07 连续生成与流匹配

从局部速度到概率演化

学习目标

- 区分向量场、轨迹与概率路径，解释随机初值和积分采样。
- 构造噪声—数据端点监督，说明训练和生成各自可用的信息。
- 推导条件平均速度及CFM/FM目标关系。
- 推导连续性、沿轨迹密度变化及所选ODE/SDE对应，核对时间方向。
- 解释Rectified Flow、重新配对及条件潜在系统的职责与边界。

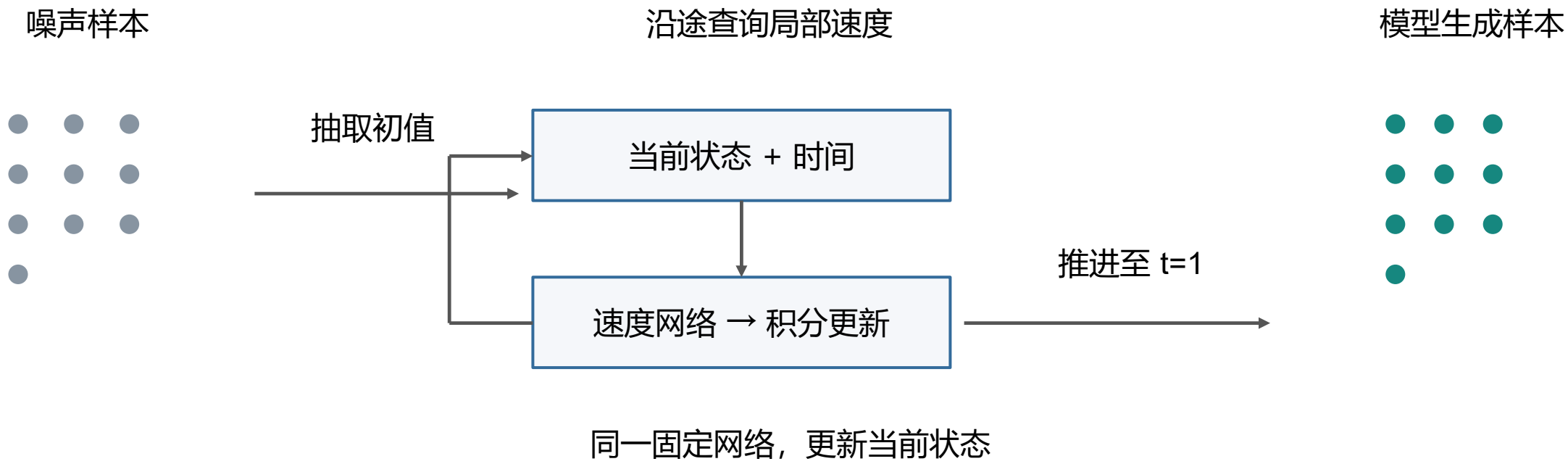
两组推导独立展开；时间不足可保留结论接口，完整材料仍保留中间理由。

1 速度场与端点监督

从局部速度建立ODE、数值采样与端点监督。

生成可以学习沿途速度，再积分到达数据

GM05、06通过逐步去噪生成样本。本讲改用连续时间描述运动：从简单噪声出发，每到一个状态，就查询此时的局部速度。



已知整条运动的终点并非生成时的输入。核心问题是：训练怎样构造速度监督，生成怎样只凭当前状态继续移动？

局部速度形成整条轨迹

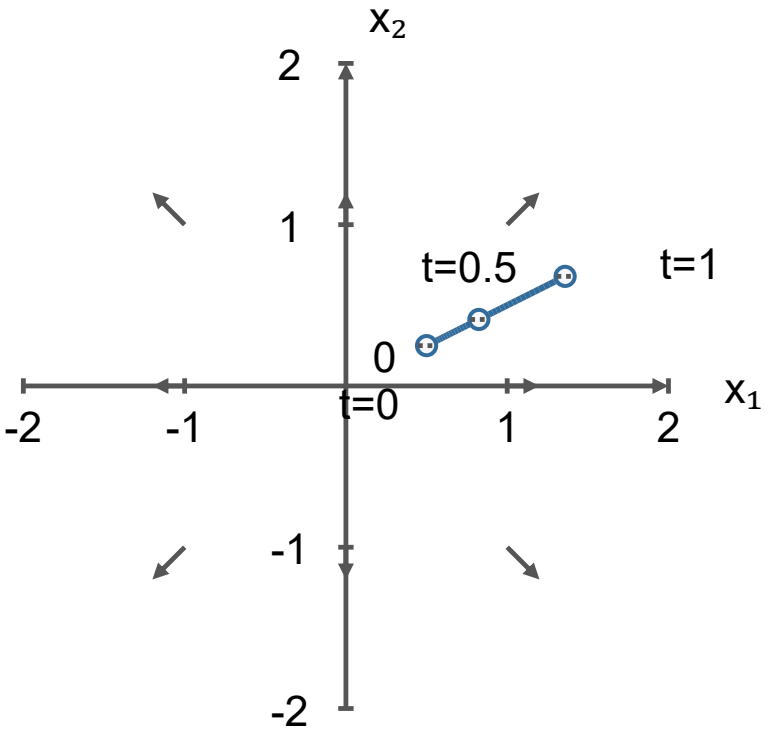
速度场 $\boldsymbol{v}(\boldsymbol{x},t)$为每个位置和时间给出瞬时运动方向及速度。
常微分方程（ODE）规定轨迹满足

$$\frac{d\boldsymbol{Z}_t}{dt} = \boldsymbol{v}(\boldsymbol{Z}_t,t), \boldsymbol{Z}_0 = \boldsymbol{z}_0.$$

轨迹记录同一初值对应的状态随时间的变化；场则在许多位置都有定义。把初值映射到时刻t的解记为**流映射** $\Phi_t(\boldsymbol{z}_0)$。

例如场关于状态全局Lipschitz、关于时间连续并满足适当增长条件时，有限时间初值问题有唯一解。足够光滑条件还可保证流映射可逆。具体网络需核对条件；Euler离散步不自动继承可逆性。

教学构造：v(x,t)=x

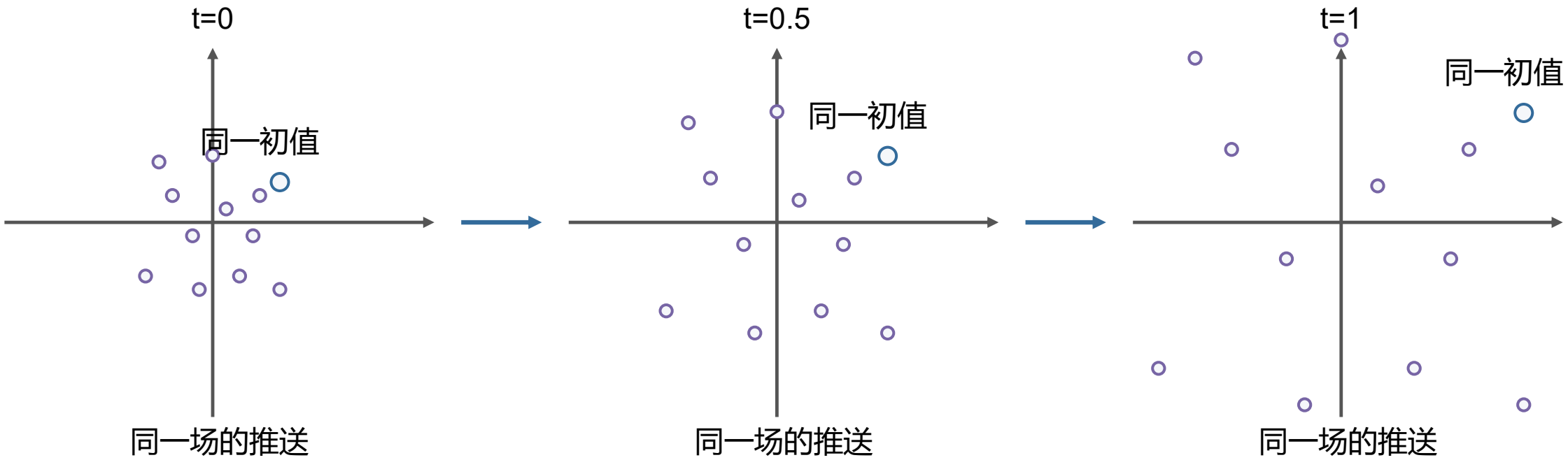


轨迹 z(t)=exp(t)(0.5,0.25)

箭头统一显示为实际速度的0.2倍

随机初值经同一个场形成一族分布

给定一个初值，ODE确定一条轨迹。随机初值经同一流映射，在每个时刻形成一个分布。



蓝点标记同一初值；坐标框间箭头仅表示时间推进

随时间排列的分布族称为**概率路径**：$p_t = (\Phi_t)_{\#} p_0$。推送表示把初值按同一流映射变换后的分布。图用扩张场演示这层关系；它不是已训练的噪声到真实数据模型。

Euler步用当前速度近似短时移动

由导数定义，在小步长h下，

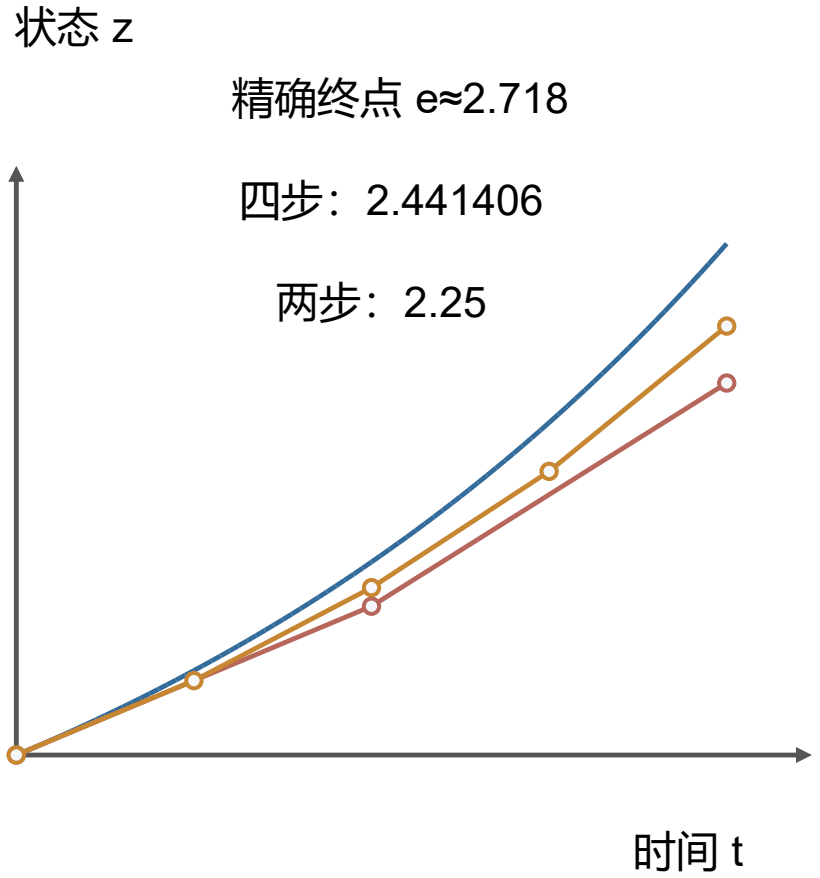
$\boldsymbol{Z}_{t+h} = \boldsymbol{Z}_t + h\boldsymbol{v}(\boldsymbol{Z}_t, t) + o(h)$.

显式Euler方法保留前两项：$\boldsymbol{z}_{k+1} = \boldsymbol{z}_k + h_k\boldsymbol{v}(\boldsymbol{z}_k, t_k)$，其中 $h_k = t_{k+1} - t_k > 0$。

取一维给定场$v(z,t) = z$、$z_0 = 1$，精确解为$Z_t = e^t$。用两个步长0.5：

$z_{0.5} = 1 + 0.5(1) = 1.5, z_1 = 1.5 + 0.5(1.5) = 2.25.$

精确终点$e \approx 2.718282$，差约0.468282。场已完全给定，仍会有离散积分误差。



同一精确场，步数改变离散误差

直线配对给出状态和速度标签

- 本讲大写粗体$\boldsymbol{X}, \boldsymbol{U}, \boldsymbol{Z}$表示d维随机向量或随机轨迹，小写粗体表示其一次取值；此处大写对象不是矩阵
- 训练中独立抽取$\boldsymbol{X}_0 \sim \mathcal{N}(0, I)$和$\boldsymbol{X}_1 \sim p_{\text{data}}$，再独立抽$t \sim \text{Uniform}(0,1)$

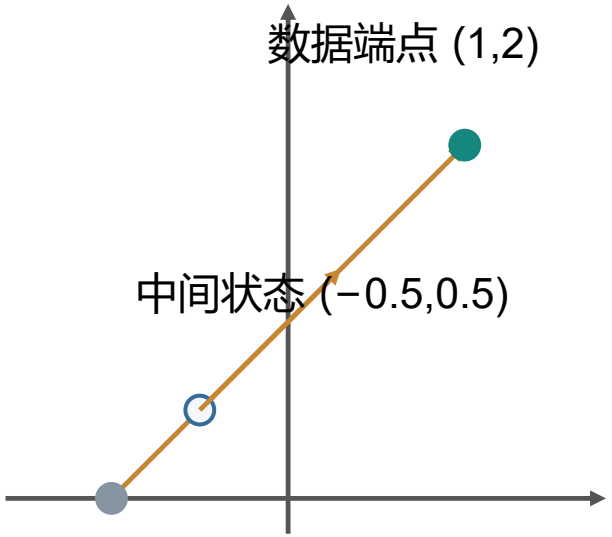
规定

$$\boldsymbol{X}_t = (1-t)\boldsymbol{X}_0 + t\boldsymbol{X}_1, \frac{d\boldsymbol{X}_t}{dt} = \boldsymbol{X}_1 - \boldsymbol{X}_0 =: \boldsymbol{U}.$$

这里导数对时间求，两个端点保持固定。基准配对独立，未声称它是最优运输。

二维构造：$\boldsymbol{x}_0 = (-1,0)^{\top}$、$\boldsymbol{x}_1 = (1,2)^{\top}$。在$t = 0.25$，$\boldsymbol{x}_t = (-0.5,0.5)^{\top}$，速度标签$\boldsymbol{U} = (2,2)^{\top}$。该标签描述已知两端的直线速度。

生成时未知数据端点，不能直接把$\boldsymbol{U}$作为可调用速度。需要学习只以当前状态和时间为输入的$\boldsymbol{v}_{\theta}(\boldsymbol{x}, t)$。



t=0.25；标签 (2,2)

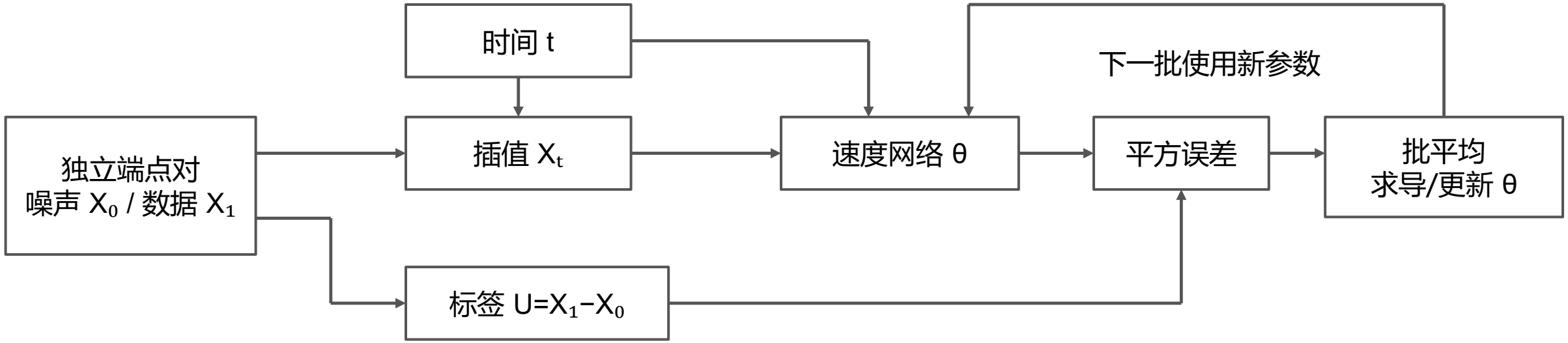
箭头为标签的0.4倍

单时间回归形成一次训练更新

训练最小化**条件流匹配**目标

$$L_{\mathrm{CFM}}(\theta) = E_{t,\boldsymbol{X}_0,\boldsymbol{X}_1}\|\boldsymbol{v}_\theta(\boldsymbol{X}_t,t) - (\boldsymbol{X}_1 - \boldsymbol{X}_0)\|^2.$$

每次取一批独立端点和时间，构造状态及标签，计算批平均损失，对共享参数θ求梯度并更新。一个训练项不需要先积分完整轨迹。

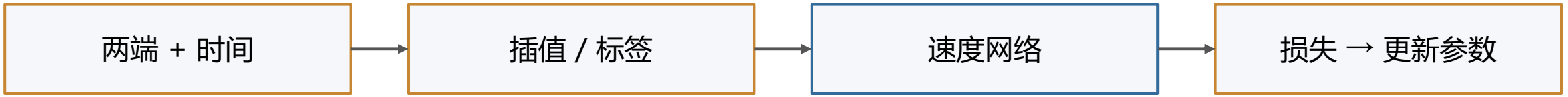


状态与标签分支

若当前预测为$(1.5,2.5)^\top$，前页标签为$(2,2)^\top$，单项损失为0.5，对预测的梯度为$(-1,1)^\top$。参数梯度还需经过网络链式法则。

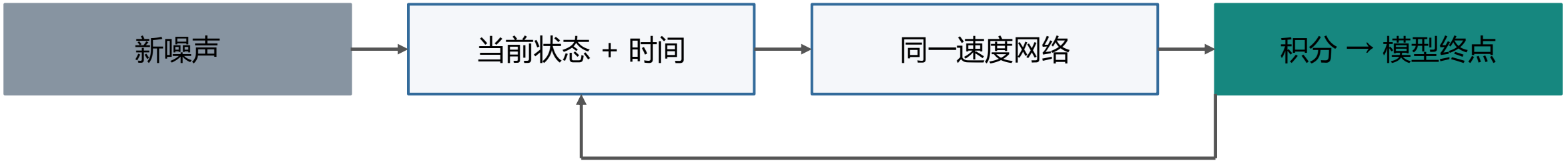
训练知道两端，生成只查询当前状态

训练：两端已知，更新参数



端点差只供给损失；网络接收状态和时间

生成：终点未知，参数固定



更新当前状态，进入下一个时间

训练直线用于构造监督。生成从新噪声出发，重复查询学得的单值速度；参数保持固定，状态随积分改变。生成轨迹通常与任一原始训练直线不同。

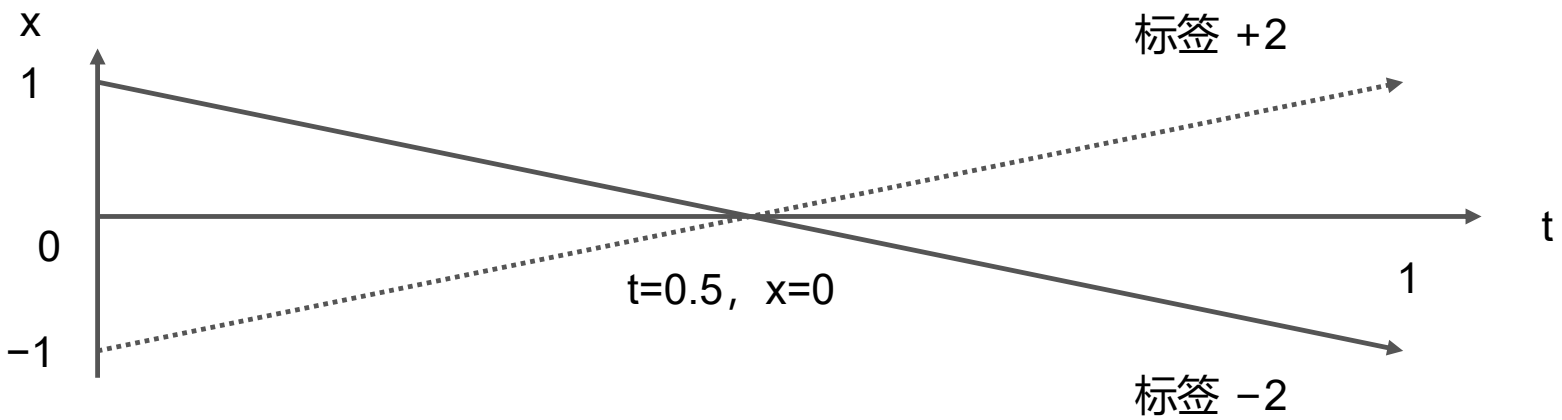
同一中间状态若收到不同速度标签，单值网络应学习哪一个？

直线训练对可以相交，生成场需给出单一速度

生成网络对每个状态只能输出一个速度。直线端点监督的理想回归为

$$\boldsymbol{v}_t^*(\boldsymbol{x}) = E[\boldsymbol{X}_1 - \boldsymbol{X}_0 \mid \boldsymbol{X}_t = \boldsymbol{x}, t].$$

不同配对直线可在同一时刻、同一位置具有不同速度；平均场在该位置给出一个速度。存在唯一解的正规ODE不会允许同一时空点同时沿不同速度继续运动；这与不同轨迹在不同时间经过同一空间位置的情况不同。

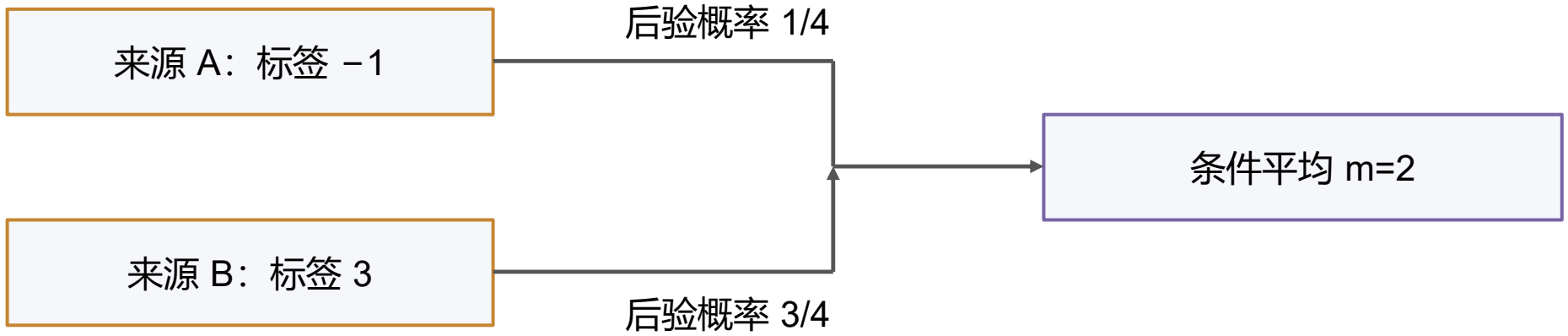


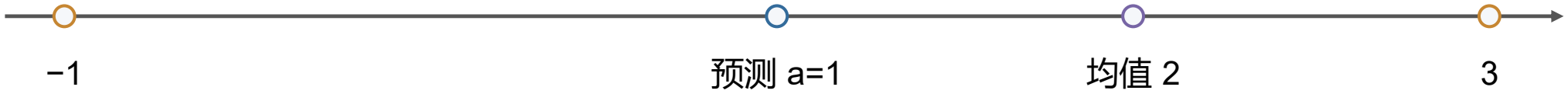
同一状态，两种配对标签；等权平均速度0

图只解释标签与状态函数的区别；这两条人工配对不是本讲独立高斯端点基准的分布示例。原始直线监督并未要求学得生成轨迹逐条复制这些配对。

后验权重决定同一状态下的平均速度

观察到当前状态后，来自不同端点的可能性可以不同。以下是固定状态与时间的一维教学构造：





预测1时：风险4 = 均值偏差1 + 条件方差3

$m = \frac{1}{4}(-1) + \frac{3}{4}(3) = 2, E[(a-U)^2 \mid S] = (a-2)^2 + 3$. 权重针对已知输入$S = (X_t, t)$。平方损失在$a = 2$最小；剩余方差来自同输入的标签差异。下一节推导这些后验权重如何由条件路径得到。

2 条件路径与边缘向量场

推导条件与边缘场；可跳至第3节结果接口。

给定数据端点得到条件高斯路径

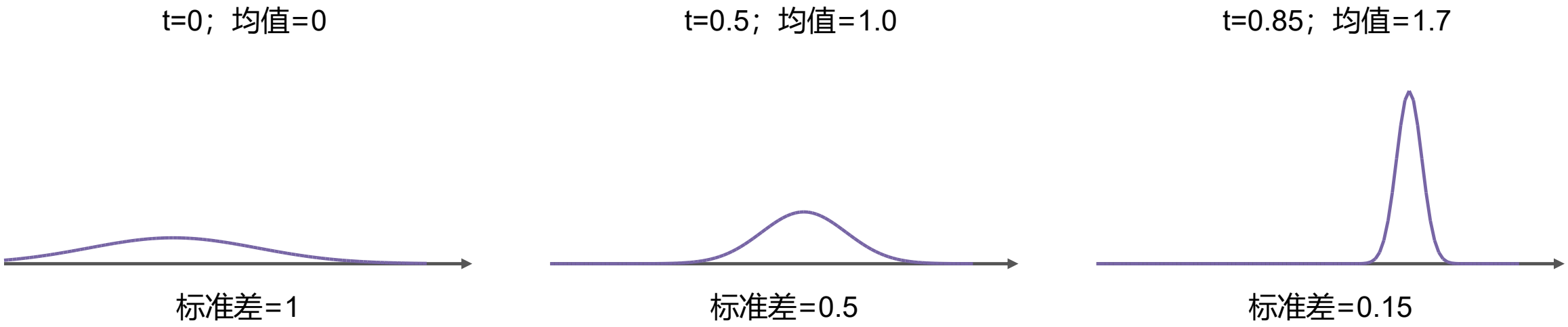
本段为独立深度内容。给定数据端点$\boldsymbol{X}_1 = \boldsymbol{y}$，标准高斯初值使

$$\boldsymbol{X}_t = (1-t)\boldsymbol{X}_0 + t\boldsymbol{y} \Longrightarrow p_t(\boldsymbol{x} \mid \boldsymbol{y}) = \mathcal{N}(t\boldsymbol{y}, (1-t)^2 I), t < 1.$$

解出$\boldsymbol{X}_0 = (\boldsymbol{X}_t - t\boldsymbol{y})/(1-t)$，代回端点差：

$$\boldsymbol{u}_t(\boldsymbol{x} \mid \boldsymbol{y}) = \boldsymbol{y} - \frac{\boldsymbol{x}-t\boldsymbol{y}}{1-t} = \frac{\boldsymbol{y}-\boldsymbol{x}}{1-t}.$$

同一条件路径上的实际速度等于此条件场。到$t = 1$，条件分布集中在端点；含$1/(1-t)$的密度/场表达不能直接代入终点。原始端点差标签仍有定义。



一维 y=2：中心移向端点，条件分布同时收缩

边缘速度平均同一位置的不同来源

混合数据端点得到边缘密度。下式以有密度的数据端点作积分表示；离散端点改用相应求和，一般情形可直接对端点分布取期望：

$$p_t(\boldsymbol{x}) = \int p_t(\boldsymbol{x} \mid \boldsymbol{y}) p_{\text{data}}(\boldsymbol{y}) d\boldsymbol{y}.$$

在$p_t(\boldsymbol{x}) > 0$处，Bayes关系给端点后验权重

$$w_t(\boldsymbol{y} \mid \boldsymbol{x}) = \frac{p_t(\boldsymbol{x}|\boldsymbol{y}) p_{\text{data}}(\boldsymbol{y})}{p_t(\boldsymbol{x})}.$$

它对$\boldsymbol{y}$积分为1。边缘目标速度是

$$\boldsymbol{v}_t^*(\boldsymbol{x}) = \int \boldsymbol{u}_t(\boldsymbol{x} \mid \boldsymbol{y}) w_t(\boldsymbol{y} \mid \boldsymbol{x}) d\boldsymbol{y} = E[\boldsymbol{U} \mid \boldsymbol{X}_t = \boldsymbol{x}, t].$$

条件期望表示：知道中间状态及时间后，按其可能来源的后验概率平均标签。它不是在全部数据端点间无条件等权平均；也不是恢复一个唯一历史端点。



权重由当前状态决定，所有来源的权重合为1

条件概率通量汇总为边缘通量

每条光滑条件路径由其场搬运概率，满足局部质量守恒关系

$$\partial_t p_t(\boldsymbol{x} \mid \boldsymbol{y}) + \nabla \cdot [p_t(\boldsymbol{x} \mid \boldsymbol{y}) \boldsymbol{u}_t(\boldsymbol{x} \mid \boldsymbol{y})] = 0.$$

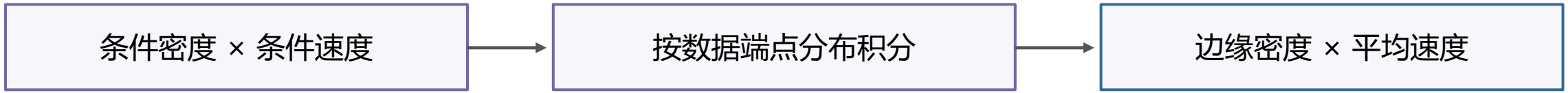
这里$\nabla \cdot$对空间坐标求散度；完整区域守恒依据在后一深度段展开。

假设密度和通量足够光滑、可积，允许时间/空间微分与端点积分交换。对端点平均：

$$\begin{aligned} \partial_t p_t &= -\nabla \cdot \int p_t(\boldsymbol{x} \mid \boldsymbol{y}) \boldsymbol{u}_t(\boldsymbol{x} \mid \boldsymbol{y}) p_{\text{data}}(\boldsymbol{y}) d\boldsymbol{y} \\ &= -\nabla \cdot [p_t(\boldsymbol{x}) \boldsymbol{v}_t^*(\boldsymbol{x})]. \end{aligned}$$

最后一步代入前页的加权平均定义。所得平均场与边缘路径满足同一连续性方程。

要用该场生成这条边缘路径，还需给定初始分布，并满足过程或方程的唯一性条件。



先合并通量，再用连续性方程确定密度变化

平方误差分解给出目标与梯度关系

令输入$S=(\boldsymbol{X}_t,t)$，$\boldsymbol{m}(S)=E[\boldsymbol{U}\mid S]=\boldsymbol{v}_t^*(\boldsymbol{X}_t)$。假设标签和预测有有限二阶矩。将$\boldsymbol{v}_\theta-\boldsymbol{U}=(\boldsymbol{v}_\theta-\boldsymbol{m})+(\boldsymbol{m}-\boldsymbol{U})$平方：

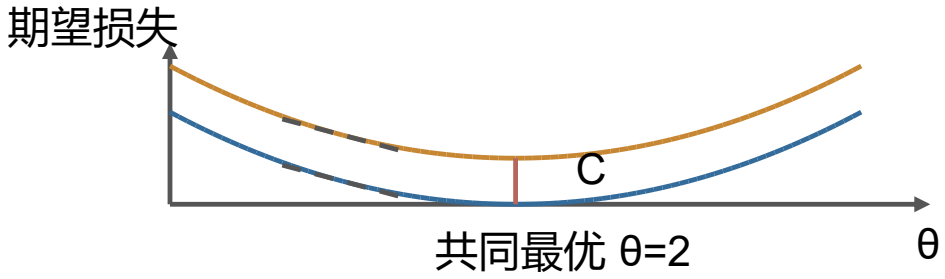
$$E[\|\boldsymbol{v}_\theta-\boldsymbol{U}\|^2\mid S]=\|\boldsymbol{v}_\theta-\boldsymbol{m}\|^2+E[\|\boldsymbol{U}-\boldsymbol{m}\|^2\mid S].$$

交叉项为零，因为在S固定时，预测及均值固定，且$E[\boldsymbol{U}-\boldsymbol{m}\mid S]=0$。再对S平均，将第一项的平均记为流匹配目标$L_{\mathrm{FM}}$：

$$L_{\mathrm{CFM}}(\theta)=L_{\mathrm{FM}}(\theta)+C,C=E\|\boldsymbol{U}-\boldsymbol{m}\|^2\geq 0.$$

配对路径固定时C与参数无关，故两目标具有相同最优参数集合；满足可微及可积控制条件以交换求导与期望后，期望参数梯度也相同。

单个抽样项的值或梯度可以不同。有限网络只能在自身函数族中优化；不保证在所有位置精确恢复目标速度。



蓝：FM；琥珀：CFM（例取 C=2）

同参数处切线平行：可定义的期望梯度相同

图为期望目标；单次抽样梯度仍可能不同

高斯端点例核对平均场与损失常数

一维构造：$X_0 \sim \mathcal{N}(0,1)$、$X_1 \sim \mathcal{N}(2,1)$独立，令$U = X_1 - X_0$。则

$$EX_t = 2t, S_t := Var(X_t) = (1-t)^2 + t^2, EU = 2, Cov(U, X_t) = 2t - 1.$$

联合高斯的条件均值给出

$$v_t^*(x) = 2 + \frac{2t-1}{S_t}(x - 2t).$$

其条件方差为$2 - (2t-1)^2/S_t = 1/S_t$。在固定$t = 1/2$时，$v^* = 2$、条件方差为2；同一中间状态仍可能有不同速度标签。

仅对这个固定时间，取常数预测$v_\theta = \theta$：

$$L_{\mathrm{FM},1/2} = (\theta - 2)^2, L_{\mathrm{CFM},1/2} = (\theta - 2)^2 + 2.$$

- 取$\theta = 1$时，FM损失为1，CFM损失为3；两者期望导数均为-2
- 若一次标签为4，该样本CFM项为9，样本导数为$2(1-4) = -6$。

单次值与相应期望3、-2不同；该例同时核对常数差、期望梯度与样本波动。

3 积分生成与误差

恢复目标接口，完成积分采样并区分三类误差。

训练目标与生成场的结论接口

训练独立抽噪声$\boldsymbol{X}_0$与数据$\boldsymbol{X}_1$，在时间t构造$\boldsymbol{X}_t = (1-t)\boldsymbol{X}_0 + t\boldsymbol{X}_1$和标签$\boldsymbol{U} = \boldsymbol{X}_1 - \boldsymbol{X}_0$。网络输入仅为当前状态与时间。

平方损失的理想回归为

$$\boldsymbol{v}_t^*(\boldsymbol{x}) = E[\boldsymbol{U} \mid \boldsymbol{X}_t = \boldsymbol{x}, t].$$

它按当前状态的可能来源平均速度。有限二阶矩下，条件标签回归与直接回归该平均场的期望损失相差参数无关常数；合法交换求导和期望后，两者期望参数梯度相同。具体分布演化还需路径/场正则与唯一性条件。

理想场从匹配的初始高斯分布出发，在正则与唯一性条件下运输规定的插值边缘。若流可延拓到终点，终端分布为数据分布。

训练后的$\boldsymbol{v}_\theta$近似该函数。生成抽$\boldsymbol{Z}_0 \sim p_0$，解

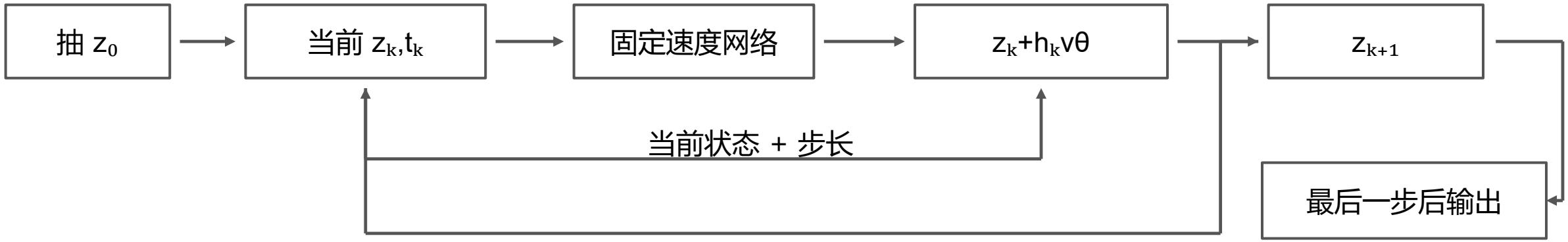
$$\dot{\boldsymbol{Z}}_t = \boldsymbol{v}_\theta(\boldsymbol{Z}_t, t), 0 \le t \le 1.$$

此时未知真实数据端点；每步使用实际积分所得当前状态。训练直线只是构造监督，通常不是网络生成的逐样本轨迹。

按时间网格使用同一速度网络

选取$0 = t_0 < t_1 < \cdots < t_K = 1$，令$h_k = t_{k+1} - t_k$。基本Euler采样算法为：

1. 抽取$\boldsymbol{z}_0 \sim p_0$，固定训练完成的参数$\theta$。
2. 对$k = 0, \ldots, K-1$，计算$\boldsymbol{v}_\theta(\boldsymbol{z}_k, t_k)$。
3. 更新$\boldsymbol{z}_{k+1} = \boldsymbol{z}_k + h_k \boldsymbol{v}_\theta(\boldsymbol{z}_k, t_k)$。
4. 输出$\boldsymbol{z}_K$；潜在生成系统还需按既定尺度解码。

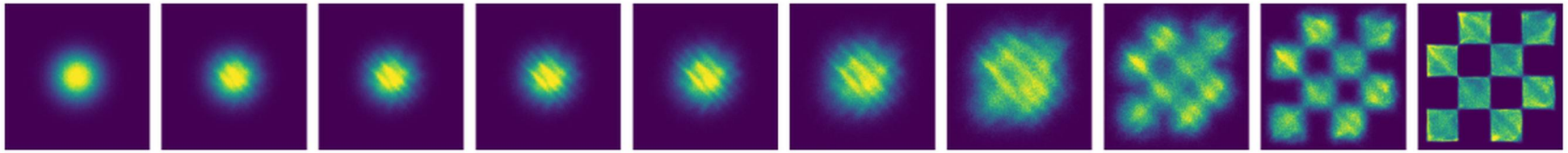


下一时间，同一固定参数网络

路径及配对设计决定要学习的运动，网络拟合决定速度近似程度，求解器决定离散误差。即使速度网络精确，有限步Euler也可能有误差；减小回归损失不单独保证所有采样设置下的生成质量。

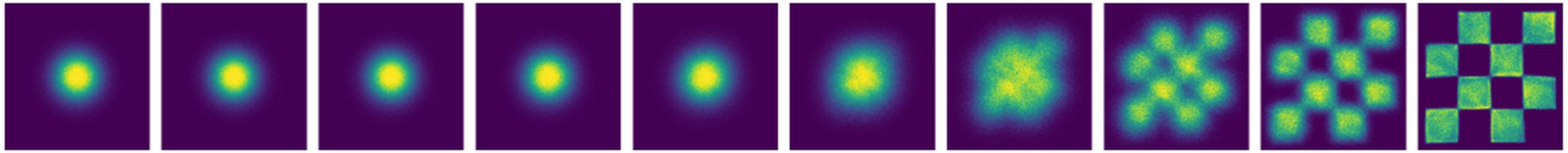
原论文密度路径显示整体分布逐渐成形

①
得分匹配
扩散路径



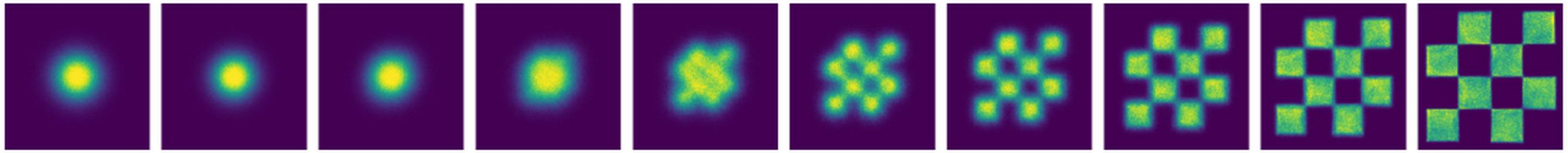
Score matching ^{w}/ Diffusion

② FM
扩散路径



Flow Matching ^{w}/ Diffusion

③ FM
OT路径



Flow Matching ^{w}/ OT

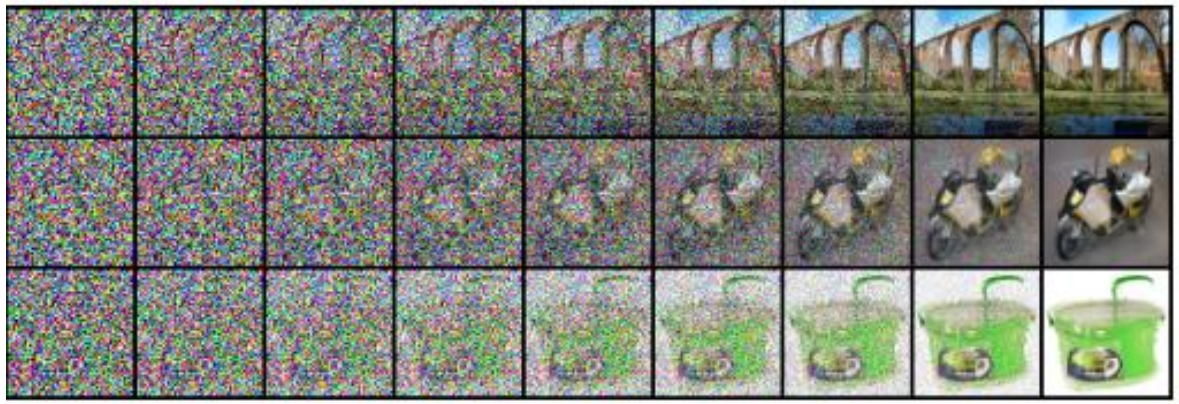
从左到右：噪声集中 → 中间分开 → 最后形成棋盘。每幅为样本总体密度，不是单样本轨迹。
OT指训练条件路径，不保证边缘运输全局最优。
作者结果，非本课实验。来源：2210.02747v2 Fig.4左。

原论文图像采样展示同一初值的中间状态

每行从左到右是一次FM/OT采样。起点为噪声。中间仍含噪，图像结构逐渐显现。终点为模型图像。三行是独立样本，不能沿纵向连接。

图中状态由学得速度场积分得到，并非已知噪声与真实图像的线性混合。训练直线标签不要求采样沿直线进行。

原例来自2210.02747v2，§6.2及Fig.6右组，为作者ImageNet64结果，不代表本课复现或通用演化速度。



Flow Matching w/ OT

4 连续生成与扩散的联系

推导连续密度及ODE/SDE；可跳至第5节接口。

区域质量变化由边界概率通量决定

前面追踪单个样本；现在解释大量样本运动为何改变整个分布。取固定空间区域A，密度p_t与场$\boldsymbol{v}_t$足够光滑，可交换时间求导与积分。**概率通量**为$p_t\boldsymbol{v}_t$，表示概率质量的局部运输。

区域内无质量产生或消失时，向外的净通量使区域质量减少：

$$\frac{d}{dt}\int_A p_t(\boldsymbol{x})d\boldsymbol{x} = -\int_{\partial A} p_t\,\boldsymbol{v}_t \cdot \boldsymbol{n}dS.$$

$\boldsymbol{n}$为外法向单位向量。

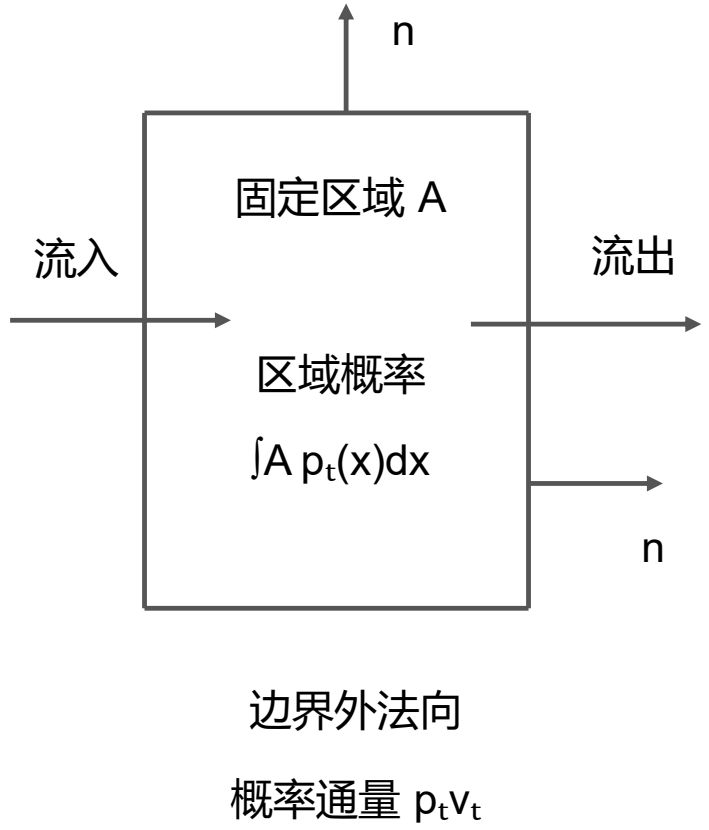
对规则边界应用散度定理，右侧等于$-\int_A \nabla\cdot(p_t\boldsymbol{v}_t)d\boldsymbol{x}$。

对任意适当区域均成立，得到

$$\partial_t p_t + \nabla\cdot(p_t\boldsymbol{v}_t) = 0.$$

这就是**连续性方程**，描述局部守恒。

总质量守恒还需边界条件：无穷远通量充分衰减，或所选边界的净通量为零。



沿轨迹的全导数消去平移项

对光滑场，

$$\nabla \cdot (p_t \boldsymbol{v}_t) = \boldsymbol{v}_t \cdot \nabla p_t + p_t \nabla \cdot \boldsymbol{v}_t.$$

连续性方程因此给出

$$\partial_t p_t + \boldsymbol{v}_t \cdot \nabla p_t = -p_t \nabla \cdot \boldsymbol{v}_t.$$

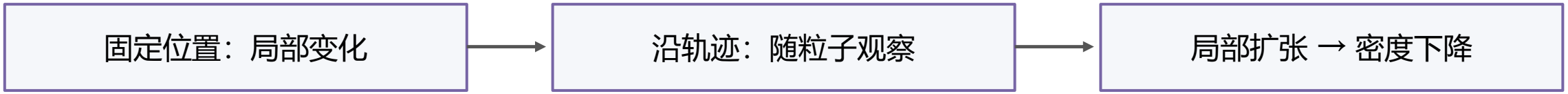
另一方面，轨迹满足$\dot{\boldsymbol{Z}}_t = \boldsymbol{v}_t(\boldsymbol{Z}_t)$。链式法则表明

$$\frac{d}{dt} p_t(\boldsymbol{Z}_t) = \partial_t p_t(\boldsymbol{Z}_t) + \nabla p_t(\boldsymbol{Z}_t) \cdot \dot{\boldsymbol{Z}}_t.$$

左侧跟随粒子移动，右侧第一项固定空间位置。合并后，在$p_t(\boldsymbol{Z}_t) > 0$处除以密度：

$$\frac{d}{dt} \log p_t(\boldsymbol{Z}_t) = -\nabla \cdot \boldsymbol{v}_t(\boldsymbol{Z}_t).$$

散度$\sum_i \partial\, v_i / \partial x_i$描述局部净扩张；正散度对应沿运动粒子观察到的密度降低。



链式法则加入平移项，再用连续性方程消去；剩余项为负散度

积分联系端点密度与局部体积

沿精确轨迹积分得到

$$\log p_1(\boldsymbol{Z}_1) = \log p_0(\boldsymbol{Z}_0) - \int_0^1 \nabla \cdot \boldsymbol{v}_t(\boldsymbol{Z}_t) dt.$$

它是GM04有限层体积修正的连续对应。成立于所定义的流及其推送密度；把学得场代入时计算的是模型密度，不是已知真实数据密度。

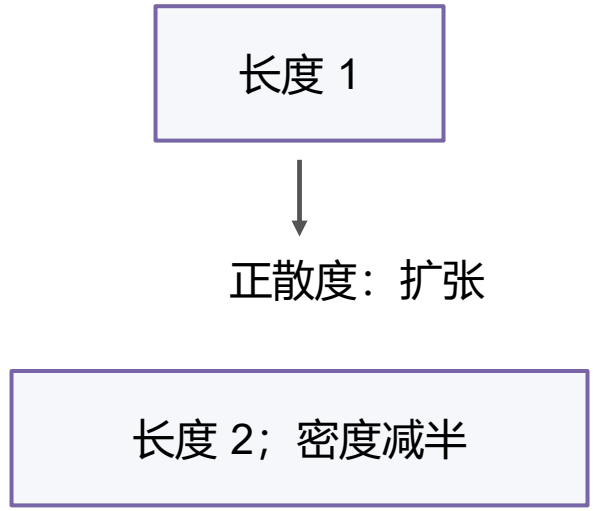
一维构造取$v(z,t) = \kappa z$。轨迹$Z_t = e^{\kappa t} Z_0$，导数$v'(z) = \kappa$，故沿轨迹

$$\log p_t(Z_t) = \log p_0(Z_0) - \kappa t.$$

变量替换也给出$p_t(z) = e^{-\kappa t} p_0(e^{-\kappa t} z)$，两条路径一致。

令$\kappa = \ln 2, Z_0 = 1, p_0 = \mathcal{N}(0,1)$。终点为2，局部长度翻倍，$\log p_1(2) = \log p_0(1) - \ln 2 \approx -2.112086$。计算高维散度及实际积分有成本，相关估计器和实现留作研究阅读。

同一小区间的概率质量保持



反向压缩：负散度，密度上升

质量 = 密度 × 长度（局部）

加噪SDE使密度同时漂移和扩散

另用加噪时间$\tau \in [0,T]$，$\tau = 0$为数据、T为噪声。选择状态无关标量扩散的Itô随机微分方程（SDE）：

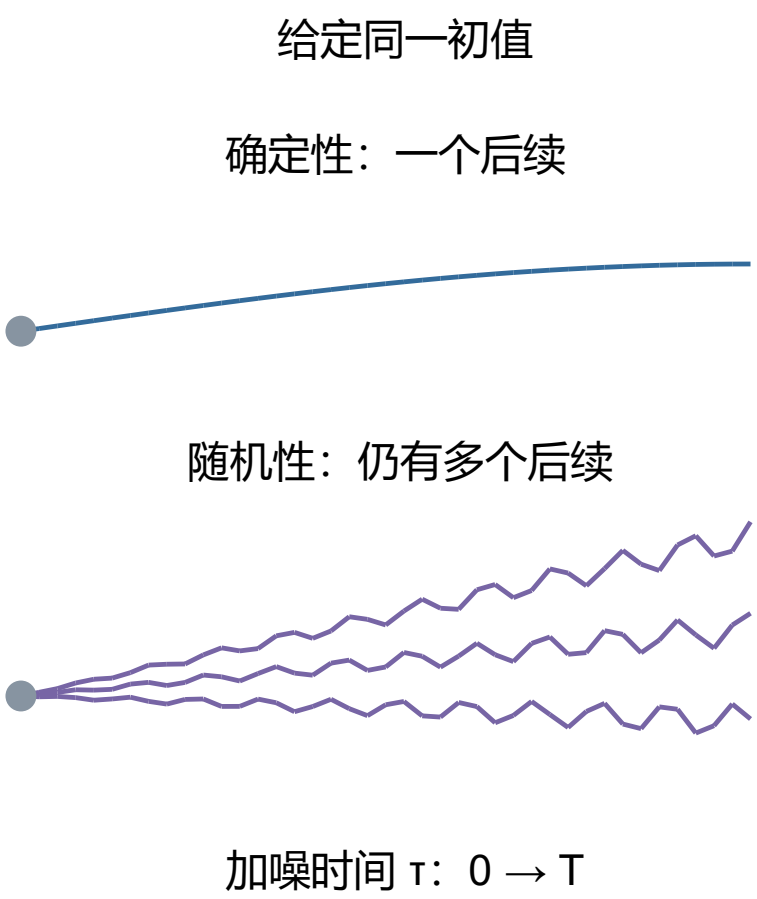
$$d\boldsymbol{Y}_\tau = \boldsymbol{f}(\boldsymbol{Y}_\tau, \tau)d\tau + g(\tau)d\boldsymbol{W}_\tau.$$

$\boldsymbol{f}$是漂移，$\boldsymbol{W}$是标准布朗运动，其独立增量在长度$\Delta\tau$上的协方差为$\Delta\tau I$。g缩放随机扰动。

在适当正则条件下，边缘密度q_τ满足**Fokker–Planck方程**：

$$\partial_\tau q_\tau = -\nabla \cdot (q_\tau \boldsymbol{f}) + \frac{1}{2} g(\tau)^2 \Delta q_\tau, \Delta q = \sum_i \partial^2 q / \partial x_i^2.$$

第一项为漂移运输，第二项为随机扩散。这里引用所选SDE的密度演化定理，不证明全部随机分析结果；要求密度和系数足够正则、边界行为适当，并保证所讨论过程/方程的存在唯一性。



路径形态为概念示意，非数值模拟

把扩散通量并入漂移得到半系数

定义得分$\boldsymbol{r}_\tau(\boldsymbol{x}) = \nabla_{\boldsymbol{x}} \log q_\tau(\boldsymbol{x})$。密度正且光滑时，$q_\tau \boldsymbol{r}_\tau = \nabla q_\tau$。因g与状态无关，

$$\frac{1}{2} g^2 \Delta q_\tau = \nabla \cdot \left(\frac{1}{2} g^2 q_\tau \boldsymbol{r}_\tau \right).$$

于是

$$\partial_\tau q_\tau = -\nabla \cdot \left[q_\tau \left(\boldsymbol{f} - \frac{1}{2} g^2 \boldsymbol{r}_\tau \right) \right].$$

这正是速度$\boldsymbol{b}_\tau^{\mathrm{PF}} = \boldsymbol{f} - \frac{1}{2} g^2 \boldsymbol{r}_\tau$的连续性方程。对应的**概率流ODE**为

$$d\boldsymbol{Y}_\tau = \left[\boldsymbol{f} - \frac{1}{2} g^2 \boldsymbol{r}_\tau \right] d\tau.$$

相同边缘分布允许不同的样本路径

- 从相同起始密度出发，使用真实得分且满足适当正则/唯一性条件，才可建立以下对应
- 满足上述条件时，ODE与SDE具有相同单时刻边缘密度；它们的轨迹和路径联合分布通常不同
- 状态相关或矩阵扩散需额外项，不能直接套用此简化式

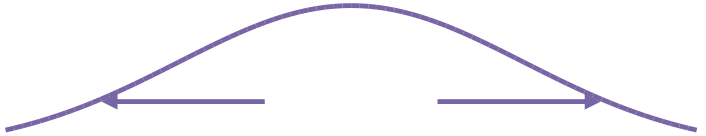
总通量可以写成

$$\boldsymbol{J}_\tau = q_\tau \boldsymbol{f} - \frac{1}{2} g^2 \nabla q_\tau = q_\tau (\boldsymbol{f} - \frac{1}{2} g^2 \boldsymbol{r}_\tau).$$

图用同一高斯密度、零漂移示例，只比较通量方向；相同边缘演化允许不同样本路径。

SDE的概率通量

漂移通量 + 扩散通量

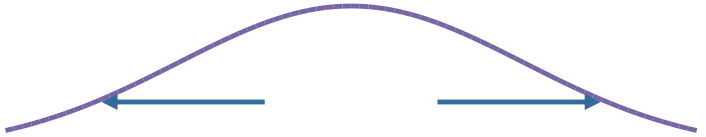


向低密度扩散

改写总通量

PF-ODE的确定运输

密度 × 修正速度



确定速度向外运输

反向SDE与生成时间一起转换

适用时间反演定理时，沿$\tau: T \to 0$运行的反向SDE为

$d\boldsymbol{Y}_\tau = [\boldsymbol{f} - g^2\boldsymbol{r}_\tau]d\tau + gd\boldsymbol{W}_\tau,$

其中$d\tau < 0$，$\boldsymbol{W}$按反向时间定义。这里得分系数为1，区别于PF-ODE的1/2。

为与本讲生成方向一致，令$t = 1 - \tau/T$，即$\tau = T(1-t)$。把状态改记为$\boldsymbol{Z}_t$，正向生成时间上的概率流ODE和反向SDE分别为：

PF-ODE: $d\boldsymbol{Z}_t = \left[-T\boldsymbol{f} + \frac{T}{2}g^2\boldsymbol{r}_\tau\right]dt,$

Reverse SDE: $d\boldsymbol{Z}_t = [-T\boldsymbol{f} + Tg^2\boldsymbol{r}_\tau]dt + \sqrt{T}gd\boldsymbol{B}_t.$

全部系数在$(\boldsymbol{Z}_t, \tau = T(1-t))$求值；$\boldsymbol{B}_t$为生成时间的标准布朗运动。

漂移因$d\tau = -Tdt$转换；布朗增量按方差T缩放，幅度是$\sqrt{T}$。空间得分只改变其时间求值，没有额外负号。初始分布应为相应终端q_T；近似先验或网络得分、离散积分带来额外误差。



高斯布朗例核对两种系数

取一维可解析构造：$T = 1, Y_0 \sim \mathcal{N}(0,1)$，加噪过程$dY_\tau = dW_\tau$。独立高斯增量使

$q_\tau = \mathcal{N}(0, 1+\tau), r_\tau(y) = -y/(1+\tau).$

其终端为已知$\mathcal{N}(0,2)$；这是教学模型，不把它当作任意真实数据加噪后的通用先验。

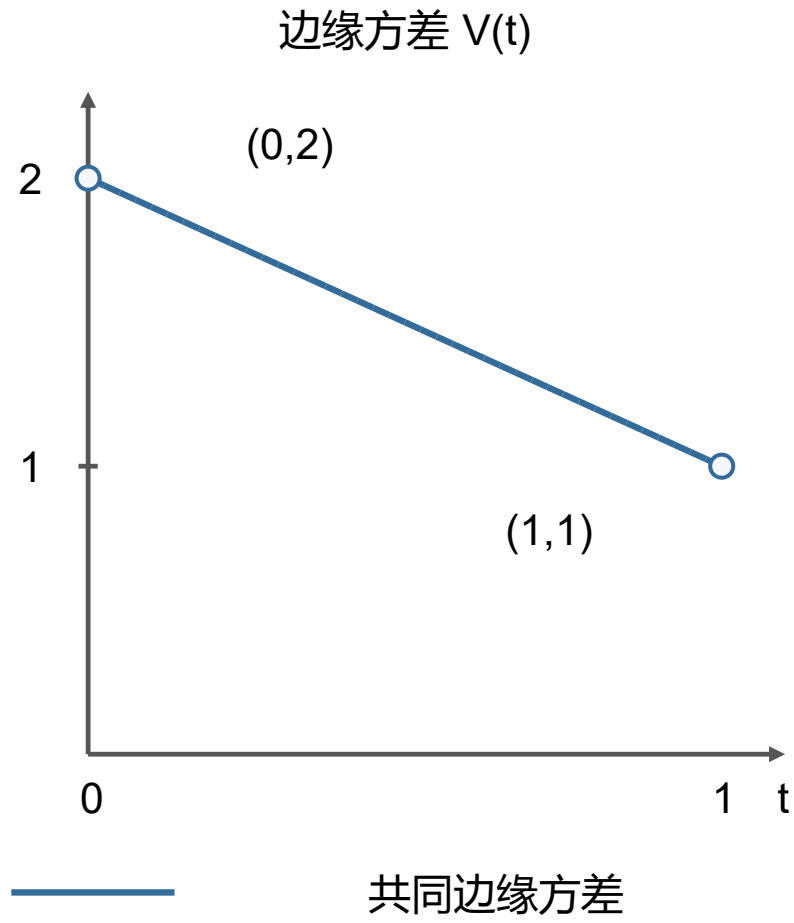
转为$t = 1 - \tau$，从$Z_0 \sim \mathcal{N}(0,2)$生成：

ODE:$dZ_t = -\frac{Z_t}{2(2-t)}dt$, SDE:$dZ_t = -\frac{Z_t}{2-t}dt + dB_t.$

ODE解为$Z_t = \sqrt{(2-t)/2}Z_0$，方差为$2-t$。SDE的方差方程为

$V'(t) = -\frac{2}{2-t}V(t) + 1.$

代入$V = 2 - t$，右侧为-1，与左侧一致。因此两者方差都从2降到1，且边缘均为零均值高斯。ODE在初值固定后无新随机性，SDE仍有布朗扰动；相同边缘没有使两条路径相同。



PF-ODE与反向SDE均满足 V(t)=2−t

同一高斯边缘；样本路径仍不同

5 连续接口、重新配对与系统

连接连续结论、重新配对与条件系统，回收本讲。

密度、得分和时间方向的使用条件

① ODE运输概率 连续性方程描述运输	② 指定扩散及其PF-ODE 条件满足时同边缘	③ 改用生成时间 须转换方向和系数

结果接口：完整证明见P25—32；核心路线按下列公式与条件使用

① 光滑场与正密度，边界适当且解唯一：

$$\partial_t p_t = -\nabla \cdot (p_t \boldsymbol{v}_t),$$
$$\frac{d}{dt}\log p_t(\boldsymbol{Z}_t) = -\nabla \cdot \boldsymbol{v}_t(\boldsymbol{Z}_t),$$
$$\dot{\boldsymbol{Z}}_t = \boldsymbol{v}_t.$$

散度$\sum_i \partial\, v_i/\partial x_i$为净扩张；首式固定位置，次式沿轨迹。

随机微分方程（SDE）为$d\boldsymbol{Y} = \boldsymbol{f}d\tau + g(\tau)d\boldsymbol{W}$，加噪$\tau: 0 \to T$；f为漂移，g为状态无关标量。$q_\tau$为边缘密度，$\boldsymbol{W}_\tau$、$\boldsymbol{B}_t$为各时间的标准布朗运动，增量协方差为时间增量乘单位阵。

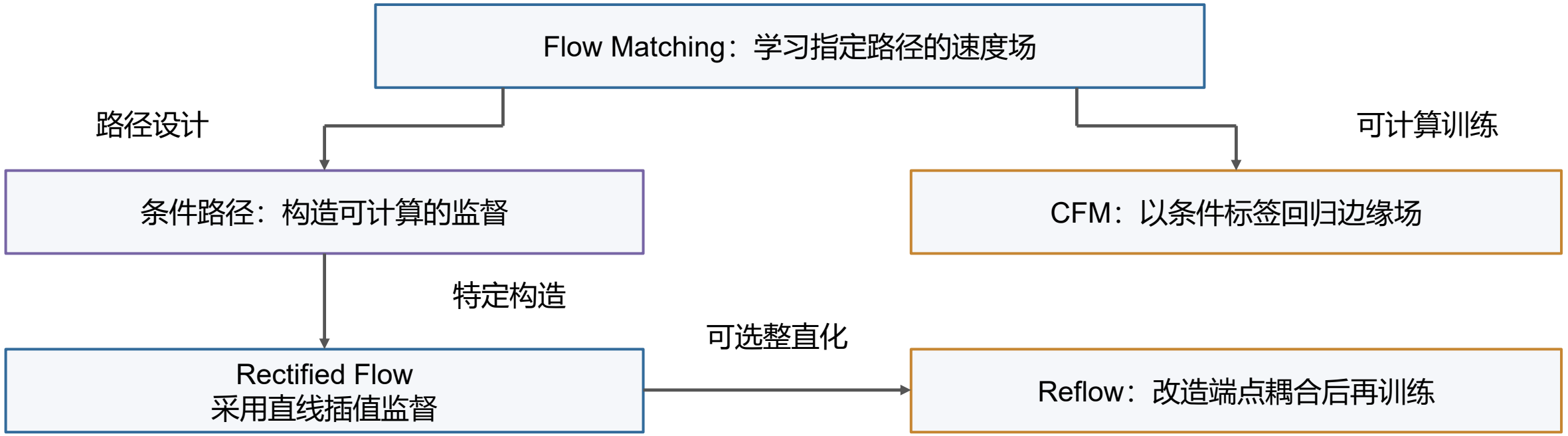
②③ 得分$\boldsymbol{r}_\tau = \nabla \log q_\tau$，生成时间$t = 1 - \tau/T$：

$$\boldsymbol{v}_t^{\mathrm{PF}} = -T\boldsymbol{f} + \frac{T}{2}g^2\boldsymbol{r}_\tau,$$
$$d\boldsymbol{Z}_t = [-T\boldsymbol{f} + Tg^2\boldsymbol{r}_\tau]dt + \sqrt{T}g d\boldsymbol{B}_t.$$

在$\tau = T(1-t)$求值，从q_T起步；真实得分、正则性、解唯一且初始分布匹配时，两过程同边缘，路径通常不同。

GM06在同一噪声尺度用$\boldsymbol{r}_\tau \approx -\boldsymbol{\epsilon}_\theta/b_\tau$，$b_\tau > 0$为扰动标准差。离散转连续须声明插值或极限；网络、先验和数值近似均影响结果。

路径、回归目标和端点耦合是不同设计选择



独立配对规定联合分布为边缘分布的乘积。条件OT路径描述给定数据端点的条件运输；全局最优运输耦合则在全部端点联合分布中优化运输代价。这三个对象具有不同职责。

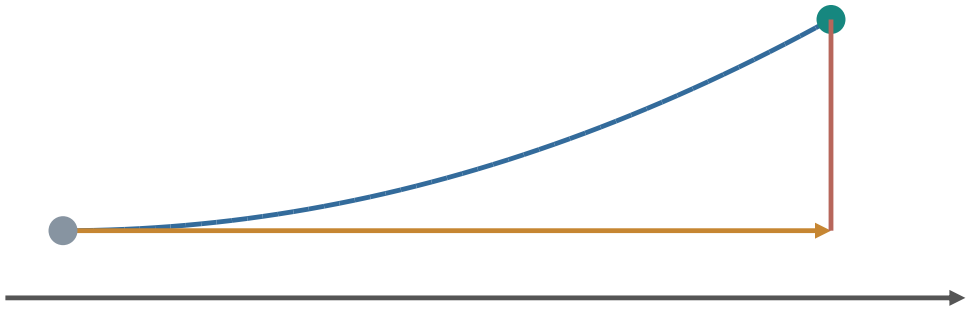
本讲直线FM路径由所选端点耦合决定；扩散PF-ODE沿指定扩散的边缘族。端点相同也不足以使中间分布相同。

依据：Lipman等§3/4；Liu等§2/3.4；Song等§4.3。

减少积分调用，需要控制轨迹的离散误差

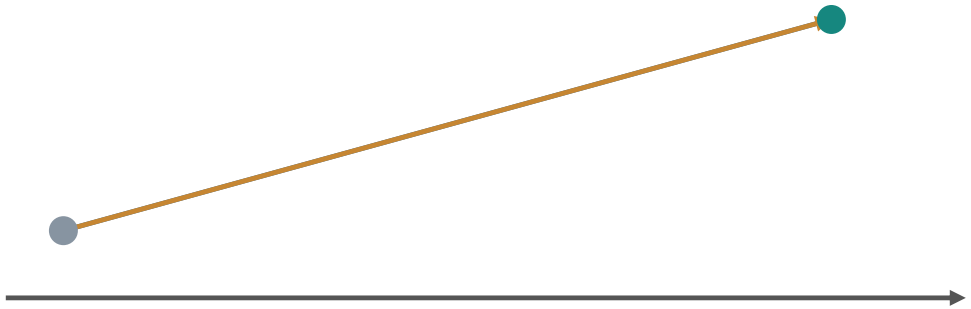
一次Euler只使用起点处的速度。速度沿实际轨迹变化较大时，粗步长难以跟随；步数、网络调用次数和端到端耗时需要分别记录。

曲线与起点切线



一步误差

理想恒速直线



一步到达

蓝：实际轨迹；琥珀：一步Euler；红：终点差

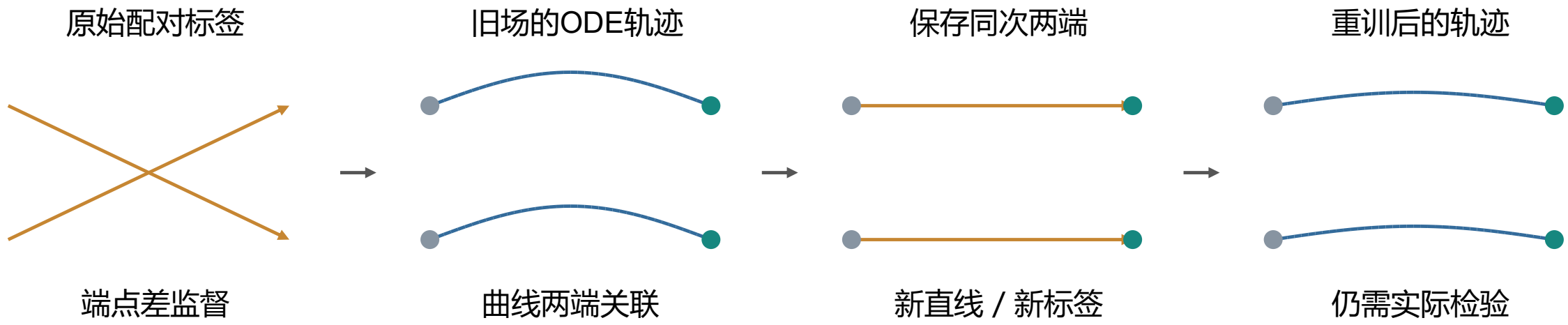
机制构造：右侧是理想条件，不是Reflow效果保证

Reflow利用已学ODE的同次起终点重建配对，再以新直线构造监督。新配对保留旧流的运输关系。理想理论支持整直化；有限模型和有限积分仍需检验。

由已学流的端点关系构造下一轮监督

Reflow（重新配对）用已学流的两端配对，再训练新场。以下据Liu等Algorithm 1的Reflow步骤译写：

1. 抽起点$\boldsymbol{z}_0 \sim p_0$，用当前已训练场积分得$\boldsymbol{z}_1$。
2. 保存同次积分的两端，保留其依赖关系。
3. 在新端点对上抽时间，构造$(1-t)\boldsymbol{z}_0 + t\boldsymbol{z}_1$及标签$\boldsymbol{z}_1 - \boldsymbol{z}_0$，重新回归速度。
4. 需要继续重新配对时，使用新场再产生端点对。



机制示意：旧场不是左图标签的解析解，末栏不是性能保证

恒速轨迹允许一步Euler精确到终点

一步Euler精确到终点的充分条件是：实际ODE轨迹上的速度恒定，且等于端点差：

$$\boldsymbol{v}(\boldsymbol{Z}_t, t) = \boldsymbol{Z}_1 - \boldsymbol{Z}_0.$$

端点直线标签虽为恒速，学得的ODE却未必具有该性质。

- 原文Algorithm 1另列可选蒸馏，拟合已有流的起终点映射；它与改变端点配对的Reflow职责不同。本讲不展开蒸馏训练。
- Reflow理论结论有其理想条件；有限网络、终点采样误差和再次拟合可能累积误差，也不普遍等于最优传输。采样调用次数的减少需结合实际场、数值误差和生成结果判断。

概念 / 条件	核心改变或约束的对象
Rectified Flow	通过直线插值监督学习流
Reflow	更新端点耦合与训练监督
Distillation	拟合已有流的起终点映射
一步Euler精确条件	实际轨迹速度恒定，且等于端点差

速度预测接入条件、潜在表示与骨干

条件生成用配对数据$(\boldsymbol{X}_1, c)$构造监督。噪声端点独立于条件。网络输入为当前状态、生成时间和条件：

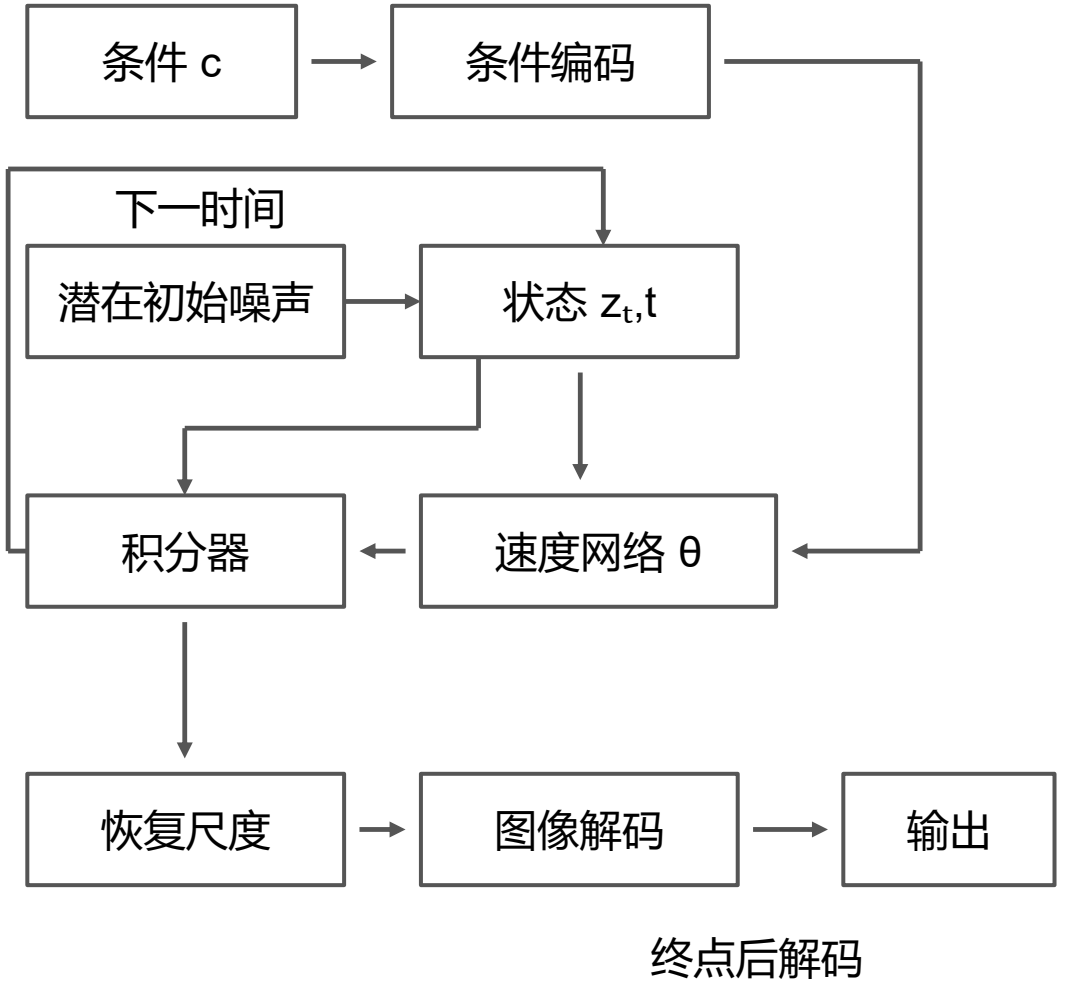
$$\boldsymbol{v}_\theta(\boldsymbol{x}_t, t, c), \boldsymbol{v}_t^*(\boldsymbol{x}_t \mid c) = E[\boldsymbol{X}_1 - \boldsymbol{X}_0 \mid \boldsymbol{X}_t = \boldsymbol{x}_t, t, c].$$

上式$\boldsymbol{v}_t^*(\boldsymbol{x}_t \mid c)$是给定c的理想速度目标。潜在状态为编码表示，速度与状态同形状。训练和生成必须使用相同尺度、条件约定。

条件表示、潜在编解码与骨干承接GM06职责。引导组合需匹配预测参数化及采样设置。

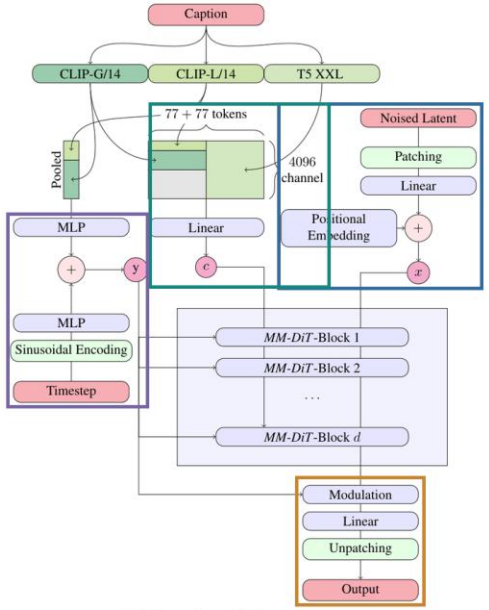
后两页据Esser等原论文Fig.2及§4，核对条件、潜在状态与骨干的实际接口；性能数值不在此展开。

分别记录求解步数、网络评估次数和端到端耗时。每步可多次调用网络，潜在解码和条件编码另有成本。GM08在统一条件下比较。



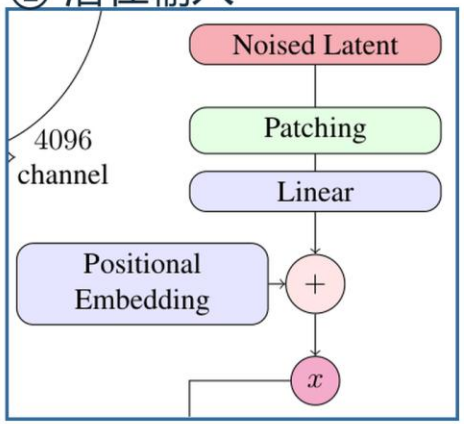
MMDiT原图把系统接口落实为网络输入

完整原图定位

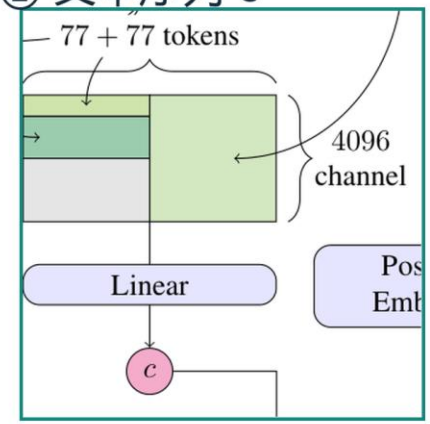


(a) Overview of all components.

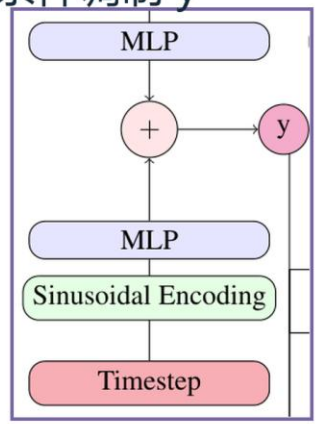
① 潜在输入



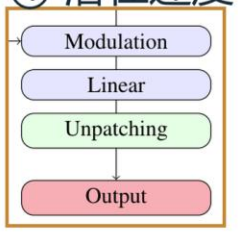
② 文本序列 c



③ 条件调制 y



④ 潜在速度输出

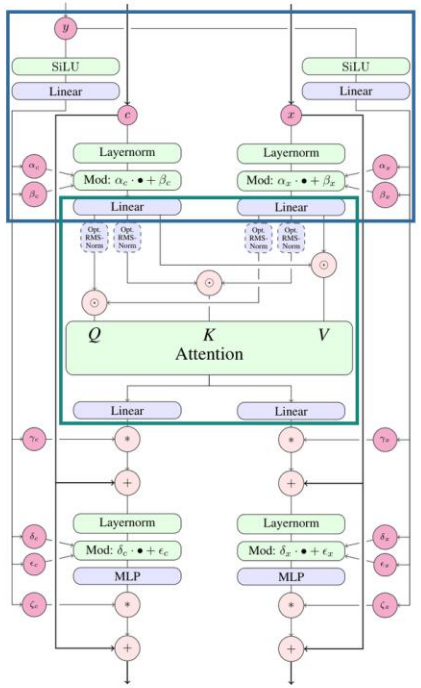


Output仍是潜在空间的速度
积分器与图像解码在骨干之外

编号与彩框为课堂阅读辅助；右侧是左侧原图的局部放大，原图标签与箭头保留。
y组合文本汇总与时间；令本讲t=1−原文时间，输出速度取反后交给外部积分器。
原图及摘录：Esser等 2403.03206v1 Fig.2(a)，§3/4；骨干不含外部积分与图像解码。

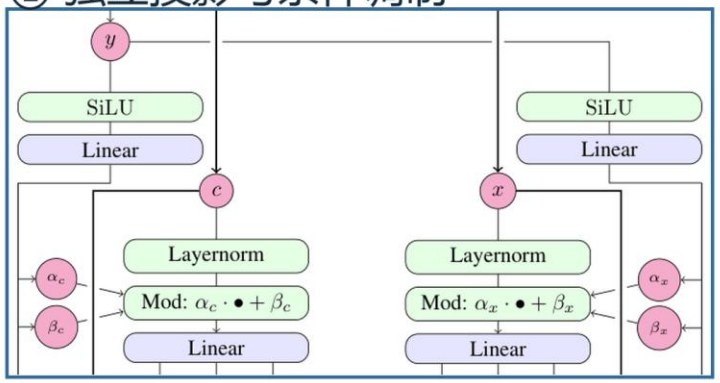
单个MMDiT块以联合注意力交换两种模态信息

完整原图定位



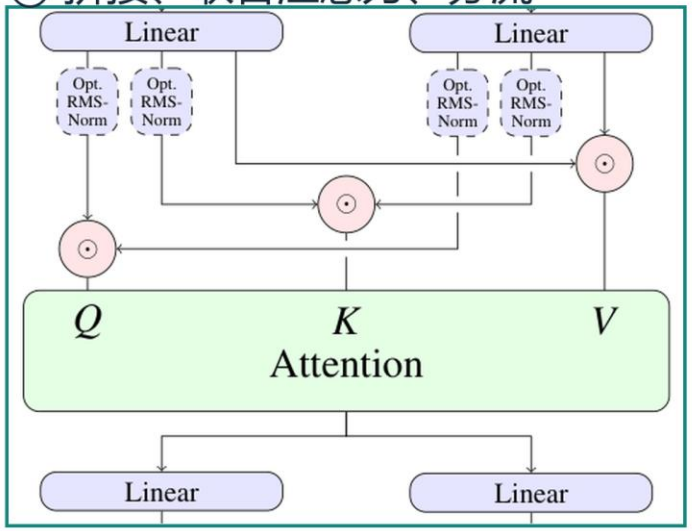
(b) One *MM-DiT* block

① 独立投影与条件调制



两种模态各用独立参数
y提供条件调制；主干内完成特征变换

② 拼接、联合注意力、分流



①两支分别归一化、调制与投影；②Q/K/V各自拼接，联合Attention后分回两支。
文本与图像双向交换信息；⊙为拼接，*为逐元素乘。单个块仍是骨干内的计算。
原图及摘录：Esser等 2403.03206v1 Fig.2(b)，§4；编号、彩框和中文标签为阅读辅助。

本讲总结

怎样让噪声连续生成？ 速度场定义局部运动，ODE积累为轨迹；随机初值经同一流映射产生随时间变化的分布，数值积分另有误差。

训练速度从哪里来？ 独立噪声—数据端点构造状态和速度标签。网络回归中间状态下的条件平均，生成时只消费当前状态、时间及已给条件。

为何能回归条件标签？ 有限二阶矩下，CFM与FM期望目标相差参数无关常数；合法求导换序后期望梯度相同，单样本梯度仍会波动。

局部运动为何改变密度？ 区域概率通量建立连续性方程；沿轨迹密度变化由负散度给出。所选SDE与概率流ODE的联系针对边缘分布，真实得分、正则条件和时间方向共同限制结论。

怎样减少调用并接入系统？ 重新配对用已学流的端点关系改造下一轮监督；直线标签不自动带来一步精确生成。条件、表示、骨干和积分器共同组成系统，成本与结果需结合设置评价。