扩散模型首先破坏高频信息，随后影响低频信息

证明：扩散模型首先破坏高频信息，随后影响低频信息

引言

在匹配流（matching flow）扩散模型中，数据通过一个正向扩散过程逐渐添加噪声，最终转化为纯噪声信号。在逆过程（去噪）中，模型学习从噪声中逐步还原原始数据。理解在正向扩散过程中不同频率成分的变化对于模型的设计和改进具有重要意义。本文将严格证明，在扩散开始时，高频信息首先被破坏，随着时间的推移，低频信息才逐渐受到影响。

1. 扩散模型的定义

考虑一个连续时间的扩散过程，初始数据为 $\mathbf{x}_0$，随着时间 $t \in [0, T]$的推移，数据演化为 $\mathbf{x}_t$。扩散过程可以被以下随机微分方程（SDE）描述：

$$d\mathbf{x}_t = f(\mathbf{x}_t, t)dt + g(t)d\mathbf{w}_t$$

其中：

- $\mathbf{w}_t$是维纳过程（布朗运动），代表随机噪声；
- $f(\mathbf{x}_t, t)$是漂移项，我们可以设 $f(\mathbf{x}_t, t) = 0$以简化分析；
- $g(t)$是时间相关的扩散系数，控制噪声的强度。

基于上述简化，扩散过程可表示为：

$d\mathbf{x}_t = g(t) d\mathbf{w}_t$

初始条件为 $\mathbf{x}_0$。

2. 频域分析

为了分析扩散过程对不同频率成分的影响，我们将信号从时域转化到频域。设信号 \mathbf{x}t *的傅里叶变换为 $\hat{\mathbf{x}}*t(\omega)，其中\omega$表示频率。傅里叶变换的线性性质和维纳过程的特性使得在频域中的分析更加简洁。

由于噪声 $d\mathbf{w}_t$是白噪声，其在频域中的功率谱密度是常数。

3. 推导过程

（1）**初始信号与噪声的表示**

正向扩散过程中，信号 $\mathbf{x}_t$可表示为初始信号与累积噪声的和：

$\mathbf{x}_t = \mathbf{x}_0 + \int_0^t g(s) d\mathbf{w}_s$

在频域中：

$\hat{\mathbf{x}}_t(\omega) = \hat{\mathbf{x}}_0(\omega) + \hat{\boldsymbol{\epsilon}}_t(\omega)$

其中， $\hat{\boldsymbol{\epsilon}}_t(\omega)$是噪声项的傅里叶变换，满足：

- **均值**： $\mathbb{E}[\hat{\boldsymbol{\epsilon}}_t(\omega)] = 0$
- **方差**： $\mathbb{E}[|\hat{\boldsymbol{\epsilon}}_t(\omega)|^2] = \int_0^t |g(s)|^2 ds$

因为 d\mathbf{w}_t 是白噪声，所以其功率谱密度是平坦的，且噪声在不同频率上是相互独立的。

（2）**信号功率谱密度的计算**

信号 $\mathbf{x}_t$的功率谱密度为：

$$S_{\mathbf{x}_t}(\omega) = \mathbb{E}[|\hat{\mathbf{x}}_t(\omega)|^2]$$

展开后：

$$S_{\mathbf{x}_t}(\omega) = |\hat{\mathbf{x}}_0(\omega)|^2 + 2\mathfrak{Re}\left(\hat{\mathbf{x}}_0(\omega)\mathbb{E}[\hat{\boldsymbol{\epsilon}}_t^*(\omega)]\right) + \mathbb{E}[|\hat{\boldsymbol{\epsilon}}_t(\omega)|^2]$$

由于 $\mathbb{E}[\hat{\boldsymbol{\epsilon}}_t(\omega)] = 0$，所以交叉项为零：

$$2\mathfrak{Re}\left(\hat{\mathbf{x}}_0(\omega)\mathbb{E}[\hat{\boldsymbol{\epsilon}}_t^*(\omega)]\right) = 0$$

因此：

$$S_{\mathbf{x}_t}(\omega) = |\hat{\mathbf{x}}_0(\omega)|^2 + \mathbb{E}[|\hat{\boldsymbol{\epsilon}}_t(\omega)|^2]$$

其中， $\mathbb{E}[|\hat{\boldsymbol{\epsilon}}_t(\omega)|^2] = \int_0^t |g(s)|^2 ds$，与频率 ω无关。

（3）**信噪比（SNR）的计算**

信噪比定义为信号功率与噪声功率之比：

$$\mathrm{SNR}(\omega) = \frac{|\hat{\mathbf{x}}_0(\omega)|^2}{\mathbb{E}[|\hat{\boldsymbol{\epsilon}}_t(\omega)|^2]} = \frac{|\hat{\mathbf{x}}_0(\omega)|^2}{\int_0^t |g(s)|^2 ds}$$

我们发现，信噪比随频率 \omega 的变化完全由初始信号的功率谱密度 $|\hat{\mathbf{x}}_0(\omega)|^2$决定，而噪声功率在所有频率上都是相同的。

（4）**频率对信噪比的影响**

对于自然信号（如图像和声音信号），一般具有低通特性，即：

$|\hat{\mathbf{x}}_0(\omega)|^2 \propto |\omega|^{-\alpha}$

其中 $\alpha > 0$表示信号的频谱衰减程度。

因此，随着频率 ω的增加， $|\hat{\mathbf{x}}_0(\omega)|^2$迅速减小。这意味着在高频处，信噪比 $\mathrm{SNR}(\omega)$较低。

当 t 增加时，分母 $\int_0^t |g(s)|^2 ds$增大，导致信噪比整体下降。

4. 信噪比达到阈值的时间与频率的关系

设定一个信噪比阈值 γ，那么对于特定频率 ω，信噪比达到阈值 γ的时间 $t_\gamma(\omega)$满足：

$$\mathrm{SNR}(\omega) = \frac{|\hat{\mathbf{x}}_0(\omega)|^2}{\int_0^{t_\gamma(\omega)} |g(s)|^2 ds} = \gamma$$

解出：

$$\int_0^{t_\gamma(\omega)} |g(s)|^2 ds = \frac{|\hat{\mathbf{x}}_0(\omega)|^2}{\gamma}$$

由于 $|\hat{\mathbf{x}}_0(\omega)|^2$随 \omega 增加而减小，因此频率越高，所需的 $t_\gamma(\omega)$越小。这表明高频成分的信噪比更快降至阈值，即高频信息先被破坏。

5. 数值示例

假设 $g(t) = \sigma$为常数，即扩散过程中的噪声强度恒定。

则：

$$\int_0^{t_\gamma(\omega)} \sigma^2 ds = \sigma^2 t_\gamma(\omega) = \frac{K|\omega|^{-\alpha}}{\gamma}$$

解出：

$$t_\gamma(\omega) = \frac{K}{\sigma^2\gamma}|\omega|^{-\alpha}$$

这清晰地表明：

- 频率 ω越大（高频）， $t_\gamma(\omega)$越小；
- 高频成分更快地达到信噪比阈值，先被噪声淹没；
- 低频成分由于初始功率较高，能够在更长时间内保持较高的信噪比。

6. 结论

通过上述严格的数学推导，我们证明了在匹配流扩散模型的正向扩散过程中，高频信息首先被噪声破坏，而低频信息在较长时间内保持相对完整。这是由于自然信号的频谱特性，初始信号的高频成分功率较低，因而在扩散过程中更快地被噪声淹没。

这种特性对于扩散模型的设计和去噪算法的优化具有重要意义。在逆扩散过程中，模型需要先重建低频结构，再逐步恢复高频细节。这与人类感知图像的过程类似：我们首先感知整体结构（低频），然后注意到细节（高频）。

参考

- Sohl-Dickstein, J., Weiss, E., Maheswaranathan, N., & Ganguli, S. (2015). Deep Unsupervised Learning using Nonequilibrium Thermodynamics.
- Ho, J., Jain, A., & Abbeel, P. (2020). Denoising Diffusion Probabilistic Models.