arXiv:2602.05559v1 [stat.CO] 5 Feb 2026

用于偏微分方程系数贝叶斯推断的分段确定性马尔可夫过程

Leon Riccius[1*]、IuriB.C.M.Rocha[1]、、JorisBierkens[2]、、Hanne Kekkonen[2]、、Frans P.vander Meer[1]

[1]代尔夫特理工大学土木工程与地球科学学院，Stevinweg 1，代尔夫特，2628 CN，荷兰。[2]代尔夫特理工大学代尔夫特应用数学研究所，Mekelweg 4，代尔夫特，2628 CD，荷兰。

*通讯作者。电子邮件：l.f.riccius@tudelft.nl；贡献作者：i.rocha@tudelft.nl；joris.bierkens@tudelft.nl；h.n.kekkonen@tudelft.nl；f.p.vandermeer@tudelft.nl；

摘要

我们提出了一个分段确定性马尔可夫过程（PDMP）采样器通用框架，可在似然函数计算开销较大的非线性逆问题中实现高效的贝叶斯推断。其关键是代理辅助稀疏化方案：代理模型提供提议事件率，而稳健的校正机制会在检测到违反上界的情况时动态调整加性偏移量，从而确保真实速率不超过上界。该框架不依赖于特定的代理模型或 PDMP。我们以 Zig-Zag采样器和弹跳粒子采样器为例进行展示，采用的代理模型包括常数、拉普拉斯和高斯过程（GP）代理模型，以及利用梯度信息和自适应细化的 GP 变体。作为一项代表性应用，我们研究了一维线弹性问题中空间变化的杨氏模量的贝叶斯推断。在不同维度下，配备基于高斯过程的代理的 PDMP 采样器，与随机游走 Metropolis 算法和无转向采样器相比，每次正向模型评估均能获得显著更高的准确度和有效样本量。弹跳粒子采样器展现出最优的整体效率和扩展性，说明所提出的 PDMP 框架在这一特定场景之外也具有应用潜力。

关键词： 贝叶斯推断、分段确定性马尔可夫过程、Zig-Zag采样器、弹跳粒子采样器、马尔可夫链蒙特卡罗、高斯过程回归、偏微分方程

1 引言

马尔可夫链蒙特卡罗（MCMC）方法是众多科学与工程学科的基石，包括统计物理学、计算生物学和机器学习。这些方法广泛用于从复杂的、通常未归一化的概率分布中进行采样，而这类分布经常出现在上述领域中。MCMC 的一个特别重要的应用是贝叶斯推断，即从后验分布中抽取样本，根据观测数据量化模型参数的不确定性。

大多数成熟的 MCMC 方法都基于随机游走 Metropolis（RWM）算法 [1]及其对非对称提议的推广形式——Metropolis-Hastings 算法 [2]。尽管这些算法因简单且能保证渐近收敛而广受欢迎，但由于其本质上具有扩散性，在高维参数空间中的采样效率较低 [3]。为克服这些局限，研究者提出了多种采样效率更高的先进方法。值得一提的例子包括一阶的 Metropolis 调整朗之万算法 [4], 哈密顿蒙特卡罗（HMC）[5],及其二阶流形感知扩展 [6]。无转向采样器（NUTS） [7] 通过无需手动调整轨迹长度即可实现大幅跃迁，进一步改进了 HMC，并广泛应用于 Stan 等概率编程框架 [8] 和 PyMC [9]。这些方法在实际工程问题上的实证评估 [6, 10, 11] 证实，它们在典型测试分布上表现出的理论优越收敛特性 [12, 13]往往也适用于现实工程应用。

上述所有方法都有一个共同特征：可逆性。具体而言，它们满足细致平衡条件，这确保了目标分布的收敛，但由于动力学呈扩散性，其采样效率受到限制 [14,ch 6.1]。近年来，人们对不可逆采样算法的兴趣日益浓厚，理论结果表明，这类算法有望实现更快的收敛和更低的渐近方差 [15, 16]。

不可逆算法大致可分为两类。一种方法是将可逆算法提升到更高维的状态空间 [17]。第二种方法则依赖于分段确定性马尔可夫过程（PDMP）。基于 PDMP 的采样器包括 Zig-Zag 采样器 [18]，弹跳粒子采样器 [19], 和 Boomerang 采样器 [20]。这些采样器以确定性方式演化，偶尔会被随机事件打断，从而使其轨迹突然改变。尽管这些算法具有良好的理论性质，例如良好的混合性、较低的自相关以及无误差子采样 [21], ，但除了少数简单的学术示例之外，其模拟成本高得令人难以承受。迄今为止，PDMP 采样器在实际问题中的应用仍然有限，例如线性贝叶斯反问题 [22]。

PDMP采样器定义了一个非齐次泊松过程。在此过程中，下一次事件发生前的等待时间服从指数分布，其速率取决于当前状态以及目标对数密度的梯度。直接模拟下一次事件时间需要求解一个积分方程，而该方程只有在少数简单分布（例如具有二次形式的高斯分布）下才有闭式解

对数密度。对于更复杂的分布，可以通过对事件率进行数值积分来获得下一次事件时间。在 [23], Zig-Zag采样器和弹跳粒子采样器采用前向欧拉格式进行离散化，本质上是在状态网格上定义了一个马尔可夫链。文献 [24]还研究了高阶离散化。虽然数值积分使 PDMP 采样能够适用于更广泛的分布，但在大多数情况下，其计算成本仍缺乏竞争力。

也可以通过泊松稀疏化间接生成事件 [21]。如果存在上界过程，即速率严格高于目标过程速率的泊松过程，则可采用稀疏化。从该上界过程生成的候选事件，会以真实速率与上界之比的概率被接受。只有在能够高效地生成上界过程中的事件时，泊松稀疏化才切实可行。在 [25]，中，通过对 Zig-Zag 采样器的切换率进行多项式拟合，求得给定时间范围内的常数上界。每当状态更新时，都会重复这一过程。

直接求逆和 [25] 中的稀疏化，都可以概括为模拟 PDMP 的纯局部方法，因为它们只考虑当前状态附近的对数密度和事件率。本文提出一种更全局的 PDMP 模拟方法，尝试寻找一个可在整个定义域内使用的事件率近似。理论上，只需知道 PDMP 沿轨迹的事件率，就足以生成下一个事件。实际上，了解定义域其余部分的事件率信息，也有助于生成该事件。

我们的方法使用代理模型构建用于泊松稀疏化的提议率，以实现近似 PDMP 采样。其一项关键特性是校正机制：在模拟过程中检测到违反上界条件时，会动态调整代理模型。这种校正机制使稀疏化过程不依赖于所用代理模型的具体形式；唯一要求是代理模型能够导出可计算的提议率。这种灵活性使我们能够探索多种代理族——从简单的常数势函数或拉普拉斯近似，到表达能力更强的高斯过程模型——并设计自适应版本，将采样过程本身产生的信息纳入其中。

我们的方法适用于定义在连续参数空间上的单峰平滑后验分布。这类分布通常出现在椭圆型偏微分方程（PDE）的背景下，例如热传导、静电学或扩散问题，以及对其系数的推断。我们在线弹性的一个简单但具有说明性的示例问题上展示并测试了该方法，在此问题中，我们根据含噪的位移观测推断参数场。我们的方法不依赖于具体选择哪种 PDMP 采样器。我们将其应用于 Zig-Zag 采样器和弹跳粒子采样器，并比较它们的性能。在所有考察的度量指标和设置下，这两种采样器均优于经过良好调优的随机游走 Metropolis 基线和广受欢迎的 NUTS，这标志着 PDMP 采样在贝叶斯推断中的应用迈出了新的一步。

本文其余部分安排如下。在第2节中，我们回顾 PDMP 采样的基础。第3 节介绍我们的代理辅助 PDMP 采样框架及代理模型的选择。第4节展示数值实验，说明我们的方法在一个 PDE 控制的贝叶斯逆问题中的性能。最后，第5 节总结本文，并展望未来的研究方向。

2 分段确定性马尔可夫过程采样器

分段确定性马尔可夫过程是连续时间马尔可夫过程，其特点是确定性轨迹会被随机事件打断。我们仅讨论用于从目标分布 $\pi(\boldsymbol{x}) \propto \exp(-\Psi(\boldsymbol{x}))$ 中进行采样的分段确定性马尔可夫过程，其中定义在 $\mathcal{X} = \mathbb{R}^d$上的 $\Psi : \mathbb{R}^d \to \mathbb{R}$ 是一个连续可微的势函数。通过为位置变量 $\boldsymbol{x}_t \in \mathcal{X}$ 增添速度变量 $\boldsymbol{v}_t \in \mathcal{V} = \mathbb{R}^d$，即可得到一类广泛使用的 PDMP 采样器，其中包括 Zig-Zag 和弹跳粒子采样器。按照 [26], 的定义，PDMP 由以下三个部分构成：

1. 确定性动力学：在事件之间，位置 $\boldsymbol{x}_t \in \mathcal{X} = \mathbb{R}^d$ 和速度$\boldsymbol{v}_t \in \mathcal{V} = \mathbb{R}^d$ 按照以下规律确定性演化：

$$\frac{\mathrm{d}\boldsymbol{x}_t}{\mathrm{d}t} = \boldsymbol{\Phi}_{\boldsymbol{x}}(\boldsymbol{x}_t, \boldsymbol{v}_t), \quad \frac{\mathrm{d}\boldsymbol{v}_t}{\mathrm{d}t} = \boldsymbol{\Phi}_{\boldsymbol{v}}(\boldsymbol{x}_t, \boldsymbol{v}_t), \tag{1}$$

其中， $\boldsymbol{\Phi}_{\boldsymbol{x}}$ 和 $\boldsymbol{\Phi}_{\boldsymbol{v}}$ 是控制演化的向量场。

2. 事件率：在切换时刻 τ，PDMP 会发生突然转变。这些事件按照一个状态依赖的非负事件率为 $\lambda(\boldsymbol{x}_t, \boldsymbol{v}_t)$ 的非齐次泊松过程发生。

3. 转移核：发生事件时，PDMP 根据转移核 $q(\boldsymbol{x}_t', \boldsymbol{v}_t' | \boldsymbol{x}_t, \boldsymbol{v}_t)$ 从 $(\boldsymbol{x}_t, \boldsymbol{v}_t)$ 转移到 $(\boldsymbol{x}_t', \boldsymbol{v}_t')$。

对 $\boldsymbol{\Phi}_{\boldsymbol{x}}$、 $\boldsymbol{\Phi}_{\boldsymbol{v}}$、 λ 和 q 作出不同选择，再结合初始条件 $(\boldsymbol{x}_0, \boldsymbol{v}_0)$，即可完整定义 PDMP。

Zig-Zag和弹跳粒子采样器都具有分段常值速度。在事件之间，动力学为

$$\frac{\mathrm{d}\boldsymbol{x}_t}{\mathrm{d}t} = \boldsymbol{v}_t, \qquad \frac{\mathrm{d}\boldsymbol{v}_t}{\mathrm{d}t} = 0. \tag{2}$$

在这种情况下，时刻 $t + s$ 的位置 $\boldsymbol{x}_{t+s}$ 很容易求得：

$$\boldsymbol{x}_{t+s} = \boldsymbol{x}_t + s \cdot \boldsymbol{v}_t. \tag{3}$$

为简洁起见，在不会引起歧义时，以下将省略时间下标。

这种简单的动力学使我们能够仅用骨架来表示 PDMP 的完整轨迹 $\{(t^k, \boldsymbol{x}^k, \boldsymbol{v}^k)\}_{k \geq 0}$, 其中包括事件时间 t^k 以及每次事件发生后紧接着的相应状态（$\boldsymbol{x}^k$, $\boldsymbol{v}^k$）。利用公式（3），即可通过线性插值求得时刻$\boldsymbol{x}_t$ 的位置 $t \in [t^k, t^{k+1}]$： $\boldsymbol{x}_t = \boldsymbol{x}^k + \boldsymbol{v}^k \cdot (t - t^k)$。

2.1 Zig-Zag采样器

Zig-Zag采样器的速度空间是离散的： $\mathcal{V} = \{-1, +1\}^d$。事件在每个坐标方向 i 上独立发生，具有状态依赖强度

$$\eta_i(\boldsymbol{x}, \boldsymbol{v}) = \max\left\{0, v_i \frac{\partial \Psi(\boldsymbol{x})}{\partial x_i}\right\}. \tag{4}$$

从而得到总事件率为

$$\lambda(\boldsymbol{x}, \boldsymbol{v}) = \sum_{i=1}^{d} \eta_i(\boldsymbol{x}, \boldsymbol{v}). \tag{5}$$

可以根据总事件率（5）模拟事件。随后，可以按各分量事件率的 p_i 比例选择要翻转的分量

$$p_i = \frac{\eta_i(\boldsymbol{x}, \boldsymbol{v})}{\lambda(\boldsymbol{x}, \boldsymbol{v})}. \tag{6}$$

或者，也可以直接根据各个分量的事件率模拟事件。下一个切换时间 τ 由所有分量的切换时间τ_i 中的最小值给出，

$$\tau = \min_{i=1,\ldots,d} \tau_i, \tag{7}$$

并且过程向前推进至 $\boldsymbol{x}' = \boldsymbol{x} + \tau \cdot \boldsymbol{v}$。

事件发生时，转移核 q 会翻转速度向量 $\boldsymbol{v}$中单个分量 v_i 的符号。事件发生后的速度 $\boldsymbol{v}'$ 为

$$\begin{aligned} v_i' &= -v_i, \\ v_j' &= v_j \quad \forall j \neq i, \end{aligned} \tag{8}$$

由此得到转移核 q:

$$q(\boldsymbol{v}'|\boldsymbol{x}', \boldsymbol{v}) = \delta\big(\boldsymbol{v} - 2\boldsymbol{e}_i \boldsymbol{e}_i^{\mathsf{T}} \boldsymbol{v}\big), \tag{9}$$

其中， $\boldsymbol{e}_i$ 是 $\mathbb{R}^d$中的第 i 个标准基向量。状态更新期间，位置保持不变。此构造得到的过程具有平稳分布$p(\boldsymbol{x}, \boldsymbol{v}) = p(\boldsymbol{x})p(\boldsymbol{v})$，其中 $p(\boldsymbol{v})$ 在 $\{-1, +1\}^d$ 上服从均匀分布，且 $p(\boldsymbol{x}) = \pi(\boldsymbol{x})$。在目标分布 [27]满足温和条件时，Zig-Zag采样器是遍历的。

Zig-Zag采样器的运动由势函数的梯度引导 $\nabla\Psi(\boldsymbol{x})$。当采样器朝概率密度较高的区域（即势函数较低的区域）移动时，事件率为零。然而，当采样器朝 $\pi(\boldsymbol{x})$的尾部（即势函数较高的区域）移动时，事件率会增加，从而更有可能遇到事件。发生此类事件时，相应的速度分量会翻转，采样器则被引导回概率密度较高的区域。图1展示了二维高斯目标分布下，Zig-Zag采样器连续两次事件的位置、速度以及由此产生的各分量事件率。

2.2 弹跳粒子采样器

弹跳粒子采样器采用连续速度空间 $\mathcal{V} = \mathbb{R}^d$。发生事件时，速度会相对于与目标势函数的等值线相切的超平面反射

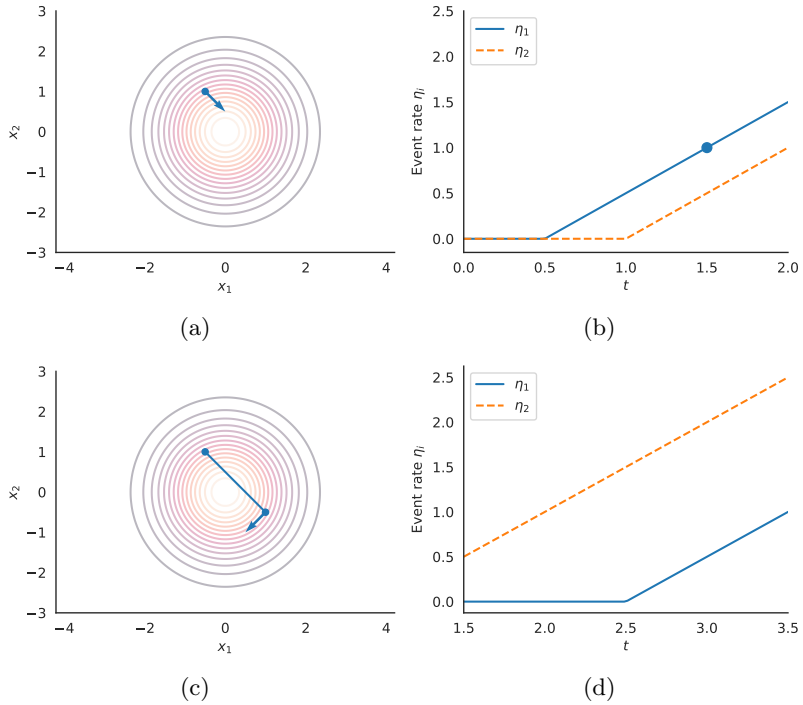


图 1：二维高斯分布上的 Zig-Zag 采样器示意图。图 (a) 和 (c) 显示了目标密度以及当前位置（点）和速度（箭头）。图 (b) 和 (d) 显示了沿各坐标方向的相应事件率。

：

$$q(\boldsymbol{v}'|\boldsymbol{x}', \boldsymbol{v}) = \delta\left(\boldsymbol{v} - 2\frac{\langle \boldsymbol{v}, \nabla\Psi(\boldsymbol{x}')\rangle}{\|\nabla\Psi(\boldsymbol{x}')\|^2}\nabla\Psi(\boldsymbol{x}')\right). \tag{10}$$

这种反射会翻转势函数梯度方向上的分量，从而使粒子返回到概率密度较高的区域。同样，位置 $\boldsymbol{x}$ 不受该事件影响。反射事件以状态依赖强度发生

$$\lambda(\boldsymbol{x}, \boldsymbol{v}) = \max\{0, \langle \boldsymbol{v}, \nabla\Psi(\boldsymbol{x})\rangle\}. \tag{11}$$

图 2 展示了二维高斯目标分布上的弹跳粒子采样器的弹跳机制和事件率。

由于该算法的原始版本 [28] 在某些情形下并非遍历的，[19] 引入了随机刷新，以确保该采样器能够遍历整个状态空间。这些随机刷新事件会以速率 $\lambda_{\text{ref}} > 0$服从指数分布；在刷新时，速度会从其平稳分布 $p(\boldsymbol{v}) = \mathcal{N}(\boldsymbol{0}, \boldsymbol{I})$中重新采样。如果 λ_{ref} 过低，采样器可能难以到达状态空间中的某些区域。

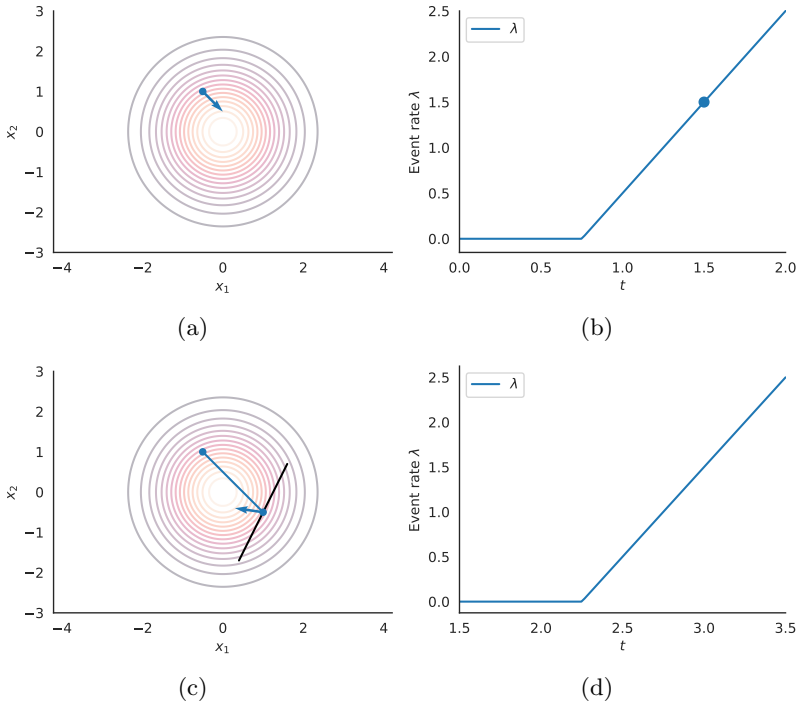


图 2：二维高斯分布上弹跳粒子采样器的示意图。图（a）和（c）展示了目标密度，以及当前的位置（点）和速度（箭头）。图（b）和（d）展示了由此产生的事件率。请注意，与 Zig-Zag采样器相比，事件率保持为零的时间更长，但斜率更陡。

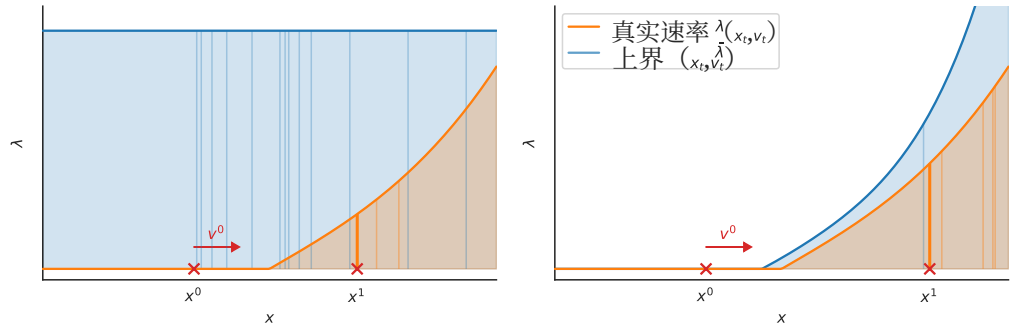
如果 λ_{ref} 过高，采样器可能会表现出扩散行为，从而降低其效率。 λ_{ref} 的最优选择取决于具体问题，通常需要通过经验进行调优。

由于具有连续的速度空间，弹跳粒子采样器理论上能够在各向同性目标分布的状态空间中找到更优的路径 [29]。然而，需要仔细调整刷新率 λ_{ref} 以确保遍历性。另一方面，Zig-Zag采样器不需要进行此类调整，并且能够开箱即用地适用于各种目标分布。

2.3 利用泊松稀疏化模拟事件

从速率函数为$\lambda(\boldsymbol{x}_t,\ \boldsymbol{v}_t)$的非齐次泊松过程中模拟事件时间可能颇具挑战。模拟这些事件时间的两种常见方法是逆变换法和稀疏化算法（两者均在 [30]中介绍）。逆变换法需要计算积分速率函数

$$\Lambda(t) = \int_0^t \lambda(\boldsymbol{x}_s, \boldsymbol{v}_s)\mathrm{d}s, \tag{12}$$



(a) 宽松上界　　(b) 紧致上界

图3：用于模拟事件时间的稀疏化过程示意图。竖线表示从上界为λ的泊松过程中抽取的候选事件时间。拒绝的事件以蓝色表示，接受的事件以橙色表示。拒绝次数越少越好。

然后通过求解来采样下一个事件时间 τ

$$\Lambda(\tau) = -\log u, \tag{13}$$

其中 $u \sim \mathcal{U}(0<style id="2">,1)$。该方法要求能够高效计算并求逆积分速率函数 $\Lambda(t)$，而对于由非线性偏微分方程控制的贝叶斯逆问题，这通常难以做到。

稀疏化算法提供了一种替代方法，无需对积分速率函数求逆。该方法通过速率函数的上界 $\bar{\lambda}$，从一个泊松过程中模拟候选切换时间 τ。候选切换时间 τ 以概率接受：

$$a(\tau) = \frac{\lambda(\boldsymbol{x}_{t+\tau}, \boldsymbol{v}_{t+\tau})}{\bar{\lambda}(\boldsymbol{x}_{t+\tau}, \boldsymbol{v}_{t+\tau})}. \tag{14}$$

如果候选事件被拒绝，我们就从时间 $t+\tau$ 重新开始生成事件，且不更新速度。

稀疏化过程如图 3 所示，分别采用较宽松（左）和较紧的上界 $\bar{\lambda}$。上界越紧（如图 3b所示），算法效率越高，因为被拒绝的候选事件更少。对于每个候选事件，都必须计算式（14），因此减少拒绝次数尤其有吸引力，特别是在计算真实事件速率 λ 的成本很高时，例如需要针对参数向量的一次实现求解偏微分方程。

2.4 仿射变换

Zig-Zag采样器和弹跳粒子采样器在对状态空间进行仿射重参数化后，均能保持其目标分布 [23]。更具体地说，令

$$\boldsymbol{\xi} = \boldsymbol{A}^{-1}(\boldsymbol{x} - \boldsymbol{b}), \tag{15}$$

对于某个可逆矩阵 $\boldsymbol{A} \in \mathbb{R}^{d\times d}$ 和向量 $\boldsymbol{b} \in \mathbb{R}^d$。在 $\boldsymbol{\xi}$ 坐标中，使用变换后的密度运行Zig-Zag采样器或弹跳粒子采样器

$$\tilde{\pi}(\boldsymbol{\xi}) \propto \pi(\boldsymbol{A}\,\boldsymbol{\xi} + \boldsymbol{b}) \tag{16}$$

然后通过以下方式将所得轨迹映射回去

$$\boldsymbol{x} = \boldsymbol{A}\,\boldsymbol{\xi} + \boldsymbol{b}, \tag{17}$$

得到原坐标中的一个过程，其不变分布为 $\pi(\boldsymbol{x})$。这种仿射等变性可以用于提高各向异性后验分布 [29] 上的采样效率。

在实践中，我们基于拉普拉斯近似构造局部预条件变换。令 $\boldsymbol{x}_{\mathrm{MAP}}$ 表示最大后验估计，并令 $\boldsymbol{H} = \nabla^2\Psi(\boldsymbol{x}_{\mathrm{MAP}})$ 表示该点处势函数的海森矩阵。我们计算乔列斯基分解 $\boldsymbol{H} = \boldsymbol{L}\,\boldsymbol{L}^{\mathsf{T}}$，并定义仿射变量变换

$$\boldsymbol{\xi} = \boldsymbol{L}^{\mathsf{T}}(\boldsymbol{x} - \boldsymbol{x}_{\mathrm{MAP}}), \tag{18}$$

或者等价地，$\boldsymbol{x} = \boldsymbol{x}_{\mathrm{MAP}} + \boldsymbol{L}^{-\mathsf{T}}\boldsymbol{\xi}$，该变换将后验分布的中心移至最大后验估计处，并根据 Ψ 的曲率在局部对其进行白化。变换后的势函数为

$$\tilde{\Psi}(\boldsymbol{\xi}) = \Psi\big(\boldsymbol{x}_{\mathrm{MAP}} + \boldsymbol{L}^{-\mathsf{T}}\boldsymbol{\xi}\big). \tag{19}$$

在上述一般映射 $\boldsymbol{x} = \boldsymbol{A}\,\boldsymbol{\xi} + \boldsymbol{b}$ 中，此选择对应于 $\boldsymbol{A} = \boldsymbol{L}^{-\mathsf{T}}$ 和$\boldsymbol{b} = \boldsymbol{x}_{\mathrm{MAP}}$。我们在变换后的$\tilde{}$空间中运行 Zig-Zag采样器或弹跳粒子采样器，其密度与 $\exp(-\Psi(\boldsymbol{\xi}))$ 成正比，得到一条轨迹 Ξ_t，然后通过 $\boldsymbol{x} = \boldsymbol{x}_{\mathrm{MAP}} + \boldsymbol{L}^{-\mathsf{T}}\boldsymbol{\xi}$映射回去。

3 使用代理模型进行泊松稀疏化

稀疏化需要真实的 PDMP事件率的上界。对于变换后的$\tilde{}$ $\tilde{}$势函数 $\Psi(\boldsymbol{\xi})$，该变换后的事件率可写为 $\lambda(\boldsymbol{\xi}, \boldsymbol{v})$。其主要思路是使用 $\bar{}$ $\tilde{}$代理模型 $\Psi(\boldsymbol{\xi})$来近似变换后的势函数 $\Psi(\boldsymbol{\xi})$。随后，我们可以$\bar{}$ $\tilde{}$近似势函数的梯度 $\nabla\Psi(\boldsymbol{\xi}) \approx \nabla\Psi(\boldsymbol{\xi})$以及过程速率 $\bar{}$ $\tilde{}\lambda(\boldsymbol{\xi}, \boldsymbol{v}) \approx \lambda(\boldsymbol{\xi}, \boldsymbol{v})$。在偏微分方程控制的反问题中，评估代理模型的成本通常比评估真实势函数的成本低几个数量级。因此，我们可以使用直接反演法，并对公式（13）进行数值求解，以

借助代理速率高效地产生事件时间候选值。随后，我们根据真实事件率 $\lambda(\boldsymbol{\xi}, \boldsymbol{v})$ 对这些候选值进行稀疏化，从而获得 PDMP 的有效事件时间。

理论上，上界过程必须满足全局约束

$$\bar{\lambda}(\boldsymbol{\xi}, \boldsymbol{v}) \geq \tilde{\lambda}(\boldsymbol{\xi}, \boldsymbol{v}), \tag{20}$$

对所有 $(\boldsymbol{\xi}, \boldsymbol{v})$ 均成立，从而保证稀疏化过程的正确性。此外，为了提高稀疏化效率，我们希望上界过程尽可能紧。

在实践中，构造既足够紧又能在全局范围内满足方程 (20) 中约束的代理模型颇具挑战。我们转而仅沿着实际生成的轨迹局部施加上界条件。这足以保证正确性，因为 PDMP 中的事件时间仅取决于沿当前路径积分得到的事件率。文献 [25], 采用了类似思路，针对给定的时间范围构造事件率的局部常数上界。我们将这一思路加以拓展，使用可适应势函数局部行为的全局代理模型。为确保稀疏化过程的正确性，我们引入一种校正机制：在模拟过程中检测到违反上界条件时，该机制会调整近似事件率 λ。

假设在某个时刻 t^k，位置为$\boldsymbol{\xi}^k$、速度为 $\boldsymbol{v}^k$ 时违反了上界条件，即 $\bar{\lambda}(\boldsymbol{\xi}^k, \boldsymbol{v}^k) < \tilde{\lambda}(\boldsymbol{\xi}^k, \boldsymbol{v}^k)$，我们可以通过在事件率上添加偏移量 γ，在局部校正代理模型。形式上，我们将校正后的事件率定义为

$$\bar{\lambda}^{\mathrm{c}}(\boldsymbol{\xi}, \boldsymbol{v}) = \max\{0, \langle \boldsymbol{v}, \nabla \bar{\Psi}(\boldsymbol{\xi}) \rangle + \gamma\}, \tag{21}$$

对于弹跳粒子采样器。我们在模拟开始时将偏移量初始化为 $\gamma = 0$。若在$(\boldsymbol{\xi}^k, \boldsymbol{v}^k)$处违反了式(20)，则将偏移量 γ 增大到足以在相应状态下满足上界条件：

$$\gamma = \max\Big\{0, \tilde{\lambda}(\boldsymbol{\xi}^k, \boldsymbol{v}^k) - \bar{\lambda}(\boldsymbol{\xi}^k, \boldsymbol{v}^k)\Big\}. \tag{22}$$

Zig-Zag采样器配备了向量偏移量 $\boldsymbol{\gamma} \in \mathbb{R}^d$，其各分量的更新方式类似于弹跳粒子采样器中的标量偏移量，并利用了Zig-Zag采样器事件率的可加结构：

$$\bar{\eta}_i^{\mathrm{c}}(\boldsymbol{\xi}, \boldsymbol{v}) = \max\left\{0, v_i \frac{\partial \bar{\Psi}(\boldsymbol{\xi})}{\partial \xi_i} + \gamma_i\right\}, \tag{23}$$

$$\bar{\lambda}^{\mathrm{c}}(\boldsymbol{\xi}, \boldsymbol{v}) = \sum_{i=1}^{d} \bar{\eta}_i^{\mathrm{c}}(\boldsymbol{\xi}, \boldsymbol{v}), \tag{24}$$

$$\gamma_i = \max\{0, \tilde{\eta}_i(\boldsymbol{\xi}^k, \boldsymbol{v}^k) - \bar{\eta}_i^{\mathrm{c}}(\boldsymbol{\xi}^k, \boldsymbol{v}^k)\}, \tag{25}$$

其中 $\bar{\eta}_i^{\mathrm{c}}$ 是由代理模型 Ψ诱导的逐分量提议率。

由于触发校正的事件无效，我们将过程回退到时间 t^{k-1}。然后，我们使用校正后的总事件率 $\bar{\lambda}^{\mathrm{c}}$ 对弹跳粒子采样器重新模拟下一个事件，或者使用其组成部分 $\bar{\eta}_i^{\mathrm{c}}$ 对Zig-Zag采样器重新模拟下一个事件。请注意

，我们还必须重新使用用于候选生成的同一个均匀随机变量的取值（见式（13））。否则，我们会产生偏向较早事件的偏差，因为当 PDMP 远离目标分布的众数时，这些校正发生得更频繁。对于均匀随机变量 u，其取值越大，这种长距离运动越有可能发生。

在其基本形式中，偏移量 γ 会随时间增大，从而使稀疏化过程的效率降低。为缓解这一问题，我们为偏移量实现了一种衰减机制。记 γ^k 为第（$k-1$）个发生于时刻 t^{k-1}的事件之后的偏移量，我们得到偏移量 γ^k ，用于下一区间 $[t^k, t^{k+1})$ ，即

$$\gamma^k = \gamma^{k-1} \cdot \exp\big(-\beta\big(t^k - t^{k-1}\big)\big), \tag{26}$$

这意味着偏移量随时间指数衰减，衰减速率为 $\beta > 0$。偏移量衰减式（26）的离散性质意味着，偏移量在事件之间保持为常数，并且在重新执行触发修正的事件时，我们不会违反式（20）。我们注意到，即便进行了这一修正，仍无法保证上界条件式（20）对轨迹上的所有（$\boldsymbol{\xi}$， $\boldsymbol{v}$）都成立，因为我们可能会遗漏事件之间速率的快速振荡。因此，我们的方法本质上是近似的。第4.3 节详细讨论了近似事件生成方案所引入的偏差。

使用校正后的代理速率进行稀疏化的完整算法详见弹跳粒子采样器的算法1 。Zig-Zag采样器的对应算法见算法 2。Zig-Zag采样器版本的主要区别在于候选事件时间的抽取方式以及校正机制的应用。我们对向量偏移量 $\boldsymbol{\gamma}$ 的每个分量采用与弹跳粒子采样器的式（26）中标量偏移量相同的衰减方式。算法1 和算法2 都会返回一条轨迹的骨架，即事件时间及其对应状态的序列 $\{(t^k,\ \boldsymbol{\xi}^k,\ \boldsymbol{v}^k)\}_{k=0}^K$。我们通常设定一个终止时间 T ，并运行算法直至 $t_K \geq T$，因为 PDMP 采样器是按模拟时间而非模拟事件收敛的 [27]。随后，只需在事件时间之间进行线性插值，即可轻松重建完整轨迹。

算法1 BPSWithSurrogate：使用校正后的代理速率进行标量稀疏化

要求： 近似提议速率 $\bar{\lambda}(\boldsymbol{\xi},\ \boldsymbol{v})$; 真实速率 $\tilde{\lambda}(\boldsymbol{\xi},\ \boldsymbol{v})$; 初始状态 $(\boldsymbol{\xi}_0,\ \boldsymbol{v}_0)$
其中 $\boldsymbol{v}_0 \sim \mathcal{N}(\mathbf{0},\ \boldsymbol{I})$; 刷新率 $\lambda_{\mathrm{ref}} > 0$; 衰减率 $\beta \geq 0$; 终止时间 T
确保： 轨迹 $\{(t^k,\ \boldsymbol{\xi}^k,\ \boldsymbol{v}^k)\}_{k\geq 0}$

$t \leftarrow 0$; $(\boldsymbol{\xi}, \boldsymbol{v}) \leftarrow (\boldsymbol{\xi}_0, \boldsymbol{v}_0)$; $k \leftarrow 0$; $\gamma \leftarrow 0$; $\Delta_t \leftarrow 0$
当 $t < T$ **时执行**
  $\gamma \leftarrow \gamma \exp(-\beta\, \Delta_t)$ ▷离散衰减
  $\bar{\lambda}^{\mathrm{c}} \leftarrow \max\{0, \langle \boldsymbol{v}, \nabla\, \bar{\Psi}(\boldsymbol{\xi})\rangle + \gamma\}$
  抽取 $u_{\mathrm{ref}} \sim \mathcal{U}(0,\ 1)$; 设置 $\tau_{\mathrm{ref}} \leftarrow -\log\, u_{\mathrm{ref}}/\lambda_{\mathrm{ref}}$ ▷ 刷新时钟
  抽取 $u_{\mathrm{b}} \sim \mathcal{U}(0,\ 1)$; 查找 $\tau_{\mathrm{b}} \geq 0$ 求解 $\int_0^{\tau_{\mathrm{b}}} \bar{\lambda}^{c}(\boldsymbol{\xi} + s\, \boldsymbol{v},\ \boldsymbol{v})\, \mathrm{d}s = -\log\, u_{\mathrm{b}}$ per 公式（13）; 设置 $\boldsymbol{\xi}_p \leftarrow \boldsymbol{\xi} + \tau_{\mathrm{b}}\, \boldsymbol{v}$; 设置事件 $\leftarrow$ 反弹
  当true **时执行** ▷重复使用 u_{b}的校正循环
    如果 $\tau_{\mathrm{b}} > \tau_{\mathrm{ref}}$ **则** ▷ 检查刷新抢占
      事件 $\leftarrow$刷新; **跳出**
    结束如果
    $\Delta_\lambda \leftarrow \tilde{\lambda}(\boldsymbol{\xi}_p, \boldsymbol{v})\ - \bar{\lambda}^{c}(\boldsymbol{\xi}_p,\ \boldsymbol{v})$
    如果 $\Delta_\lambda < 0$ **则**
      中断 ▷界限下的有效反弹候选
    else
      $\gamma \leftarrow \gamma + \Delta_\gamma$; $\bar{\lambda}^{c} \leftarrow$最大值$\{0, \langle \boldsymbol{v}, \nabla\, \bar{\Psi}(\boldsymbol{\xi})\rangle + \gamma\}$
      通过求解 $\int_0^{\tau_{\mathrm{b}}} \bar{\lambda}^{c}(\boldsymbol{\xi} + s\, \boldsymbol{v},\ \boldsymbol{v})\, \mathrm{d}s = -\log\, u_{\mathrm{b}}$, 重新计算 τ_{b}
      设置 $\boldsymbol{\xi}_p \leftarrow \boldsymbol{\xi} + \tau_{\mathrm{b}}\, \boldsymbol{v}$
    结束条件判断
  结束循环
  如果事件为刷新 **，则**
    $t \leftarrow t + \tau_{\mathrm{ref}}$; $\boldsymbol{\xi} \leftarrow \boldsymbol{\xi} + \tau_{\mathrm{ref}}\, \boldsymbol{v}$; $\Delta_t \leftarrow \tau_{\mathrm{ref}}$
    重采样 $\boldsymbol{v} \sim \mathcal{N}(\mathbf{0},\ \boldsymbol{I})$ ▷ 刷新
    继续
    $k \leftarrow k + 1$; $t^k \leftarrow t$; $\boldsymbol{\xi}^k \leftarrow \boldsymbol{\xi}$; $\boldsymbol{v}^k \leftarrow \boldsymbol{v}$
  else
    $t \leftarrow t + \tau_{\mathrm{b}}$; $\boldsymbol{\xi} \leftarrow \boldsymbol{\xi}_p$; $\Delta_t = \tau_{\mathrm{b}}$
    如果 $u_{\mathrm{acc}} \leq \tilde{\lambda}(\boldsymbol{\xi}_p,\ \boldsymbol{v})/\bar{\lambda}^{c}(\boldsymbol{\xi}_p,\ \boldsymbol{v})$ **则** ▷接受反弹
      从 $u_{\mathrm{acc}} \sim \mathcal{U}(0,\ 1)$ 中抽取
      更新 $\boldsymbol{v}$ ，按照公式（10）进行镜面反射
      $k \leftarrow k + 1$; $t^k \leftarrow t$; $\boldsymbol{\xi}^k \leftarrow \boldsymbol{\xi}$; $\boldsymbol{v}^k \leftarrow \boldsymbol{v}$;
    结束如果
  结束如果
结束循环
返回轨迹 $\{(t^k, \boldsymbol{\xi}^k, \boldsymbol{v}^k)\}_{k\geq 0}$

算法 2 ZZSWithSurrogate：逐分量稀释，校正后的代理速率

要求： 近似势能 $\bar{\Psi}(\boldsymbol{\xi})$；真实分量速率 $\{\tilde{\eta}_i(\boldsymbol{\xi}, \boldsymbol{v})\}_{i=1}^{d}$；初始状态 $(\boldsymbol{\xi}^0, \boldsymbol{v}^0)$，带有 $\boldsymbol{v}_0 \in \{-1, +1\}^d$；衰减率 $\beta \geq 0$；终止时间 T

确保： 轨迹 $\{(t^k, \boldsymbol{\xi}^k, \boldsymbol{v}^k)\}_{k \geq 0}$

$t \leftarrow 0$; $(\boldsymbol{\xi}, \boldsymbol{v}) \leftarrow (\boldsymbol{\xi}^0, \boldsymbol{v}^0)$; $k \leftarrow 0$; $\Delta_t \leftarrow 0$

初始化向量偏移 $\boldsymbol{\gamma} \leftarrow \mathbf{0} \in \mathbb{R}^d$

当 $t < T$ **时**

  $\boldsymbol{\gamma} \leftarrow \boldsymbol{\gamma} \exp(-\beta\, \Delta_t)$

    $\bar{\eta}_i^c \leftarrow \max\{0, v_i \partial \bar{\Psi}(\boldsymbol{\xi}) / \partial \xi_i + \gamma_i\}$ 对于 $i = 1, \ldots, d$

  对 $i = 1, \ldots, d$ **执行** ▷ 抽取相互独立的候选时刻

    抽取 $u_i \sim \mathcal{U}(0,\ 1)$

    求解 $\tau_i \geq 0$ 以得到 $\int_0^{\tau_i} \bar{\eta}_i^c(\boldsymbol{\xi} + s\, \boldsymbol{v},\ \boldsymbol{v})\, \mathrm{d}s = -\log u_i$，根据方程（13）

  结束循环

  当 条件为真**时执行** ▷选取最早的候选时刻；若违反上界则进行校正

    $i^* \leftarrow \arg\min_i \tau_i$; $\tau \leftarrow \tau_{i^*}$; $\boldsymbol{\xi}_p \leftarrow \boldsymbol{\xi} + \tau\, \boldsymbol{v}$

    $\Delta_\lambda \leftarrow \tilde{\eta}_{i^*}(\boldsymbol{\xi}_p,\ \boldsymbol{v}) - \bar{\eta}_{i^*}^c(\boldsymbol{\xi}_p,\ \boldsymbol{v})$

    如果 $\Delta_\lambda \leq 0$，**则**

      跳出 ▷ 满足界限的有效候选项

    else

      $\gamma_{i^*} \leftarrow \gamma_{i^*} + \Delta_\lambda$；更新 $\bar{\eta}_{i^*}^c \leftarrow \max\{0, v_{i^*} \partial\, \bar{\Psi}(\boldsymbol{\xi}) / \partial \xi_{i^*} + \gamma_{i^*}\}$

      求解 τ_{i^*} $\int_0^{\tau_{i^*}} \bar{\eta}_{i^*}^c(\boldsymbol{\xi} + s\, \boldsymbol{v},\ \boldsymbol{v})\, \mathrm{d}s = -\log u_{i^*}$ ▷重用 u_{i^*}

    结束如果

  结束循环

  $t \leftarrow t + \tau$; $\boldsymbol{\xi} \leftarrow \boldsymbol{\xi}_p$; $\Delta_t \leftarrow \tau$

  从 $u_{\mathrm{acc}} \sim \mathcal{U}(0,1)$ 中抽取

  如果 $u_{\mathrm{acc}} \leq \tilde{\eta}_{i^*}(\boldsymbol{\xi}_p, \boldsymbol{v}) / \bar{\eta}_{i^*}^c(\boldsymbol{\xi}_p, \boldsymbol{v})$ **则** ▷接受事件

    翻转分量 i^*: $v_i \leftarrow -v_i$

    $k \leftarrow k + 1$; $t^k \leftarrow t$; $\boldsymbol{\xi}^k \leftarrow \boldsymbol{\xi}$; $\boldsymbol{v}^k \leftarrow \boldsymbol{v}$

  结束如果

结束循环

返回 轨迹 $\{(t^k, \boldsymbol{\xi}^k, \boldsymbol{v}^k)\}_{k \geq 0}$

3.1 代理无关设计与代理族

校正机制的一项关键优势是，无论代理模型如何，都能保证结果正确。唯一的要求是代理模型能够导出可计算的提议率 $\lambda(\boldsymbol{\xi},\ \boldsymbol{v})$；当该提议率不是上界时，方程（$\gamma$ 21）中的局部偏移量会恢复结果的正确性。因此，即使使用极其粗糙的 $\Psi\boldsymbol{\xi}$代理模型，$\equiv c$ 这种方法也能奏效。例如，取一个常势函数 $\nabla\Psi \equiv \mathbf{0}$ $\gamma = 1$，则 λ初始提议率为零。在这种情况下，我们必须以一个很小的非零λ偏移量（例如），才能从生成候选事件；首次检测到违例后，偏移量便会增大，之后过程即可正确运行。虽然这种基线方法效率低下——偏移量增长很快，稀疏化接受率很低，而且需要频繁计算λ——但它可用作合理性检查和下界基准。

我们考虑以下代理族，按表达能力和成本递增的顺序排列：

- 常势函数（基线）：$\Psi(\boldsymbol{\xi}) \equiv c$。无需训练；正确性完全通过偏移量保证。
- 拉普拉斯（在变换坐标中是平凡的）：经过仿射变换后，拉普拉斯模型退化为标准正态势能，即 $\bar{\Psi}(\boldsymbol{\xi}) = \frac{1}{2}\boldsymbol{\xi}^\top\boldsymbol{\xi} + c$，其梯度为 $\boldsymbol{\xi}$，海森矩阵为 $\boldsymbol{I}$，并带有任意常数 c。这为众模附近提供了低成本且具有稳定作用的基线。

- 拉普拉斯 $+$ 高斯过程残差：对差异 $r(\boldsymbol{\xi}) = \Psi(\boldsymbol{\xi}) - \Psi_{\mathrm{L}}(\boldsymbol{\xi})$ 建模，并使用由 N_0 个 $r(\boldsymbol{\xi})$ 的真实实现预训练的高斯过程。该组合模型能够捕捉非二次结构，同时保持良好的局部行为。

- 带梯度观测的拉普拉斯 $+$ 高斯过程：在可获得的情况下，使用导数核，将 Ψ的梯度纳入高斯过程训练集。
- 自适应高斯过程（稀疏化驱动）：从 N_0 个初始训练点开始，并在里程碑规模 $N_0\,2^n$ 下扩展训练集，针对 $n = 1, \ldots, 5$，使用从稀疏化/模拟中收集的数据（例如校正位置或接受路径上的位置）。这一过程会逐步收紧局部速率提议。

尽管我们承认，关于更复杂的高斯过程主动学习方案已有大量研究，但我们采用自适应高斯过程模型的简单启发式方法，来展示自适应代理在此情境中的潜力。在这里，我们将模拟期间收集的所有 数据纳入训练集，以扩充训练数据。因此，我们将该模型视为主动学习性能的上限。

我们对所有代理模型采用简单且一致的流程：首先，按照算法 3 所述，应用第 2.4 节中的仿射变换对白化状态进行处理，使得拉普拉斯近似在变换坐标中服从标准正态分布。在这些坐标中，拉普拉斯势能是平凡的：$\bar{\Psi}_{\mathrm{L}} = \frac{1}{2}\boldsymbol{\xi}^\top\boldsymbol{\xi} + c$，其梯度为 $\nabla\bar{\Psi}_{\mathrm{L}}(\boldsymbol{\xi}) = \boldsymbol{\xi}$，海森矩阵为 $\nabla^2\bar{\Psi}_{\mathrm{L}}(\boldsymbol{\xi}) = \boldsymbol{I}$，参见算法 4。对于所有基于高斯过程的代理，我们通过抽取 N_0 个样本 $\{\boldsymbol{\xi}_i\}_{i=1}^{N_0} \sim \mathcal{N}(\mathbf{0}, \boldsymbol{I})$、计算 $\tilde{\Psi}(\boldsymbol{\xi}_i)$，并使用 残差 $r(\boldsymbol{\xi}_i) = \Psi(\boldsymbol{\xi}_i) - \Psi_{\mathrm{L}}(\boldsymbol{\xi}_i)$ 训练高斯过程来进行初始化，如算法 5 所示。对于观测梯度的变体，我们还会纳入导数观测 $(\boldsymbol{\xi}_i) = \nabla\Psi(\boldsymbol{\xi}_i) - \nabla\Psi_{\mathrm{L}}(\boldsymbol{\xi}_i)$。对于自适应方法，只要数据点总数不超过 1000，就会在每个里程碑更新高斯过程的超参数，因为计算成本会随数据点数量呈立方增长，而且在当前数据规模下，预计进一步改进的幅度较小。

算法 3 ComputeAffineTransformation：用于白化拉普拉斯近似的仿射变换 要求：目标势函数 $\Psi(\boldsymbol{x})$；初始猜测 $\boldsymbol{x}^{(0)}$；容差 ε确保：最大后验估计 $\boldsymbol{x}_{\mathrm{MAP}}$ 和 Cholesky 因子 $\boldsymbol{L}$，使得 $\boldsymbol{H} = \boldsymbol{L}\boldsymbol{L}^{\top} = \nabla^2\Psi(\boldsymbol{x}_{\mathrm{MAP}})$；变换坐标为 $\boldsymbol{\xi} = \boldsymbol{L}^{\top}(\boldsymbol{x} - \boldsymbol{x}_{\mathrm{MAP}})$，如第 2.4 节所示 1：使用基于梯度的优化器（例如 SciPy BFGS），从 $\boldsymbol{x}^{(0)}$ 开始寻找 $\boldsymbol{x}_{\mathrm{MAP}} \leftarrow \arg\min_{\boldsymbol{x}} \Psi(\boldsymbol{x})$；当 $\|\nabla\Psi(\boldsymbol{x})\| < \varepsilon$ 时停止 2：在众数处计算解析 Hessian：$\boldsymbol{H} \leftarrow \nabla^2\Psi(\boldsymbol{x}_{\mathrm{MAP}})$ 3：进行乔列斯基分解：$\boldsymbol{H} = \boldsymbol{L}\boldsymbol{L}^{\top}$4：返回 $\boldsymbol{x}_{\mathrm{MAP}}, \boldsymbol{L}$

算法 4 BuildLaplaceSurrogate：变换坐标中的二次代理。要求：目标势函数 $\Psi(\boldsymbol{x})$；算法 3 的设置 确保：$\boldsymbol{\xi}$ 空间中的拉普拉斯代理 Ψ_{L} ({v5)；梯度 $\nabla\Psi_{\mathrm{L}}(\boldsymbol{\xi})$；参数 $\boldsymbol{x}_{\mathrm{MAP}}, \boldsymbol{L}$

1 $\boldsymbol{x}_{\mathrm{MAP}}$ $\boldsymbol{L} \leftarrow$ ComputeAffineTransformation Ψ ▷ 3 :

2 $\tilde{\Psi}$

, ( ) 算法：通过 ($\boldsymbol{\xi}$) 隐式定义变换后的势函数：$= \Psi\big(\boldsymbol{x}_{\mathrm{MAP}} + \boldsymbol{L}^{-\top}\boldsymbol{\xi}\big)$3：定义 $\bar{\Psi}_{\mathrm{L}}(\boldsymbol{\xi}) \leftarrow \frac{1}{2}\boldsymbol{\xi}^{\top}\boldsymbol{\xi} + \tilde{\Psi}(0)$ 4：定义 $\nabla\Psi_{\mathrm{L}}(\boldsymbol{\xi}) \leftarrow \boldsymbol{\xi}$ 5：返回 $\Psi_{\mathrm{L}}, \nabla\Psi_{\mathrm{L}}, \boldsymbol{x}_{\mathrm{MAP}}, \boldsymbol{L}$

算法 5 TrainGPSurrogate：变换坐标中拉普拉斯之上的高斯过程残差 要求：目标势函数 $\Psi(\boldsymbol{x})$；初始设计规模 N_0；核函数 k_{ϑ}，初始值 $\boldsymbol{\vartheta}^{(0)}$；includeGrad $\in$ {true,false} 确保：代理均值 $\Psi_{\mathrm{GP}}(\boldsymbol{\xi}) = \Psi_{\mathrm{L}}(\boldsymbol{\xi}) + m(\boldsymbol{\xi})$ 及其梯度；参数 $\boldsymbol{x}_{\mathrm{MAP}}, \boldsymbol{L}$

1 Ψ_{L} $\nabla\Psi_{\mathrm{L}}$ $\boldsymbol{x}_{\mathrm{MAP}}$ $\boldsymbol{L} \leftarrow$ BuildLaplaceSurrogate Ψ ▷ 4 : ,

2 $\mathbf{0}$ $\boldsymbol{I}$ ▷

, , ( ) 算法：在变换空间中抽取 $\{\boldsymbol{\xi}_i\}_{i=1}^{N_0} \sim \mathcal{N}(\,,\,)$ 个训练输入 3: for $i = 1, \ldots, N_0$ do 4: $r_i \leftarrow r(\boldsymbol{\xi}_i) := \Psi(\boldsymbol{\xi}_i) - \Psi_{\mathrm{L}}(\boldsymbol{\xi}_i)$ ▷ 变换坐标中的残差 5: if includeGrad then 6: $\boldsymbol{g}_{\xi,i} \leftarrow \nabla_{\boldsymbol{\xi}}\Psi(\boldsymbol{\xi}_i)$ ▷ 通过链式法则得到变换坐标中的梯度 7: $\boldsymbol{g}_i \leftarrow \boldsymbol{g}_{\xi,i} - \nabla\Psi_{\mathrm{L}}(\boldsymbol{\xi}_i)$ ▷ 残差梯度 8: end if 9: end for 10: 组装数据集 $\mathcal{D} = \{(\boldsymbol{\xi}_i, r_i)\}$（若可用，还包括梯度残差） 11: 通过最大化 $\mathcal{L}(\boldsymbol{\vartheta})$ (33) 拟合高斯过程先验 $f \sim \mathcal{GP}(0, k_{\vartheta})$；ˆ 12: 通过 (30) 定义 $m(\boldsymbol{\xi}) \leftarrow \mathbb{E}[f(\boldsymbol{\xi}) \mid \mathcal{D}, \boldsymbol{\vartheta}]$ 以及 $\nabla m(\boldsymbol{\xi})$ 13: 定义复合代理 $\Psi_{\mathrm{GP}}(\boldsymbol{\xi}) = \Psi_{\mathrm{L}}(\boldsymbol{\xi}) + m(\boldsymbol{\xi})$，其中 $\nabla\Psi_{\mathrm{GP}}(\boldsymbol{\xi}) = \nabla\Psi_{\mathrm{L}}(\boldsymbol{\xi}) + \nabla m(\boldsymbol{\xi})$ 14: return $\Psi_{\mathrm{GP}}, \nabla\Psi_{\mathrm{GP}}, \boldsymbol{x}_{\mathrm{MAP}}, \boldsymbol{L}$

3.2 高斯过程代理

我们用高斯过程对变换坐标上残差 $r(\boldsymbol{\xi}) = \tilde{\Psi}(\boldsymbol{\xi}) - \bar{\Psi}_{\mathrm{L}}(\boldsymbol{\xi})$中的光滑标量函数进行建模。令 $\{(\boldsymbol{\xi}_i, y_i)\}_{i=1}^{N}$ 表示训练输入和观测值，其中 $y_i = f(\boldsymbol{\xi}_i) + \epsilon_i$ 带有独立噪声 $\varepsilon_i \sim \mathcal{N}(0, \sigma_{\mathrm{n}}^2)$。我们设定常数均值[1] 高斯过程先验 $f \sim \mathcal{GP}(m, k_{\vartheta})$。令 $\boldsymbol{\xi}, \boldsymbol{\xi}' \in \mathbb{R}^d$ 表示两个输入位置。我们使用各向异性平方指数核

$$k_{\vartheta}(\boldsymbol{\xi}, \boldsymbol{\xi}') = \sigma_{\mathrm{f}}^2 \exp\left(-\frac{1}{2} \sum_{j=1}^{d} \frac{(\xi_j - \xi_j')^2}{\ell_j^2} \right), \qquad \boldsymbol{\vartheta} = \{ m, \sigma_{\mathrm{f}}^2, \ell_1, \dots, \ell_d, \sigma_{\mathrm{n}}^2 \}. \tag{27}$$

对角的、与维数相关的长度尺度 $\boldsymbol{\ell} = \{\ell_1, \dots, \ell_d\}$ 体现了各向异性，这是必要的，因为第2.4 节中的仿射变换仅在后验众数附近对白化目标。结合信号方差 σ_{f}^2、观测噪声 σ_{n}^2和均值参数 m，由此得到 $d+3$ 个超参数，其中 d 为 $\boldsymbol{\xi}$的维数。

我们收集核矩阵 $\boldsymbol{K} \in \mathbb{R}^{N \times N}$，其元素为 $[\boldsymbol{K}]_{ij} = k_{\vartheta}(\boldsymbol{\xi}_i, \boldsymbol{\xi}_j)$，并令 $\boldsymbol{K}_y = \boldsymbol{K} + \sigma_{\mathrm{n}}^2 \boldsymbol{I}$。对于测试位置 $\boldsymbol{\xi}_*$，我们定义交叉协方差向量$\boldsymbol{K}_* = [k_{\vartheta}(\boldsymbol{\xi}_*, \boldsymbol{\xi}_1)\text{.....}, k_{\vartheta}(\boldsymbol{\xi}_*, \boldsymbol{\xi}_N)]^{\mathsf{T}}$ 以及观测值 $\boldsymbol{y} = [y_1, \dots, y_N]^{\top}$。高斯过程后验均值和方差如下：

$$m_*(\boldsymbol{\xi}_*) = m + \boldsymbol{K}_*^{\mathsf{T}} \boldsymbol{K}_y^{-1} (\boldsymbol{y} - m\mathbf{1}), \tag{28}$$

$$s_*^2(\boldsymbol{\xi}_*) = k_{\vartheta}(\boldsymbol{\xi}_*, \boldsymbol{\xi}_*) - \boldsymbol{K}_*^{\mathsf{T}} \boldsymbol{K}_y^{-1} \boldsymbol{K}_*, \tag{29}$$

我们评估 $m_*(\boldsymbol{\xi})$ 作为代理均值，并在需要时评估其梯度，以近似目标分布的势函数。

关于输入的梯度可以用闭式形式获得。记 $\boldsymbol{\alpha} = \boldsymbol{K}_y^{-1}(\boldsymbol{y} - m\mathbf{1})$，预测均值的梯度为

$$\nabla_{\boldsymbol{\xi}} m_*(\boldsymbol{\xi}) = \sum_{i=1}^{N} \alpha_i \nabla_{\boldsymbol{\xi}} k_{\vartheta}(\boldsymbol{\xi}, \boldsymbol{\xi}_i) \tag{30}$$

$$\nabla_{\boldsymbol{\xi}} k_{\vartheta}(\boldsymbol{\xi}, \boldsymbol{\xi}_i) = k_{\vartheta}(\boldsymbol{\xi}, \boldsymbol{\xi}_i) \mathrm{diag}(\boldsymbol{\ell})^{-1} (\boldsymbol{\xi}_i - \boldsymbol{\xi}), \tag{31}$$

注意，常数均值参数 m 不依赖于 $\boldsymbol{\xi}$，因此不会对 $\nabla_{\boldsymbol{\xi}} m_*(\boldsymbol{\xi})$ 产生贡献。如果纳入导数观测，我们将通过对核函数求导得到的联合协方差加入高斯过程：$\mathrm{cov}(\partial_p f(\boldsymbol{\xi}), f(\boldsymbol{\xi}')) = \partial_{\boldsymbol{\xi}_p} k_{\vartheta}(\boldsymbol{\xi}, \boldsymbol{\xi}')$，以及

$$\mathrm{cov}(\partial_p f(\boldsymbol{\xi}), \partial_q f(\boldsymbol{\xi}')) = \partial_{\boldsymbol{\xi}_p} \partial_{\boldsymbol{\xi}_q'} k_{\vartheta}(\boldsymbol{\xi}, \boldsymbol{\xi}') = k_{\vartheta}(\boldsymbol{\xi}, \boldsymbol{\xi}') \left(\frac{\delta_{pq}}{\ell_p^2} - \frac{(\boldsymbol{\xi}_p - \boldsymbol{\xi}_p')(\boldsymbol{\xi}_q - \boldsymbol{\xi}_q')}{\ell_p^2 \, \ell_q^2} \right), \tag{32}$$

其中 δ_{pq} 为克罗内克δ，∂_p 为关于第 p 个输入维度的导数。

[1]虽然常数均值不会影响高斯过程后验均值的梯度，但可以改善超参数估计和数值稳定性 [31]。

超参数估计通过最大化边际似然来进行。 $\boldsymbol{\vartheta}$ 的对数边际似然为

$$\mathcal{L}(\boldsymbol{\vartheta}) = \log p(\boldsymbol{y}|\boldsymbol{\xi}_1, \ldots, \boldsymbol{\xi}_N, \boldsymbol{\vartheta}) = -\frac{1}{2}(\boldsymbol{y} - m\mathbf{1})^\top \boldsymbol{K}_y^{-1}(\boldsymbol{y} - m\mathbf{1}) - \frac{1}{2}\log|\boldsymbol{K}_y| - \frac{N}{2}\log(2\pi), \tag{33}$$

在实际计算中，我们通过乔列斯基分解 $\boldsymbol{K}_y = \boldsymbol{L}\boldsymbol{L}^\mathsf{T}$ 计算 $\mathcal{L}$ 及其梯度，从而得到 $\boldsymbol{\alpha} = \boldsymbol{L}^{-\mathsf{T}}\boldsymbol{L}^{-1}(\boldsymbol{y} - m\mathbf{1})$ 和 $\log|\boldsymbol{K}_y| = 2\sum_i \log(\boldsymbol{L}_{ii})$。关于核函数和噪声超参数的梯度具有标准形式

$$\frac{\partial \mathcal{L}}{\partial \vartheta_j} = \frac{1}{2}\,\mathrm{tr}\left(\left(\boldsymbol{\alpha}\boldsymbol{\alpha}^\mathsf{T} - \boldsymbol{K}_y^{-1}\right)\frac{\partial \boldsymbol{K}_y}{\partial \vartheta_j}\right), \tag{34}$$

with

$$\frac{\partial \boldsymbol{K}_y}{\partial \sigma_\mathrm{n}^2} = \boldsymbol{I}, \tag{35}$$

以及核函数特定的导数 $\partial[\boldsymbol{K}_y]_{pq}/\partial\vartheta_j = \partial k_\vartheta(\boldsymbol{\xi}_p, \boldsymbol{\xi}_q)/\partial\vartheta_j$，其中

$$\frac{\partial k_\vartheta(\boldsymbol{\xi}, \boldsymbol{\xi}')}{\partial \sigma_\mathrm{f}^2} = \frac{k_\vartheta(\boldsymbol{\xi}, \boldsymbol{\xi}')}{\sigma_\mathrm{f}^2}, \tag{36}$$

$$\frac{\partial k_\vartheta(\boldsymbol{\xi}, \boldsymbol{\xi}')}{\partial \ell_j} = k_\vartheta(\boldsymbol{\xi}, \boldsymbol{\xi}')\,\frac{\left(\xi_j - \xi_j'\right)^2}{\ell_j^3}. \tag{37}$$

关于均值参数的导数可由二次项 $\partial\mathcal{L}/\partial m = \mathbf{1}^\mathsf{T}\,\boldsymbol{K}_y^{-1}(\boldsymbol{y} - m\mathbf{1})$ 得到。

4 数值实验

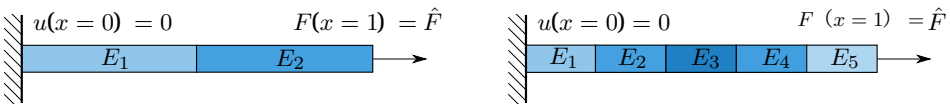
我们首先在第 4.1 节介绍实验设置，包括偏微分方程问题、性能指标和实现细节。首先聚焦于 Zig-Zag 采样器，我们将在第 4.3 节分析所提出的基于代理模型的稀疏化方法在何种情况下能够实现正确收敛。第 4.4 节随后比较不同的代理模型，并在第 4.5 节分析自适应高斯过程版本。最后，第 4.6节将 Zig-Zag 和弹跳粒子采样器与无转向采样器进行基准比较 [7]。

4.1 测试案例

为展示所提出的 PDMP 采样方法，我们引入一个可扩展的案例研究。该问题的计算成本较低，因此可以直接与使用随机游走 Metropolis 算法（RWM）进行的长时间参考运行结果进行比较。同时，该案例研究的灵感来自其与实际工程问题的相似性。其非线性参数到观测量映射会产生非高斯后验分布，非常适合用于测试 PDMP 采样器和代理模型辅助框架的能力。

具体而言，我们考虑一根杨氏模量随空间变化且具有不确定性的一维线弹性杆的力学平衡问题，该模量将通过

位移测量进行推断。长度为 L 的杆 $= 1$具有恒定的横截面积 A $= 1$ 如图4所示。



(a) d $= 2$ (b) d $= 5$

图4：杆问题，其中(a)和(b)分别采用2维和5维离散化表示杨氏模量场。

ˆ 杆的左端固定，右端承受集中载荷 $F = 1$ 。假设材料为线弹性材料，杆上位移 $u(x)$ 由以下边值问题控制：

$$\frac{\mathrm{d}}{\mathrm{d}x}\left(E(x)\frac{\mathrm{d}u(x)}{\mathrm{d}x}\right) = 0, \quad x \in (0,1), \tag{38}$$

$$u(0) = 0, \tag{39}$$

$$E(1)\frac{\mathrm{d}u(1)}{\mathrm{d}x} = 1, \tag{40}$$

其中，杨氏模量 $E(x)$ 取决于位置 x。我们将该参数场 $E(x)$ 视为不确定的。然而，在许多情况下，我们可以获得关于该参数场的一些先验知识。

在此，我们采用一个平滑且具有空间相关性的先验场。由于杨氏模量具有正值约束，我们将其建模为对数高斯随机场。这样的场可以表示为高斯过程：

$$\log E(x) \sim \mathcal{GP}(\mu_E(x), k_E(x,x')), \tag{41}$$

$$k_E(x,x') = \sigma_E^2 \exp\left(-\frac{(x-x')^2}{2l_E^2}\right). \tag{42}$$

其三个参数分别为均值函数 $\mu_E(x)$、信号方差 σ_E^2，以及平方指数协方差函数 $k_E(x, x')$ 的长度尺度 l_E 。在实践中，这些参数可以根据数据进行估计，例如使用最大似然估计；也可以根据专家知识设定，或在层次贝叶斯模型中将其视为超参数。为进行测试，我们设定 $\mu_E = 1$、 $\sigma_E = 1$ 和 $l_E = 0.3$，从而得到一个变化适中的场。

从理论上讲，我们希望推断连续场 $E(x)$。然而在实践中，我们需要将该场离散化，以便用有限数量的参数 $\boldsymbol{\theta}$ 来表示它。我们

使用分段常数基函数 ϕ_i 对该场进行离散化

$$\log E(x;\boldsymbol{\theta}) = \sum_{i=1}^{d} \theta_i \phi_i x, \tag{43}$$

$$\phi_i(x) = \begin{cases} 1, & \text{if } \dfrac{i-1}{d} \le x < \dfrac{i}{d}, \\ 0, & \text{elsewhere.} \end{cases}. \tag{44}$$

从而得到维度为 d 的参数向量 $\boldsymbol{\theta} \in \mathbb{R}^d$。通过将高斯过程投影到分段常数基函数上，我们得到参数 $p(\boldsymbol{\theta}) = \mathcal{N}(\mu_{\boldsymbol{\theta}}, \mathrm{Cov}(\boldsymbol{\theta}))$ 的先验分布。对于 $M_{ij} = \langle \phi_i, \phi_j \rangle$，一般情况下参数 $\boldsymbol{\theta}$ 的协方差矩阵为

$$\mathrm{Cov}(\theta_i, \theta_j) = \boldsymbol{M}^{-1} \int_0^1 \int_0^1 \phi_i(x) k_E(x, x') \phi_j(x') \,\mathrm{d}x \,\mathrm{d}x' \boldsymbol{M}^{-\mathsf{T}}. \tag{45}$$

$$\tag{46}$$

由于基函数 ϕ_i 彼此不重叠，且基在 $M_{ij} = 1/d\, \boldsymbol{I}$ 下正交，协方差可简化为

$$\mathrm{Cov}(\theta_i, \theta_j) = d^2 \int_{(i-1)/d}^{i/d} \int_{(j-1)/d}^{j/d} k_E(x, x') \,\mathrm{d}x \,\mathrm{d}x'. \tag{47}$$

对于底层过程的常数均值函数 $\mu_E(x) = 1$，系数 θ_i 的均值为

$$\mathbb{E}[\theta_i] = d \int_{(i-1)/d}^{i/d} E(x) \,\mathrm{d}x = 1. \tag{48}$$

结合特征向量 $\boldsymbol{\phi}(x) = [\phi_1(x) \ldots, \phi_d(x)]^{\mathsf{T}}$，参数场离散化对数的协方差为

$$\mathrm{Cov}(\log E(x;\boldsymbol{\theta}), \log E(x';\boldsymbol{\theta})) = \boldsymbol{\phi}(x)^{\mathsf{T}} \,\mathrm{Cov}(\boldsymbol{\theta})\, \boldsymbol{\phi}(x'), \tag{49}$$

当 $d \to \infty$ 时，它收敛到高斯过程的原始协方差函数 $k_E(x, x')$。

我们承认，分段常数基函数并不适合表示平滑的高斯过程。对于大多数应用而言，切比雪夫多项式或 Karhunen-Lo`eve 基函数等其他基函数会更加合适。然而，分段常数基函数对我们的研究有一项关键优势：

它们便于解析求解正问题，其表达式如下：

$$u(x;\boldsymbol{\theta}) = \left(\sum_{i=1}^{\lfloor xd \rfloor} \frac{\frac{1}{d}}{\exp\theta_i} + \frac{x - \frac{\lfloor xd \rfloor}{d}}{\exp\theta_{\lfloor xd \rfloor + 1}} \right), \quad x \in [0,1]. \tag{50}$$

其中，$\lfloor \cdot \rfloor$ 表示向下取整函数。该解连续且呈分段线性，在各线段边界处存在拐点。最值得注意的是，位移是潜变量 $\boldsymbol{\theta}$ 的非线性函数，因此构成一个非线性逆问题。由于存在解析解，我们可以几乎不耗费计算成本地评估正向模型并进行推断，从而能够专注于分析采样算法本身。

首先，我们从先验 $\boldsymbol{\theta}^* \sim p(\boldsymbol{\theta})$ 中抽取一个样本作为真实值，随后尝试根据含噪位移测量值 $\tilde{\boldsymbol{u}}$ 对其进行推断。我们在 m 个等距传感器位置$x_i = i/(m+1)$ 处评估正向模型，并添加标准差为 $\sigma_{\text{obs}} = 0.025$ 的独立高斯噪声，从而生成这些位移测量值 $\tilde{u}_i$。因此，我们采用高斯似然模型

$$p(\tilde{\boldsymbol{u}}|\boldsymbol{\theta}) = \mathcal{N}(\tilde{\boldsymbol{u}}|\boldsymbol{u}(\boldsymbol{\theta}), \sigma_{\text{obs}}^2 \boldsymbol{I}), \tag{51}$$

其中，$\tilde{\boldsymbol{u}} = [\tilde{u}_1, \ldots, \tilde{u}_m]^\mathsf{T}$ 为含噪测量值，而$\boldsymbol{u}(\boldsymbol{\theta}) = [u(x_1;\ \boldsymbol{\theta}).....,\ u(x_m;\ \boldsymbol{\theta})]^\mathsf{T}$ 为传感器位置处的模型预测值。遵循 [32]，，我们将传感器数量设为 $m = \lfloor 3/4 d \rfloor$。以这种方式固定观测数与参数之比，有助于在不同参数维数下保持问题难度一致 d。

最后，我们将贝叶斯推断问题总结如下

$$p(\boldsymbol{\theta}|\tilde{\boldsymbol{u}}) \propto p(\tilde{\boldsymbol{u}}|\boldsymbol{\theta})\, p(\boldsymbol{\theta}), \tag{52}$$

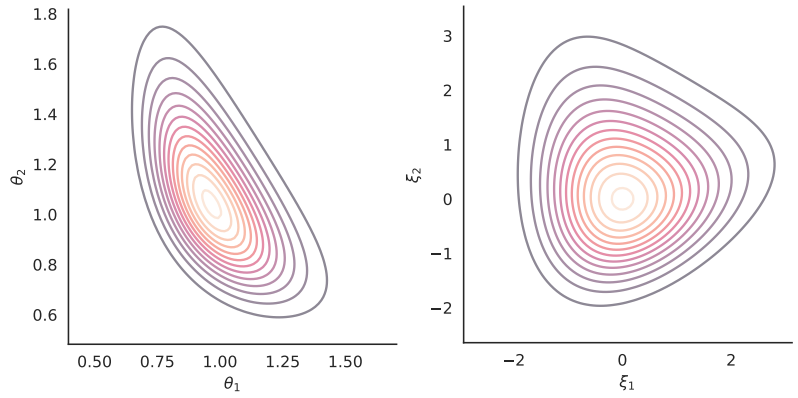
其中，先验 $p(\boldsymbol{\theta})$ 是根据上述推导得到的高斯分布。现在，我们可以使用第 2 节介绍的 PDMP方法估计后验分布，即 $\pi(\boldsymbol{\theta}) = p(\boldsymbol{\theta}|\tilde{\boldsymbol{u}})$。二维情形的后验分布如图 5 所示。

4.2 评估指标

首先，对于每种设置 $d = 2, \ldots, 10$，我们使用包含 10^7 个样本的长时间随机游走 Metropolis 运行计算参考解。令 $\mu_i^\star$ 和 $\sigma^2{}_{,i}{}^\star$ 分别表示该参考后验分布各坐标的后验均值和方差。对于任意方法 $\mathcal{M}$，令 $\hat{\mu}_i^{(\mathcal{M})}$ 和 $\hat{\sigma}^2{}_{,i}{}^{(\mathcal{M})}$ 表示相应的估计值。然后，我们评估以下度量指标：

后验均值误差：

$$\text{RMSE}_{\text{mean}}^{(\mathcal{M})} = \sqrt{\frac{1}{d} \sum_{i=1}^{d} \left(\hat{\mu}_i^{(\mathcal{M})} - \mu_i^\star \right)^2} \tag{53}$$



(a) (b)

图 5：二维问题在原始空间 (a) 和仿射空间 (b) 中的后验分布。

后验方差误差：

$$\mathrm{RMSE}_{\mathrm{var}}^{(\mathcal{M})} = \sqrt{\frac{1}{d}\sum_{i=1}^{d}\left(\hat{\sigma}_i^{2,(\mathcal{M})} - \sigma_i^{2,\star}\right)^2} \tag{54}$$

瓦瑟斯坦距离： 我们计算参考解 $\hat{p}^*$ 与方法 $\mathcal{M}$ 得到的近似 $\hat{p}^{(\mathcal{M})}$ 之间的二阶瓦瑟斯坦距离（WD）。

$$W_2^2(\hat{p}^*, \hat{p}^{(\mathcal{M})}) = \min_{\pi \in \Pi(\hat{p}^*, \hat{p}^{(\mathcal{M})})} \iint \|\boldsymbol{\xi} - \boldsymbol{\xi}'\|^2 \mathrm{d}\pi(\boldsymbol{\xi}, \boldsymbol{\xi}') \tag{55}$$

其中， $\Pi(\hat{p}^*, \hat{p}^{(\mathcal{M})})$ 表示这两个分布的所有耦合集合。

每次模型评估的有效样本量（ESS）： 我们计算每次模型评估的有效样本量，以评估 PDMP 方法的采样效率。ESS 的计算公式为

$$\mathrm{ESS} = \frac{N}{1 + 2\sum_{k=1}^{\infty}\rho_k}, \tag{56}$$

其中， N 是样本总数， ρ_k 是滞后 k 处的自相关。为考虑计算成本，我们将 ESS 除以模型评估次数 N_{eval} 进行归一化。

备注。我们在仿射空间中计算所有度量指标，以获得标准化结果。如 [27]所示，分段确定性马尔可夫过程的后验均值和方差可以通过解析方式计算。我们将这些量显示为偏微分方程模型评估次数的函数 N_{eval}，

因为我们假设这些评估是本问题中的计算瓶颈。这样，我们便能在同等条件下比较不同方法的收敛情况。因此，我们需要跟踪分段确定性马尔可夫过程采样器执行的模型评估次数。马尔可夫链通常以抽取的样本数量来评估；对于随机游走 Metropolis 采样器而言，该数量与模型评估次数成正比。

在 [27], 中，还提供了一种无需离散化、直接从PDMP骨架计算有效样本量的方法。然而，我们发现，在本文考察的较短运行中，根据离散化过程计算的有效样本量结果更加稳定、准确。因此，我们将PDMP近似至时间 T，并将其离散化为 N个在时间上等间隔的样本，即 $\boldsymbol{\xi}_i = \boldsymbol{\xi}(i \cdot T/N)$，其中 $i = 1, \ldots\ldots, N$。

同样，由于该指标没有解析解，我们对分段确定性马尔可夫过程进行离散化，以计算经验WD。对于较大的样本量和维度，精确计算经验WD的代价高得难以承受。因此，我们采用 [33], 中所述的去偏辛钦散度，作为一种高效的近似。为保证数值稳定性，同时将偏差控制在较低水平，我们将正则化参数设为 $\varepsilon = 0.02$ 。

与后验均值和方差类似，我们也报告了WD随模型评估次数的变化情况。经验WD是真实WD的有偏且带噪声的估计量，用于衡量PDMP与参考后验之间的差异；样本量较小时尤为如此。在实验中，我们不对偏差进行校正，但通过为所有方法使用相同数量的样本来确保可比性。这意味着，在进行 N 次模型评估后计算WD时，我们将分段确定性马尔可夫过程离散化为 N个点。我们通过从参考后验中多次随机抽取 N 个样本来降低方差。

我们对每个数值实验使用50个不同的随机种子重复50次，然后报告四项度量指标各自的平均值。对于每个随机种子，采样算法的初始位置从后验分布的拉普拉斯近似中抽取。需要注意的是，我们同样在仿射空间中运行马尔可夫链蒙特卡罗算法，以使所有算法处于同等条件。这样，无论采用哪种算法，我们都预期预烧阶段较短。我们仅针对每种设定获取一次仿射变换，并将其用于所有方法和随机种子。

我们为所有实验提供了一个 RWM基线，这不要与采用 10^7 样本的参考解混淆。众所周知，RWM 的性能在很大程度上取决于提议分布的选择。我们首先采用标准高斯提议，以体现后验分布在仿射空间中的各向同性。在每 100 次迭代后，如果接受率低于 0.2，就将当前提议协方差乘以 0.9；如果接受率高于 0.25，则乘以 1.1。迭代 1000 次后，我们将提议协方差矩阵设为迄今收集到的样本的经验协方差，并乘以在 [12]中推导出的最优因子 $2.38^2/d$ 。随后，我们像之前一样每 100 次迭代重新缩放提议协方差，并在 2000 次迭代后保持不变，以满足渐消适应 [34]。我们也使用这一版本的 RWM算法来计算参考解，不过采用了不同的随机种子，并进行了长得多的运行，同时舍弃最初的 5000 个样本作为预热期。

4.3 Zig-Zag采样器的收敛

首先，我们想阐明 Zig-Zag采样器何时会收敛到真实的后验分布。随着我们不再局限于那些已得到充分研究、且已知能够收敛的过程模拟方法，这一点尤为重要。可能存在若干违反收敛保证假设的情形：

1. 代理模型无法为稀疏化生成严格的上界过程。2. 我们不断调整偏移量 γ 以恢复上界性质（尽管在后期调整幅度较小）。3. 我们不断减小偏移量，以弥补代理模型与真实模型之间可能存在的较大差距。4. 当发现代理模型不是有效的上界且偏移量 γ 增大时，我们会撤销并重新执行相关事件。

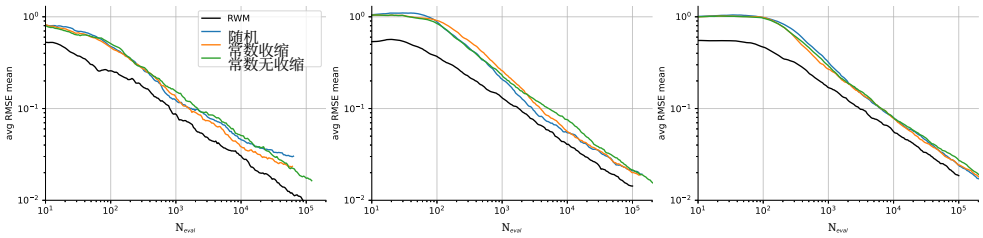
我们比较以下三种情形，以研究这些潜在违规对采样器性能和准确度的影响：

- **常势函数：**我们使用常数替代模型，并在需要时增大偏移量 γ。此外，偏移量会随时间呈指数收缩 t。这检验了第 2、3 和 4 点。
- **无收缩的常势函数：**我们使用常数替代模型，并在需要时增大偏移量 γ，但不允许其收缩。这检验了第 2 和 4 点。
- **随机梯度：** 我们使用一种代理模型，为 Zig-Zag采样器的每次迭代生成一个随机值。对于势函数梯度的每个分量，我们从 $\mathcal{U}(-0.5,\ 0。5)$ 中采样一个值。必要时，我们增大偏移量 γ，并允许其缩小。这将测试采用非静态上界过程的极限，即测试第 1 点和第 2 点。

这两个常数替代模型代表了代理建模中最朴素的方法，即简单地使用一个常数值作为速率函数的上界。从理论上讲，这种模型只适用于有界域，或者适用于定义在无界域上的少数分布，例如柯西分布。例如，高斯分布会产生无界的速率函数。然而在实践中，Zig-Zag采样器的所有实现都具有有限的运行时间，因此沿任意模拟轨迹，速率函数都是有界的。我们希望通过这个常数模型找到速率函数的最大值。

随机代理模型代表一种高度不准确的模型，无法提供太多有关速率函数的有用信息。此外，它是一个不断变化的上界的极端情形，有助于单独考察随时间篡改上界过程所产生的影响。

在图 6中，我们比较了三种情形下后验均值（上图）和方差（下图）估计的平均均方根误差，维度分别为 $d = 2$，$d = 5$ 和 $d = 10$。图 7还报告了每次模型评估对应的平均有效样本量。首先要指出的是，如图 6a 至 6c所示，三种情形都收敛到了真实的后验分布，这一点可由后验均值估计的均方根误差看出。我们还可以从图 6d 至 6f中观察到后验方差的收敛，尽管



(a) $d = 2$ (b) $d = 5$ (c) $d = 10$

(d) $d = 2$ (e) $d = 5$ (f) $d = 10$

图6：均值（第一行）和方差（第二行）的均方根误差随模型评估次数的变化。所有模型均收敛，但表现不如基线RWM采样器。

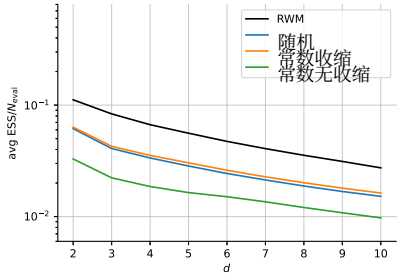


图7：每次模型评估的平均有效样本量随维数的变化。

采用收缩的模型在后期出现了轻微的向上曲率。这是一项重要结果，因为它表明Zig-Zag采样器能够适应可能违反收敛保证假设的情况；这也与 [25], 一致，即为每个事件找到一个新的上界。

需要注意的第二点是，就每次模型评估所对应的有效样本量而言，带收缩的常数模型表现最佳；如图7所示，它略优于带收缩的随机模型。在这一指标上，这两个模型都明显优于不带收缩的常数模型。常数模型位居首位并不令人意外，因为在这三种模型中，它应当产生最紧的上界。随机模型位居第二则不那么直观。这一发现表明了两点：i)

虽然模型中的随机性可能导致偏移量增加，但它也可能偶然地在过程速率较低的区域产生更紧的上界。ii) 对于这样一个朴素的代理模型而言，偏移量的收缩十分重要。更普遍地说，我们发现，随着代理模型准确度的提高，收缩参数的影响 β 会减弱，详见A中的详细讨论。分析的关键结论是，只要代理模型足够准确，Zig-Zag采样器就能稳健地应对收缩参数的选择。

最后，显然这三种模型的准确性都不足，因为在所有维度下，Zig-Zag采样器在每次模型评估的均方根误差和有效样本量方面均不如随机游走 Metropolis采样器。这清楚地表明，我们需要更好的代理模型。

4.4 替代模型比较

现在，我们从收敛率和采样效率两方面比较不同代理模型的性能。我们考虑常数替代模型、拉普拉斯近似，以及两种高斯过程替代模型：一种仅观测函数值，另一种还观测梯度（梯度高斯过程）。两种高斯过程模型的训练数据均为每个潜在空间维度 25 个点。所有模拟中，我们将收缩参数设为 $\beta = 2 \times 10^{-2}$。

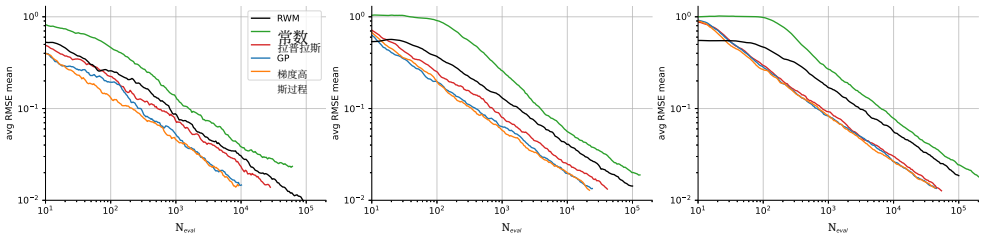
图8 展示了在 $d = 2$、 $d = 5$和 $d = 10$ 维度下，后验均值（上排）和方差（下排）估计的平均均方根误差。除常数模型外，对于所有代理模型，Zig-Zag采样器在所有度量指标和维数下均优于随机游走 Metropolis采样器。即使对于拉普拉斯模型，这一结论也成立；考虑到它对速率函数的近似相当粗糙，这一点尤为显著。其余模型则形成了稳定的性能梯队：梯度高斯过程表现最佳，其次是高斯过程、拉普拉斯模型，最后是常数模型。不过，梯度高斯过程仅以微弱优势领先，这表明在这些设置中，梯度提供的额外信息并非关键。事实上，随着维数增加，非恒定模型之间的差距逐渐缩小，这表明随着底层场的离散化变得更精细，目标分布会变得更加各向同性。

图9展示了所有代理模型每次模型评估的平均有效样本量随维度的变化。模型排名与均方根误差结果一致，梯度高斯过程领先全场，其次是高斯过程、拉普拉斯、RWM采样器，最后是常数模型。

4.5 高斯过程代理模型的自适应细化

代理模型的最后一项扩展，是在PDMP模拟过程中对代理模型进行自适应细化。由于需要在每个提议的事件时间评估模型，我们可以利用这些点的额外信息来更新代理模型。我们在图10中比较了预训练高斯过程替代模型与自适应高斯过程版本（自适应高斯过程）。

模拟开始时，自适应高斯过程和高斯过程曲线完全重合，因为自适应高斯过程模型与高斯过程模型使用相同的训练数据进行初始化。在模拟的后续阶段，自适应细化开始显现，两个模型的均方根误差曲线



(a) $d = 2$ (b) $d = 5$ (c) $d = 10$

(d) $d = 2$ (e) $d = 5$ (f) $d = 10$

图 8：代理模型的比较，包括均值的均方根误差（第一行）和方差的均方根误差（第二行）。更灵活的拉普拉斯、高斯过程和梯度高斯过程模型足够准确，优于 RWM 采样器。

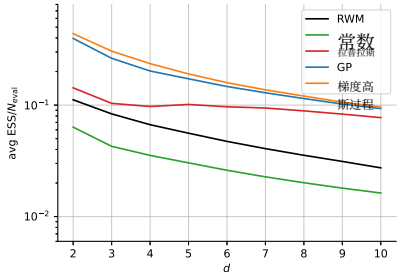


图 9：所有代理模型的每次模型评估平均有效样本量随维数的变化。除常数代理外，其他模型均优于 RWM 采样器。

逐渐出现差异。与直觉相反，自适应高斯过程并未始终优于高斯过程模型。事实上，在所有三项度量指标中，两条曲线都多次交叉。自适应高斯过程模型每次模型评估的有效样本量略高于高斯过程模型，但优势微乎其微，如图11所示。

我们假设，在模拟过程中添加的额外训练点位于初始训练集已充分覆盖的区域，因此不会为模型提供太多新信息。对于这个非线性程度适中的问题，通过拉普拉斯近似进行初始采样可获得高质量的训练点；但存在强非线性时，情况可能有所不同。因此我们假设

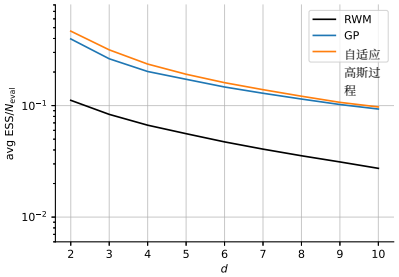
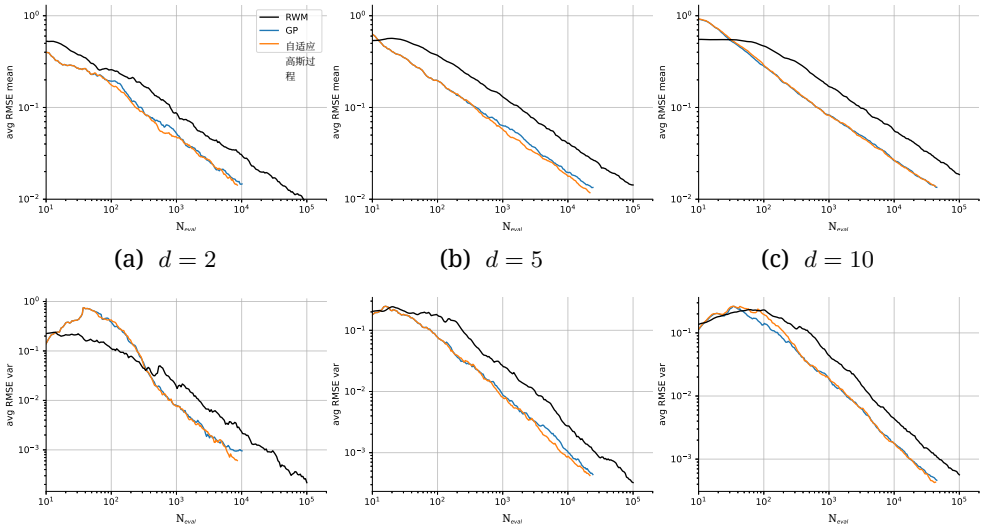


图 11：GP 和自适应高斯过程模型中，每次模型评估的平均有效样本量随维数的变化。

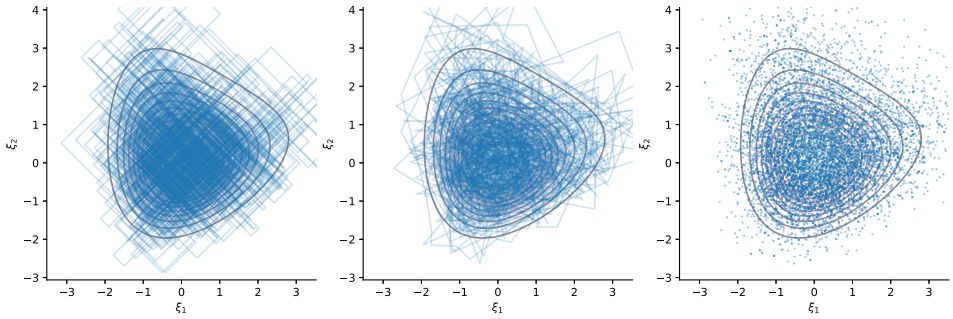
自适应细化的益处高度依赖于具体问题，并且要在更具挑战性的场景中充分发挥其潜力，还需要更复杂的自适应采样策略。



(a) $d = 2$ (b) $d = 5$ (c) $d = 10$

(d) $d = 2$ (e) $d = 5$ (f) $d = 10$

图10：预训练高斯过程与自适应高斯过程在均值的均方根误差（第一行）和方差的均方根误差（第二行）方面的比较。在模拟过程中增加更多训练点并不能持续提升性能。



(a) ZZS (b) BPS (c) NUTS

图 12: (a) Zig-Zag采样器和 (b) 弹跳粒子采样器的路径，以及 (c) 来自 NUTS 的样本。

4.6 Zig-Zag采样器 vs 弹跳粒子采样器 vs 无转向采样器

最后，我们将 Zig-Zag采样器的性能与弹跳粒子采样器和 NUTS 进行比较。这种比较是必要的，因为后者被认为是连续分布领域最先进的 HMC 算法。对于 Zig-Zag采样器和弹跳粒子采样器，我们使用每个潜在空间维度 25 个点的高斯过程代理模型。对于 NUTS，我们采用自己的算法实现，并使用双重平均调整步长，如 [7]中所述。我们对 Zig-Zag采样器和 NUTS 进行冷启动，也就是说，在推断之前不对超参数进行任何调整。对于弹跳粒子采样器，我们使用第B节中认为合适的刷新率，即 $\lambda_{\text{ref}} = 10^{-1}$。我们在图12中展示了 Zig-Zag采样器、弹跳粒子采样器和 NUTS 的路径。

弹跳粒子采样器在后验均值估计方面明显优于Zig-Zag采样器，如图13a 至13c所示。两种PDMP算法在该指标的所有维度上均优于 NUTS：假设模型评估预算为 1000 次，Zig-Zag采样器和弹跳粒子采样器的均方根误差均约为 0.04，而 NUTS 在 $d = 2$ 下仅达到约 0.085。在 $d = 10$ 维度下，弹跳粒子采样器的均方根误差约为 0.04，NUTS 为 0.2，Zig-Zag采样器介于两者之间，约为 0.08。NUTS 在所有维度上的性能最差，均方根误差约为 0.2。

观察后验方差估计时，情况略有不同，如图 13d 至 13f所示：在所有维度上，Zig-Zag采样器均优于弹跳粒子采样器和 NUTS。弹跳粒子采样器仍优于 NUTS，但其方差估计的准确度与 RWM算法相当。

由于没有任何算法在两项度量指标上都占据优势，我们还考察了不同采样器所得估计值与参考解之间的 WD。通过检查图 13g 至 13i中的 WD，我们看到，两个 PDMP 在 $d = 2$ 和 $d = 5$ 维度上的表现同样出色，其中在 $d = 10$维度上，弹跳粒子采样器略占优势。

情况下，NUTS 在所有维度上的表现再次最差。不过，需要注意的是，NUTS 在这一比较中处于结构性劣势：在给定的N_{eval} 模型评估预算下，与 PDMP 采样器相比，它返回的后验样本更少（因为每条被接受的轨迹都包含多个蛙跳步）。因此，WD 是根据更少的样本计算得出的，并表现出更高的有限样本偏差。

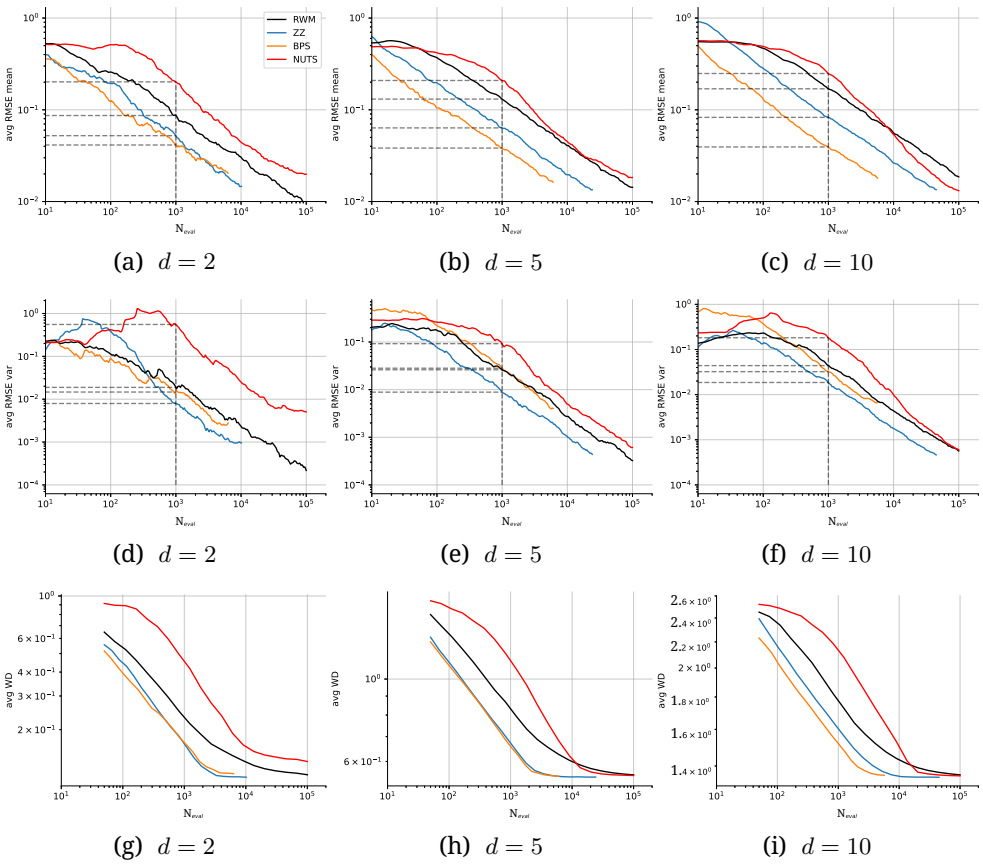
从每次模型评估所得的有效样本量来看，情况也大致相同，如图 14 所示：弹跳粒子采样器表现最佳，其次是 Zig-Zag采样器，最后是 NUTS。只有当 $d > 6$ 时，NUTS 才优于 RWM采样器，这得益于它在维数增加时良好的扩展性。Zig-Zag采样器的斜率与 RWM算法相近。而弹跳粒子采样器的斜率更接近 NUTS，但偏移量更大，这表明它在维数增加时的扩展性优于 Zig-Zag采样器。

乍看之下，NUTS采样器在这项比较中的性能不佳，令人意外，毕竟它以先进的马尔可夫链蒙特卡罗算法著称。对于这一结果，有几种可能的解释：i）仿射变换使分布更加各向同性，从而提升了 Zig-Zag采样器和 RWM算法相对于 NUTS 的性能。ii）该分布仅呈现中等程度的非线性。iii）NUTS采样器的冷启动导致超参数未能达到最优，进而影响其性能。在许多实际应用中，用户会在正式推断之前投入 $> 10^3$ 次迭代来调整 NUTS 的超参数。当涉及耗费较高的偏微分方程模型评估时，这种前期投入通常并不可行。简而言之，NUTS 在此处考察的设置中并不突出，尽管它在许多其他场景中表现出色。

5 结论

在本工作中，我们提出并分析了一种代理模型辅助框架，用于在由偏微分方程（PDEs）控制的非线性贝叶斯反问题中模拟分段确定性马尔可夫过程（PDMP）采样器。我们聚焦于Zig-Zag采样器（Zig-Zag采样器）和弹跳粒子采样器（弹跳粒子采样器），设计了一种全局稀疏化策略：该策略使用势函数的代理模型，并配合一种简单而有效的校正机制来处理由此产生的事件率。这一构造使该方法不依赖于特定的代理模型族：每当代理模型在局部低估真实速率时，便增大偏移量，以恢复有效稀疏化所需的上界性质。

我们的数值实验——线弹性中系数场的推断——表明，该方法能够得到既稳健又高效的分段确定性马尔可夫过程采样器。即便是非常粗糙的代理模型，例如常数模型或白化坐标下的纯拉普拉斯近似，也可以即时校正，并且在经验上观察到它们收敛到参考后验。在拉普拉斯近似之上使用表达能力更强的高斯过程（GP）代理模型，进一步收紧了速率近似，并全面提高了采样效率。尽管高斯过程模型带来了明显收益，但在我们的基准问题中，引入导数观测和自适应细化仅带来了有限的额外性能提升。



(a) $d = 2$ (b) $d = 5$ (c) $d = 10$

(d) $d = 2$ (e) $d = 5$ (f) $d = 10$

(g) $d = 2$ (h) $d = 5$ (i) $d = 10$

图 13：Zig-Zag采样器、弹跳粒子采样器和 NUTS 在均值的均方根误差（上排）、方差的均方根误差（中排）和 WD（下排）方面的比较。在所有维度上，就后验均值和方差估计以及 WD 而言，两种分段确定性马尔可夫过程均优于 NUTS。

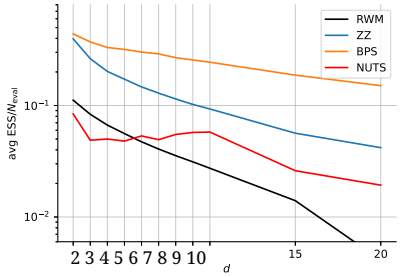


图 14：Zig-Zag采样器、弹跳粒子采样器、NUTS 和随机游走 Metropolis 的每次模型评估平均有效样本量随维数变化的情况。弹跳粒子采样器效率最高，其次是 Zig-Zag采样器、NUTS 和随机游走 Metropolis。

最后，我们比较了代理辅助框架中的 Zig-Zag采样器与弹跳粒子采样器。在后验均值估计误差方面，弹跳粒子采样器更胜一筹；在方差估计方面，Zig-Zag采样器表现更佳；而在瓦瑟斯坦距离方面，两种算法总体不相上下。弹跳粒子采样器在每次模型评估中获得的有效样本量更高，这一优势在高维情形下尤为明显。在所有设置和考察的各项度量指标上，两种 PDMP 采样器的表现都优于经过良好调优的随机游走 Metropolis 基准方法以及广受欢迎的无转向采样器。

本研究仅考察了由具有解析正向映射的椭圆型偏微分方程产生的光滑单峰后验分布，且参数维数适中。所用高斯过程代理模型的一个局限是，其训练成本随设计点数量呈立方增长，这限制了它们在超高维情形下的适用性。此外，尽管我们的实证结果支持在所提出的校正方案下，代理辅助 PDMP 采样器能够收敛，但对于自适应非严格上界情形下的收敛性进行理论分析，仍是一个尚待解决的问题。

未来的研究方向包括将该方法扩展到更贴近实际的偏微分方程控制的反问题、更高维的参数空间，以及可能存在多峰后验分布的情形。在代理模型方面，稀疏或低秩高斯过程构造或神经网络代理模型有望在保持全局精度的同时，缓解立方级计算复杂度。最后，对所提出的代理模型辅助 PDMP 采样器的收敛性质进行严谨的理论分析，将为实际应用提供宝贵的见解和指导。尽管如此，本研究首次成功地将 PDMP 采样应用于 PDE控制的贝叶斯反问题，证明了这类方法在广泛的科学与工程应用中具有可行性。

CRediT 作者贡献声明。 L. Riccius：概念构思、方法设计、软件开发、验证、调查、可视化、撰写初稿。I.B.C.M. Rocha：概念构思、方法设计、监督、撰写、审阅与编辑。J. Bierkens：方法设计、撰写、审阅与编辑。H. Kekkonen：撰写、审阅与编辑。F.P. van der Meer：概念构思、方法设计、监督、撰写、审阅与编辑。

致谢。本研究由代尔夫特理工大学 AILabs 项目通过 SLIMM 实验室提供支持。

代码可用性。本研究所用代码可在以下地址获取：github.com/SLIMM-Lab/pdmp.git。

附录 A 收缩参数的影响 β

为完整起见，我们研究收缩参数 β 对式（26）中不同代理模型下 Zig-Zag采样器性能的影响。我们纳入了常数替代模型、拉普拉斯近似，以及两种高斯过程替代模型，其每个潜在空间维度分别使用 25 个和 100 个训练点。对于每种代理模型（在图 A1中以一行表示），我们报告了在一系列收缩参数

取值 β下的后验均值估计的平均均方根误差。实验分别在 $d=2$、 $d=5$ 和 $d=10$ 维度下进行，并在图A1中按列展示。

首先可以观察到，当收缩参数足够小时，即 $\beta \leq 2\times10^{-2}$，Zig-Zag采样器在所有代理模型下都会收敛。其次，使用高斯过程代理模型时，Zig-Zag采样器对所有 β 值都能收敛，因为这类模型在逼近速率函数方面最为灵活。对于拉普拉斯模型，Zig-Zag采样器在 $\beta \geq 2$时无法收敛。在这些情况下，均方根误差在进行数百次模型评估后便停滞不前，表明模型偏差持续存在。常数替代模型的表现最差。它不仅在所有维度下均无法对$\beta \leq 2\times10^{-2}$ 收敛，实际上还会发散。由于偏移值过大引发数值问题，导致所有随机种子对应的PDMP模拟都提前终止，因此省略了 $\beta \geq 2\times10^{-2}$的结果。

为更好地理解收缩参数较大时常数代理模型为何失效，我们考察了Zig-Zag采样器的轨迹。我们选取二维情形的一个有代表性的实现，其中 $\beta = 2\times10^{-1}$。经过预期的初始阶段后，采样器在此期间探索分布的高密度区域，随后突然漂移至定义域的左下角，如图 A2a所示。图 A2b 中的特写显示，采样器围绕一条梯度为零的点构成的直线振荡，并在 ξ_1 方向上稳定地向负方向移动。图 A2c 展示了偏离发生前后的时间区间：两个速度分量都定期翻转，直到约 $t=50$，之后 v_2 保持在 -1不变，而 v_1 则以越来越快的频率持续翻转。

这种行为可以追溯到偏移动力学。 v_1分量中的偏移量较大，会导致频繁发生事件；而 v_2 分量中的偏移量较小，在 β较大时会迅速收缩至零。与此同时，随着我们进一步远离高密度区域， v_1 分量中的偏移量会继续增大。因此，提议机制无法在 v_2 方向生成事件，导致采样器未经校正便向下漂移。由于偏移量只有在相应方向发生事件后才会更新，而这种事件始终不会发生，采样器最终发散。

当收缩参数足够小时，PDMP 会在初始阶段调整至合适的偏移量，这表现为图中均方根误差曲线起始处的平台。A1a 至A1c。确定该偏移量后，它基本保持稳定，仅随时间略有下降，从而使采样器得以缓慢收敛。对于更准确、因而也更稳健的高斯过程代理模型，取 β 较大值的运行也会收敛。在这些情况下，分布尾部发生低概率事件后，偏移量能够迅速恢复似乎是有利的，因为这会迫使偏移量发生跳变。

附录 B 弹跳粒子采样器刷新率

在测试弹跳粒子采样器之前，我们需要确定合适的刷新率 λ_{ref}，它决定了速度从其先验分布中重新采样的频率。较高的刷新率会使这类更新更加频繁，从而确保采样器具有遍历性。然而，这也会使过程更加扩散，更接近随机游走，因此效率更低。最优

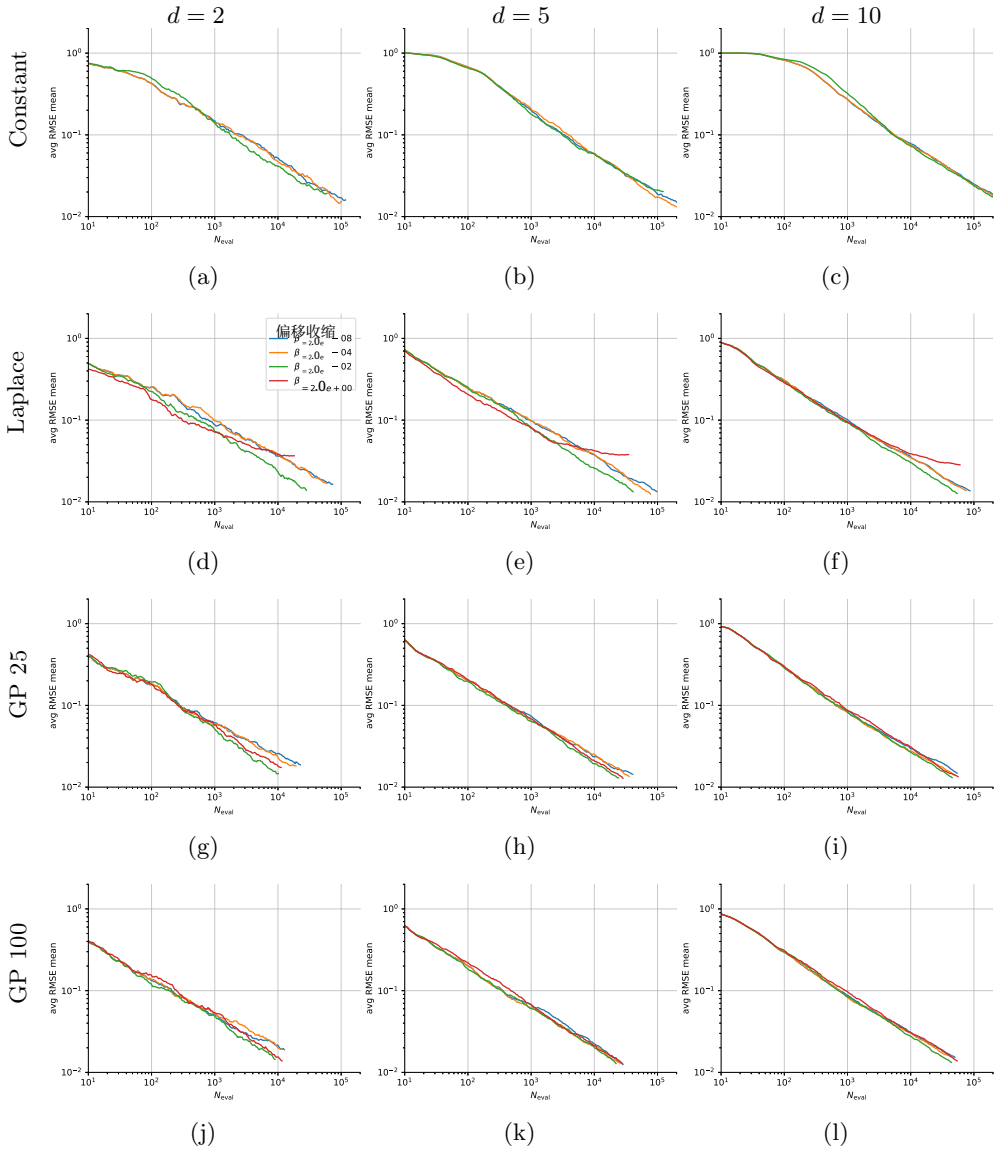
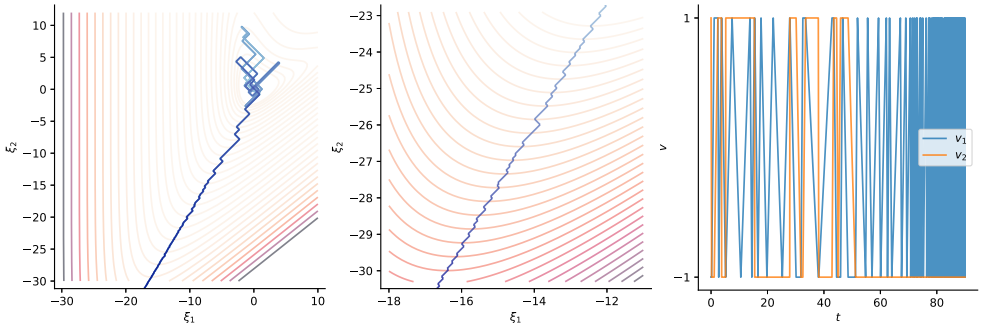


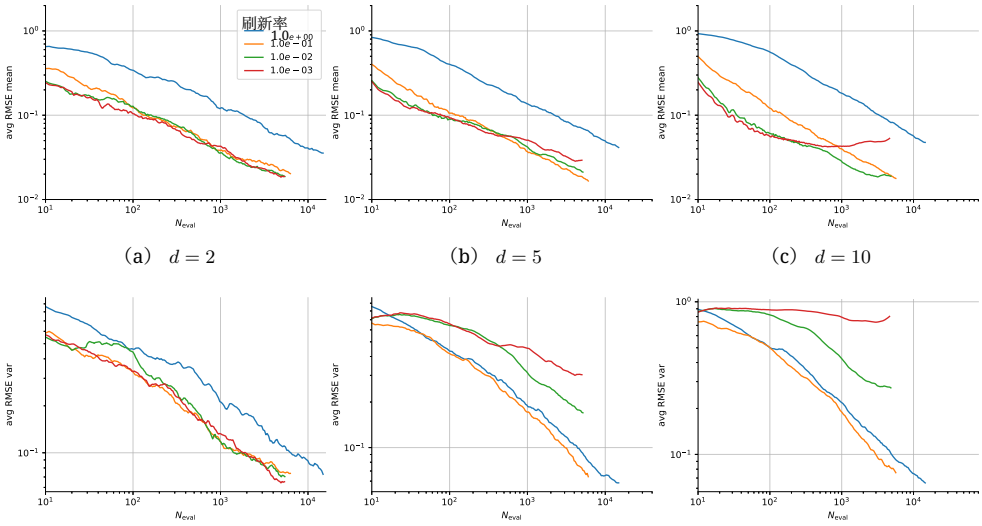
图 A1：不同收缩参数和代理模型下，后验均值估计的平均均方根误差随模型评估次数的变化。

刷新率因此需要在这两种效果之间取得平衡。我们使用每个潜在空间维度含 25 个点的高斯过程代理模型，研究刷新率对弹跳粒子采样器性能的影响。我们考察的刷新率为 $\lambda_{\text{ref}} \in \{10^{-3}, 10^{-2}, 10^{-1}, 1\}$。



(a) (b) (c)

图 A2：检查采用收缩参数 $\beta = 2 \cdot 10^{-1}$ 的常数替代模型的 Zig-Zag采样器轨迹。（a）轨迹和对数密度。（b）采样器在尾部的行为。（c）采样器的速度。（a）和（b）中的时间由从浅到深的颜色渐变表示。



（a） $d = 2$ （b） $d = 5$ （c） $d = 10$

(d) $d = 2$ (e) $d = 5$ (f) $d = 10$

图 B3：不同弹跳粒子采样器刷新率下均值的均方根误差（上排）和方差（下排）的比较。适中的刷新率在这两项度量指标上都表现最佳。

图 B3 中的结果表明，刷新率会显著影响弹跳粒子采样器的性能。刷新率较高时，图 B3顶行所示的后验均值估计收敛较慢。对于 $\lambda_{\text{ref}} \leq 10^{-1}$，均值的准确度表现相对稳定，

这表明继续降低刷新率的收益会逐渐减少。不过，观察图 B3底行中的方差，会发现情况有所不同：刷新率极低会导致后验方差近似较差，这表明缺少刷新会损害采样器的遍历性。我们可以看到，刷新率设得过高会导致探索缓慢，而设得过低则会导致覆盖度不足；不过，最佳范围似乎在 $\lambda_{\text{ref}} = 10^{-1}$附近。

参考文献

[1] Metropolis, N., Rosenbluth, A.W., Rosenbluth, M.N., Teller, A.H., Teller, E.：快速计算机上的状态方程计算。《化学物理学杂志》21，1087–1092（1953）https://doi.org/10.1063/1.1699114[2] Hastings, W.K.：使用马尔可夫链的蒙特卡罗采样方法及其应用。《生物计量学》57，97（1970）https://doi.org/10.2307/2334940[3] Roberts, G.O., Rosenthal, J.S.：各种 Metropolis-Hastings 算法的最优尺度调整。https://doi.org/10.1214/ss/1015346320 16，351–367（2001）https: //doi.org/10.1214/SS/1015346320[4] Roberts, G.O., Stramer, O.：朗之万扩散与 Metropolis-Hastings 算法。《应用概率的方法与计算》2002 4:4 4，337–357（2002）https://doi.org/10.1023/A:1023562417138[5] Duane, S., Kennedy, A.D., Pendleton, B.J., Roweth, D.：混合蒙特卡罗。《物理快报B辑》195，216–222（1987）https://doi.org/10.1016/0370-2693(87) 91197-X[6] Girolami, M.,Calderhead, B.：黎曼流形朗之万和哈密顿蒙特卡罗方法。《皇家统计学会杂志》B辑：统计方法学 73，123–214（2011）https://doi.org/10.1111/j.1467-9868.2010.00765. x[7] Hoffman, M.D., Gelman, A.：无转向采样器：自适应设定哈密顿蒙特卡罗的路径长度。《机器学习研究杂志》15，1593–1623（2014）[8] Carpenter, B., Gelman, A., Hoffman, M.D., Lee, D., Goodrich, B., Betancourt, M., Brubaker, M., Guo, J., Li, P., Riddell, A.：Stan：一种概率编程语言。《统计软件杂志》76（2017）https://doi.org/10.18637/jss.v076.i01[9] Salvatier, J., Wiecki, T.V., Fonnesbeck, C.：使用 PyMC3 在 Python 中进行概率编程。《PeerJ计算机科学》2，55（2016）https://doi.org/10.7717/peerj-cs.55

[10] Chong, A., Lam, K.P.：用于建筑模拟的建筑能源模型贝叶斯校准马尔可夫链蒙特卡罗算法比较，2017年会议。载于：《建筑模拟会议论文集》第2卷，第582–591页（2017）。https://doi.org/ 10.26868/25222708.2017.336 . https://doi.org/10.26868/25222708.2017.336[11] Goodman, J., Weare, J.：具有仿射不变性的系综采样器。《应用数学与计算科学通讯》5，65–80（2010）https://doi.org/10.2140/camcos.2010.5.65[12] Gelman, A., Gilks, W.R., Roberts, G.O.：随机游走 Metropolis 算法的弱收敛与最优尺度调整。《应用概率年刊》7，110–120（1997）https://doi.org/10.1214/aoap/1034625254[13] Roberts, G.O., Rosenthal, J.S.：朗之万扩散离散近似的最优尺度调整。《皇家统计学会会刊B辑》：统计方法学 60，255–268（1998）https://doi.org/10.1111/1467-9868.00123[14] Grimmett, G., Stirzaker, D.：《概率与随机过程》，第4版。牛津大学出版社，牛津（2016）。https://academic.oup.com/book/52807[15] Chen, T.-L., Hwang, C.-R.：加速可逆马尔可夫链。《统计学与概率学通讯》83，1956–1962（2013）https://doi.org/10.1016/j.spl.2013.05. 002[16] Bierkens, J.：非可逆 Metropolis-Hastings 算法。《Statistics and Computing》26，1213–1228（2016）https://doi.org/10.1007/S11222-015-9598-X/TABLES/2[17] Ma, Y.-A., Fox, E.B., Chen, T., Wu, L.：构建高效不可逆马尔可夫链蒙特卡罗采样器的统一框架（2016）[18] Bierkens, J., Roberts, G.：Curie–Weiss模型中提升型 Metropolis–Hastings 算法的分段确定性尺度极限。《应用概率年刊》27，846–882（2017）https://doi.org/10.1214/16-AAP1217[19] Bouchard-Cˆot´e, A., Vollmer, S.J., Doucet, A.：弹跳粒子采样器：一种非可逆无拒绝马尔可夫链蒙特卡罗方法。《美国统计学会会刊》113，855–867（2018）https://doi.org/10.1080/01621459.2017.1294075[20]Bierkens, J., Grazzi, S., Kamatani, K., Roberts, G.：回旋镖采样器。载于：Daum´e, A.H., Singh（编），第37届国际机器学习会议论文集，第908–918页。PMLR，维也纳（2020）。https://proceedings.mlr.press/v119/bierkens20a.html[21] Bierkens, J., Fearnhead, P., Roberts, G.：用于大数据贝叶斯分析的之字形过程与超高效采样。《统计学年刊》47，1288–1320（2019）https://doi.org/10.1214/18-AOS1715

[22] Ke, X., Fan, Y., Zhou, Q.：用于大规模线性逆问题的融合 $L_{1/2}$ 先验及吉布斯弹跳粒子采样器（2025）[23] Bertazzi, A., Bierkens, J.：分段确定性蒙特卡罗算法的自适应方案。《伯努利》28（2022）https://doi.org/10.3150/21-BEJ1423[24] Pagani, F., Chevallier, A., Power, S., House, T., Cotter, S.：Nuzz：适用于一般模型的数值 Zig-Zag。Statistics and Computing 34, 1–26（2024）https: // doi.org/10.1007/S11222-023-10363-8/FIGURES/2[25] Corbella, A., Spencer, S.E.F., Roberts, G.O.：实践中的自动 Zig-Zag 采样。Statistics and Computing 32, 1–16（2022）[26] Fearnhead, P., Bierkens, J., Pollock, M., Roberts, G.O.：用于连续时间蒙特卡罗的分段确定性马尔可夫过程。统计科学 33, 386–412（2018）https://doi.org/10.1214/18-STS648[27] Bierkens, J., Roberts, G.O., Zitt, P.-A.：Zig-Zag 过程的遍历性。《应用概率年刊》29, 2266–2301（2019）https://doi.org/10.1214/ 18-AAP1453[28] Peters, E.A.J.F., With, G.D.：针对一般势函数的无拒绝蒙特卡罗采样。《物理评论 E - 统计、非线性与软物质物理》85, 026703（2012）https://doi.org/10.1103/PHYSREVE.85.026703/FIGURES/5/ MEDIUM[29] Bierkens, J., Kamatani, K., Roberts, G.O.：针对各向异性目标分布的分段确定性蒙特卡罗算法的尺度性质。《伯努利》31（2025）https://doi.org/10.3150/ 24-BEJ1807[30] Lewis, P.A.W., Shedler, G.S.：通过稀疏化模拟非齐次泊松过程。《海军研究后勤季刊》26, 403–413（1979）https://doi.org/10.1002/NAV.3800260304[31] Kennedy, M.C., O'Hagan, A.：计算机模型的贝叶斯校准。《皇家统计学会杂志》B辑：统计方法学 63, 425–464（2001）https://doi.org/10.1111/1467-9868.00294[32] Riccius, L., Rocha, I.B.C.M., Bierkens, J., Kekkonen, H., Meer, F.P.：结合主动学习与马尔可夫链蒙特卡罗采样，以高效校准力学性质。《欧洲力学杂志：A辑／固体》，106015（2026）https: // doi.org/10.1016/J.EUROMECHSOL.2026.106015[33] Feydy, J., S´ejourn´e, T., Vialard, F.X., Amari, S.I., Trouv´e, A., Peyr´e, G.：利用辛克霍恩散度在最优传输与最大均值差异之间插值。AISTATS 2019——第22届人工智能与统计学国际会议（2018）

[34] Andrieu, C., Thoms, J.：自适应马尔可夫链蒙特卡罗教程。Statistics and Computing18，343–373（2008）https://doi.org/10.1007/s11222-008-9110-y