

什么是谱偏差？

上海师范大学数理学院

Outline

现象	2
研究方法	3
F-Principle 的原因	4
激活函数正则性	4
损失函数中的频率权重	5
混叠效应	6
深度频率原理	7
神经网络的积分方程性质	8
利用 F-Principle	9
克服 Spectral Bias	10

谱偏差(spectral bias)是指在深度神经网络的训练过程中，神经网络容易拟合到数据的低频成分，难以捕捉高频成分的现象。这也被叫做频率原理(F-Principle)。

例如在拟合函数 $f(x) = \sin x + \sin 5x$ 时，DNN往往先学习到低频的 $\sin x$ 部分，而高频的 $\sin 5x$ 需要更多的训练时间或者根本无法被有效拟合。

谱偏差现象具有普遍性，无论DNN的过参数化如何，或者是在低维还是高维场景，都被观察到谱偏差现象。

研究方法

在低维场景，许志钦团队使用合成数据，设计目标函数 $f(x) = \sin x + \sin 3x + \sin 5x$, 使用离散傅里叶变换(DFT)计算了目标函数和网络输出的频率成分的相对误差进行实验。结果发现：网络的低频成分快速收敛，高频成分收敛很慢。

此外，在DNN拟合“摄影师图像”实验中，DNN先拟合轮廓后拟合细节，进一步验证了F-Principle.

在高维场景，离散傅里叶变换受到维数灾难影响。研究人员提出了以下两种方法，验证谱偏差现象在高维场景下依然存在：

-投影法：取输入空间第一主成分 p_1 , 将输入 x_i 投影到 $p_1 \cdot x_i$ 进行一维傅里叶变换。

-滤波法：利用高斯卷积 $y_{\text{low}, \sigma}^i = (y * G_\sigma)^i$ 保留频率低于 $\frac{1}{\sigma}$ 的成分，模糊高频成分，进一步使用均方误差分析。

F-Principle 的原因

激活函数正则性

\tanh 等激活函数的 Fourier 谱随 k 指数衰减。

研究者使用带参数 a 的里克函数（Ricker function）作为激活函数。 a 取值越小， Ricker function 的傅里叶谱从越高频率开始衰减。

实验结果表明：

- 当激活函数的傅里叶谱从低频开始衰减时，低频分量的收敛速度显著快于高频分量；
- 当激活函数的傅里叶谱从高频开始衰减时，我们无法观察到任何频率分量的优先收敛现象。这与理论分析完全一致。

F-Principle 的原因

损失函数中的频率权重

研究者使用两种损失函数进行对比实验，一种是常用均方损失 L_{nograd} ，另一种是引入梯度损失的损失函数 $L_{\text{grad}} = L_{\text{nograd}} + \frac{1}{n} \sum_i^n (\nabla_x f_\theta(x_i) - \nabla_x f^*(x_i))^2$.

实验结果表明：在引入梯度信息的损失函数 L_{grad} 作用下，高频分量的收敛速度显著加快。

这一现象的原因在于： $\nabla_x f_\theta(x)$ 的傅里叶变换等于 $ik\hat{f}_\theta(k)$ ，即梯度信息在频域中放大了高频成分的权重。

频率收敛行为是激活函数与损失函数共同作用的结果。在 PINN 中，损失函数中引入了偏微分方程的梯度信息，此时谱偏差现象未必一定成立。

F-Principle 的原因

混叠效应

根据香农采样定理，采样频率应当大于信号中最高频率的两倍，否则会产生混叠效应。

也就是说，当输入数据的采样频率较低时，高频成分会被错误地映射为低频成分，从而导致神经网络更容易学习到这些错误的低频成分，而忽略了真实的高频信息。

深度神经网络在训练过程中，由于频率原理，往往会优先拟合这些由混叠效应产生的虚假低频分量，以匹配训练样本，而真实的高频分量则会被牺牲，最终导致实验中观测到的泛化性能下降。

F-Principle 的原因

深度频率原理

Xu 和 Zhou 提出了深度频率原理，以理解深度在加速训练中的效应。对于深度神经网络，第 1 个隐藏层的有效目标函数可以通过以下方式理解：其输入是第(1-1)层的输出。第 1 层部分的任务是学习从第(1-1)层的输出到真实标签的映射。因此，第 1 个隐藏层的有效目标函数由第(1-1)层的输出和真实标签组成。

Xu 和 Zhou 通过实证研究发现了一个深度频率原理：在训练过程中，更深隐藏层的有效目标函数偏向于更低的频率。由于 F-原理，这项实证研究为理解深度为何能加速训练提供了一个理论基础。

F-Principle 的原因

神经网络的积分方程性质

神经网络的梯度流可被描述为积分方程，而积分方程具有更高的正则性，进而使其在傅里叶域中呈现出快速衰减的特性。与之形成对比的是求导的微分方程性质会降低正则性。

因此传统方法，例如 Jacobian 法更早学习到的是目标的高频成分。许志钦等人开发了一种混合策略：

- 先用 DNN 进行优化，快速获得一个低频成分已基本正确的初始解。
- 再将此 DNN 解作为初始猜测，代入雅可比法等传统迭代法继续求解，以快速补充高频成分。

利用 F-Principle

神经网络能够有效的本质是归纳偏差(inductive bias)。F-Principle 作为神经网络的基本归纳偏差，可以降低对噪声的过拟合，从而提升泛化能力。

在实际应用中，当训练数据被噪声污染时，通常会采用早停策略以避免模型过拟合。基于频率原则的视角，早停策略能够防止模型拟合训练数据中由噪声主导的高频分量，从而自然地得到泛化性能优良的模型。

克服 Spectral Bias

但在某些应用中，我们希望神经网络能够更好地捕捉高频信息。最简单的方法是 PhaseDNN: PhaseDNN 的核心步骤如下：先对数据执行离散时间变换 (DTT)，再将频域数据划分为若干等频谱长度的子带；对每个子带数据，乘以其中心频率 ω 对应的调谐因子 $e^{-i\omega x}$ ，实现高频谱域向低频谱域的迁移；随后通过傅里叶逆变换转换至时域，交由神经网络完成低频特征学习；最后将学习后的特征映射回原始高频域，并对各子带结果进行求和，得到目标函数。

这一方法的缺点是：傅里叶级数的项数会随数据维度的增加呈指数级增长，这导致相位深度神经网络不可避免地面临维数灾难问题。

稍后我们会介绍另一种克服谱偏差的方法——随机傅里叶特征。