

# A Review of Unify Diffusion Model Theory Based on EDM

Created by Zeng zhiwen

Reference: [Elucidating the Design Space of Diffusion-Based Generative Models](#)

## 前向：通用加噪公式与SDE的相互转化

从一个分布  $\mathbf{x}_0$  到另一个分布  $\mathbf{x}_t$  的桥梁，也即流（Flow）：

$$p(\mathbf{x}_t|\mathbf{x}_0) = \mathcal{N}(\mathbf{x}_t; s(t)\mathbf{x}_0, s^2(t)\sigma^2(t)\mathbf{I}) \quad (1)$$

将公式（1）进行重参数化采样，这里的  $\sigma(t)$  为噪声强度的相对系数， $s(t)\sigma(t)$  为绝对噪声强度（即噪声实际标准差）：

$$\mathbf{x}_t = s(t)\mathbf{x}_0 + s(t)\sigma(t)\epsilon \quad (2)$$

- $s(t)\mathbf{x}_0$  为信息部分：

$$\text{信号功率} = \mathbb{E}[\|s(t)\mathbf{x}_0\|^2] = s(t)^2 \cdot \mathbb{E}[\|\mathbf{x}_0\|^2] = s(t)^2 \cdot \alpha \cdot n$$

- $s(t)\sigma(t)\epsilon$  为噪声部分：

$$\text{噪声功率} = \mathbb{E}[\|s(t)\sigma(t)\epsilon\|^2] = s(t)^2\sigma(t)^2 \cdot \mathbb{E}[\|\epsilon\|^2] = s(t)^2\sigma(t)^2 \cdot n$$

- 信噪比：

$$SNR(t) = \frac{\alpha}{\sigma^2(t)}$$

为了适配EDM统一框架输入：

$$\hat{\mathbf{x}}_t = \frac{\mathbf{x}_t}{s(t)} = \mathbf{x}_0 + \sigma(t)\epsilon$$

最终单步递推公式（见 [#证明1](#)）：

$$\mathbf{x}_t = \frac{s(t)}{s(t-1)}\mathbf{x}_{t-1} + s(t)\sqrt{\sigma(t)^2 - \sigma(t-1)^2}\epsilon_t$$

具体可转化为各模型形式，如DDPM：

$$\mathbf{x}_t = \sqrt{1 - \beta_t}\mathbf{x}_{t-1} + \sqrt{\beta_t}\epsilon$$

其中， $s(t) = \sqrt{\bar{\alpha}_t}$ ， $\bar{\alpha}_t = \prod_{s=1}^t (1 - \beta_s)$ ， $\sigma(t) = \sqrt{\frac{1 - \bar{\alpha}_t}{\bar{\alpha}_t}}$

对应的随机微分方程（SDE），形式如下：

$$d\mathbf{x}_t = f(t)\mathbf{x}_t dt + g(t)dw_t \quad (3)$$

反之，递推式可通过求取极限（见 [#证明2](#)）得到公式（3），对应系数如下：

$$s(t) = e^{\int_0^t f(r)dr} \quad (4)$$

$$f(t) = \frac{s'(t)}{s(t)}$$

$$\sigma^2(t) = \int_0^t \frac{g^2(r)}{s^2(r)} dr \quad (5)$$

$$g(t) = s(t) \sqrt{2 \cdot \sigma(t) \sigma'(t)}$$

公式 (2) 与公式 (3) 是等价的，公式 (4) 和 (5) 是沟通二者的桥梁，即起点分布相同条件下（即原图像  $\mathbf{x}_0$ ），针对特定加噪时间点  $t$ ，公式 (3) SDE的解  $\mathbf{x}_t$  与公式 (1) 采样获得的  $\mathbf{x}_t$  在分布上是一致的。特别地，EDM框架令  $s(t) = 1$ ， $\delta(t) = t$ ，保证了时间和噪声水平完全等价。

## 模型：通用模型框架

EDM概括了所有扩散模型中，神经网络部分的模型框架：

$$D_\theta(\hat{\mathbf{x}}; \sigma) = C_{skip}(\sigma)\hat{\mathbf{x}} + C_{out}(\sigma)F_\theta(C_{in}(\sigma)\hat{\mathbf{x}}; C_{noise}(\sigma)) \quad (6)$$

- $D_\theta(\hat{\mathbf{x}}; \sigma)$  是接收一个规范化后的噪声图片（即原始图片  $\mathbf{x}_0$  直接添加  $\sigma$  水平的噪声，不进行尺度缩放得到的  $\hat{\mathbf{x}}$ ），以及我们为其指定的噪声水平  $\sigma$ ，输出降噪后的“纯净图像”
  - 但是直接训练一个纯净网络效果不佳，因此  $F_\theta$ （残差）才是真正的网络组成。
  - $\hat{\mathbf{x}} = \frac{\mathbf{x}}{s(t)}$ ，用于统一所有模型的输入尺度
- $C_{skip}, C_{out}$ ：纯净去噪网络  $D_\theta(\hat{\mathbf{x}}; \sigma)$  的输出由噪声图片  $\hat{\mathbf{x}}$  和模型  $F_\theta$  输出加权组成。
- $C_{in}$ ：用于适配不同网络对标准输入  $\hat{\mathbf{x}}$  的系数，如  $s(t)$ 。
- $C_{noise}$ ：EDM要求模型框架  $D_\theta(\hat{\mathbf{x}}; \sigma)$  的输入为  $\sigma$ ，但不同模型真正输入  $F_\theta$  的参数可能为  $\sigma$  的函数，因此需要作变换。

## VP

$$D_\theta(\hat{\mathbf{x}}; \sigma) = \underbrace{1 \cdot \hat{\mathbf{x}}}_{c_{skip}} \underbrace{- \sigma}_{c_{out}} \cdot F_\theta\left(\underbrace{\frac{1}{\sqrt{\sigma^2+1}}}_{c_{in}} \cdot \hat{\mathbf{x}}; \underbrace{(M-1)\sigma^{-1}(\sigma)}_{c_{noise}}\right)$$

## VE

$$D_\theta(\hat{\mathbf{x}}; \sigma) = \underbrace{1 \cdot \hat{\mathbf{x}}}_{c_{skip}} + \underbrace{\sigma}_{c_{out}} \cdot F_\theta\left(\underbrace{1 \cdot \hat{\mathbf{x}}}_{c_{in}}; \underbrace{\log\left(\frac{1}{2}\sigma\right)}_{c_{noise}}\right)$$

注意：这里的 VE  $\hat{x} = x$

## 训练：通用训练框架

训练的过程，是针对从原始图像集合中采样得到的真实图像  $\mathbf{x}_0$ ，进行一次  $\sigma$  级别的噪声添加，得到  $\mathbf{x}_0 + \mathbf{n}$ ，随后可构造损失函数：

$$\mathcal{L}_{diff} = \mathbb{E}_{\sigma, \mathbf{n}, \mathbf{x}_0} \left[ \lambda(\sigma) \|D_\theta(\hat{\mathbf{x}}; \sigma) - \mathbf{x}_0\|_2^2 \right] \quad (7)$$

$$\mathcal{L}_{diff} = \underbrace{\mathbb{E}_{\sigma, \mathbf{n}, \mathbf{x}_0}}_{P_{train}} \left[ \underbrace{\lambda(\sigma) C_{out}^2(\sigma)}_{\text{损失权重}(\sigma)} \underbrace{\|F_\theta(C_{in}(\sigma) \cdot (\mathbf{x}_0 + \mathbf{n}); C_{noise}(\sigma))\|_2^2}_{\text{模型输出}} - \underbrace{\frac{1}{C_{out}(\sigma)} \|\mathbf{x}_0 - C_{skip}(\sigma) \cdot (\mathbf{x}_0 + \mathbf{n})\|_2^2}_{\text{训练目标}} \right]$$

- $\sigma \sim P_{train}$ ，即前向噪声采样分布，由各个模型决定
- $\mathbf{n} \sim \mathcal{N}(0, \sigma^2 \mathbf{I})$ ， $\mathbf{x}_0 \sim P_{data}$
- $Var(\mathbf{x}_0) = \sigma_{data}^2$
- $C_{in}$ ：保证神经网络的输入保持单位方差（式117）

$$c_{in}(\sigma) = \frac{1}{\sqrt{\sigma^2 + \sigma_{data}^2}}$$

- $C_{out}, C_{skip}$ : 保证训练目标保持方差恒为1, 同时让  $C_{out}^2$  被最小化, 防止模型误差被放大 (式138、131):

$$C_{skip}(\sigma) = \frac{\sigma_{data}^2}{\sigma_{data}^2 + \sigma^2}$$

$$C_{out}(\sigma) = \frac{\sigma \cdot \sigma_{data}}{\sqrt{\sigma^2 + \sigma_{data}^2}}$$

- $\lambda(\sigma)$ : 保证损失权重  $w(t) = 1$  (式 144):

$$\lambda = \frac{\sigma^2 + \sigma_{data}^2}{(\sigma \cdot \sigma_{data})^2}$$

- 当初始化神经网络权重为0 (即输出恒0), 方差暂时固定某个初始值时, 有:

$$\mathbb{E}(\mathcal{L}) = 1$$

- $\sigma$ 调度: EDM认为, 损失函数在加噪水平很低或很高情况下, 损失函数均难以下降, 因此训练噪声调度的选择如下:

$$\ln(\sigma) \sim \mathcal{N}(P_{mean}, P_{std}^2)$$

其中  $P_{mean} = -1.2, P_{std} = 1.2$

- 对于 VP, 其训练调度为面向  $t$  的均匀分布:

$$\beta(t) = (\beta_{max} - \beta_{min})t + \beta_{min}, \quad t \sim \mathcal{U}(\epsilon_t, 1)$$

- 对于 VE, 其训练调度与EDM同样面对  $\sigma$  进行, 但是为均匀调度:

$$\ln(\sigma) \sim \mathcal{U}(\ln(\sigma_{min}), \ln(\sigma_{max}))$$

## VP

$$\underbrace{\mathbb{E}_{\sigma^{-1}(\sigma) \sim \mathcal{U}(\epsilon_t, 1)}}_{p_{train}} \mathbb{E}_{\mathbf{x}_0, \mathbf{n}} \left[ \underbrace{\frac{1}{\sigma^2}}_{\text{损失权重}} \left\| D_{\theta}(\mathbf{x}_0 + \mathbf{n}; \sigma) - \mathbf{x}_0 \right\|_2^2 \right]$$

## VE

$$\underbrace{\mathbb{E}_{\ln(\sigma) \sim \mathcal{U}(\ln(\sigma_{min}), \ln(\sigma_{max}))}}_{p_{train}} \mathbb{E}_{\mathbf{x}_0, \mathbf{n}} \left[ \underbrace{\frac{1}{\sigma^2}}_{\text{损失权重}} \left\| D_{\theta}(\mathbf{x}_0 + \mathbf{n}; \sigma) - \mathbf{x}_0 \right\|_2^2 \right]$$

- 二者的损失权重均随噪声减小而增大, 即后期精细降噪过程给与更多关注

$$\hat{\mathbf{x}} = \mathbf{x}_0 + \mathbf{n} = \frac{\mathbf{x}}{s(t)}$$

损失类型	数学形式	适用场景
残差损失	$\ D_{\theta}(\hat{\mathbf{x}}) - \mathbf{x}_0\ ^2$	直接去噪
噪声损失	$\ F_{\theta}(\mathbf{x}) - \epsilon\ ^2$	DDPM 类模型
分数匹配	$\ N_{\theta}(\mathbf{x}) - \nabla \log p(\mathbf{x})\ ^2$	基于分数的生成模型

- $D_{\theta}$  是逐步去噪后的结果, 可能导致误差积累, 而直接预测比较简单的单步噪声分布  $F_{\theta}$  更为容易学习。

## 反向：通用推理过程

## 通用 概率流常微分方程 PFODE

► 通用反向：对于任意一个扩散模型加噪SDE（公式(3)），通过福克普朗克方程，可进一步推导出一个常微分方程（ODE），也叫概率流常微分方程（PFODE）：

$$d\mathbf{x}_t = \left[ f(t)\mathbf{x}_t - \frac{1}{2}g^2(t) \nabla_{\mathbf{x}_t} \log p_t(\mathbf{x}_t) \right] dt \quad (8)$$

注意，这里的  $p_t(\mathbf{x}_t)$  可以描述为：

$$\begin{aligned} p_t(x) &= \int_{\mathcal{R}^d} p_{0t}(x|x_0) p_{data}(x_0) dx_0 \\ &= s(t)^{-d} [p_{data} * \mathcal{N}(0, \sigma^2(t)\mathbf{I})](\mathbf{x}/s(t)) \\ &= s(t)^{-d} p(\mathbf{x}/s(t); \sigma(t)) \end{aligned}$$

- \* 表示卷积操作
- $\mathbf{x}/s(t)$  表示分布在此处的取值

$$\begin{aligned} \nabla_{\mathbf{x}} \log p_t(\mathbf{x}_t) &= \nabla_{\mathbf{x}} \log s(t)^{-d} + \nabla_{\mathbf{x}} \log [p_t(\frac{\mathbf{x}_t}{s(t)}; \sigma(t))] = \nabla_{\mathbf{x}} \log [p_t(\frac{\mathbf{x}_t}{s(t)}; \sigma(t))] \\ &= \frac{1}{s(t)} \nabla_{\hat{\mathbf{x}}} \log p(\hat{\mathbf{x}}; \sigma(t)) = \frac{1}{s(t)\sigma^2(t)} (D_{\theta}(\hat{\mathbf{x}}; \sigma(t)) - \hat{\mathbf{x}}) \end{aligned} \quad (10)$$

分数函数的方向由当前噪声图谱  $\hat{\mathbf{x}}$ （低概率密度）指向神经网络预测的真实的分布  $\mathcal{D}_{\theta}(\hat{\mathbf{x}}; \sigma(t))$ （高概率密度）

► 通用反向：因此，在确定起点  $\mathbf{x}_0$ （前向）或  $\mathbf{x}_N$ （逆向）前提下，式（8）解的分布  $p(\mathbf{x}_t)$ ，即 $\mathbf{x}_t$ 的边缘概率密度与加噪过程SDE求解得到的分布是完全相同：

$$d\mathbf{x}_t = \left[ \left( \frac{\dot{\sigma}(t)}{\sigma(t)} + \frac{\dot{s}(t)}{s(t)} \right) \mathbf{x}_t - \frac{\dot{\sigma}(t)s(t)}{\sigma(t)} D_{\theta}(\frac{\mathbf{x}_t}{s(t)}; \sigma(t)) \right] dt \quad (11)$$

说明：针对PFODE，当dt取反时，便可实现前向加噪和后向加噪的切换，因此式（8）、式（9）、式（11）都可以兼顾前向和反向的描述，但**不可用PFODE实现图像加噪**，因为PFODE在给定起点时，其终点是确定的，于是变形成非分布的一对一输入输出样本匹配对。扩散模型建模的是真实分布与完全噪声分布之间的关系，必须通过随机采样配对实现，因此不能用PFODE实现图像加噪，而是仅能用于反向采样去噪

## 通用确定性采样

公式（11）的  $\mathcal{D}_{\theta}$  可用神经网络模拟，具体为公式（6），随后通过使用ODE求解器，如一阶Euler，二阶Heun，在给定起点  $\mathbf{x}_N$  下，逐步采样获得生成图像。**注意：训练过程的时间步和采样过程的时间步定义不同**，EDM采样过程的噪声水平定义为：

$$\sigma_{i < N} = \left( \sigma_{max}^{\frac{1}{\rho}} + \frac{i}{N-1} (\sigma_{min}^{\frac{1}{\rho}} - \sigma_{max}^{\frac{1}{\rho}}) \right)^{\rho} \text{ and } \sigma_N = 0 \quad (12)$$

## Algorithm 1 Deterministic sampling using Heun's 2<sup>nd</sup> order method with arbitrary $\sigma(t)$ and $s(t)$ .

```

1: procedure HEUNSAMPLER( $D_\theta(x; \sigma)$ ,  $\sigma(t)$ ,  $s(t)$ ,  $t_i \in \{0, \dots, N\}$ )
2:   sample  $x_0 \sim \mathcal{N}(0, \sigma^2(t_0) s^2(t_0) \mathbf{I})$ 
3:   for  $i \in \{0, \dots, N-1\}$  do
4:      $d_i \leftarrow \left( \frac{\dot{\sigma}(t_i)}{\sigma(t_i)} + \frac{\dot{s}(t_i)}{s(t_i)} \right) x_i - \frac{\dot{\sigma}(t_i) s(t_i)}{\sigma(t_i)} D_\theta \left( \frac{x_i}{s(t_i)}; \sigma(t_i) \right)$ 
5:      $x_{i+1} \leftarrow x_i + (t_{i+1} - t_i) d_i$ 
6:     if  $\sigma(t_{i+1}) \neq 0$  then
7:        $d'_i \leftarrow \left( \frac{\dot{\sigma}(t_{i+1})}{\sigma(t_{i+1})} + \frac{\dot{s}(t_{i+1})}{s(t_{i+1})} \right) x_{i+1} - \frac{\dot{\sigma}(t_{i+1}) s(t_{i+1})}{\sigma(t_{i+1})} D_\theta \left( \frac{x_{i+1}}{s(t_{i+1})}; \sigma(t_{i+1}) \right)$ 
8:        $x_{i+1} \leftarrow x_i + (t_{i+1} - t_i) \left( \frac{1}{2} d_i + \frac{1}{2} d'_i \right)$ 
9:   return  $x_N$ 

```

▷ Generate initial sample at  $t_0$   
 ▷ Solve Eq. 4 over  $N$  time steps  
 ▷ Evaluate  $dx/dt$  at  $t_i$   
 ▷ Take Euler step from  $t_i$  to  $t_{i+1}$   
 ▷ Apply 2<sup>nd</sup> order correction unless  $\sigma$  goes to zero  
 ▷ Eval.  $dx/dt$  at  $t_{i+1}$   
 ▷ Explicit trapezoidal rule at  $t_{i+1}$   
 ▷ Return noise-free sample at  $t_N$

## 通用 随机微分方程

► 通用：逆向随机形式 SDE 为：

$$dx = [f(t)x_t - g(t)^2 \nabla_x \log p_t(x)] dt + g(t) d\bar{w}, \quad (a)$$

SDE适用于前向和反向过程，因此这里特别分别给出了其在扩散模型中的前向、反向过程定义。而PFODE在扩散模型领域只适用于反向过程，见式 (8)。

当  $g(t) = 0$  时，方程变为  $dx = f(x, t) dt$ ，此时不能称为扩散模型的确定性采样过程，因为扩散项以及被消除了，不再是扩散模型的范畴。真正确定性逆向采样参见式 (8)。

思考：既然  $-g(t)^2 \nabla_x \log p_t(x)$  已经能够为逆向采样过程提供降噪方向正确的、降噪强度确定的保证了，为什么还要后面的随机项？保持生成多样性，避免坍缩到单一模式。

#猜想：变成随机后二分之一没了，但多了一个随机项

## 非通用 随机性采样

随机性采样过程方法众多，甚至和逆向SDE公式本身“关系不大”。EDM论文也表示它设计的机性采样过程不是一种通用的SDE求解器，而是一种面向扩散模型问题的垂类SDE求解器。EDM设计的随机性采样过程非常简单，其核心就是在确定性采样的基础上增加了“回退”操作，也即先对样本额外加噪，再采用ODE求解器采样获得下一个时间点的图像。这种回退操作可以有效修正前面迭代步骤产生的误差，所以通常相比PFODE的生成效果更好，但同时也要花费更多的采样步数。EDM提出的SDE采样器(求解器)基本算法流程如图所示：

Algorithm 2 Our stochastic sampler with  $\sigma(t) = t$  and  $s(t) = 1$ .

```

1: procedure STOCHASTICSAMPLER( $D_\theta(x; \sigma)$ ,  $t_i \in \{0, \dots, N\}$ ,  $\gamma_i \in \{0, \dots, N-1\}$ ,  $S_{\text{noise}}$ )
2:   sample  $x_0 \sim \mathcal{N}(0, t_0^2 \mathbf{I})$ 
3:   for  $i \in \{0, \dots, N-1\}$  do
4:     sample  $\epsilon_i \sim \mathcal{N}(0, S_{\text{noise}}^2 \mathbf{I})$ 
5:      $\hat{t}_i \leftarrow t_i + \gamma_i \epsilon_i$ 
6:      $\hat{x}_i \leftarrow x_i + \sqrt{t_i^2 - \hat{t}_i^2} \epsilon_i$ 
7:      $d_i \leftarrow (\hat{x}_i - D_\theta(\hat{x}_i; \hat{t}_i)) / \hat{t}_i$ 
8:      $x_{i+1} \leftarrow \hat{x}_i + (\hat{t}_{i+1} - \hat{t}_i) d_i$ 
9:     if  $\hat{t}_{i+1} \neq 0$  then
10:       $d'_i \leftarrow (x_{i+1} - D_\theta(x_{i+1}; \hat{t}_{i+1})) / \hat{t}_{i+1}$ 
11:       $x_{i+1} \leftarrow \hat{x}_i + (\hat{t}_{i+1} - \hat{t}_i) (\frac{1}{2} d_i + \frac{1}{2} d'_i)$ 
12:   return  $x_N$ 

```

令  $\epsilon_i \sim \mathcal{N}(0, 2)$   
 $\hat{t}_i = t_i + \gamma_i \epsilon_i$   
 $\hat{t}_i = t_i + (\sqrt{2}-1) t_i = \sqrt{2} t_i$   
 $\hat{x}_i = x_i + \sqrt{2 t_i^2 - t_i^2} \epsilon_i$   
 $= x_i + t_i \epsilon_i$   
 $\mathcal{N}(x, \frac{t_i^2}{2})$   
 $\mathcal{N}(0, \frac{t_i^2}{2})$   
 加噪不算太大  
 $x_i$  原有噪声

最大化噪声水平  
 (完全高斯随机噪声)  
 告诉模型每一轮降噪的力度

$N$  轮降噪，每一轮对应一种降噪的力度  
 扩散模型的采样过程本质是在噪声空间和数据空间之间插值。若新增噪声不足 ( $S_{\text{noise}} = 1.0$ )，去噪器可能因噪声过强而丢失低频信息，可能过早“确信”某些区域已干净，导致信息丢失。

加噪回退，在不考虑  $S_{\text{noise}}$  情况下，所添加的噪声水平不超过当前图像噪声 (当前指定模型降噪的力度)  
 ▷ Select temporarily increased noise level  $\hat{t}_i$   
 ▷ Add new noise to move from  $t_i$  to  $\hat{t}_i$   
 ▷ Evaluate  $dx/dt$  at  $\hat{t}_i$   
 ▷ Take Euler step from  $\hat{t}_i$  to  $t_{i+1}$   
 ▷ Apply 2<sup>nd</sup> order correction

- 特点：EDM采取加噪后立即利用最新状态计算梯度并更新，而欧拉离散化方法采取先更新再叠加噪声，没有立刻更新梯度基准，导致离散误差在  $\Delta t$  较大时增大。

## VP (DDPM / DDIM)

VP(Variance Perserving)，噪声调度满足 **信号与噪声的方差总和恒定**，前向公式：

$$\begin{aligned} \mathbf{x}_t &= \sqrt{1 - \beta_t} \mathbf{x}_{t-1} + \sqrt{\beta_t} \epsilon \\ \mathbf{x}_t &= \sqrt{\bar{\alpha}_t} \mathbf{x}_0 + \sqrt{1 - \bar{\alpha}_t} \epsilon \\ \hat{\mathbf{x}}_t &= \frac{\mathbf{x}_t}{\sqrt{\bar{\alpha}_t}} = \mathbf{x}_0 + \frac{\sqrt{1 - \bar{\alpha}_t}}{\sqrt{\bar{\alpha}_t}} \epsilon \\ d\mathbf{x} &= -\frac{1}{2} \beta(t) \mathbf{x} dt + \sqrt{\beta(t)} d\mathbf{w} \end{aligned} \quad (16)$$

对离散递推式取极限可以直接导出连续式（16）。

VP一步加噪式中的  $\bar{\sigma} = \sqrt{1 - \bar{\alpha}_t}$  表示原框架下的图谱噪声，其融合了尺度缩放项  $s(t)$ ，而在同一框架下， $\sigma(t)$  才是前向过程所叠加的噪声标准差：

$$\begin{aligned} \bar{\sigma}(t) &= \sqrt{1 - \bar{\alpha}_t} \\ s(t) &= \sqrt{\bar{\alpha}_t} \\ \sigma(t) &= \frac{\sqrt{1 - \bar{\alpha}_t}}{\sqrt{\bar{\alpha}_t}} \end{aligned}$$

满足

$$s(t) = \frac{1}{\sqrt{\sigma^2(t) + 1}}$$

上面的符号与式（2）对应。

漂移项  $f(t)$ :

$$f(t) = -\frac{1}{2} \beta(t)$$

扩散项  $g(t)$ :

$$g(t) = \sqrt{\beta(t)} \quad (18)$$

式（17）、式（18）带入式（b）得到**VP逆向SDE**过程:

$$\begin{aligned} d\mathbf{x} &= \frac{1}{2} \beta(t) \sigma(t)^2 \nabla_{\mathbf{x}} \log p(\mathbf{x}) dt + \sqrt{\beta(t)} d\mathbf{w} \\ &= \left[ -\frac{1}{2} \beta(t) \mathbf{x} - \beta(t) \nabla_{\mathbf{x}} \log p(\mathbf{x}) \right] dt + \sqrt{\beta(t)} d\mathbf{w} \end{aligned} \quad (19)$$

## VE (SMLD)

VE(Variance Exploding)，VE过程的噪声调度允许 **噪声方差无限增长**，前向公式：

$$\begin{aligned} \mathbf{x}_t &= \mathbf{x}_{t-1} + \sqrt{\sigma^2(t) - \sigma^2(t-1)} \epsilon \\ \mathbf{x}_t &= \mathbf{x}_0 + \sigma(t) \epsilon \\ d\mathbf{x} &= \sqrt{\frac{d\sigma^2(t)}{dt}} d\mathbf{w} \end{aligned} \quad (20)$$

其中， $\bar{\sigma}(t) = s(t) \cdot \sigma(t) = \sigma(t)$

扩散项：

$$g(t) = \sqrt{\frac{d\bar{\sigma}^2(t)}{dt}} = \sqrt{\frac{d\sigma^2(t)}{dt}} \quad (22)$$

漂移项:

$$f(t) = 0$$

带入式（a）得到与song等人定义一致的**VE**逆向过程：

$$\begin{aligned} d\mathbf{x} &= [-g(t)^2 \nabla_x \log p_t(\mathbf{x})]dt + g(t)d\mathbf{w} \\ &= \left[-\frac{d\sigma^2(t)}{dt} \nabla_x \log p_t(\mathbf{x})\right]dt + \sqrt{\frac{d\sigma^2(t)}{dt}}d\mathbf{w} \end{aligned} \tag{23}$$

## 附录

### 【补充1】 如何理解 VP 方差保持与 VE 方差爆炸

考虑**VP**递推：

$$x_t = \sqrt{\alpha_t}x_{t-1} + \sqrt{1 - \alpha_t}\epsilon$$

由方差性质：

$$Var(x_t) = \alpha_t Var(x_{t-1}) + (1 - \alpha_t) \cdot 1$$

由数学归纳法，假设  $Var(x_{t-1}) = 1$ ，有：

$$Var(x_t) = \alpha_t + 1 - \alpha_t = 1$$

因此方差是保持的。

而**VE**中， $Var(x_t) = 1 - \alpha_t = \beta_t$ ，由于噪声水平逐步增大，因此方差是爆炸式增大的。

### 【补充2】 DDPM与VP的关系

**DDPM**是**VP**的离散化形式

模型	前向加噪公式
VP SDE	$d\mathbf{x} = -\frac{1}{2}\beta(t)\mathbf{x}dt + \sqrt{\beta(t)}d\mathbf{w}$
DDPM	$\begin{aligned} \mathbf{x}_t &= \sqrt{1 - \beta_t}\mathbf{x}_{t-1} + \sqrt{\beta_t}\epsilon \\ \mathbf{x}_t &= \sqrt{\bar{\alpha}_t}\mathbf{x}_0 + \sqrt{1 - \bar{\alpha}_t}\epsilon \end{aligned}$

模型	反向采样公式
VP SDE	$d\mathbf{x} = [-\frac{1}{2}\beta(t)\mathbf{x} - \beta(t)\nabla_x \log p_t(x)]dt + \sqrt{\beta(t)}d\bar{\mathbf{w}}$
DDPM	$\begin{aligned} \mathbf{x}_{t-1} &= \frac{1}{\sqrt{\alpha_t}} \left( \mathbf{x}_t - \beta_t \frac{\epsilon_{\theta}(\mathbf{x}_t, t)}{\sqrt{1 - \bar{\alpha}_t}} \right) + \sqrt{\beta_t}\epsilon_t \\ \mathbf{x}_{t-1} &= \frac{1}{\sqrt{1 - \beta_t}} \left( \mathbf{x}_t + \beta_t \frac{D_{\theta}(\frac{\mathbf{x}_t}{s(t)}; \sigma(t)) - \frac{\mathbf{x}_t}{s(t)}}{s(t)\sigma^2(t)} \right) + \sqrt{\beta_t}\epsilon_t \end{aligned}$

**VP**中的变量定义与转化

- $\alpha_t = 1 - \beta_t$
- $\bar{\alpha}_t = \prod_{s=1}^t (1 - \beta_s)$

在实践中，有如下近似：

$$\log \bar{\alpha}_t = \sum_{i=1}^t \log \alpha_i \approx \int_0^t \log(1 - \beta_s) ds$$

这里的  $\log(1 - \beta_s) \approx -\beta_s$ ，表示在0处的泰勒展开  
因此

$$\bar{\alpha}_t \approx \exp\left(-\int_0^t \beta_s ds\right)$$

$\beta_t$  在实践中被定义为时间线性函数

$$\begin{aligned} \beta(t) &= (\beta_{\max} - \beta_{\min})t + \beta_{\min} \\ \int_0^t \beta_s ds &= \int_0^t (\beta_{\min} + s(\beta_{\max} - \beta_{\min})) ds = \beta_{\min} t + \frac{t^2}{2}(\beta_{\max} - \beta_{\min}) \end{aligned}$$

故

$$\bar{\alpha}_t = \exp(-\beta_{\min} t + \frac{t^2}{2}(\beta_{\max} - \beta_{\min}))$$

3.  $\bar{\sigma}(t) = s(t)\sigma(t) = \sqrt{1 - \bar{\alpha}_t}$ ，即原框架下  $t$  时刻的图像噪声水平

4.  $s(t) = \sqrt{\bar{\alpha}_t}$ ，与式 (2) 对应

5.  $\sigma(t) = \frac{\sqrt{1 - \bar{\alpha}_t}}{\sqrt{\bar{\alpha}_t}}$ ，与式 (2) 对应

## 【其他】

### 1. EDM前向

$$\begin{aligned} \mathbf{x}_t &= \mathbf{x}_{t-1} + \sqrt{\sigma^2(t) - \sigma^2(t-1)}\epsilon \\ \mathbf{x}_t &= \mathbf{x}_0 + \sigma(t)\epsilon \\ d\mathbf{x} &= \sqrt{\frac{d\sigma^2(t)}{dt}} d\mathbf{w} \end{aligned}$$

与VE一致，但噪声分布不同：

$$\ln(\sigma(t)) = \ln(t) \sim \mathcal{N}(P_{\text{mean}}, P_{\text{std}}^2)$$

### 2. Wiener过程

$w_t \sim \mathcal{N}(0, t)$  是一个布朗运动 (Wiener) 过程

- 独立增量性： $w_{t+\Delta t} - w_t$  与  $w_t$  独立
- 布朗运动的增量服从正态分布： $w_{t+\Delta t} - w_t \sim \mathcal{N}(0, \Delta t)$
- 令  $\Delta t \rightarrow 0, d_w \sim \mathcal{N}(0, d_t)$ 
  - 布朗运动无穷小增量的平方  $d_w^2 = d_t$  为确定性质量
  - 重参数化展开： $d_w = \sqrt{d_t} \cdot \epsilon, \epsilon \sim \mathcal{N}(0, 1)$

### 3. EDM论文相关

Song et al. present a stochastic differential equation (SDE) that **maintains** the desired distribution  $p$  as sample  $\mathbf{x}$  evolves over time

若一个SDE的解  $\mathbf{x}_t$  的边际分布  $p_t$  满足：

$$\lim_{t \rightarrow \infty} p_t(\mathbf{x}) = p(\mathbf{x}) \quad \text{且} \quad \text{一旦达到 } p \text{ 后, 分布不再随时间变化}$$



则称  $p_t$  是该SDE的**不变分布**（或稳态分布）。此时，SDE“保持”了分布  $p_t$ 。

- **正向过程：**

从数据分布  $p_{data}$  出发，通过SDE逐渐将数据破坏为噪声分布纯粹的高斯分布  $\epsilon$ 。

- **逆向过程：**

从噪声  $\epsilon$  出发，通过SDE将样本演化回  $p_{data}$ 。

To specify the ODE, we must first choose a schedule  $\sigma(t)$  that defines the desired noise level at time  $t$ .

在PFODE中， $\sigma(t)$  **直接表示  $t$  时刻数据的噪声水平（累积结果）**，而非单步添加量。这样一来，在前向加噪训练时，针对某一时刻  $t$  噪声水平  $\sigma(t)$ ，直接向  $\mathbf{x}_0$  添加  $\mathcal{N} \sim (0, \sigma^2(t))$  的高斯随机噪声即可。在反向降噪采样时，也可以直接告诉神经网络当前图像的噪声水平  $\sigma(t)$ ，从而做出**相应力度**的降噪操作。

The score function has the remarkable property that it does not depend on the generally intractable normalization constant of the underlying density function  $p(\mathbf{x}; \sigma)$

**Score Function**它不依赖于概率密度函数  $p(\mathbf{x}; \sigma)$  的归一化常数（normalization constant）。

假设概率密度函数可以分解为：

$$p(x; \sigma) = \frac{1}{Z(\sigma)} \tilde{p}(x; \sigma)$$

其中：

- $\tilde{p}(x; \sigma)$  是未归一化的概率密度（可能难以计算积分）。
- $Z(\sigma)$  是归一化常数（通常难以计算，尤其是高维数据）。

取对数后：

$$\log p(x; \sigma) = \log \tilde{p}(x; \sigma) - \log Z(\sigma)$$

计算梯度时：

$$\nabla_x \log p(x; \sigma) = \nabla_x \log \tilde{p}(x; \sigma) - \underbrace{\nabla_x \log Z(\sigma)}_{=0}$$

由于  $Z(\sigma)$  不依赖  $x$ ，其梯度为零，因此：

$$\nabla_x \log p(x; \sigma) = \nabla_x \log \tilde{p}(x; \sigma)$$

**结论：** Score 函数仅依赖于未归一化的  $\tilde{p}(x; \sigma)$ ，与  $Z(\sigma)$  无关！

Excessive Langevin-like addition and removal of noise results in gradual **loss of detail** in the generated images. There is also a drift toward oversaturated colors at very low and high noise levels. We suspect that practical denoisers induce a slightly nonconservative vector field.

**引入随机性（SDE，朗之万噪声步骤）**虽然能修正早期采样误差，但会导致**细节丢失**和在极端噪声水平下的颜色过饱和。

原因可能在于：

- Denoiser的过渡去噪移除了比理论值更多的噪声，破坏了朗之万扩散所需的保守向量场；
- $\mathcal{L}^2$  损失使得模型倾向于预测均值，忽略极端边缘细节

解决方案在于：

- 限制噪声添加的时机范围  $t_i \in [S_{t_{min}}, S_{t_{max}}]$
- 使得每次添加随机噪声的水平  $S_{noise}$  略微大于1抵消细节损失

- 确保每次新增噪声的强度不超过当前图像的噪声水平，防止过度破坏结构

## 【证明1】前向离散一步加噪式（2）转化为离散单步递推式

离散一步加噪形式：

$$\mathbf{x}_t = s(t)\mathbf{x}_0 + s(t)\sigma(t)\epsilon$$

假设：

$$\mathbf{x}_t = \alpha_t \mathbf{x}_{t-1} + \beta_t \epsilon_t$$

1. 期望匹配：

$$\mathbb{E}[\mathbf{x}_t] = s(t)\mathbf{x}_0 = \alpha_t \mathbb{E}[\mathbf{x}_{t-1}] = \alpha_t s(t-1)\mathbf{x}_0$$

因此：

$$\alpha_t = \frac{s(t)}{s(t-1)}$$

2. 方差匹配：

$$\text{Var}(\mathbf{x}_t) = s(t)^2 \sigma(t)^2 = \alpha_t^2 \text{Var}(\mathbf{x}_{t-1}) + \beta_t^2$$

代入  $\text{Var}(\mathbf{x}_{t-1}) = s(t-1)^2 \sigma(t-1)^2$ ：

$$\beta_t^2 = s(t)^2 \sigma(t)^2 - \left( \frac{s(t)}{s(t-1)} \right)^2 s(t-1)^2 \sigma(t-1)^2 = s(t)^2 (\sigma(t)^2 - \sigma(t-1)^2)$$

因此：

$$\beta_t = s(t) \sqrt{\sigma(t)^2 - \sigma(t-1)^2}$$

3. 最终单步递推公式：

$$\mathbf{x}_t = \frac{s(t)}{s(t-1)} \mathbf{x}_{t-1} + s(t) \sqrt{\sigma(t)^2 - \sigma(t-1)^2} \epsilon_t \quad (24)$$

## 【证明2】前向加噪离散形式到连续形式的转化

原始递推式如下：

$$\mathbf{x}_t = \frac{s(t)}{s(t-1)} \mathbf{x}_{t-1} + s(t) \sqrt{\sigma(t)^2 - \sigma(t-1)^2} \epsilon_t$$

两边减去  $\mathbf{x}_{t-1}$ ，并写为极限形式：

$$\mathbf{x}_t - \mathbf{x}_{t-\Delta t} = \left( \frac{s(t)}{s(t-\Delta t)} - 1 \right) \mathbf{x}_{t-\Delta t} + s(t) \sqrt{\sigma(t)^2 - \sigma(t-\Delta t)^2} \epsilon_t.$$

当时间步长  $\Delta t = t - (t-1) \rightarrow 0$  时：

(1) 对  $s(t-\Delta t)$  进行泰勒展开

$$\begin{aligned} s(t-\Delta t) &\approx s(t) - s'(t)\Delta t + \mathcal{O}(\Delta t^2) \\ \frac{s(t)}{s(t-\Delta t)} &\approx \frac{s(t)}{s(t) - s'(t)\Delta t} \approx 1 + \frac{s'(t)}{s(t)} \Delta t \\ \left( \frac{s(t)}{s(t-\Delta t)} - 1 \right) &\approx \frac{s'(t)}{s(t)} \Delta t \end{aligned}$$

当 $\Delta t \rightarrow 1$ 令：

$$f(t) = \left( \frac{s(t)}{s(t-1)} - 1 \right) \approx \frac{s'(t)}{s(t)}$$

(2)对 $\sigma^2(t - \Delta t)$  泰勒展开

$$\sigma(t - \Delta t)^2 \approx \sigma(t)^2 - \frac{d}{dt}[\sigma(t)^2]\Delta t$$

因此

$$\sigma(t)^2 - \sigma(t - \Delta t)^2 \approx 2\sigma(t)\sigma'(t)\Delta t$$

当 $\Delta t \rightarrow 1$

$$\sigma(t)^2 - \sigma(t-1)^2 \approx 2\sigma(t)\sigma'(t)$$

因此

$$g(t) = s(t)\sqrt{\sigma(t)^2 - \sigma(t-1)^2}\epsilon_t = s(t)\sqrt{2\sigma(t)\sigma'(t)}$$

### 【证明3】VE 前向离散形式推导

$$dx = \sqrt{\frac{d^2\sigma(t)}{dt}}dw_t$$

由欧拉离散化 $x_t = x_{t-1} + dx(t-1)$ ：

$$x_t = x_{t-1} + \sqrt{\frac{\sigma^2(t) - \sigma^2(t-\Delta t)}{\Delta t}}\sqrt{\Delta t}\epsilon$$

令 $\Delta t = 1$ 得：

$$x_t = x_{t-1} + \sqrt{\sigma^2(t) - \sigma^2(t-1)}\epsilon$$

由迭代求和可得：

$$x_t = x_0 + \sigma(t)\epsilon$$

特别地， $\sigma(t) = \sigma_{\min}(\frac{\sigma_{\max}}{\sigma_{\min}})^t$ ，其中 $t \sim \mathcal{U}(0, 1)$

也可由式（24）通式带入相关项得到。

### 【证明4】VP 前向离散形式推导

$$d\mathbf{x}_t = -\frac{1}{2}\beta(t)\mathbf{x}_tdt + \sqrt{\beta(t)}dw_t$$

欧拉离散化：

$$\mathbf{x}_t = \mathbf{x}_{t-1} + (-\frac{1}{2}\beta(t)\mathbf{x}_{t-1}\Delta t + \sqrt{\beta(t)}\sqrt{\Delta t}\epsilon)$$

令 $\Delta t = 1$ ，得：

$$\mathbf{x}_t = (1 - \frac{1}{2}\beta(t))\mathbf{x}_{t-1} + \sqrt{\beta(t)}\epsilon$$

泰勒展开近似有： $\sqrt{1 - \beta(t)} = 1 - \frac{1}{2}\beta(t)$  得：

$$\mathbf{x}_t = \sqrt{1 - \beta(t)}\mathbf{x}_{t-1} + \sqrt{\beta(t)}\epsilon$$

递推求和得,  $\sigma(t)^2 = 1 - e^{-\int_0^t \beta(s) ds}$ :

$$\mathbf{x}_t = \sqrt{\bar{\alpha}_t} \mathbf{x}_0 + \sqrt{1 - \bar{\alpha}_t} \epsilon$$

特别地,  $\beta(t) = (\beta_{max} - \beta_{min})t + \beta_{min}$ , 其中,  $t \sim \mathcal{U}(\epsilon_t, 1)$

给定VP逆向SDE:

$$d\mathbf{x} = \left[ -\frac{1}{2}\beta(t)\mathbf{x} - \beta(t)\nabla_x \log p_t(\mathbf{x}) \right] dt + \sqrt{\beta(t)} d\bar{\mathbf{w}}$$

(1) 忽略扩散项

$$d\mathbf{x} = \left[ -\frac{1}{2}\beta(t)\mathbf{x} - \beta(t)\nabla_x \log p_t(\mathbf{x}) \right] dt$$

(2) 欧拉离散化

对时间 (t) 离散化, 步长  $\Delta t$ , 逆向时间从 (t) 到 (t-1):

$$\mathbf{x}_{t-1} = \mathbf{x}_t + \left[ -\frac{1}{2}\beta_t \mathbf{x}_t - \beta_t \nabla_x \log p_t(\mathbf{x}_t) \right] (-\Delta t)$$

令  $\Delta t = 1$

$$\mathbf{x}_{t-1} = \mathbf{x}_t + \frac{1}{2}\beta_t \mathbf{x}_t + \beta_t \nabla_x \log p_t(\mathbf{x}_t)$$

(3) 得分函数替换 (残差形式):

$$\nabla_x \log p_t(\mathbf{x}_t) \approx -\frac{\epsilon_\theta(\mathbf{x}_t, t)}{\sqrt{1 - \bar{\alpha}_t}} = \frac{1}{s(t)\sigma^2(t)} (D_\theta(\frac{\mathbf{x}_t}{s(t)}; \sigma(t)) - \frac{\mathbf{x}_t}{s(t)})$$

- 这里  $\epsilon_\theta$  表示噪声方向
- 前者为噪声预测形式, 后者为残差形式。
- $\sqrt{1 - \bar{\alpha}_t} = \bar{\sigma}_t$
- $\mathbf{x}_t$  表示原模型输入  $F_\theta$  的形式

直观理解: 分母  $\sqrt{1 - \bar{\alpha}_t} = s(t)\sigma(t)$ , 分子量纲为  $\sigma(t) \cdot \epsilon_\theta$ , 因此需要再除一份  $\sigma(t)$

(3') 得分函数替换 (噪声形式):

$$\begin{aligned} \mathbf{x}_{t-1} &= \mathbf{x}_t + \frac{1}{2}\beta_t \mathbf{x}_t - \beta_t \frac{\epsilon_\theta(\mathbf{x}_t, t)}{\sqrt{1 - \bar{\alpha}_t}} \\ \mathbf{x}_{t-1} &= \left(1 + \frac{1}{2}\beta_t\right) \mathbf{x}_t - \beta_t \frac{\epsilon_\theta(\mathbf{x}_t, t)}{\sqrt{1 - \bar{\alpha}_t}} \end{aligned}$$

此处  $1 + \frac{1}{2}\beta_t$  泰勒展开近似

$$\mathbf{x}_{t-1} \approx \frac{1}{\sqrt{\alpha_t}} \left( \mathbf{x}_t - \sqrt{\alpha_t} \beta_t \frac{\epsilon_\theta(\mathbf{x}_t, t)}{\sqrt{1 - \bar{\alpha}_t}} \right)$$

由于  $\beta_t$  近似0,  $\alpha_t = 1 - \beta_t$  近似1, 因此  $\sqrt{\alpha_t} \beta_t \approx \beta_t$

(4) 加入随机噪声

噪声预测形式:

$$\mathbf{x}_{t-1} = \frac{1}{\sqrt{\alpha_t}} \left( \mathbf{x}_t - \beta_t \frac{\epsilon_\theta(\mathbf{x}_t, t)}{\sqrt{1 - \bar{\alpha}_t}} \right) + \sqrt{\beta_t} \epsilon_t$$

残差预测形式:

$$\mathbf{x}_{t-1} = \frac{1}{\sqrt{1-\beta_t}} \left( \mathbf{x}_t + \beta_t \frac{D_\theta(\frac{\mathbf{x}_t}{s(t)}; \sigma(t)) - \frac{\mathbf{x}_t}{s(t)}}{s(t)\sigma^2(t)} \right) + \sqrt{\beta_t} \epsilon_t$$

## 【证明6】VE 反向形式离散化推导过程

由VE反向连续SDE形式：

$$d\mathbf{x} = \sqrt{\frac{d\sigma^2(t)}{dt}} d\mathbf{w}_t$$

可知

$$g(t) = \sqrt{\frac{d\sigma^2(t)}{dt}}$$

对VE反向SDE连续形式欧拉离散化：

$$\begin{aligned} d\mathbf{x} &= [-g(t)^2 \nabla_x \log p_t(\mathbf{x})] dt + g(t) d\mathbf{w} \\ \mathbf{x}_{t-1} &= \mathbf{x}_t + d\mathbf{x}_t = \mathbf{x}_t + \frac{d^2\sigma_t}{dt} \cdot \frac{\epsilon_\theta(\mathbf{x}_t, t)}{\bar{\sigma}_t} + \sqrt{\frac{d^2\sigma_t}{dt}} \sqrt{dt} \epsilon_t \end{aligned}$$

进一步，令 $\Delta t = 1$ ：

$$\mathbf{x}_{t-1} = \mathbf{x}_t + \frac{\sigma_t^2 - \sigma_{t-1}^2}{\bar{\sigma}_t} \epsilon_\theta(\mathbf{x}_t, t) + \sqrt{\sigma_t^2 - \sigma_{t-1}^2} \epsilon_t$$

由于 $\bar{\sigma}(t) = \sigma(t)$ ，因此：

噪声预测形式：

$$\mathbf{x}_{t-1} = \mathbf{x}_t + \frac{\sigma_t^2 - \sigma_{t-1}^2}{\sigma_t} \epsilon_\theta(\mathbf{x}_t, t) + \sqrt{\sigma_t^2 - \sigma_{t-1}^2} \epsilon_t$$

残差预测形式：

$$\mathbf{x}_{t-1} = \mathbf{x}_t - (\sigma_t^2 - \sigma_{t-1}^2) \frac{1}{s(t)\sigma^2(t)} (D_\theta(\frac{\mathbf{x}_t}{s(t)}; \sigma(t)) - \frac{\mathbf{x}_t}{s(t)}) + \sqrt{\sigma_t^2 - \sigma_{t-1}^2} \epsilon_t$$

由于 $s(t) = 1$ ，有：

$$\mathbf{x}_{t-1} = \mathbf{x}_t - \frac{(\sigma_t^2 - \sigma_{t-1}^2)}{\sigma_t^2} (D_\theta(\mathbf{x}_t; \sigma_t) - \mathbf{x}_t) + \sqrt{\sigma_t^2 - \sigma_{t-1}^2} \epsilon_t$$

## 实践

### VP

- 训练
  - 时间调度： $t \sim \mathcal{U}(\epsilon_t, 1)$
  - 噪声调度： $\sigma(t) = \sqrt{e^{\frac{1}{2}\beta_d t^2 + \beta_{\min} t} - 1}$
  - 损失权重： $\lambda(\sigma) = \frac{1}{\sigma^2}$
- 采样
  - 时间调度： $t \sim 1 + \frac{i}{N-1}(\epsilon_s - 1)$
  - 噪声调度： $\sigma(t) = \sqrt{e^{\frac{1}{2}\beta_d t^2 + \beta_{\min} t} - 1}$
  - 缩放系数： $s(t) = \frac{1}{\sqrt{1-\sigma^2}}$

## VE

- 训练
  - 时间调度:  $t \sim \mathcal{U}(\ln(\sigma_{min}), \ln(\sigma_{max}))$
  - 噪声调度:  $\sigma(t) = \sigma_{min}(\frac{\sigma_{max}}{\sigma_{min}})^t$ , 人为定义
  - 损失权重:  $\lambda(\sigma) = \frac{1}{\sigma^2}$
- 采样
  - 时间调度:  $t \sim \sigma_{max}^2 (\frac{\sigma_{min}^2}{\sigma_{max}^2})^{\frac{i}{N-1}}$
  - 噪声调度:  $\sqrt{t}$ , 理论值
  - 缩放系数:  $s(t) = 1$

## EDM

- 训练
  - 时间调度: 无
  - 噪声调度:  $\ln(\sigma) \sim \mathcal{N}(P_{mean}, P_{std}^2)$
  - 损失权重:  $\frac{\sigma^2 + \sigma_{data}^2}{(\sigma \cdot \sigma_{data})^2}$
- 采样
  - 时间调度:  $(\sigma_{max}^{\frac{1}{\rho}} + \frac{i}{N-1}(\sigma_{min}^{\frac{1}{\rho}} - \sigma_{max}^{\frac{1}{\rho}}))^{\rho}$
  - 噪声调度:  $\sigma(t) = t$
  - 缩放系数:  $s(t) = 1$

VE、EDM采用的是基于sigma的均匀分布采样训练，需要避免极端噪声

VP采用的是基于时间均匀分布的采样训练，可以覆盖全部时间范围