



北京大学

# 本科生毕业论文

题目：利用贝叶斯神经网络代理模型  
进行失效概率估计

Title: Estimating failure probability  
with Bayesian neural networks  
as surrogate models

姓名：严格

学号：1700010674

院系：数学科学学院

专业：信息与计算科学

指导教师：杨超

二〇二一年六月



## 版权声明

任何收存和保管本论文各种版本的单位和个人，未经本论文作者同意，不得将本论文转借他人，亦不得随意复制、抄录、拍照或以任何方式传播。否则一旦引起有碍作者著作权之问题，将可能承担法律责任。



## 摘要

失效概率估计问题是工程等领域应用中经常遇到的问题。在估计失效概率时，由于判断系统是否失效的开销往往很大，如何降低计算成本是一个重要问题。本文结合了代理模型方法<sup>[1]</sup>和贝叶斯网络求解偏微分方程的方法<sup>[2]</sup>来求解失效概率问题，并利用贝叶斯神经网络的方差信息对代理模型方法<sup>[1]</sup>进行了改进。此外，本文探索了使用重要性采样法来降低所需的采样数量的可能性。在一个涉及二维泊松方程的测试问题上，本文进行了相应的数值实验。

关键词：失效概率，代理模型，贝叶斯神经网络



Estimating failure probability with Bayesian neural networks as surrogate models

Yan Ge (Information and computing mathematics)

Directed by Yang Chao

**ABSTRACT**

Estimating failure probability is a kind of problems occurs frequently in engineering and other aspects. Because it is often computationally expensive to determine the state of the system when estimating failure probability, it is important to reduce the cost of computation. This article combines surrogate method<sup>[1]</sup> and method to solve partial differential equations via Bayesian neural networks<sup>[2]</sup> to solve failure probability problems. This article exploits the variation information provided by Bayesian neural networks to refine the surrogate algorithm in Li et al. [1]. In addition, this article explores the possibility to reduce sample number via importance sampling. This article conducts numerical experiments on a test problem involving a 2-dimensional Poisson equation.

**KEY WORDS:** failure probability, surrogate model, Bayesian neural networks





## 目录

<b>第一章 引言</b>	<b>1</b>
1.1 引言 . . . . .	1
1.2 问题描述 . . . . .	2
1.3 论文结构 . . . . .	2
<b>第二章 代理模型</b>	<b>5</b>
2.1 校正边界样本 . . . . .	5
2.2 迭代方法 . . . . .	6
<b>第三章 用神经网络模型估计偏微分方程的解</b>	<b>7</b>
3.1 编码器-译码器网络 . . . . .	7
3.2 贝叶斯网络 . . . . .	8
<b>第四章 失效概率估计</b>	<b>11</b>
4.1 用贝叶斯网络作为代理模型 . . . . .	11
4.2 利用方差信息 . . . . .	11
4.3 重要性采样 . . . . .	12
<b>第五章 数值实验与结果</b>	<b>15</b>
5.1 数值实验设计 . . . . .	15
5.2 代理模型训练 . . . . .	15
5.3 Monte Carlo 方法 . . . . .	15
5.4 重要性采样方法 . . . . .	18
<b>第六章 总结和讨论</b>	<b>19</b>
<b>参考文献</b>	<b>21</b>
<b>致谢</b>	<b>23</b>
<b>北京大学学位论文原创性声明和使用授权说明</b>	<b>25</b>



## 第一章 引言

### 1.1 引言

失效概率估计在工程应用、风险管理等领域都具有重要作用。例如，在结构设计中，运用失效概率估计的方法，可以帮助评估结构的可靠性。从数学的角度来看，计算失效概率的问题，可以看作一个多变量的数值积分问题：给定系统参数  $\eta \in \mathbf{R}^M$  的分布  $p_\eta$ ，我们希望求解系统状态处于某个失效区域  $F = \{\eta | f(\eta) \geq 0\}$  的概率，即计算积分

$$P_f = \int_{\mathbf{R}^M} I_{\{f(\eta) \geq 0\}} p_\eta d\eta. \quad (1.1)$$

由于梯形法等传统的数值积分计算方法的计算量随维数指数增长，在问题的维数  $M$  相对较高时，蒙特卡洛 (Monte Carlo) 算法是求解这一问题的常用方法。然而，使用 Monte Carlo 算法进行采样时，每次采样都要计算  $f(\eta)$  以确定系统是否失效，而  $f(\eta)$  的计算开销通常较大。例如，需要求解一个偏微分方程。而 Monte Carlo 法为了得到合理的精度，采样数量往往要达到  $10^6$  数量级以上，这就使得总的计算开销非常巨大。

为了减低计算开销，J. Li et al. [1] 提出，通过代理模型的方法来计算失效概率。代理模型是对  $f(\eta)$  进行低成本的近似计算的模型。Li 等希望用一个计算代价相对低的代理模型来取代计算代价相对高的  $f(\eta)$  的精确计算，以达到降低计算开销的目的。之后，J. Li 等<sup>[3]</sup> 又提出将代理模型与重要性采样法结合，以降低方差，减少采样数目，从而节约计算成本。

在 Li et al. [1] 的代理模型方法的基础上，本文主要做了三方面的工作：第一，Li et al. [1] 中选用的代理模型为 generalized polynomial chaos(gPC)，而这一模型在  $f(\eta)$  涉及到一些复杂的偏微分方程时并不适用。因而，本文将最近提出的基于贝叶斯神经网络的偏微分方程近似求解方法作为代理模型，结合 Li et al. [1] 中的算法求解失效概率问题；第二，本文利用贝叶斯神经网络的方差信息，对 Li et al. [1] 中的算法进行了改进；第三，本文探索了运用重要性采样方法减少采样次数的可能性。

## 1.2 问题描述

本文关注如下的失效概率问题。考虑一个二维的稳态流体问题，渗透系数为一高斯随机场  $A$ 。由达西定律 (Darcy's law)，可以得到以下的偏微分方程，

$$\begin{aligned} u &= A\nabla p \\ \nabla \cdot u &= f. \end{aligned}$$

整理得，

$$\nabla \cdot (A\nabla p) = f. \quad (1.2)$$

这是一个关于压力  $p$  的二维泊松方程。问题的研究区域为  $[0, 1] \times [0, 1]$ 。边界条件设置为混合边界条件，上下边界为 Neumann 边界条件，左右边界为 Dirichlet 边界条件：

$$\begin{cases} p(x, y) = 0, & x = 0, 1; \\ p_y(x, y) = 0, & y = 0, 1. \end{cases} \quad (1.3)$$

右端项为  $f(x, y) = 1$ 。设随机场  $A$  的均值  $a_0(x) = 1$ ，标准差  $\sigma = 0.25$ ，协方差函数

$$\text{Cov}(x, y) = \sigma^2 \exp\left(-\frac{1}{l_c}(|x_1 - y_1| + |x_2 - y_2|)\right). \quad (1.4)$$

其中， $x = (x_1, x_2)$ ， $y = (y_1, y_2)$ ， $l_c$  设置为 0.8。

对于随机场  $A$ ，采用 Karhunen-Loève (KL) 展开的方式，将其离散化。

$$A(x, y) = a_0(x) + \sum_{i=1}^M \sqrt{\lambda_i} a_i(x, y) \eta_i \quad (1.5)$$

其中， $\lambda_i$  是协方差矩阵的特征值， $a_i$  是相应的特征函数， $\eta_i$  是随机变量。本文设置  $\eta_i$  互相独立地服从  $[-1, 1]$  上的均匀分布。

当给定点  $(x_0, y_0)$  处的压力  $p$  大于给定点阈值  $p_{max}$  时，我们称系统失效。这样一来，失效区域  $F = \{\eta | f(\eta) \geq 0\}$ ，其中

$$f(\eta) = p(x_0, y_0) - p_{max}. \quad (1.6)$$

这样，问题便转化为一个积分问题：求

$$P_f = \int_{R^M} I_{\{p(x_0, y_0) \geq p_{max}\}} p_\eta d\eta. \quad (1.7)$$

## 1.3 论文结构

本文之后的结构如下：第二章与第三章介绍本文应用到的两个重要算法：第二章介绍用代理模型解失效概率问题的方法；第三章介绍用贝叶斯神经网络近似求解偏微

分方程的方法。第四章介绍本文的主要贡献，即如何结合前两章所介绍的方法来求解失效概率问题和在其基础上进行的改进，以及对重要性采样方法的探索。第五章为数值实验及结果，第六章为总结和讨论。



## 第二章 代理模型

对于失效概率的计算，造成计算开销巨大的重要原因是失效条件  $f(\eta) \geq 0$  的判断中， $f(\eta)$  的计算往往需要求解偏微分方程，因而计算代价很大。为了解决这一问题，Li et al. [1] 提出，对于失效概率问题，可以通过引入代理模型  $\hat{f}(\eta)$  来近似计算  $f(\eta)$ ，从而降低失效条件计算的开销。代理模型是对  $f(\eta)$  进行近似的模型，其需要满足两个要求：第一，其对  $f(\eta)$  的近似要有一定的精度；第二，其计算代价应该相对直接精确计算  $f(\eta)$  较小。

考虑本文第一章中所描述的问题，可用一个代理模型  $\hat{f}$  来近似计算压强  $p$ ，

$$\hat{p} = \hat{f}(\eta). \quad (2.1)$$

进而，在使用 Monte Carlo 等方法计算式 1.7 时，就可用  $I_{\{\hat{p}(x_0, y_0) \geq p_{max}\}}$  来近似  $I_{\{p(x_0, y_0) \geq p_{max}\}}$ ，从而减少计算成本。

### 2.1 校正边界样本

Li et al. [1] 提出，即使代理模型精度很高，直接使用代理模型来代替精确计算仍然可能带来问题。不妨考虑以下的例子：假设代理模型计算的结果

$$\hat{p} = \hat{f}(\eta) = p + \epsilon. \quad (2.2)$$

设  $p_{max} = 1$ ，真实解  $p(\eta)$  服从  $[0, 1]$  上的均匀分布，此时失效概率的真实解为 0，但用代理模型计算得到的失效概率将为  $\epsilon$ 。这个例子虽然特殊，但其提示我们即使代理模型的误差很小，其也有可能对于靠近失效区域边界的样本做出错误的判断。此时，直接使用代理模型进行计算，得到的失效概率仍可能有数量级上的误差。

为了尽可能减少代理模型与精确解的误差对我们估计失效区域造成的影响，Li et al. [1] 提出，在靠近失效区域边界处的样本，采用精确模拟的结果来代替代理模型的计算结果。这一做法类似于预估-校正的思想：先用不精确的代理模型估计  $f(\eta)$ ，再对位于边界值附近的样本通过精确求解的方式进行校正。具体地说，可以引入一个函数

$$\tilde{p}(\eta, p_{max}, \delta) = \begin{cases} f(\eta), & |F(\eta) - p_{max}| \leq \delta; \\ \hat{f}(\eta), & \text{else.} \end{cases} \quad (2.3)$$

来近似  $p(\eta)$ 。其中， $\delta$  是设定的校正阈值， $f(\eta)$  是精确解， $\hat{f}(\eta)$  是使用的代理模型。即，对  $|p - p_{max}| > \delta$  的样本直接使用代理模型的结果；在  $|p - p_{max}| \leq \delta$  时，采用精确

计算的结果来校正代理模型，从而防止对失效与否的误判。通过混合使用代理模型和精确计算，一方面减少了精确计算  $f(\eta)$  的次数，降低了计算成本；另一方面，对处于失效区域边界附近的样本，防止了代理模型的误差导致的误判。

## 2.2 迭代方法

之后，Li et al. [3] 又进一步发展了 Li et al. [1] 中的方法。前述的方法在决定哪些样本需要精确计算时，需要恰当地设置参数  $\delta$  的值。若  $\delta$  取得太大，需要精确计算的样本数量就会上升，从而增大计算开销。而  $\delta$  取得过小，又会使部分被代理模型错误判断的点得不到校正。为了避免  $\delta$  选取中的潜在的困难，Li 等提出了一种基于迭代的方法。将其运用在第一章所介绍的问题上，可表述如下

---

### Algorithm 1 Li et al. [3] 求解失效概率问题的代理模型迭代算法

---

输入：随机生成的大小为  $K$  的样本集  $\{\eta_i\}_{i=1}^K$ ，迭代步长  $S$ ，代理模型  $\hat{f}$ ，精确求解模型  $f$

输出：失效概率  $P_f$

用代理模型计算： $p_i \leftarrow \hat{f}(\eta_i)$ 。

排序：将  $\{p_i\}$  按  $|p_i - p_{max}|$  从小到大排列得到数组  $p$ ，相应地将样本排列为数组  $\eta$ 。

$q \leftarrow 1$

**while**  $P_f = |\{p[i] \mid p[i] \geq p_{max}\}| / K$  未收敛 **do**

    用精确计算校正边界样本： $p[q : q + S] \leftarrow f(\eta[q : q + S])$ 。

$q \leftarrow q + S$

**end while**

---

算法通过迭代的方式，按照  $|p - p_{max}|$  从小到大的顺序，逐步校正样本直至计算的失效概率收敛。这一算法不需要设置参数  $\delta$ ，从而避免了潜在的问题。



### 第三章 用神经网络模型估计偏微分方程的解

神经网络作为一种强大的函数拟合模型，在图像分类、图像识别、自然语言处理等许多领域受到了广泛地应用。最近，有许多工作将神经网络模型运用到偏微分方程的数值求解中来，如 Raissi et al. [4], Zhu et al. [2], Zhu et al. [5] 等。在运用第二章中介绍的方法求解本文所述的失效概率问题时，需要寻找一个代理模型来近似求解方程1.2。因此，我们考虑选取一个基于神经网络近似求解偏微分方程的模型来作为这个代理模型。本文选择了 Zhu et al. [2] 中提出的基于贝叶斯网络的模型，下面对其作一个简要的介绍。

Zhu et al. [2] 中考虑了一个二维随机偏微分方程的求解。假设偏微分方程中包含一随机场  $X(s, \omega), s \in S, \omega \in \Omega$ , 其中  $s$  为空间坐标,  $\Omega$  是样本空间。现在的问题是，对于生成的某一随机样本  $x(s)$ , 希望求解相应的  $y = y(s, x(s))$ 。

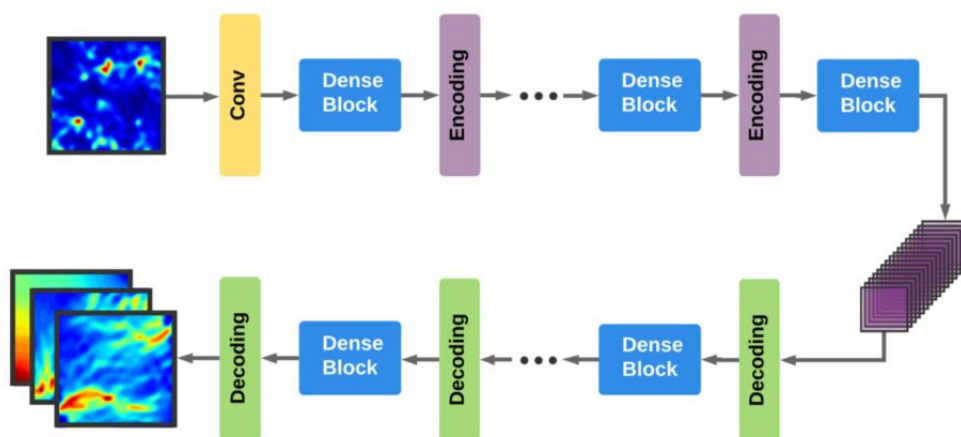
对于二维问题，Zhu et al. [2] 用  $W \times H$  的均匀矩形网格对  $x, y$  进行离散。离散后的问题转化为，已知  $x(s)$  在二维网格格点上的值  $X \in R^{W \times H}$ , 希望预测偏微分方程的解  $y$  在格点上的值  $Y \in R^{W \times H}$ 。Zhu et al. [2] 提出了基于深度连接的全卷积编码器-译码器网络 (DenseED) 来进行预测。这一网络架构由两部分构成，第一部分为编码器网络，将二维的输入  $X$  映射到一个低维的向量  $Z = Encoder(X)$ ; 第二部分为译码器网络，将  $Z$  恢复到二维预测值  $Y: Y = Decoder(z)$ 。

#### 3.1 编码器-译码器网络

在编码器网络上，Zhu et al. [2] 使用的是基于深度连接的神经网络 (DenseNet)<sup>[6]</sup> 的全卷积网络架构。DenseNet<sup>[6]</sup> 是一种在图像识别等领域使用的卷积神经网络 (CNN)，其在残差神经网络 (Resnet)<sup>[7]</sup> 和高速公路神经网络 (Highway Network)<sup>[8]</sup> 的基础上做了进一步拓展。

一般的卷积神经网络中，输入  $x$  通过神经网络的每一层而得到输出  $y$ , 即  $x_{l+1} = h_l(x_l)(0 < l \leq M), x_0 = x, y = x_M$ 。卷积神经网络的每一层由卷积层，非线性函数 (常用的如线性整流函数 (ReLU)<sup>[9]</sup>) 以及批正则化层 (Batch Normalization)<sup>[10]</sup> 组成。

在此基础上，DenseNet<sup>[6]</sup> 增加了跨层之间的连接，将每一层的输出作为之后各层的输入，即  $x_{l+1} = h_l([x_l, x_{l-1}, \dots, x_0])$ 。在应用到编码器网络或译码器网络上时，由于编码器网络需要进行下采样，译码器网络需要进行上采样，从而改变输入的空间尺寸大小，而不同大小的  $x_i$  无法拼接起来。为此，需要将网络分成不同的稠密模块 (Dense Block)，每个模块中输入的空间尺寸相同并添加有跨层连接，模块之间进行下采样 (编


 图 3.1 DenseED<sup>[2]</sup> 的网络结构。本图引自 Zhu et al. [2]

码器网络)或是上采样(译码器网络),在减少特征层的数目的同时改变空间尺寸。具体地而言,编码器网络把通常的神经网络层中的步长为 1 的卷积替换为一个步长大于 1 的卷积层,从而实现下采样。而译码器网络则把卷积替换为转置卷积(ConvT)<sup>[11]</sup>来实现上采样。

### 3.2 贝叶斯网络

一般的确定性神经网络中,参数  $\theta$  是一个确定的值,其输出  $y = f(x, \theta)$ 。而贝叶斯神经网络将网络的参数  $\theta$  当作一个随机变量,同时在输出上添加随机的噪声,即  $y = f(x, \theta_w) + n(\theta_n)$ 。对于贝叶斯神经网络,需要给定参数  $\theta = \{\theta_w, \theta_n\}$  的先验分布,然后利用训练数据,计算相应的后验分布。Zhu et al. [2] 选择 t 分布作为参数  $\theta$  的先验分布,即  $p(\theta|\alpha) = N(\theta|0, \alpha^{-1}I)$ ,  $p(\alpha) = \text{Gamma}(\alpha|a_0, b_0)$ 。

在计算参数的后验分布时,Zhu et al. [2] 采用了粒子近似的方式。具体地说,选取一组参数值  $\{\theta_i\}_{i=1}^S$ , 有经验测度  $\mu_S(d\theta) = \frac{1}{S} \sum_{i=1}^S \delta(\theta - \theta_i)d\theta$ 。Zhu et al. [2] 希望将测度  $\mu_S$  通过某个变换  $T$  后,使其与后验分布  $p(\theta|D)$  的交叉熵最小。为了达到这一目的,Zhu et al. [2] 使用了随机变分梯度下降(SVGD)方法<sup>[12]</sup>。具体的算法如下:

**Algorithm 2** 用 SVGD 算法做贝叶斯推断<sup>[2]</sup>

---

输入：初始样本  $\{\theta_0^i\}_{i=1}^S$ , 目标函数  $\nabla \log p(\theta)$ , 核函数  $k(\theta, \theta')$ , 步长  $\{\epsilon_t\}$

输出：样本  $\{\theta^i\}_{i=1}^S$ , 其经验分布作为参数后验分布的近似

**for** 第 t 个迭代步 **do**

$$\theta_{t+1}^i \leftarrow \theta_t^i + \epsilon_t \phi(\theta_t^i);$$

$$\phi(\theta_t^i) = \frac{1}{S} \sum_{j=1}^S [k(\theta_t^j, \theta_t^i) \nabla_{\theta_t^j} \log p(\theta_t^j) + \nabla_{\theta_t^j} k(\theta_t^j, \theta_t^i)]$$

**end for**

---



## 第四章 失效概率估计

本章将结合第三章中介绍的近似计算偏微分方程解的贝叶斯网络模型和第二章中介绍的代理模型方法，对第一章中所描述的失效概率问题进行求解。

### 4.1 用贝叶斯网络作为代理模型

第三章中介绍的贝叶斯神经网络方法<sup>[2]</sup>具有两个特点：第一，其能对我们关心的偏微分方程的解做一个近似；第二，其计算的时间开销相对于传统的有限元方法等要小很多。因此，可以将其作为代理模型  $\hat{f}(\eta)$ ，应用第二章中的算法1来计算失效概率。

### 4.2 利用方差信息

除了直接应用算法1以外，注意到贝叶斯神经网络方法<sup>[2]</sup>不但给出了解的估计，还对解的不确定性给出了一定的量度，即方差信息。因此，本文希望利用这一方差信息，来对算法1做改进。

考虑一个简化的情形：假设

$$p = p + N, \quad (4.1)$$

$N = N(0, \sigma^2)$  是一个高斯噪声，那么我们就得到此时代理模型得到的失效情况与实际解不一致的概率

$$P = 1 - T\left(\frac{|p - p_{max}|}{\sigma}\right), \quad (4.2)$$

其中， $T$  是高斯分布的累计分布函数。这一简化的情形启示我们， $|p - p_{max}|/\sigma$  可能是一个关键的量。这个值越低，意味着代理模型对失效与否给出了错误的判断的可能性越大。因此， $|p - p_{max}|/\sigma$  低的样本应当优先进行精确校正。

遵循这一思路，本文对算法1进行了改进。对于第一章中所述的问题，1中原本按照  $|p - p_{max}|$  来排序。本文将算法1中的排序一步改为按照  $|p - p_{max}|/\sigma$  排序。优先校正  $|p - p_{max}|/\sigma$  较小的那些样本。改进后的算法见算法3。

---

**Algorithm 3** 利用方差信息改进的算法1

---

输入：随机生成的大小为  $K$  的样本集  $\{\eta_i\}_{i=1}^K$ ，迭代步长  $S$ ，代理模型  $\hat{f}$ ，精确求解模型  $f$

输出：失效概率  $P_f$

用代理模型计算： $p_i, \sigma_i \leftarrow \hat{f}(\eta_i)$ 。

排序：将  $\{p_i\}$  按  $|p_i - p_{max}|/\sigma_i$  从小到大排列得到数组  $p$ ，相应地将样本排列为数组  $\eta$ 。

$q \leftarrow 1$

**while**  $P = |\{p[i] \mid p[i] \geq p_{max}\}|/K$  未收敛 **do**

    用精确计算校正边界样本： $p[q : q + S] \leftarrow f(\eta[q : q + S])$

$q \leftarrow q + S$

**end while**

---

### 4.3 重要性采样

在用蒙特卡洛方法进行数值积分计算时，如何降低方差是影响性能的重要因素。重要性采样方法是一种降低方差的重要方法。积分

$$I = \int_{\Omega} f(x) dx = \int_{\Omega} \frac{f(x)}{p(x)} p(x) dx. \quad (4.3)$$

故通过随机生成  $M$  个服从概率密度函数  $p(x)$  的样本，计算

$$\hat{I} = \frac{1}{M} \sum_{i=1}^M \frac{f(x_i)}{p(x_i)}. \quad (4.4)$$

可以得到积分的数值近似。对于重要性采样方法，其方差

$$\begin{aligned} V_{IS} &= \int_{\Omega} \left( \frac{f(x)}{p(x)} - I \right)^2 p(x) dx, \\ &= \int_{\Omega} \frac{f^2(x)}{p(x)} dx - I^2. \end{aligned} \quad (4.5)$$

由(4.5)可知，当

$$p(x) = p^* = f(x)/I \quad (4.6)$$

时，方差达到最小值 0。但由于我们事先并不知道积分值  $I$ ，我们无法直接使用  $p^*$  作为重要性采样中的  $p(x)$ 。因此，如何构造一个好的分布  $p(x)$  使得方差(4.5)尽可能小，是重要性采样方法的关键。Zhu et al. [2] 中采用了交叉熵方法<sup>[13]</sup>。对于两个概率分布  $p_1$  和  $p_2$ ，他们的交叉熵（又称 KL 散度）如下

$$D(p_1|p_2) = \int p_1(x) \ln \frac{p_1(x)}{p_2(x)} dx. \quad (4.7)$$

交叉熵方法要求从一族概率分布  $p(x; \nu)$  中寻找与分布  $p^*$  最“接近”的分布，通过最小化其与  $p^*$  之间的交叉熵的方式。具体地，交叉熵方法希望求

$$\nu^* = \operatorname{argmin}_{\nu} D(p^* | p(x; \nu)). \quad (4.8)$$

利用(4.7)和(4.6),(4.8)等价于求

$$\nu^* = \operatorname{argmax}_{\nu} \int_{\Omega} f(x) \ln p(x; \nu) dx. \quad (4.9)$$

在分布族  $p(x, \nu)$  的选择上，一种常见的选取是高斯混合分布。然而这一选择在维数变高时会面临“维数灾难”的问题，因而对于本文所述的问题并不适用。为了解决维度过高带来的问题，本文尝试采用了一种基于深度神经网络的生成模型 KRnet<sup>[14]</sup> 来作为重要性采样的分布族。由于本文的参数范围为有限区间，本文在 KRnet<sup>[14]</sup> 之后又加入了一个 Sigmoid 函数和线性变换

$$x^i = 2\operatorname{Sigmoid}(x_{KR}^i) - 1. \quad (4.10)$$

这样，输出的结果  $x$  就落在区间  $[-1, 1]^M$  上。

由于深度神经网络在训练样本数量较少时面临着过拟合的问题，本文参考 Wan et al. [15] 中的做法，加入了正则项

$$R = \beta \nabla_x \log \frac{f(x)}{p(x)}. \quad (4.11)$$

综合起来，损失函数为：

$$F(x) = - \int_{\Omega} f(x) \ln p(x; \nu) dx + \beta \nabla_x \log \frac{f(x)}{p(x)}. \quad (4.12)$$

优化式4.12的关键在于求前半部分的积分，这可以通过再次使用重要性采样法进行。即

$$\int_{\Omega} f(x) \ln p(x; \nu) dx = \int_{\Omega} \frac{f(x)}{p(x; \mu)} \ln p(x; \nu) p(x; \mu) dx \quad (4.13)$$

对于第一章中的问题， $f(x) = I_{\{p(\eta) \geq p_{max}\}} p_{\eta}$ 。为了求解这一优化问题，本文借鉴了 Li et al. [3] 中的多层适应性迭代<sup>[13]</sup> 算法，将算法中所用到的分布族由高斯分布改为 KRnet<sup>[14]</sup>，并加入正则项，得到以下的算法。

**Algorithm 4** 利用多层适应性迭代<sup>[13]</sup>的交叉熵方法<sup>[3]</sup>

输入：参数  $\rho \in [10^{-2}, 10^{-1}]$ , 正整数  $M_s$ , 初始的 KRnet 模型  $p_{\text{KR}}(\eta; \nu_0)$ , 迭代总步数  $T$

输出：KRnet 模型  $p_{\text{KR}}(\eta; \nu)$

初始化  $q$ : 从初始分布  $p_{\text{KR}}(\eta; \nu_0)$  采样  $M_s$  个样本  $\{\eta_0^i\}_{i=1}^M$ , 选择最小的  $q_0 \geq 0$ , 使得满足  $p(\eta_0^i) \geq p_{\text{max}} - q_0$  的样本数量大于  $\rho M_s$ 。  $q \leftarrow q_0$ 。

**for** 第  $t$  个迭代步 **do**

更新 KRnet 模型: 用  $p_{\text{KR}}(\eta; \nu_{t-1})$  作为式4.13中的  $p(x; \mu)$  进行重要性采样, 对式4.12进行优化, 得到  $p_{\text{KR}}(\eta; \nu_t)$ 。本步可利用上一步更新  $q$  时采样计算的样本, 以减少采样与计算的次数。

更新  $q$ : 从  $p_{\text{KR}}(\eta; \nu_t)$  采样  $M_s$  个样本  $\{\eta_0^i\}_{i=1}^M$ , 选择最小的  $q \geq 0$ , 使得满足  $p(\eta_0^i) \geq p_{\text{max}} - q$  的样本数量大于  $\rho M_s$ 。

**end for**



## 第五章 数值实验与结果

### 5.1 数值实验设计

我们采用第一章中所介绍的流体问题作为测试问题。对于问题中的这一泊松方程，我们用  $64 \times 64$  的均匀网格对问题进行离散，用有限元方法求解作为问题的精确解。在  $(x_0, y_0)$  与  $p_{max}$  的选取上，为了简便起见， $(x_0, y_0)$  均选取在网格点上，利用 Monte Carlo 方法进行  $10^5$  次模拟后，选择  $p_{max}$  使得失效概率  $P_f = 10^{-3}$ 。对于每一给定的维数  $M$ ，选取 5 组  $(x_0, y_0)$  与  $p_{max}$  进行实验。

### 5.2 代理模型训练

在代理模型方面，我们选用了第二章中所介绍的 Zhu et al. [2] 的贝叶斯神经网络作为代理模型。我们随机生成了 512 个样本，利用有限元方法进行求解，得到结果后作为贝叶斯神经网络的训练集。另外，我们随机生成了 512 个样本作为测试集。训练参数方面，学习率选择为 0.0025，噪声学习率为 0.01，批尺寸为 32，进行 300 个周期的训练。对于误差，采用 RMSE 进行衡量：

$$RMSE(y, \hat{y}) = \sqrt{\frac{1}{N} \sum_{i=1}^N \|y_i - \hat{y}_i\|_2^2} \quad (5.1)$$

其中， $N$  是测试集的大小， $y_i, \hat{y}_i$  分别是第  $i$  个测试数据的真实值和神经网络计算的值。训练过程中网络在训练集和测试集上的 RMSE 如图 5.1 与 5.2 所示。可以看到，在 300 个周期的训练后，贝叶斯神经网络模型在测试集上达到了较好的精度。

### 5.3 Monte Carlo 方法

本节中，应用第二章所介绍的算法 1 (A) 和第四章中介绍的利用方差信息改进的算法 3 (B)，分别在维数  $M = 8$  和维数  $M = 48$  的情况下，按照前述方法各选取 5 组测试点进行计算。在参数的选取上，步长  $S = 100$ 。收敛的判断方法是，连续 3 个迭代步计算的概率  $P_f$  没有改变。数值结果见表 5.1 与表 5.2 所示。

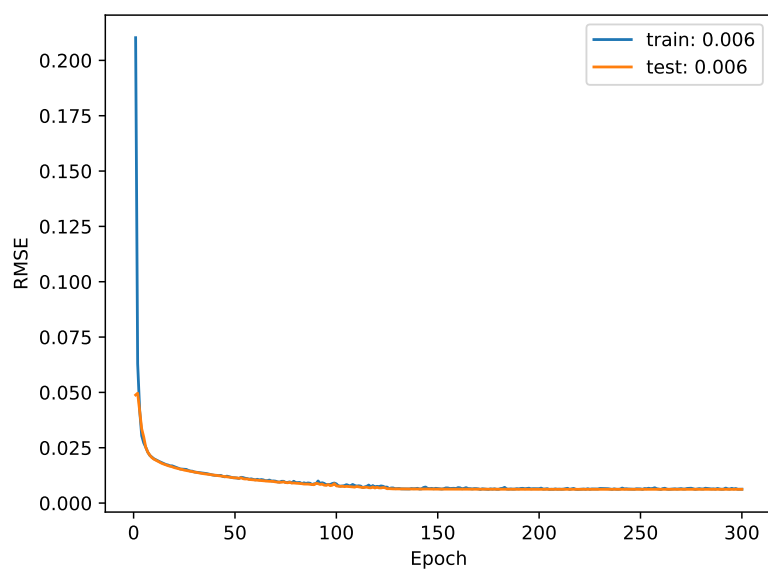


图 5.1 维数  $M = 8$  时贝叶斯神经网络<sup>[2]</sup> 在训练集和测试集上的 RMSE 随训练周期变化的图像。

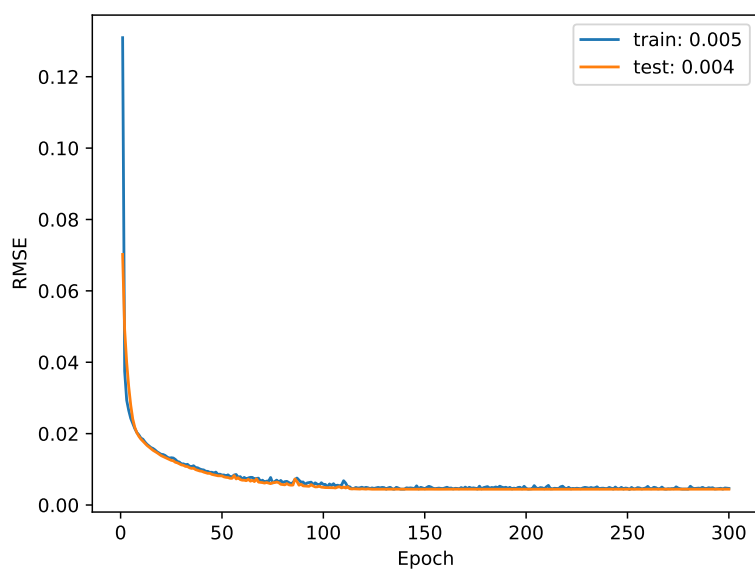


图 5.2 维数  $M = 48$  贝叶斯神经网络<sup>[2]</sup> 在训练集和测试集上的 RMSE 随训练周期变化的图像。

测试点序号	$(x_0, y_0)$	$p_{max}$	相对误差 (A)	相对误差 (B)	精确计算次数 (A)	精确计算次数 (B)
1	$(\frac{7}{64}, \frac{52}{64})$	0.1426	4%	15%	4100	2100
2	$(\frac{9}{64}, \frac{24}{64})$	0.1159	11%	15%	800	1200
3	$(\frac{10}{64}, \frac{60}{64})$	0.1212	2%	3%	1200	1600
4	$(\frac{12}{64}, \frac{32}{64})$	0.1385	4%	6%	1000	1000
5	$(\frac{14}{64}, \frac{4}{64})$	0.0905	1%	7%	1300	1500
平均值			4.4%	9.2%	1700	1480

表 5.1 维数  $M = 8$  时, 算法1 (A) 和算法3 (B) 在五个测试点上的相对误差及精确计算次数比较

测试点序号	$(x_0, y_0)$	$p_{max}$	相对误差 (A)	相对误差 (B)	精确计算次数 (A)	精确计算次数 (B)
1	$(\frac{7}{64}, \frac{52}{64})$	0.1435	16%	0%	3500	3200
2	$(\frac{9}{64}, \frac{24}{64})$	0.1169	12%	26%	1400	1900
3	$(\frac{10}{64}, \frac{60}{64})$	0.1223	5%	4%	1600	1500
4	$(\frac{12}{64}, \frac{32}{64})$	0.1391	1%	10%	1800	1900
5	$(\frac{14}{64}, \frac{4}{64})$	0.0920	16%	3%	1200	1500
平均值			10%	8.6%	1900	2000

表 5.2 维数  $M = 48$  时, 算法1 (A) 和算法3 (B) 在五个测试点上的相对误差及精确计算次数比较

## 5.4 重要性采样方法

由于重要性采样方法在维度过高时会遇到“维数灾难”带来的困难，本文仅在维数  $M = 8$  的问题上应用了重要性采样方法。在采样分布  $\text{KRnet}^{[14]}$  的设置上，本文设置维数分割步长  $n_s = 2$ ，不使用旋转层，深度  $D = 4$ ，宽度  $W = 24$ 。有关这些参数含义的详细介绍可参考 Tang et al. [14]。在训练上，初始时先在一个由  $[-1, 1]^8 \in \mathbb{R}^8$  上的均匀分布采样出来的，大小为 1028 的数据集上，采用学习率  $l = 0.001$ ，批尺寸 256，正则项系数  $\beta_0 = 5$ ，进行了 10000 个周期的训练，作为算法4中的初始模型。接下来，取参数  $\rho = 0.05$ ，迭代总步数  $T = 5$ 。每步采样并计算  $M_s = 5000$  个样本，采用学习率  $l = 0.001$ ，批尺寸 64，正则项系数  $\beta = 1$ ，进行 10000 个周期的训练。训练完后，利用其作为重要性采样中的采样分布，做  $10^4$  次采样，在五个样本点上的结果如表5.3。表中  $\sigma_{MC} \approx 3.16 \times 10^{-4}$ ，是用 Monte Carlo 方法进行  $10^4$  次采样计算失效概率的标准差。

测试点序号	$(x_0, y_0)$	$p_{max}$	计算结果	相对误差	标准差 $\sigma$	$\sigma/\sigma_{MC}$
1	$(\frac{7}{64}, \frac{52}{64})$	0.1435	$1.20 \times 10^{-3}$	20%	$1.22 \times 10^{-4}$	38.5%
2	$(\frac{9}{64}, \frac{24}{64})$	0.1169	$1.08 \times 10^{-3}$	8%	$0.85 \times 10^{-4}$	27.0%
3	$(\frac{10}{64}, \frac{60}{64})$	0.1223	$1.02 \times 10^{-3}$	2%	$1.26 \times 10^{-4}$	40.0%
4	$(\frac{12}{64}, \frac{32}{64})$	0.1391	$1.01 \times 10^{-3}$	1%	$0.80 \times 10^{-4}$	25.4%
5	$(\frac{14}{64}, \frac{4}{64})$	0.0920	$0.96 \times 10^{-3}$	4%	$1.03 \times 10^{-4}$	32.7%

表 5.3 问题维数  $M = 8$  时，算法4在五个测试点上的计算结果及标准差

## 第六章 总结和讨论

本文的主要工作是：第一，将 Li et al. [1] 的代理模型方法与 Zhu et al. [2] 的贝叶斯神经网络预测偏微分方程解的方法结合起来，运用到失效概率问题上；第二，利用贝叶斯神经网络的方差信息，对 Li et al. [1] 中的算法进行了改进。第三，探索在相对高维问题上运用生成模型进行重要性采样的可能性。

数值实验的结果表明：第一，利用代理模型的 Monte Carlo 方法无论在相对低的 8 维问题上还是在相对较高的 48 维问题上，都能用直接 Monte Carlo 方法 1% – 2% 的精确模拟数量，达到大体相当的精度；第二，利用方差信息改进的算法，在一些样本点上能起到降低精确模拟样本数量的作用；第三，用重要性采样方法，将标准差降低到了使用 Monte Carlo 方法的 20% – 40%，从而能一定程度上降低采样数量。

在实验过程中，仍然发现了一些问题。例如，利用重要性采样方法时，训练 KRnet<sup>[14]</sup> 带来了相当大的额外的计算开销。在未来的工作上，可以考虑：第一，对生成模型进行改进，减少训练带来的额外的计算代价，并尝试向更高的维度上进行推广。第二，利用类似于 Li et al. [3] 中的方法，将代理模型方法与重要性采样方法，以达到更好的效果；第三，就对贝叶斯神经网络的方差利用上，本文的方法是相对较为简单的，可以进一步探索，是否有更好的方法利用贝叶斯网络提供的方差信息。



## 参考文献

- [1] LI J, XIU D. Evaluation of failure probability via surrogate models[J]. *J. Comput. Phys.*, 2010, 229: 8966-8980.
- [2] ZHU Y, ZABARAS N. Bayesian deep convolutional encoder–decoder networks for surrogate modeling and uncertainty quantification[J/OL]. *Journal of Computational Physics*, 2018, 366: 415-447. <https://www.sciencedirect.com/science/article/pii/S0021999118302341>. DOI: <https://doi.org/10.1016/j.jcp.2018.04.018>.
- [3] LI J, XIU D. An efficient surrogate-based method for computing rare failure probability[J]. *J. Comput. Phys.*, 2011, 230: 8683-8697.
- [4] RAISSI M, PERDIKARIS P, KARNIADAKIS G. Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations[J]. *J. Comput. Phys.*, 2019, 378: 686-707.
- [5] ZHU Y, ZABARAS N, KOUTSOURELAKIS P S, et al. Physics-constrained deep learning for high-dimensional surrogate modeling and uncertainty quantification without labeled data[J/OL]. *Journal of Computational Physics*, 2019, 394: 56-81. <https://www.sciencedirect.com/science/article/pii/S0021999119303559>. DOI: <https://doi.org/10.1016/j.jcp.2019.05.024>.
- [6] HUANG G, LIU Z, van der MAATEN L, et al. Densely Connected Convolutional Networks[C]// *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. [S.l. : s.n.], 2017.
- [7] HE K, ZHANG X, REN S, et al. Deep residual learning for image recognition[C]// *Proceedings of the IEEE conference on computer vision and pattern recognition*. [S.l. : s.n.], 2016: 770-778.
- [8] SRIVASTAVA R K, GREFF K, SCHMIDHUBER J. Training very deep networks[J]. *ArXiv preprint arXiv:1507.06228*, 2015.
- [9] GLOROT X, BORDES A, BENGIO Y. Deep sparse rectifier neural networks[C]// *Proceedings of the fourteenth international conference on artificial intelligence and statistics*. [S.l. : s.n.], 2011: 315-323.
- [10] IOFFE S, SZEGEDY C. Batch normalization: Accelerating deep network training by reducing internal covariate shift[C]// *International conference on machine learning*. [S.l. : s.n.], 2015: 448-456.
- [11] TEAM T T D, AL-RFOU R, ALAIN G, et al. Theano: A Python framework for fast computation of mathematical expressions[J]. *ArXiv preprint arXiv:1605.02688*, 2016.
- [12] LIU Q, WANG D. Stein Variational Gradient Descent: A General Purpose Bayesian Inference Algorithm[Z]. 2019. arXiv: 1608.04471 [stat.ML].
- [13] DE BOER P T, KROESE D P, MANNOR S, et al. A tutorial on the cross-entropy method[J]. *Annals of operations research*, 2005, 134(1): 19-67.
- [14] TANG K, WAN X, LIAO Q. Deep density estimation via invertible block-triangular mapping[J]. *Theoretical and Applied Mechanics Letters*, 2020, 10(3): 143-148.

- [15] WAN X, WEI S. Coupling the reduced-order model and the generative model for an importance sampling estimator[J]. *Journal of Computational Physics*, 2020, 408: 109281.



## 致谢

感谢杨超老师对本文的指导。感谢唐科军学长为本文提供了有限元方法的相关代码，以及在本文写作过程中为我提供的讨论与帮助。



# 北京大学学位论文原创性声明和使用授权说明

## 原创性声明

本人郑重声明：所提交的学位论文，是本人在导师的指导下，独立进行研究工作所取得的成果。除文中已经注明引用的内容外，本论文不含任何其他个人或集体已经发表或撰写过的作品或成果。对本文的研究做出重要贡献的个人和集体，均已在文中以明确方式标明。本声明的法律结果由本人承担。

论文作者签名： 严格

日期：2021年5月28日

## 学位论文使用授权说明

本人完全了解北京大学关于收集、保存、使用学位论文的规定，即：

- 按照学校要求提交学位论文的印刷本和电子版；
- 学校有权保留学位论文的印刷本和电子版，并提供目录检索与阅览服务，在校园网上提供服务；
- 学校可以采用影印、缩印、数字化或其它复制手段保存论文；

论文作者签名： 严格 导师签名： 杨超

日期：2021年5月28日