h}_1 = (1,-1)^{\mathsf{T}}$，再得到$\boldsymbol{x} = (2,-1)^{\mathsf{T}}$。

逆计算必须从最后一层开始：

1. $h_{1,2} = x_2 = -1$，$h_{1,1} = (x_1 - x_2)/3 = 1$。

2. $z_1 = h_{1,1} = 1$，$z_2 = (h_{1,2} - z_1^2)e^{-az_1} = -1$。

链式法则使logdet累加

01 矩阵次序：第一层导数在右

$$\boldsymbol{J}_g(\boldsymbol{z}) = \boldsymbol{J}_{g_2}(\boldsymbol{h}_1)\boldsymbol{J}_{g_1}(\boldsymbol{z}), \boldsymbol{h}_1 = g_1(\boldsymbol{z}).$$

使用列向量，右侧矩阵先作用于小位移，次序一般不能交换。

02 代入同一例，记 $d_0 = 2 - 2\ln 2$：

$$\boldsymbol{J}_g = \begin{pmatrix} 3 & 1 \\ 0 & 1 \end{pmatrix}\begin{pmatrix} 1 & 0 \\ d_0 & 2 \end{pmatrix} = \begin{pmatrix} 3 + d_0 & 2 \\ d_0 & 2 \end{pmatrix}.$$

$$det\boldsymbol{J}_g = 2(3 + d_0) - 2d_0 = 3\times 2 = 6.$$

03 行列式相乘，logdet相加

$$\log|det\boldsymbol{J}_g| = \log|det\boldsymbol{J}_{g_2}(\boldsymbol{h}_1)| + \log|det\boldsymbol{J}_{g_1}(\boldsymbol{z})|.$$

行列式是标量；推广到 K 层，即逐层累加。

完整密度由基密度与局部修正组成

01 倒序求逆：先 g_2^{-1}，再 g_1^{-1}。

$$\boldsymbol{x} = (2,-1)^{\top} \longrightarrow \boldsymbol{h}_1 = (1,-1)^{\top} \longrightarrow \boldsymbol{z} = (1,-1)^{\top}.$$

02 基密度逐层修正

$$p_Z = \frac{e^{-1}}{2\pi}, p_{H_1} = \frac{p_Z}{2} \approx 0.029275, p_X = \frac{p_{H_1}}{3} \approx 0.009758.$$

$$\log p_\theta(2,-1) = -\ln(2\pi) - 1 - \ln 2 - \ln 3 \approx -4.629637.$$

总修正因子 $2\times3 = 6$，与复合雅可比一致。

03 一般密度求值

1. 倒序求逆，保存各层状态。
2. 在每层输入计算生成方向logdet。
3. 在恢复的 $\boldsymbol{z}$ 求 $\log p_Z$，减去全部logdet。

合法变量替换保持归一化；是否拟合数据仍取决于学习。

固定观测下参数仍影响基密度

01 固定观测，改变第一层参数 a

$\boldsymbol{x} = (2, -1)^{\mathsf{T}}$，第二层固定。先逆第二层得到 $h_{1,1} = 1, h_{1,2} = -1$。

02 逆映射求值点随参数移动

$$z_1 = 1, z_2(a) = (-1 - 1)e^{-a} = -2e^{-a}.$$

基分布形式不变，基密度的求值点仍随 a 改变。

03 同一个目标包含基项与体积项

$$J(a) = -\log p_\theta(2, -1) = \ln(2\pi) + \frac{1}{2}(1 + 4e^{-2a}) + a + \ln 3.$$

前两项为基密度负对数，后两项为两层生成方向logdet。

04 两项梯度共同求导

$$J'(a) = -4e^{-2a} + 1, a = \ln 2 : -1 + 1 = 0, J \approx 4.629637.$$

基项导数为 -1，第一层logdet导数为 1。这是人工例的驻点，不能据此判断一般训练收敛或生成质量。

图像耦合怎样选择条件块

二维向量的“保留一块、更新另一块”，可以按空间位置或通道划分图像张量。

棋盘划分：空间位置

青色：保留条件

金色：更新分量

通道划分：前半 / 后半

前2通道保留，后2通道更新

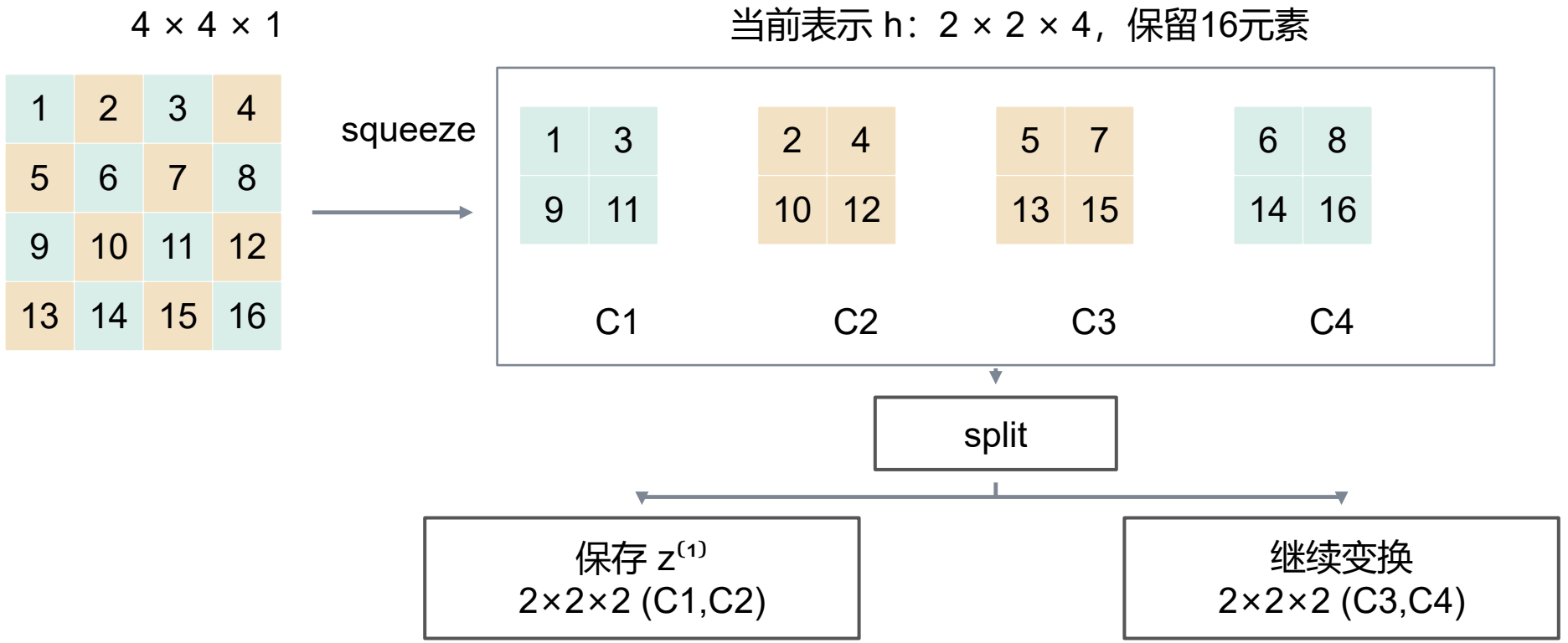
下一层交换掩码，两块都可改变

条件网络读取保留块，计算更新块的 s, t，无需求子网络的逆。交换掩码后，两块都能改变。

据Real NVP §3.4的教学图，非实验输出。

squeeze重排与多尺度变量保留

空间尺寸缩小，可以由元素重排实现，总信息维数保持不变。



$HWC = (H/2)(W/2)4C$，重排logdet为0；两分支均保留。教学图据Real NVP §3.4、§3.6。

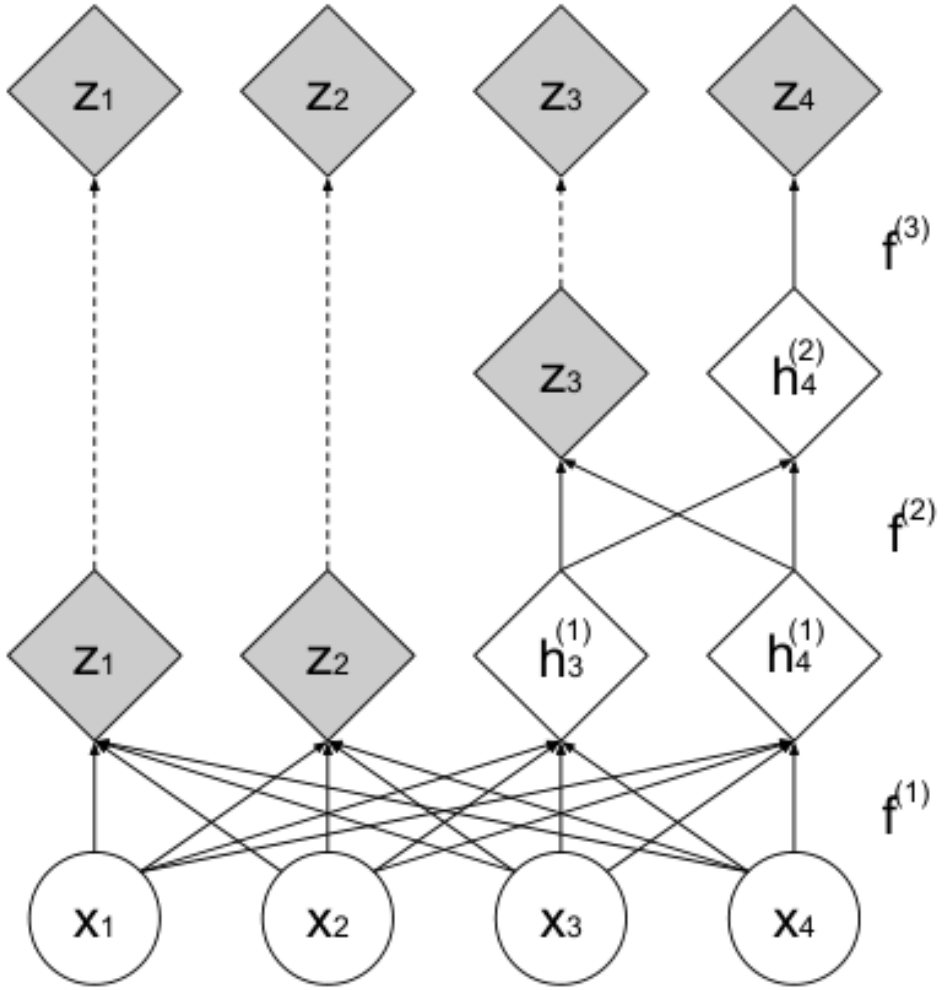
原论文怎样组织多尺度的可逆表示

前面的交替耦合已使各分量参与更新。原Fig.4(b)进一步组织多尺度表示：沿原文 $f: x \to z$ 向上读，逐级变换后分出部分变量，其余变量继续变换。

“分出”仍保存在最终 $z = (z^{(1)}, \ldots, z^{(L)})$ 中，并没有丢弃。生成沿 $g = f^{-1}$ 反向读取全部潜变量块，才能恢复数据。

squeeze将 $s \times s \times c$ 重排为 $(s/2) \times (s/2) \times 4c$，元素总数不变；与交替掩码耦合组成图像架构。

来源：Dinh等，Real NVP v3，Fig.4(b)与§3.6，PDF第6页。该图是多尺度分出示意；中间表示不是可直接观看的图片，也不是完整卷积层表。



3 似然学习与建模边界

用密度接口学习参数，比较训练与生成，说明数据条件和连续流接口。

从数据密度形成参数学习目标

密度公式已给出如何评价观测。记有限训练数据为$\mathcal{D} = \{\boldsymbol{x}^{[n]}\}_{n=1}^{N}$。

每组参数都已定义归一化密度；学习要选择使训练观测整体具有较高密度的参数。最大似然等价于最小化平均负对数密度。

$$J(\theta) = \frac{1}{N}\sum_{n=1}^{N}\left[-\log p_Z(\boldsymbol{z}^{[n]}) + \sum_{k=1}^{K}\log\left|\, det \boldsymbol{J}_{g_k}(\boldsymbol{h}_{k-1}^{[n]})\right|\right], \boldsymbol{z}^{[n]} = g_\theta^{-1}(\boldsymbol{x}^{[n]}).$$

对默认标准高斯，$-\log p_Z(\boldsymbol{z}^{[n]}) = \frac{d}{2}\ln(2\pi) + \frac{1}{2}\|\boldsymbol{z}^{[n]}\|^2$。固定观测$\boldsymbol{x}^{[n]}$并改变$\theta$时，逆映射通常把它送到不同的$\boldsymbol{z}^{[n]}$，因此这个平方范数也会改变。基分布的函数形式固定，不表示基密度项对参数为常数。

logdet项同时记录这些参数对应的体积变化，也参与优化；删除它一般会改变目标。深入节点中的固定观测演算给出两项共同求导的数值核对。

训练更新与生成使用的执行路径

一次训练更新按以下顺序执行：

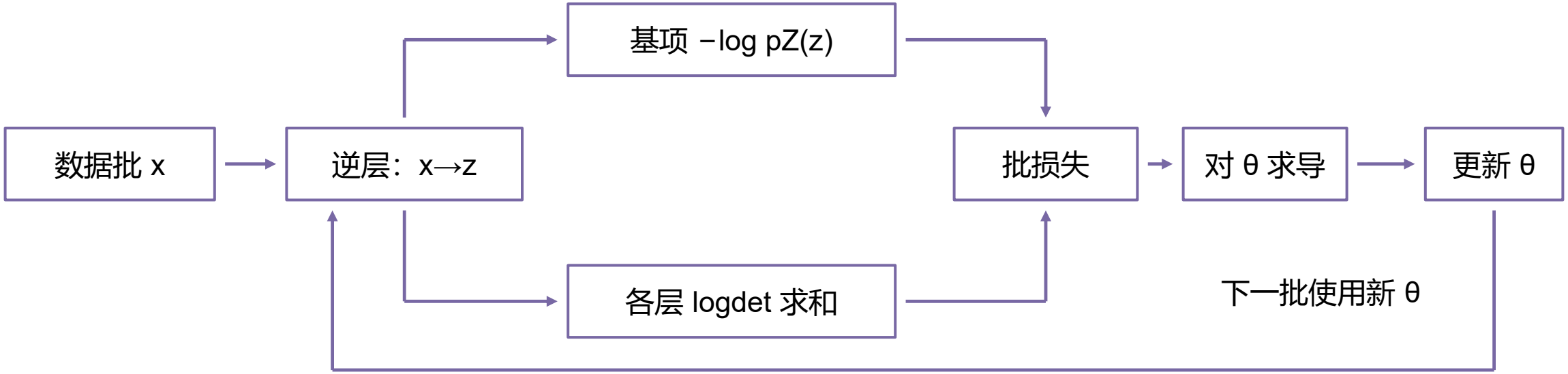
1. 抽取数据批
2. 倒序运行逆层，恢复基变量与中间状态
3. 计算基密度与各层logdet，平均为批损失
4. 对参数求梯度并更新

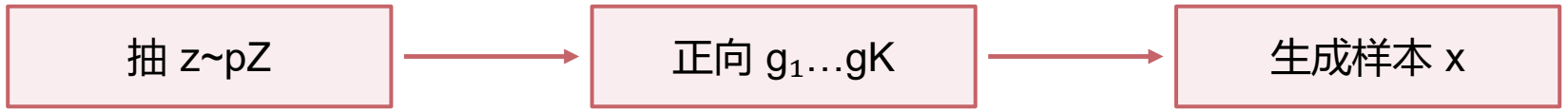
梯度下降写为$\theta \leftarrow \theta - \eta \nabla_\theta \hat{J}_B$，其中$\hat{J}_B$是数据批平均，$\eta > 0$。生成不需要给定待重建样本，也不需要近似后验抽样。

可逆性给出双向定义，实际高效性还依赖求逆、行列式及子网络的计算成本。有限批优化不保证每次提高总体似然或取得全局最优。

训练与生成的流程图

训练用观测数据形成损失并更新参数；生成用固定模型将基分布样本变为新样本。





使用：参数固定；没有损失或更新

原论文插值的开始、中间和最后都是生成终点

训练后的固定模型，能把怎样的潜变量位置映为完整图像?



从左到右 $\phi = 0, \pi/4, \pi/2, 3\pi/4, \pi$，固定 $\phi' = \pi/4$。每格都重新执行完整 $x_\phi = g(z_\phi)$，参数不变。

横向改变的是潜位置；这些图不表示流层间状态或训练进展。下一页分开潜位置选择和层计算。

来源：Dinh等，Real NVP v3，Fig.13，PDF第19页，第2行前5列，CelebA结果摘录，本课未训练。

固定模型怎样产生这些插值终点

外层选潜位置，内层执行完整流。 教学整理据Real NVP式(19)，原文无编号Algorithm。

1. 固定参数 θ、四个编码 $z^{(1)}, \ldots, z^{(4)}$ 与 $\phi' = \pi/4$。令 $a = (z^{(1)} + z^{(2)})/\sqrt{2}$，$b = (z^{(3)} + z^{(4)})/\sqrt{2}$。
2. 依次取 $\phi = 0, \pi/4, \pi/2, 3\pi/4, \pi$，计算

$$z_\phi = \cos(\phi)a + \sin(\phi)b.$$

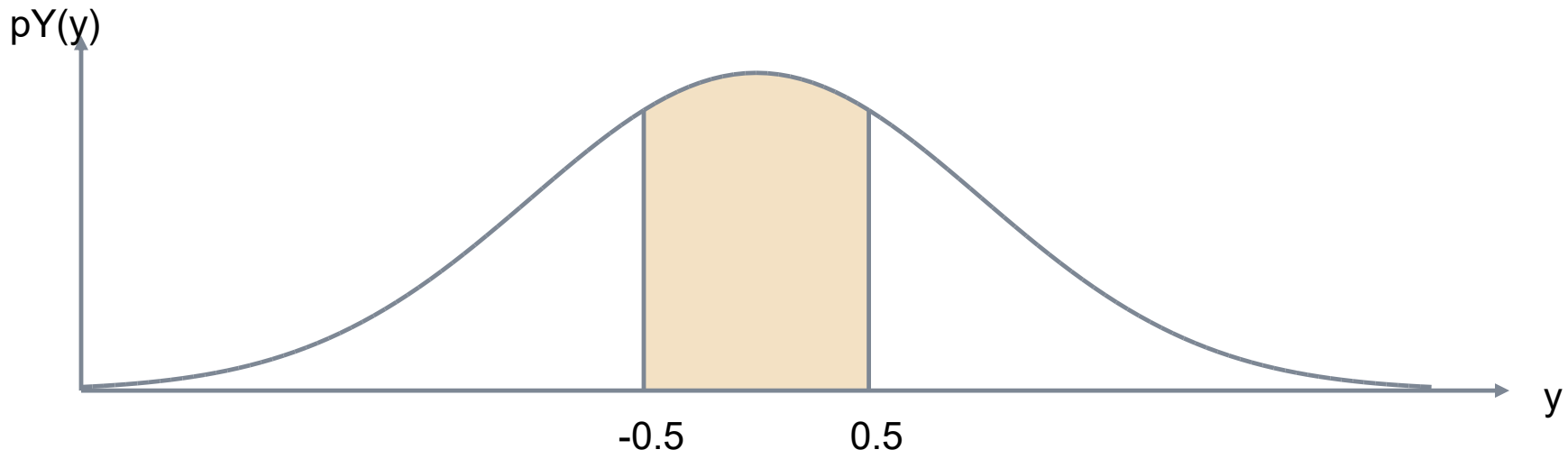
3. 对每个 z_ϕ 重新从 $h_0 = z_\phi$ 开始，依次执行 $h_k = g_k(h_{k-1})$，直到输出 $x_\phi = h_K$。显示图像前，还需逆预处理。

开始、中间、最后的潜变量分别为 $a, b, -a$。三角组合不同于两点线性插值，外观不保证单调变化。

普通生成抽取 $z \sim p_Z$；此处按路径选 z_ϕ，两者都执行固定的完整 g。

连续密度与离散像素概率分别定义

连续密度值描述单位体积的概率强度；一个离散像素取值对应一个量化区间的概率质量。



量化值 v=0 对应阴影面积

曲线高度为密度，区间积分才是概率

$$P(V = v) = \int_{v-1/2}^{v+1/2} p_Y(y)dy.$$

连续变量单点概率为0。多维像素使用量化单元积分；比较似然须统一单位、连续化与预处理。

若 $r = h(y)$ 可逆，则 $p_Y(y) = p_R(h(y))|h'(y)|$；不能把两个变量的密度值直接比较。

分布变换与训练方法分别比较

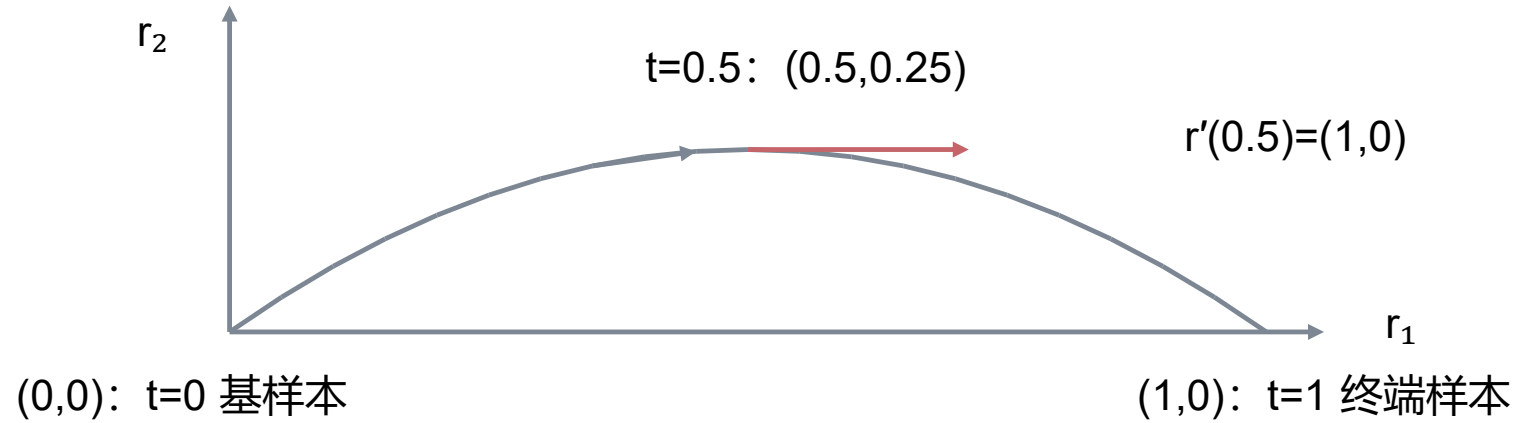
模型接口	观测密度直接可算?	本专题学习信号
VAE	一般需边缘化潜变量	对数证据下界，结合近似推断
GAN	通常无显式可算密度	判别器提供对抗信号
本讲正规化流	连续、同维可逆设定下可算	变量替换给出可计算模型似然

比较限于常见设定：VAE由先验及条件分布生成，GAN由生成映射诱导分布，Flow可逆地推送基分布。精确似然可计算，不意味着视觉质量、模式覆盖或总体建模效果必然更好。

有限层变换与连续轨迹

有限层状态 h_k 由层号 k 索引；连续状态 $r(t)$ 由时间 t 索引。





人为轨迹 r(t)=(t,t(1−t))，非学得场

Flow Matching学习连续流的速度场，训练目标由GM07单独定义。连续密度变化、条件 / 边缘场和ODE/SDE关系不由本讲逐层logdet公式直接替代。

本讲总结

1. **怎样同时定义生成和密度？** 从简单基分布抽样，经同维可逆 g_θ 生成；给定观测先求逆，再用雅可比绝对行列式修正密度。
2. **为什么耦合层支持可计算似然？** 保留条件块可重算 s, t，仿射步骤可反解；三角雅可比只需对角项，复合后logdet逐层累加。
3. **怎样用数据学习参数？** 固定观测，对含基密度与体积项的负对数似然求导，更新模型参数。

连续密度与离散概率需要区间积分联系；潜空间插值的每格都是完整生成终点。

后续将有限层变换推进到连续时间轨迹，并用Flow Matching学习速度场。