arXiv:2302.03110v1 [stat.AP] 6 Feb 2023

基于压力数据的高维非线性贝叶斯孔弹性场推断

MINA KARIMI*、MEHRDAD MASSOUDI†、KAUSHIK DAYAL‡和MATTEO POZZI§

即将发表于《固体数学与力学》(https://doi.org/10.1177/10812865221140840)

摘要。 我们通过贝叶斯推断，研究由偏微分方程（PDE）支配的大规模逆问题的求解方法。贝叶斯框架为根据含噪测量推断不确定参数提供了统计基础。为量化后验不确定性，我们采用马尔可夫链蒙特卡罗（MCMC）方法生成样本。为提高这些方法在高维情形下的效率，我们利用目标势函数的梯度和海森矩阵的局部信息，并采用哈密顿蒙特卡罗（HMC）。我们的目标应用是根据孔隙压力观测值推断土壤渗透率场，并使用非线性偏微分方程孔隙力学模型预测渗透率对应的压力。我们比较了不同采样方法在该问题及其他情形下的性能。我们还研究了分布的维数和非高斯性对不同采样方法性能的影响。

关键词。 哈密顿蒙特卡罗、高维推断、马尔可夫链蒙特卡罗、孔弹性模型

1. 引言与背景。 科学和工程中的许多问题都可以用偏微分方程（PDEs）建模。受偏微分方程约束的反问题是一类具有挑战性的问题，在研究包括岩土力学工程、医学工程和天体物理学在内的许多物理系统时发挥着重要作用，例如。 [34, 28, 6, 33, 1, 45, 14, 10, 21, 19, 20]。反问题旨在根据观测值推断物理系统的未知模型参数，而这些观测值可能是有限的、与参数存在间接关系，并且受到噪声影响。当参数域为高维，且参数与观测值之间的关系由复杂的力学模型定义时，求解反问题的计算成本很高。处理这类问题的一种方法是通过最小化一个用于量化模型预测值与观测值之间差异的函数，来确定一个点估计量，例如负对数似然函数。然而，在这种表述下，解可能并不唯一，且点估计可能无法提供有关后验知识的充分信息。克服病态性并降低解对测量噪声敏感性的一种经典方法，是添加正则化项，以强化解的连续性和光滑性[34, 12, 42]。在贝叶斯框架中，未知参数被视为随机变量，并为其指定先验概率，而似然函数则依赖于正向模型和含噪测量值 [28, 45]。先验和似然共同构成后验分布。用于后验评估的一种点估计量是最大后验估计（MAP）；但也可以表示后验不确定性：当后验分布无法以精确形式导出时，可通过样本或近似来描述后验分布。

本文采用贝叶斯推断框架，推断基于偏微分方程的正向孔隙力学模型中的不确定参数。为对贝叶斯逆问题进行数值求解，必须将无限维参数空间离散化为有限维区域。在无限维逆问题的研究中，人们已经探讨了不同的离散化方法。有限元离散化方法使用有限个连续拉格朗日基函数来近似无限维参数空间 [6, 45]。Karhunen-Loeve（K-L）展开则提供了另一种方法，即用特征值表示无限维参数集，并允许在有限项后截断，从而实现近似降维[43, 44]。

我们讨论基于马尔可夫链蒙特卡罗（MCMC）的采样方法在高维参数空间中表征后验分布的有效性。对于我们的高维应用而言，标准的 Metropolis-Hastings算法计算成本过高 [16]。为提高采样速度，人们已作出许多努力

*卡内基梅隆大学土木与环境工程系（minakari@andrew.cmu.edu）。
†国家能源技术实验室，宾夕法尼亚州匹兹堡，邮编 15236-0940。
‡卡内基梅隆大学土木与环境工程系；卡内基梅隆大学非线性分析中心、数学科学系；卡内基梅隆大学机械工程系；卡内基梅隆大学斯科特能源创新研究所。§卡内基梅隆大学土木与环境工程系；卡内基梅隆大学斯科特能源创新研究所。

针对高维逆问题，例如开发降阶模型， [2, 8, 15, 32],利用局部梯度信息的哈密顿蒙特卡罗（HMC）方法， [4], 利用海森信息的黎曼流形哈密顿蒙特卡罗 [26, 7], 以及基于海森矩阵的马尔可夫链蒙特卡罗方法 [35, 28, 6, 33]。本文研究基于海森矩阵的马尔可夫链蒙特卡罗方法和基于海森矩阵的 HMC 方法在高维问题中的效率。

在无限维逆问题中，利用海森信息可以显著提升马尔可夫链蒙特卡罗方法的性能。Qi 和 Minka [35] 提出了一种基于海森矩阵的 Metropolis-Hasting（HMH）算法。他们通过利用局部梯度和海森信息近似高斯分布，构造了一种自适应提议密度。Martin等人 [28] 将 HMH 方法应用于高维逆问题，通过近似低秩局部海森矩阵来确保矩阵正定，并由此提出了他们称为随机牛顿马尔可夫链蒙特卡罗（SNMCMC）的方法。Petra等人 [33]通过使用最大后验点处的海森矩阵改进了 SNMCMC 方法，以降低其计算成本。

HMC方法利用局部梯度信息，在参数空间中进行长距离移动，因此比常规马尔可夫链蒙特卡罗方法更快地探索后验分布 [4]。通过选择合适的步长、样本总数和质量矩阵，HMC算法可用于在高维参数空间中高效采样 [16]。已有多项研究提出通过利用海森信息调整质量矩阵来改进HMC算法 [13, 16, 26, 48, 49]。Lee 和 Vempala [26] 证明，将局部海森信息用作质量矩阵的黎曼流形哈密顿蒙特卡罗（RMHMC）能够提高HMC的收敛率；Zhang 和 Sutton [49] 则通过使用近似的BFGS海森矩阵作为质量矩阵，开发了一种改进的HMC方法。Bui-Thanh 和 Girolami [7] 研究了一种采用精确费 Fisher信息矩阵及其低秩近似的 RMHMC方法。

我们重点研究近期针对高维逆问题提出的海森信息辅助的马尔可夫链蒙特卡罗和 HMC 方法的性能。我们比较 Petra等人提出的 SN-MAP方法 [33], 采用最大后验点处海森信息的 Metropolis-调整朗之万算法（MALA） [36] 算法，以及 Bui-Thanh 和 Girolami 提出的 H-HMC方法 [7]。我们通过若干数值算例，考察问题的维数和非高斯特性如何影响不同海森信息辅助方法的性能。我们不研究采用局部海森信息的海森信息辅助方法，因为这类方法计算开销很大，而且每次迭代都需要计算二阶导数信息。

高维逆问题有许多应用 [1, 7, 28, 33]。我们的分析旨在推断与多种能源活动相关的注入地点附近区域的土壤渗透率，例如废水注入、二氧化碳封存、地热能开发和水力压裂。推断这些参数对于估算储层容量、预测注入中心周围的孔隙压力和应力，以及评估滑移和注入诱发地震事件的危害至关重要 [41, 38, 24, 39, 46, 47, 25, 23, 27, 40, 5]。我们根据收集到的压力测量值推断未知的渗透率场。在本研究中，我们采用贝叶斯框架，根据含噪且稀疏的测量值推断非线性孔隙力学模型中未知的孔弹性性质。我们使用非精确牛顿求解器确定最大后验估计，然后从该点出发，采用马尔可夫链蒙特卡罗方法生成样本。我们将所建立的模型用于求解大规模非线性反问题，并确定深部地下层的未知性质。

组织结构。第2节介绍高维逆问题的一般贝叶斯公式，并讨论先验的选择。第3 节回顾并讨论用于探索后验分布的标准采样方法和加速采样方法。第4 节给出正向孔隙力学模型的控制方程。第5 节展示识别最大后验点以及从先验分布和后验分布生成样本的数值结果。

2. 高维逆问题的贝叶斯框架。 我们遵循贝叶斯框架推断场 $\theta(x)$，其中 x 是参考区域 Ω上的空间坐标。关于该场的先验知识由函数 $\pi(\theta) = \log(p(\theta))$建模，并假设 $p(\theta)$服从高斯分布，因此：

$$\pi(\theta) = -\frac{1}{2}||\mathcal{A}(\theta - m_\pi)||^2 + \text{const.} \tag{2.1}$$

其中 $m_\pi(x)$ 是先验均值函数， $\mathcal{A}$ 是类拉普拉斯算子，将在第 2.1节中定义。对于任意函数 $f \in L^2(\Omega)$， $L^2(\Omega)$ 属于平方可积函数空间，其范数定义为：

$$||f(x)||^2 = \int_\Omega f^2(x)\,\mathrm{d}x \tag{2.2}$$

现有一个包含 n 含噪观测的数据集，这些观测值列于向量 $\boldsymbol{y} = [y_1, y_2, \cdots, y_n]^\top$ 中。观测值 i 与场的关系由方程 $y_i = Q_\theta(x_i,\ t_i) + \epsilon_i$ 给出，其中函数 Q_θ 根据参数场 θ、时间 t 和位置 x 预测待测量的值，该函数通过求解前向预测模型计算得到。噪声项列于向量 $\boldsymbol{\epsilon} = [\epsilon_1, \epsilon_2, \cdots, \epsilon_n]^\top$ 中，并建模为高斯随机变量；该变量与 θ 相互独立，均值为零，协方差矩阵为 $\boldsymbol{\Gamma}_\epsilon$，即 $\boldsymbol{\epsilon} \sim \mathcal{N}(\mathbf{0},\ \boldsymbol{\Gamma}_\epsilon)$。

从概率角度看，预测值与测量值之间的一致性由对数似然函数$l(\theta) = \log\big(p(\boldsymbol{y}|\theta)\big)$建模：

$$l(\theta) = -\frac{1}{2}||\boldsymbol{\Gamma}_\epsilon^{-1/2}(\boldsymbol{Q}_\theta - \boldsymbol{y})||^2 + \text{const.} \tag{2.3}$$

其中 $\boldsymbol{Q}_\theta = [Q_\theta(x_1,\ t_1)\ \cdots,\ Q_\theta(x_n,\ t_n)]^\top$ 是预测值向量。本文中，对于每个正定对称矩阵 $\boldsymbol{G}$，矩阵 $\boldsymbol{G}^{-1/2}$ 是该矩阵逆矩阵的一种分解，因此 $\boldsymbol{G}^{-1} = \boldsymbol{G}^{-1/2}\boldsymbol{G}^{-\top/2}$（例如，该分解可以基于特征值分析或乔列斯基方法）。对于列向量，范数定义为 $||\boldsymbol{v}||^2 = \boldsymbol{v}^\top\boldsymbol{v}$。如果 $\boldsymbol{\Gamma}_\epsilon = \sigma_\epsilon^2\boldsymbol{I}$（其中 $\boldsymbol{I}$是单位矩阵），那么 $\boldsymbol{\Gamma}_\epsilon^{-1/2}$可以替换为 σ_ϵ^{-1}，见方程（2.3）。

根据贝叶斯法则，对数后验密度 $\omega(\theta) = \log\big(p(\theta|\boldsymbol{y})\big)$ 为：

$$\omega(\theta) = l(\theta) + \pi(\theta) + \text{const.} \tag{2.4}$$

受上述方程启发，我们将目标函数定义为 $\mathcal{J}(\theta) = -\omega(\theta) + $ 常数：

$$\mathcal{J}(\theta) = \frac{1}{2}||\boldsymbol{\Gamma}_\epsilon^{-1/2}(\boldsymbol{Q}_\theta - \boldsymbol{y})||^2 + \frac{1}{2}||\mathcal{A}(\theta - m_\pi)||^2 \tag{2.5}$$

最大后验点为 $\theta_{\text{MAP}} = \text{argmin}\,\big(\mathcal{J}(\theta)\big)$，它是使后验密度达到最大的参数场。为确定$\theta_{\text{MAP}}$，我们最小化 $\mathcal{J}$：最小二乘方法会最小化第一项（基于似然），而第二项（基于先验知识）通常起到正则化的作用。

2.1. 先验模型的选择。 选择 $p(\theta)$至关重要，因为它能够在贝叶斯推断框架下纳入关于参数场 θ的先验知识。本文使用连续高斯测度 $\mathcal{N}(m_\pi, \mathcal{C}_\pi)$，而不采用离散化高斯分布和协方差矩阵 [6, 7]。这里， $\mathcal{C}_\pi$ 是协方差算子，其定义如下：

$$(\mathcal{C}_\pi\phi)(\boldsymbol{x}) = \int_\Omega \boldsymbol{c}(\boldsymbol{x}, \boldsymbol{y})\boldsymbol{\phi}(\boldsymbol{y})\,\mathrm{d}\boldsymbol{y} \tag{2.6}$$

其中， $\boldsymbol{c}(\boldsymbol{x}, \boldsymbol{z})$是协方差函数：

$$\boldsymbol{c}(\boldsymbol{x}, \boldsymbol{z}) = \mathbb{E}\left[\big(\theta(\boldsymbol{x}) - m_\pi(\boldsymbol{x})\big)\big(\theta(\boldsymbol{z}) - m_\pi(\boldsymbol{z})\big)\right] \tag{2.7}$$

给定测度函数 μ_π，算子定义如下：

$$\mathcal{C}_\pi = \int_\Omega (\theta - m_\pi) \otimes (\theta - m_\pi)\mu_\pi(\,\mathrm{d}\theta) = \mathbb{E}[(\theta - m_\pi) \otimes (\theta - m_\pi)] \tag{2.8}$$

算子 $\mathcal{C}_\pi$ 具有快速求值的特点，并能在高维空间中实现简单的离散化 [6]。如附录A 所述，先验函数 π （方程（2.1））写作如下：

$$\pi(\theta) = -\frac{1}{2}||\mathcal{C}_\pi^{-1/2}(\theta - m_\pi)||^2 + \text{const.} \tag{2.9}$$

通常，协方差算子被视为类拉普拉斯算子 $\mathcal{A}$的分数次幂，即 $\mathcal{C}_\pi = \mathcal{A}^{-\alpha}$， $\alpha > 0$其中 $\mathcal{A}$ 是正定、自伴且可逆的 [11, 31, 9]。这里我们选择协方差算子为 $\mathcal{C}_\pi = \mathcal{A}^{-2}$，使得 $\mathcal{C}_\pi^{-1/2} = \mathcal{A}$，而算子 $\mathcal{A}$ 作用于 θ的方式，即 $\mathcal{A}\theta$ 由以下椭圆边值问题（$\mathcal{A} = -\gamma \boldsymbol{\Psi} \Delta + \delta \boldsymbol{I}$）定义：

$$
\begin{aligned}
-\gamma \operatorname{div}(\boldsymbol{\Psi} \nabla \theta) + \delta \theta &= s \ \text{ in } \Omega \\
\gamma (\nabla \theta) \cdot \boldsymbol{n} &= 0 \ \text{ on } \partial \Omega
\end{aligned} \tag{2.10}
$$

其中， Δ 是拉普拉斯算子， γ 和 δ 是正的（超）参数， s 是白噪声场，随机场 θ可表示为随机噪声的线性函数，即 $\mathcal{A}^{-1}s$。这里， $\boldsymbol{n}$ 是 $\partial\Omega$上的单位法向量。对称正定二阶张量 $\boldsymbol{\Psi}$ 定义了各向异性，并指明了参数空间 θ 中变分较大的方向$\hat{\boldsymbol{m}}$ 。随后，可根据单位向量$\hat{\boldsymbol{m}}$将 $\boldsymbol{\Psi}$ 定义如下：

$$
\boldsymbol{\Psi} = a \left(\boldsymbol{I} - (1 - b/a) \hat{\boldsymbol{m}} \otimes \hat{\boldsymbol{m}}^\top \right) \tag{2.11}
$$

在二维情形下， $\boldsymbol{\Psi}$ 可表示为：

$$
\boldsymbol{\Psi} = \begin{bmatrix} a \sin(\beta)^2 + b \cos(\beta)^2 & (b-a) \sin(\beta) \cos(\beta) \\ (b-a) \sin(\beta) \cos(\beta) & b \sin(\beta)^2 + a \cos(\beta)^2 \end{bmatrix} \tag{2.12}
$$

其中 a， b 和 β 均为常数。和 δ 越大，方差就越小（方程（2.10））。根据上述公式，先验项的连续形式可写为：

$$
\mathcal{A}\theta = \begin{cases} -\gamma \operatorname{div}(\boldsymbol{\Psi} \nabla \theta) + \delta \theta & \text{in } \Omega \\ \gamma (\nabla \theta) \cdot \boldsymbol{n} & \text{on } \partial \Omega \end{cases} \tag{2.13}
$$

2.2. 场的离散化。我们采用具有连续拉格朗日基函数$\{\phi_j\}_{j=1}^m$的有限元离散化，并使用相应的节点值 $\{\boldsymbol{x}_j\}_{j=1}^m$，以获得该问题的有限维近似。因此，离散场可表示为系数向量 $\boldsymbol{\psi} = [\psi_1, \ldots, \psi_m]^\top$。节点系数向量之间的内积可由正定对称矩阵 $\boldsymbol{N}$加权，其元素定义如下：

$$
N_{ij} = \int_\Omega \phi_i(\boldsymbol{x}) \phi_j(\boldsymbol{x}) \, \mathrm{d}\boldsymbol{x} \tag{2.14}
$$

然后，对于每一对函数 θ_1， $\theta_2 \in L^2(\Omega)$，我们可以将加权内积定义为离散形式 $(\theta_1, \ \theta_2)_{L^2(\Omega)} \approx (\boldsymbol{\psi}_1, \ \boldsymbol{\psi}_2)_{\boldsymbol{N}} = \boldsymbol{\psi}_1^\top \boldsymbol{N} \boldsymbol{\psi}_2$。拉普拉斯算子 $\mathcal{A}$的矩阵表示可定义为$\boldsymbol{A} = \boldsymbol{N}^{-1} \boldsymbol{K}$ [33]，其中 $\boldsymbol{K}$的元素 $\{i, j\}$ 为：

$$
K_{ij} = \int_\Omega \left[\gamma \boldsymbol{\Psi} \nabla \phi_i(\boldsymbol{x}) \cdot \nabla \phi_j(\boldsymbol{x}) + \delta \phi_i(\boldsymbol{x}) \phi_j(\boldsymbol{x}) \right] \mathrm{d}\boldsymbol{x} \tag{2.15}
$$

因此，对数先验项可以近似为：

$$
\pi(\theta) \cong -\frac{1}{2} ||\boldsymbol{A}^{-1/2} \, \boldsymbol{\psi}||^2 + \text{const.} \tag{2.16}
$$

目标函数可以用 $\boldsymbol{\psi}$ 表示为：

$$
\mathcal{J}(\boldsymbol{\psi}) = \frac{1}{2} ||\boldsymbol{\Gamma}_\epsilon^{-1/2} (\boldsymbol{Q}_{\boldsymbol{\psi}} - \boldsymbol{y})||^2 + \frac{1}{2} ||\boldsymbol{A}^{-1/2} \, \boldsymbol{\psi}||^2 \tag{2.17}
$$

$_m$其中， $\boldsymbol{Q}_{\boldsymbol{\psi}}$ 为预测向量，用于重建场 $\theta = \sum \psi_j \phi_{j_=1}$。为了最小化关于 θ的这个非线性函数，通常可以使用拟牛顿方法。

3. 高维反问题的采样方法。 本节回顾从概率密度函数中抽取样本的一些关键方法。我们讨论 Metropolis-Hastings 马尔可夫链蒙特卡罗方法和基于海森矩阵的马尔可夫链蒙特卡罗方法，并比较它们在高维参数空间中采样时的计算成本和效率。

3.1. Metropolis-Hastings 方法。 Metropolis-Hastings 马尔可夫链蒙特卡罗（MH-MCMC）方法（算法3.1）依赖于提议密度 q，用于沿着一条链对随机变量 $\boldsymbol{z}$进行采样，其中心位于当前位置$\boldsymbol{c}$。最常用的提议（对数）密度之一是各向同性高斯分布：

$$q(\boldsymbol{c},\boldsymbol{z}) = -\frac{1}{2}||\Delta t^{-1}(\boldsymbol{c}-\boldsymbol{z})||^2 + \text{const.} \tag{3.1}$$

其中 Δt 是固定步长。

算法 3.1 Metropolis-Hastings 马尔可夫链蒙特卡罗

输入：初始参数 $\boldsymbol{\psi}_0$1：计算 $\mathcal{J}_0 = \mathcal{J}(\boldsymbol{\psi}_0)$，2：对于 $k = 0, \cdots, N$，执行 3：从提议密度 $q(\boldsymbol{\psi}_k, \cdot)$ 中抽取样本 $\boldsymbol{z}$ 4：计算 $\mathcal{J}_z = \mathcal{J}(\boldsymbol{z})$、$\Delta q = q(\boldsymbol{z}, \boldsymbol{\psi}_k) - q(\boldsymbol{\psi}_k, \boldsymbol{z})$，5：计算 $\alpha = -\mathcal{J}_z + \mathcal{J}_k + \Delta q$6：计算 $a = \min[1, \exp(\alpha)]$7：从 $\mathcal{U}([0,1))$ 中抽取 u 8：如果 $u < a$，则 9：接受：设定 $\boldsymbol{\psi}_{k+1} = \boldsymbol{z}, \mathcal{J}_{k+1} = \mathcal{J}_z$10：否则 11：拒绝：设定 $\boldsymbol{\psi}_{k+1} = \boldsymbol{\psi}_k, \mathcal{J}_{k+1} = \mathcal{J}_k$12：结束 if 13：结束 for 输出：随机游走：$\{\boldsymbol{\psi}_k\}_{k=1}^N$

在由偏微分方程控制的高维问题中，场变量线性组合的后验标准差会随组合方向不同而显著变化，这给步长的选择带来了一些挑战。选择较大的步长会因某些方向的标准差较小而导致接受率偏低；而选择与最小标准差一样小的步长则会导致混合缓慢（并且由于每次迭代都需要求解正向模型，整个模拟过程在计算上会非常昂贵）。

在接下来的几节中，我们将讨论一些利用目标函数 $\mathcal{J}$ 的局部梯度和海森信息的加速采样方法。

3.2. 基于海森矩阵的马尔可夫链蒙特卡罗方法。本节回顾 Ghattas 及其合作者提出的三种利用海森矩阵信息的马尔可夫链蒙特卡罗方法 [28, 6, 33]:

方程（2.5）的局部二次型，可基于当前样本 $\boldsymbol{\psi}_k$ 附近的牛顿法，通过拉普拉斯近似（二阶泰勒级数）近似为如下形式 [29]:

$$\mathcal{J}(\boldsymbol{\psi}) \approx \tilde{\mathcal{J}}(\boldsymbol{\psi}) = \frac{1}{2}||\boldsymbol{H}_k^{1/2}(\boldsymbol{\psi}-\boldsymbol{\psi}_k)||^2 + \boldsymbol{g}_k^\top(\boldsymbol{\psi}-\boldsymbol{\psi}_k) + \mathcal{J}(\boldsymbol{\psi}_k) \tag{3.2}$$

其中 $\boldsymbol{g}_k = \nabla\mathcal{J}(\boldsymbol{\psi}_k)$ 和 $\boldsymbol{H}_k = \nabla^2\mathcal{J}(\boldsymbol{\psi}_k)$ 分别是目标函数的梯度和海森矩阵，˜而 $\mathcal{J}$是近似函数。我们注意到，最大后验点处的海森矩阵为正定矩阵。然而，我们无法保证任意一点处的矩阵均为正定矩阵。因此，为了确保˜海森矩阵为正定矩阵，我们可以采用低秩近似 $\boldsymbol{H}_k$ (附录 B)，该近似基于特征值分解计算，并截去负特征值和零特征值。展开方程（3.2），再加上并减去项

$\frac{1}{2}\boldsymbol{g}_k^T\tilde{\boldsymbol{H}}_k^{-1}\boldsymbol{g}_k$，可得：

$$\tilde{\mathcal{J}}(\boldsymbol{\psi}) = \frac{1}{2}(\boldsymbol{\psi}-\boldsymbol{\psi}_k)^\top\tilde{\boldsymbol{H}}_k(\boldsymbol{\psi}-\boldsymbol{\psi}_k) + \boldsymbol{g}_k^\top(\boldsymbol{\psi}-\boldsymbol{\psi}_k) + \mathcal{J}(\boldsymbol{\psi}_k) + \frac{1}{2}\boldsymbol{g}_k^T\tilde{\boldsymbol{H}}_k^{-1}\boldsymbol{g}_k - \frac{1}{2}\boldsymbol{g}_k^T\tilde{\boldsymbol{H}}_k^{-1}\boldsymbol{g}_k \tag{3.3}$$

因此，方程（3.3）可重写为 [28]:

$$\tilde{\mathcal{J}}(\boldsymbol{\psi}) = \frac{1}{2}||\tilde{\boldsymbol{H}}_k^{1/2}\,(\boldsymbol{\psi}-\boldsymbol{\psi}_k+\tilde{\boldsymbol{H}}_k^{-1}\boldsymbol{g}_k)||^2 + \text{const.} \tag{3.4}$$

其中，项 $\tilde{\boldsymbol{H}}_k^{-1}\boldsymbol{g}_k$ 可解释为牛顿步。受此近似启发，随机牛顿马尔可夫链蒙特卡罗（SN-MCMC）方法采用以下提议对数密度：

$$q(\boldsymbol{\psi}_k,\boldsymbol{z}) = -\frac{1}{2}||\tilde{\boldsymbol{H}}_k^{1/2}\,(\boldsymbol{z}-\boldsymbol{\psi}_k+\tilde{\boldsymbol{H}}_k^{-1}\boldsymbol{g}_k)||^2 + \text{const.} \tag{3.5}$$

在实现该方法时，我们需要在采样过程的每次迭代中求解正问题，并计算每次迭代的局部海森矩阵、局部梯度和近似海森矩阵。

由于在每个样本点重新计算海森矩阵和近似海森矩阵的开销很大，Bui等人 [6, 33]提出了一种改进方法，该方法始终使用最大后验点处的海森矩阵。因此，在基于MAP海森矩阵的随机牛顿马尔可夫链蒙特卡罗（SN-MAP）方法中，我们将提议对数密度函数重写如下：

$$q(\boldsymbol{\psi}_k,\boldsymbol{z}) = -\frac{1}{2}||\boldsymbol{H}_{\text{MAP}}^{1/2}\,(\boldsymbol{z}-\boldsymbol{\psi}_k+\boldsymbol{H}_{\text{MAP}}^{-1}\,\boldsymbol{g}_k)||^2 + \text{const.} \tag{3.6}$$

因此，在每次迭代中采用这种方法时，我们需要运行正问题，并计算每个采样点处的局部梯度，但只需在最大后验点处计算一次海森矩阵。

与 MH-MCMC 相比，使用局部近似的提议密度可以提高收敛速度。然而，当目标分布与高斯分布相差较大时，通过添加项（$\boldsymbol{H}^{-1}$ 最大后验估计$\boldsymbol{g}_k$）来改变提议密度的均值（见方程（3.2））， 可能会降低收敛速度，或者导致采样过程陷入低概率区域。为克服这些问题，我们可以定义一个较小的随机学习率$\gamma \le 1$，用项（$\gamma\boldsymbol{H}^{-1}$ 最大后验估计$\boldsymbol{g}_k$）替换项（$\boldsymbol{H}^{-1}$ 最大后验估计$\boldsymbol{g}_k$），或者使用信赖域方法 [35]。

Bui等人 [6, 33], 提出了另一种利用海森矩阵的马尔可夫链蒙特卡罗方法，称为基于MAP点高斯提议的独立采样（IS-MAP）。该方法通过采用以最大后验点为中心的提议密度，避免在每个样本处重新计算局部梯度：

$$q(\boldsymbol{\psi}_{\text{MAP}},\boldsymbol{z}) = -\frac{1}{2}||\boldsymbol{H}_{\text{MAP}}^{1/2}\,(\boldsymbol{z}-\boldsymbol{\psi}_{\text{MAP}})||^2 + \text{const.} \tag{3.7}$$

其中， $\boldsymbol{\psi}_{\text{MAP}}$是最大后验点处的参数值。因此，通过避免在每个采样点处重新计算局部海森矩阵和梯度，我们只需计算一次最大后验点处的海森矩阵，并在每个采样点求解正问题。请注意，在此方法中，每一步生成的样本彼此独立，因此样本生成比其他任何马尔可夫链蒙特卡罗方法都更简单。然而，接受准则仍采用 MH-MCMC 的传统准则。该方法可与经典的拒绝采样方法联系起来。

在梅特罗波利斯-调整朗之万算法（MALA）方法中， [18, 36, 37], 提议密度定义为 s:

$$q(\boldsymbol{\psi}_k,\boldsymbol{z}) = -\frac{1}{2}||(2\tau\boldsymbol{B})^{-1/2}\,(\boldsymbol{z}-\boldsymbol{\psi}_k+\tau\boldsymbol{B}\boldsymbol{g}_k)||^2 + \text{const.} \tag{3.8}$$

其中， $\boldsymbol{B}$ 是正定预条件矩阵， τ 是固定时间步长，必须取为 $0<\tau\le 1$。如果我们使用最大后验估计处的局部海森信息的逆矩阵作为预条件矩阵，则提议密度变为：

$$q(\boldsymbol{\psi}_k,\boldsymbol{z}) = -\frac{1}{2}||\big(\boldsymbol{H}_{\text{MAP}}/(2\tau)\big)^{1/2}\,(\boldsymbol{z}-\boldsymbol{\psi}_k+\tau\boldsymbol{H}_{\text{MAP}}^{-1}\boldsymbol{g}_k)||^2 + \text{const.} \tag{3.9}$$

方程（τ 3.8）和（3.9）中的标量参数可以起到与较小学习率 γ类似的作用，防止算法陷入低概率区域，并为非高斯分布带来更好的结果。

3.3. 哈密顿蒙特卡洛方法。HMC 是一种依赖局部梯度信息的高效采样方法。在 HMC 算法（算法3.2）中，我们定义一个辅助动量值，然后根据哈密顿系统的微分方程更新位置和动量。对于一般系统，位置是我们关注的变量（例如需要推断的参数）。势能定义为第 2 节的目标函数，即目标分布 [17, 30] 的负对数。哈密顿函数可定义为:

$$\mathcal{H}(\boldsymbol{\psi}, \boldsymbol{p}) = \mathcal{K}(\boldsymbol{\psi}, \boldsymbol{p}) + \mathcal{J}(\boldsymbol{\psi}) \tag{3.10}$$

其中， $\mathcal{K}(\boldsymbol{\psi}, \boldsymbol{p})$是动能， $\boldsymbol{p}$ 是动量向量。为了更新变量 $\boldsymbol{p}$ 和 $\boldsymbol{\psi}$，我们需要计算这些变量相对于“伪时间”变量 t的导数。因此，哈密顿方程为:

$$\frac{d\boldsymbol{\psi}}{dt} = \frac{\partial \mathcal{H}}{\partial \boldsymbol{p}} = \frac{\partial \mathcal{K}}{\partial \boldsymbol{p}} + \frac{\partial \mathcal{J}}{\partial \boldsymbol{p}} = \frac{\partial \mathcal{K}}{\partial \boldsymbol{p}} \tag{3.11}$$

$$\frac{d\boldsymbol{p}}{dt} = -\frac{\partial \mathcal{H}}{\partial \boldsymbol{\psi}} = -\frac{\partial \mathcal{K}}{\partial \boldsymbol{\psi}} - \frac{\partial \mathcal{J}}{\partial \boldsymbol{\psi}} = -\frac{\partial \mathcal{J}}{\partial \boldsymbol{\psi}} \tag{3.12}$$

文献中提出了多种不同的 $\mathcal{K}$ 函数。最常见的选择之一是欧几里得-高斯动能 [4]:

$$\mathcal{K}(\boldsymbol{\psi}, \boldsymbol{p}) = \frac{1}{2}||\boldsymbol{M}^{-1/2}\boldsymbol{p}||^2 + \frac{1}{2}\log|\boldsymbol{M}| + \text{const.} \tag{3.13}$$

其中， $\boldsymbol{M}$ 是一个称为“质量矩阵”的正定矩阵。质量矩阵的选择会影响收敛速度和采样过程的性能。

3.3.1. 基于海森矩阵的 HMC 方法。事实证明，在质量矩阵取后验协方差 [13]的逆时，线性问题的收敛速度会更快。在更一般的情况下，黎曼流形 HMC 方法采用黎曼-高斯动能 [3]，将质量矩阵设为 $\boldsymbol{M} = \boldsymbol{C}(\boldsymbol{\psi})$，其中 $\boldsymbol{C}(\boldsymbol{\psi})$是一个关于感兴趣参数的函数，需要根据局部信息 [4]进行更新。

当我们考虑 $\boldsymbol{C}(\boldsymbol{\psi}) = \nabla^2\mathcal{J}(\boldsymbol{\psi})$ 时，哈密顿更新方程将产生类似于牛顿法 [26]，的更新步骤，从而显著改善采样效果。对于高维逆问题，使用黎曼流形哈密顿蒙特卡罗的计算开销很大，因为我们需要在每次迭代时计算后验分布的局部海森矩阵。

为降低计算成本，我们可以使用最大后验点处后验分布的海森矩阵作为质量矩阵[7]，，从而避免在每次迭代时计算海森矩阵。

3.4. SN-MCMC、MALA 和 H-HMC 采样方法的比较。本节比较依赖海森信息的 MCMC 采样方法在高维参数空间中的表述，并讨论它们的相似之处与差异。

3.4.1. SN-MAP。根据方程 (3.6)，SN-MAP 方法的更新步骤为（此处我们直接将下一个候选点表示为 $\boldsymbol{\psi}_{k+1}$，而不是 $\boldsymbol{z}$）：

$$\boldsymbol{\psi}_{k+1} = \boldsymbol{\psi}_k - \boldsymbol{H}_{\text{MAP}}^{-1}\boldsymbol{g}_k + \boldsymbol{H}_{\text{MAP}}^{-1/2}\boldsymbol{u} \tag{3.14}$$

其中 $\boldsymbol{u}$ 是从多元标准高斯分布（即协方差矩阵为单位协方差矩阵、均值向量为零均值向量）中抽取的随机样本。对于 MCMC 方法，如上所述，接受准则基于对数比 α。对于 MH-MCMC（见算法3.1），其过程如下:

$$\alpha = \mathcal{J}(\boldsymbol{\psi}_k) - \mathcal{J}(\boldsymbol{\psi}_{k+1}) + \Delta q \tag{3.15}$$

算法 3.2 哈密顿蒙特卡罗

输入：初始参数 ψ_0**1：对于** $k=0$、$\cdots$、N，**执行 2：从提议密度** $\mathcal{N}(\mathbf{0}, \boldsymbol{M})$**中抽取动量样本** $\boldsymbol{p}_k$ **3：计算** $\mathcal{H}(\psi_k, \boldsymbol{p}_k)$、$\frac{\partial \mathcal{J}}{\partial \psi_k}$、$\frac{\partial \mathcal{J}}{\partial \psi_{k+1}}$ **4：更新** ψ **（基于蛙跳算法）5：** $\boldsymbol{p}_{k+\frac{1}{2}} = \boldsymbol{p}_k - \frac{\Delta t}{2}\frac{\partial \mathcal{J}}{\partial \psi_k}$ **6：** $\psi_{k+1} = \psi_k + \Delta t \overline{\boldsymbol{M}^{-1}} \overline{\boldsymbol{p}_{k+\frac{1}{2}}}$ **7：** $\boldsymbol{p}_{k+1} = \boldsymbol{p}_{k+\frac{1}{2}} - \frac{\Delta t}{2}\frac{\partial \mathcal{J}}{\partial \psi_{k+1}}$ **8：计算** $\mathcal{H}(\bar{\psi}_{k+1}, \boldsymbol{p}_{k+1})$ **9：计算** $\alpha = \mathcal{H}(\psi_k, \boldsymbol{p}_k) - \mathcal{H}(\psi_{k+1}, \boldsymbol{p}_{k+1})$ **10：计算** $a = \min[1, \exp(\alpha)]$ **11：从** $\mathcal{U}([0,1])$**中抽取** u **12：如果** $u < a$，**则 13：接受：** $\overline{\psi_{k+1}}$ **14：否则 15：拒绝：设置** $\psi_{k+1} = \psi_k$ **16：结束 if 17：结束 for**

方程（3.15）的扩展形式，利用方程（3.6）可得：

$$\begin{aligned}\alpha =& \mathcal{J}(\psi_k) - \mathcal{J}(\psi_{k+1}) - \frac{1}{2}||\boldsymbol{H}_{\mathrm{MAP}}^{1/2}(\psi_k - \psi_{k+1} + \boldsymbol{H}_{\mathrm{MAP}}^{-1}\boldsymbol{g}_{k+1})||^2 \\ &+ \frac{1}{2}||\boldsymbol{H}_{\mathrm{MAP}}^{1/2}(\psi_{k+1} - \psi_k + \boldsymbol{H}_{\mathrm{MAP}}^{-1}\boldsymbol{g}_k)||^2 \\ =& \mathcal{J}(\psi_k) - \mathcal{J}(\psi_{k+1}) + (\psi_{k+1} - \psi_k)^\top(\boldsymbol{g}_k + \boldsymbol{g}_{k+1}) + \frac{1}{2}||\boldsymbol{H}_{\mathrm{MAP}}^{-1/2}\boldsymbol{g}_k||^2 - \frac{1}{2}||\boldsymbol{H}_{\mathrm{MAP}}^{-1/2}\boldsymbol{g}_{k+1}||^2\end{aligned}$$

3.4.2. MALA。根据方程（3.9），可推导出 MALA 的更新步骤：

$$\psi_{k+1} = \psi_k - \tau \boldsymbol{H}_{\mathrm{MAP}}^{-1}\boldsymbol{g}_k + \sqrt{(2\tau)}\boldsymbol{H}_{\mathrm{MAP}}^{-1/2}\boldsymbol{u} \tag{3.16}$$

其中，$\boldsymbol{u}$ 如上所定义。考虑到 $\sqrt{(2\tau)} = \overline{\Delta t,}$ 上述方程可重写为 s:

$$\psi_{k+1} = \psi_k - \frac{\Delta t^2}{2}\boldsymbol{H}_{\mathrm{MAP}}^{-1}\boldsymbol{g}_k + \Delta t \boldsymbol{H}_{\mathrm{MAP}}^{-1/2}\boldsymbol{u} \tag{3.17}$$

考虑方程（3.9），MALA方法接受准则的对数比值展开形式可推导为：

$$\begin{aligned}\alpha =& \mathcal{J}(\psi_k) - \mathcal{J}(\psi_{k+1}) - \frac{1}{4\tau}||\boldsymbol{H}_{\mathrm{MAP}}^{1/2}(\psi_k - \psi_{k+1} + \tau\boldsymbol{H}_{\mathrm{MAP}}^{-1}\boldsymbol{g}_{k+1})||^2 \\ &+ \frac{1}{4\tau}||\boldsymbol{H}_{\mathrm{MAP}}^{1/2}(\psi_{k+1} - \psi_k + \tau\boldsymbol{H}_{\mathrm{MAP}}^{-1}\boldsymbol{g}_k)||^2 \\ =& \mathcal{J}(\psi_k) - \mathcal{J}(\psi_{k+1}) + \frac{1}{2}(\psi_{k+1} - \psi_k)^\top(\boldsymbol{g}_k + \boldsymbol{g}_{k+1}) + \frac{\tau}{4}||\boldsymbol{H}_{\mathrm{MAP}}^{-1/2}\boldsymbol{g}_k||^2 - \frac{\tau}{4}||\boldsymbol{H}_{\mathrm{MAP}}^{-1/2}\boldsymbol{g}_{k+1}||^2\end{aligned} \tag{3.18}$$

3.4.3. H-HMC。类似地，利用最大后验点处的海森信息和蛙跳算法3.2，H-HMC方法的更新步骤可写为：

$$\begin{aligned} \boldsymbol{p}_{k+\frac{1}{2}} &= \boldsymbol{p}_k - \frac{\Delta t}{2}\boldsymbol{g}_k \\ \boldsymbol{\psi}_{k+1} &= \boldsymbol{\psi}_k + \Delta t \boldsymbol{H}_{\text{MAP}}^{-1}\boldsymbol{p}_{k+\frac{1}{2}} \\ \boldsymbol{p}_{k+1} &= \boldsymbol{p}_{k+\frac{1}{2}} - \frac{\Delta t}{2}\boldsymbol{g}_{k+1} \end{aligned} \tag{3.19}$$

其中，$\boldsymbol{p}_k$ 是一个来自均值为零、协方差为 $\boldsymbol{H}_{\text{MAP}}$的高斯分布的随机样本。因此，更新步骤如方程（3.17）所示。对于 H-HMC 方法，结合方程（3.10），对数比的扩展形式为：

$$\alpha = \mathcal{J}(\boldsymbol{\psi}_k) + \frac{1}{2}||\boldsymbol{H}_{\text{MAP}}^{-1/2}\boldsymbol{p}_k||^2 - \mathcal{J}(\boldsymbol{\psi}_{k+1}) - \frac{1}{2}||\boldsymbol{H}_{\text{MAP}}^{-1/2}\boldsymbol{p}_{k+1}||^2 \tag{3.20}$$

根据蛙跳算法（方程（3.19））， 可定义：

$$\begin{aligned} \boldsymbol{p}_k &= \frac{(\boldsymbol{\psi}_{k+1} - \boldsymbol{\psi}_k)^\top}{\Delta t}\boldsymbol{H}_{\text{MAP}} + \frac{\Delta t}{2}\boldsymbol{g}_k \\ \boldsymbol{p}_{k+1} &= \frac{(\boldsymbol{\psi}_{k+1} - \boldsymbol{\psi}_k)^\top}{\Delta t}\boldsymbol{H}_{\text{MAP}} - \frac{\Delta t}{2}\boldsymbol{g}_{k+1} \end{aligned} \tag{3.21}$$

将方程（3.21）代入方程（3.20），可得：

$$\alpha = \mathcal{J}(\boldsymbol{\psi}_k) - \mathcal{J}(\boldsymbol{\psi}_{k+1}) + \frac{1}{2}(\boldsymbol{\psi}_{k+1} - \boldsymbol{\psi}_k)^\top(\boldsymbol{g}_k + \boldsymbol{g}_{k+1}) + \frac{\Delta t^2}{8}||\boldsymbol{H}_{\text{MAP}}^{-1/2}\boldsymbol{g}_k||^2 - \frac{\Delta t^2}{8}||\boldsymbol{H}_{\text{MAP}}^{-1/2}\boldsymbol{g}_{k+1}||^2$$

这与 MALA 方法的接受准则（方程（3.18））相似。

下面总结基于海森矩阵的采样方法比较中的要点。上述分析表明，在 MALA 方法中适当选择正定矩阵 $\boldsymbol{B}$，并在 HMC 方法中适当选择质量矩阵 $\boldsymbol{M}$ 后，这两种方法的更新步骤和接受系数相同。在 HMC 方法中选择质量矩阵 $\boldsymbol{M}$以及在 MALA 方法中选择正定矩阵 $\boldsymbol{B}$，分别对 HMC 和 MALA 方法的调参及性能起着关键作用。在上述分析中，我们提出利用最大后验点处的海森信息来调整 HMC 和 MALA 方法，与标准采样方法相比，这可以显著提升这些方法的性能。对这三种基于海森矩阵的采样方法进行比较可知，尽管 SN-MAP 方法在高维高斯分布上的表现极佳，但采用可变步长的 H-HMC 和 MALA 方法在探索非高斯参数空间时可能表现更好。第5节将对这些方法进行数值比较。

4. 孔弹性正演模型。本节基于孔弹性的连续介质能量表述，介绍正向模型 [22], ，用于预测 p 土壤中的水孔隙压力 κ随渗透率变化的情况，并考虑可变形介质中的流体输运。按照第 2节的符号，观测值 i是含噪压力测量值 $y_i = p^i_{obs}$，不确定场为 $\theta(x) = -\log_e \kappa(x)$，其中 κ 的单位为 m^2。下文将介绍该模型的控制偏微分方程。

固体基质不可压缩的多孔介质，其控制方程可写为：

$$\frac{\mathrm{d}\,\mathrm{tr}(\boldsymbol{\varepsilon})}{\mathrm{d}t} + \frac{\mathrm{d}p}{\mathrm{d}t}\frac{\phi_f}{K_f} = -\frac{1}{\rho_f}\,\mathrm{div}\,\boldsymbol{Q} \tag{4.1}$$

$$\mathrm{div}\,\boldsymbol{\sigma} + \boldsymbol{b} = \boldsymbol{0} \tag{4.2}$$

其中，ϵ、ϕ_f、K_f和 γ分别为应变张量、流体体积分数、流体体积模量和流体的动力黏度。

通量向量 $\boldsymbol{Q}$定义为:

$$\boldsymbol{Q} = -\frac{\kappa}{\gamma}\rho_f \left(\nabla p + P_f \boldsymbol{g}\right) \tag{4.3}$$

其中 ρ_f和 P_f 分别为真实流体密度和流体密度，二者的关系为 $P_f = \phi_f \rho_f$。在方程 (4.2) 中， $\boldsymbol{\sigma}$为柯西应力张量，其定义如下:

$$\boldsymbol{\sigma} = (1-\phi_f)\big(2\mu\boldsymbol{\varepsilon} + \lambda\,\mathrm{tr}(\boldsymbol{\varepsilon})\boldsymbol{I}\big) - \phi_f p \boldsymbol{I} \tag{4.4}$$

其中 μ和 λ为拉梅常数，体力定义为 $\boldsymbol{b} := (P_s + P_f)\boldsymbol{g}$。其中 P_s为固相密度。

5. 数值结果与讨论。 本节展示第2-3节所述贝叶斯公式的应用结果，并通过不同实例比较基于海森矩阵的采样方法在处理孔隙压力数据以推断渗透率方面的表现。我们在第 5.1节比较高斯目标分布上的采样方法，在第 5.2节比较非高斯目标分布上的采样方法，并在第 4 节定义的、似然函数中嵌入耦合孔隙弹性模型的后验分布上进行比较，相关内容见第 5.3节。后一个分析通过 *FEniCS*实现，并使用 *dolfin-adjoint* 软件包中的拟牛顿求解器求解非线性最小化问题。

5.1. 高斯后验分布。在本例中，我们考虑一个解析高斯后验分布，其协方差矩阵为 $\boldsymbol{\Sigma}$，均值为 $\boldsymbol{m}$。我们采用了四种采样方法：(i) MH-MCMC、(ii) HMC、(iii) SN-MCMC 和 (iv) H-HMC。SN-MCMC方法的接受率为1，因为根据方程（3.5）和（3.15），可以证明，当分布为高斯分布时， q（$\boldsymbol{\psi}_k$， $\boldsymbol{z}$）与 ω（$\boldsymbol{z}$）相同。因此，对数比为:

$$\alpha = \omega(\boldsymbol{z}) - \omega(\boldsymbol{\psi}_k) + q(\boldsymbol{z}, \boldsymbol{\psi}_k) - q(\boldsymbol{\psi}_k, \boldsymbol{z}) = 0 \tag{5.1}$$

接受系数变为:

$$a = \min\Big\{1, \exp(\alpha)\Big\} = 1 \tag{5.2}$$

图1 比较了这四种采样方法在一维和高维高斯分布下的自相关与滞后的关系。对于一维分布，我们假设 $m = 0.5$ 和 $\sigma^2 = 2.0$。我们生成了2，000 个样本，所有采样过程均从 $\theta = 0.5$开始，MH-MCMC、HMC和H-HMC方法的步长为 $\Delta t = 1$。

对于高维问题，我们假设区域 Ω 离散化为 936 个网格点。均值和协方差矩阵由第5.3节求解的问题推导得出（协方差矩阵为最大后验点处海森矩阵的逆矩阵）。我们生成 10,000 个样本，且所有采样过程均从最大后验点开始（即 $\boldsymbol{\psi} = \boldsymbol{m}$）。MH-MCMC方法的步长为 $\Delta t = 0.01$，HMC方法的步长为 $\Delta t = 0.1$，H-HMC方法的步长假设为$\Delta t = 0.3$；采用这些步长时，MH-MCMC、HMC和H-HMC方法的接受率分别约为 85%、94%和 92%。此外，SN-MAP方法的接受率为 1（如方程（5.2）所示）。

我们注意到，在标准 HMC 和 MH-MCMC 采样方法中，为了保持较高的接受率并避免采样陷入低概率区域，需要采用较小的步长，这样效率低下且计算成本高昂。然而，使用 H-HMC 等加速采样方法时，可以在保持较高接受率的同时选用更大的步长，从而显著减少探索参数空间所需的样本数。

如图 1所示，对于基于海森矩阵的采样方法（SN-MCMC 和 H-HMC），无论在一维还是高维情形下，自相关都迅速趋于零，这表明与 MH-MCMC 和 HMC 方法相比，其收敛速度更快。此外，对比高维情形下的 H-HMC 和 HMC 方法可知，H-HMC 方法可以采用更大的步长而不降低接受率，从而加快收敛速度。此外，对比一维和

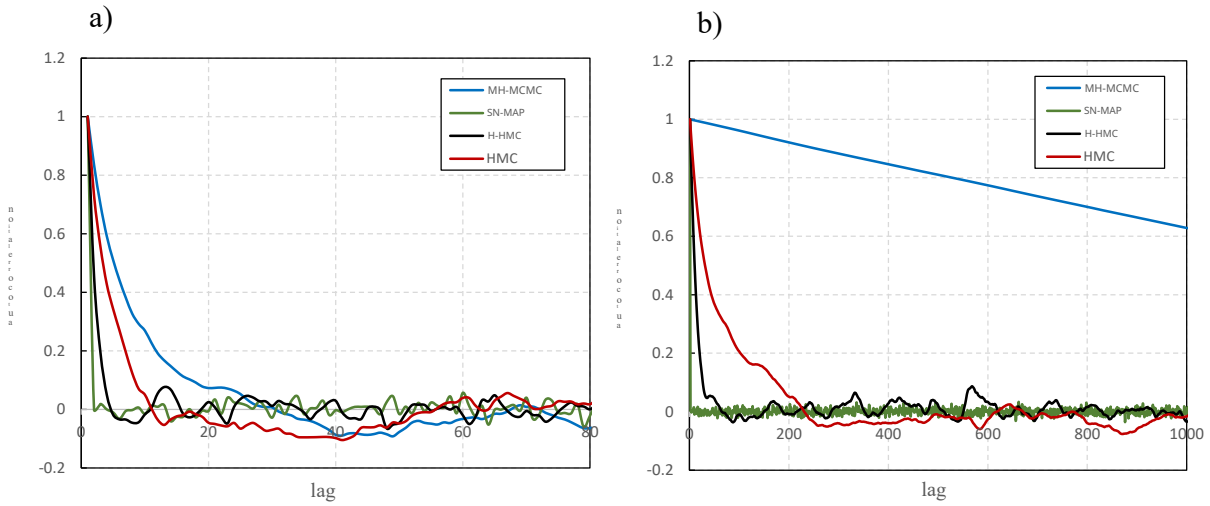

图1：a）一维情形和 b）高维情形下，不同采样方法的自相关与滞后关系；自相关是根据区域上的平均值计算的。

高维情形下的自相关函数可知，MH-MCMC 方法的性能会随着问题维数的增加而显著下降。对于高维情形，图 2展示了基于区域中虚线所示截面的 95% 可信区间和样本均值，结果基于 10,000 个样本。结果表明，SN-MAP 和 H-HMC 方法的可信区间和样本均值与解析解吻合良好，二者与解析解几乎无法区分。然而，MH-MCMC 和 HMC 等标准采样方法需要更多样本才能达到同等精度（图 2）。

$_\infty$表1a 通过相关时间总结了分析结果，相关时间 τ定义为 $\tau = 1 + 2\sum \rho_{t\circ}$ 其中 ρ_t 是$_{t=1}$滞后等于 t时两条链之间的相关性。 N_{eff} 是有效样本量，其定义为(N/τ)；SE 是标准误，定义为 $\tilde{\sigma}/N_{\text{eff}}$ ，其中 $\tilde{\sigma}$ 是样本链标准差。该表展示了

方法	τ	N_{eff}	SE
HMC	1149.3	9	0.0036
H-HMC	62.5	160	0.00027
SN-MAP	1	10,000	1.4×10^{-5}

(a) 高维正态分布不同采样方法的收敛性分析总结。

案例 #	σ_ϵ	测量值。
1	0.5 兆帕	60
2	2 兆帕	60
3	4 兆帕	60
4	4 兆帕	20
5	10 兆帕	20

（b）具有不同不确定性水平的案例，噪声水平（σ_ϵ）和测量值数量。

表1：高维正态示例的汇总分析。

SN-MAP采样方法的标准误最低，是针对高维正态分布效率最高的方法。在本分析中，我们不考虑MH-MCMC方法，因为经过10,000次迭代后，该方法的采样链仍远未收敛。

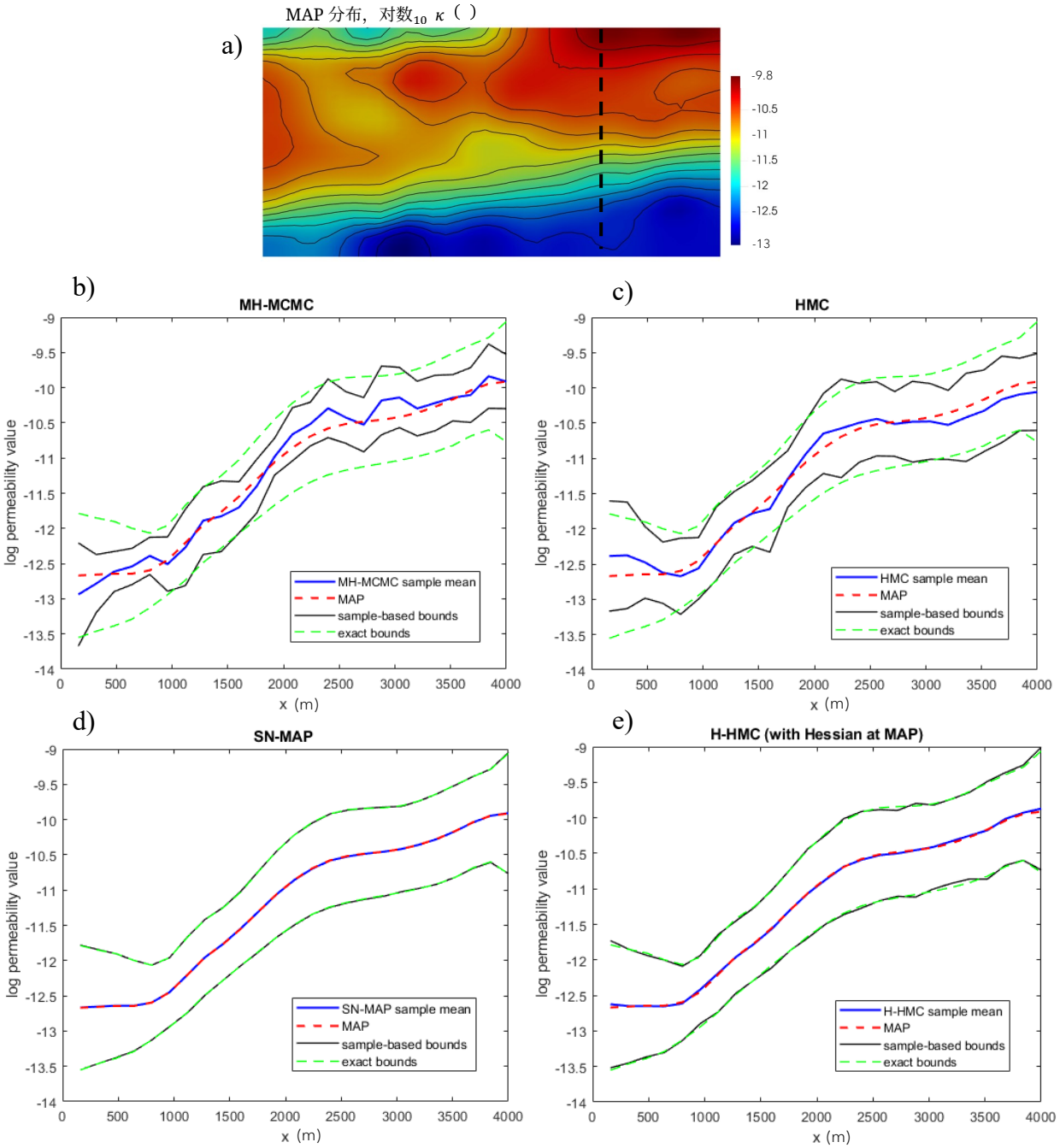

图2：a）最大后验点，b）MH-MCMC 的区间，c）HMC 的区间，d）SN-MAP 的区间，e）H-HMC方法的区间，适用于高斯后验分布。

表 1b 展示了 5 个不同案例，其不确定性水平各不相同，用于研究测量次数和噪声水平对可信区间幅度的影响。我们通过减少

测量值并提高噪声水平。对于每个案例，计算最大后验点及其处的海森矩阵。然后，考虑在最大后验点处近似后验分布的正态分布，我们使用 H-HMC方法抽取 10,000 个样本，并针对每个案例，计算区域内四个随机点处的 95% 可信区间。结果见图 3。正如预期，结果表明，随着不确定性水平提高，可信区间的振幅也随之增大（图 3）。

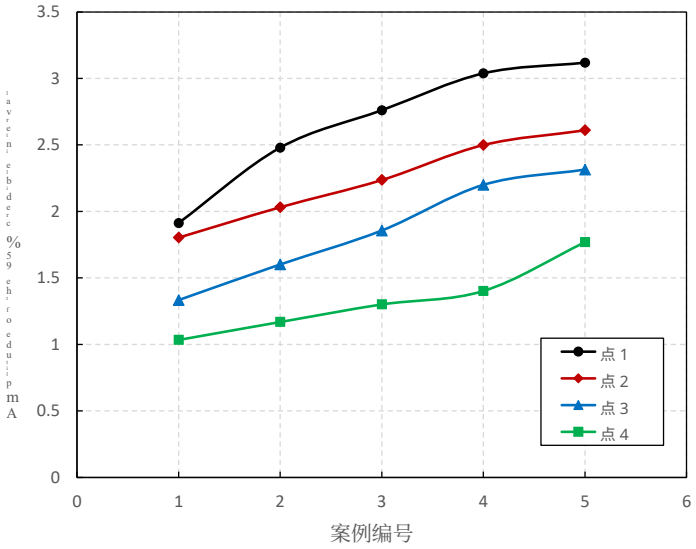

图 3：使用 H-HMC方法得到的高斯后验分布的 95%可信区间振幅与测量不确定性的关系。测点 1：(1371.4，0)，测点 2：(0，160)，测点 3：(914.3，320)，测点 4：(4114.3，1280)。

5.2. 近高斯后验分布。本节将采样方法应用于对数正态目标分布。目标函数定义如下：

$$\mathcal{J}(\boldsymbol{\psi}) = \frac{1}{2}||\boldsymbol{\Lambda}^{1/2}\big(\log(\boldsymbol{\psi}) - \boldsymbol{m}_l\big)||^2 - \log(\Pi_{i=1}^{N}\psi_i^{-1}) \tag{5.3}$$

其中 $\boldsymbol{\Lambda} = \boldsymbol{\Sigma}_l^{-1}$。 $\boldsymbol{\Sigma}_l$和 $\boldsymbol{m}_l$分别是对数正态分布的协方差矩阵和均值向量输入。目标函数的局部梯度和海森矩阵可推导如下：

$$\frac{\partial \mathcal{J}(\boldsymbol{\psi})}{\partial \psi_j} = \psi_j^{-1}\Big[\boldsymbol{n}_j^T\boldsymbol{\Lambda}\big(\log(\boldsymbol{\psi}) - \boldsymbol{m}_l\big) + 1\Big] \tag{5.4}$$

其中 $\boldsymbol{n}_j = \{n_j^i\}_{i=1}^N$ 是一个分量为 $n_j^i = 0, \quad i \neq j$ 和 $n_j^i = 1, \quad i = j$的向量。

$$\frac{\partial^2 \mathcal{J}(\boldsymbol{\psi})}{\partial \psi_j \partial \psi_i} = \begin{cases} -\psi_j^{-2}\big[\boldsymbol{n}_j^T\boldsymbol{\Lambda}\big(\log(\boldsymbol{\psi}) - \boldsymbol{m}_l\big) + 1 + \Lambda_{jj}\big] & \text{if } i = j \\ \psi_j^{-1}\big[\Lambda_{ji}\psi_i^{-1}\big] & \text{if } i \neq j \end{cases} \tag{5.5}$$

根据方程（5.4），最大后验点为 $\boldsymbol{\psi}_{\text{MAP}} = \exp(\boldsymbol{m}_l - \boldsymbol{\Sigma}_l\mathbf{1})$。我们讨论一维和高维对数正态分布的标准采样方法与加速采样方法的性能。在一维情形下，假设 $m_l = 0.5$ 且 $\sigma_l^2 = 2.0$。我们生成4,000 个样本，MH-MCMC、HMC 和 H-HMC 方法的步长为 $\Delta t = 1$。所有采样过程均从对数正态分布的众数$\exp(m_l - \sigma_l^2)$开始。

作为第二种设定，我们考虑高维对数正态分布。其区域和分布假设与第5.1节相同，最大后验点 $\boldsymbol{\psi}_{\text{MAP}} = \exp(\boldsymbol{m}_l - \boldsymbol{\Sigma}_l \mathbf{1})$ 假设与前一个示例的最大后验点相近（图2）。同样，MH-MCMC 方法的步长为 $\Delta t = 0.01$，HMC 方法的步长为 $\Delta t = 0.1$，H-HMC 方法的步长为 $\Delta t = 0.3$。MH-MCMC、HMC 和 H-HMC 方法分别在获得10,000 个样本后的对应接受率为82%、94% 和92%。此外，SN-MAP 方法的接受率约为8.3%，显著低于高维高斯分布的 SN-MAP 采样方法的接受率。

图4 展示了一维和高维情形下自相关函数随滞后的变化。结果表明，四种方法在一维分布中的表现相当接近。此外，如图 4 所示，在高维分布中，利用最大后验点处海森信息的 SN-MAP方法和 H-HMC方法，其收敛速度显著快于标准 MH-MCMC方法。我们还考虑了另一个示例，从对数正态分布中抽取样本，并假设最大后验点为$\boldsymbol{\psi}_{\text{MAP}} = \exp(\boldsymbol{m})$。其中， $\boldsymbol{m}$ 假设与正态分布示例的最大后验点相似。结果已在附录 E中报告。

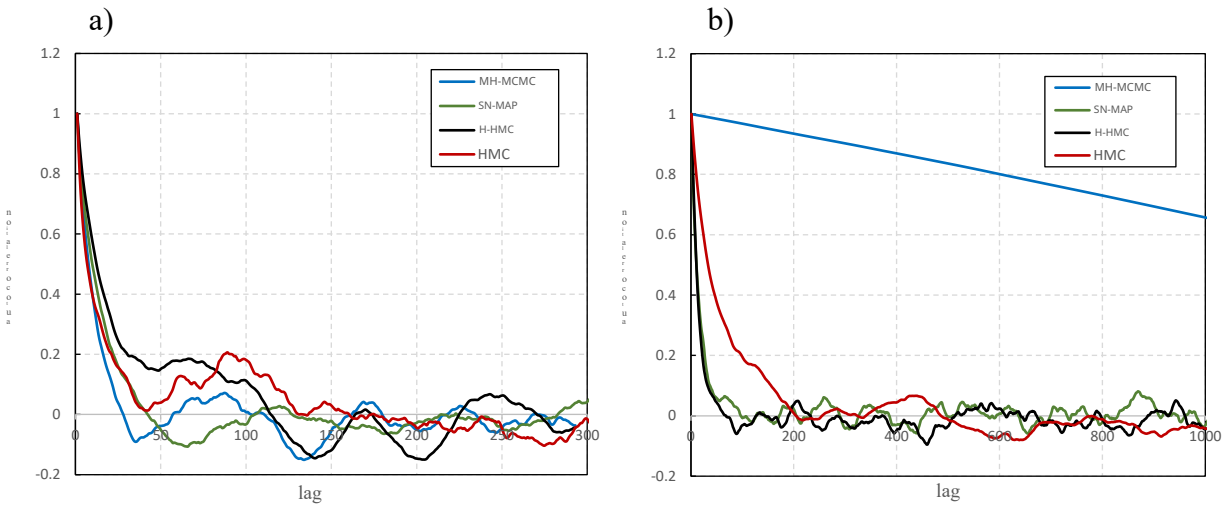

图 4：不同采样方法的自相关与滞后关系：a）一维情形，b）高维情形；自相关是根据空间区域上的平均值计算的。

图 5 展示了 MH-MCMC、HMC、SN-MAP 和 H-HMC 方法沿虚线的 95% 可信区间和样本均值。结果表明，HMC 方法比 MH-MCMC 方法收敛更快，并能更快地探索参数空间。H-HMC 和 SN-MAP 方法的收敛速度显著快于标准采样方法。经过 10,000 个样本后，H-HMC 和 SN-MAP 方法的样本可信区间均与精确可信区间吻合良好。

表 2a 展示了高维对数正态分布的收敛性分析摘要，该分析基于使用 MH-MCMC、HMC 和 H-HMC 采样方法获得的10,000 个样本。表 2a 的结果显示， N_{eff} 与对高维正态分布采用类似采样方法的结果（表 1a）相比，其数值有所降低，这证明从对数正态分布中采样更加困难。

此外，表 1a 和 2a 中的 SE 值比较表明，尽管 SN-MAP 方法对正态分布效果良好且估计误差最低，但对于对数正态分布，使用 H-HMC 方法效率更高，且估计误差更低。

表2b展示了 正态分布与对数正态分布之间的 Kullback-Leibler散度（Kullback–Leibler 散度）值，以及 d

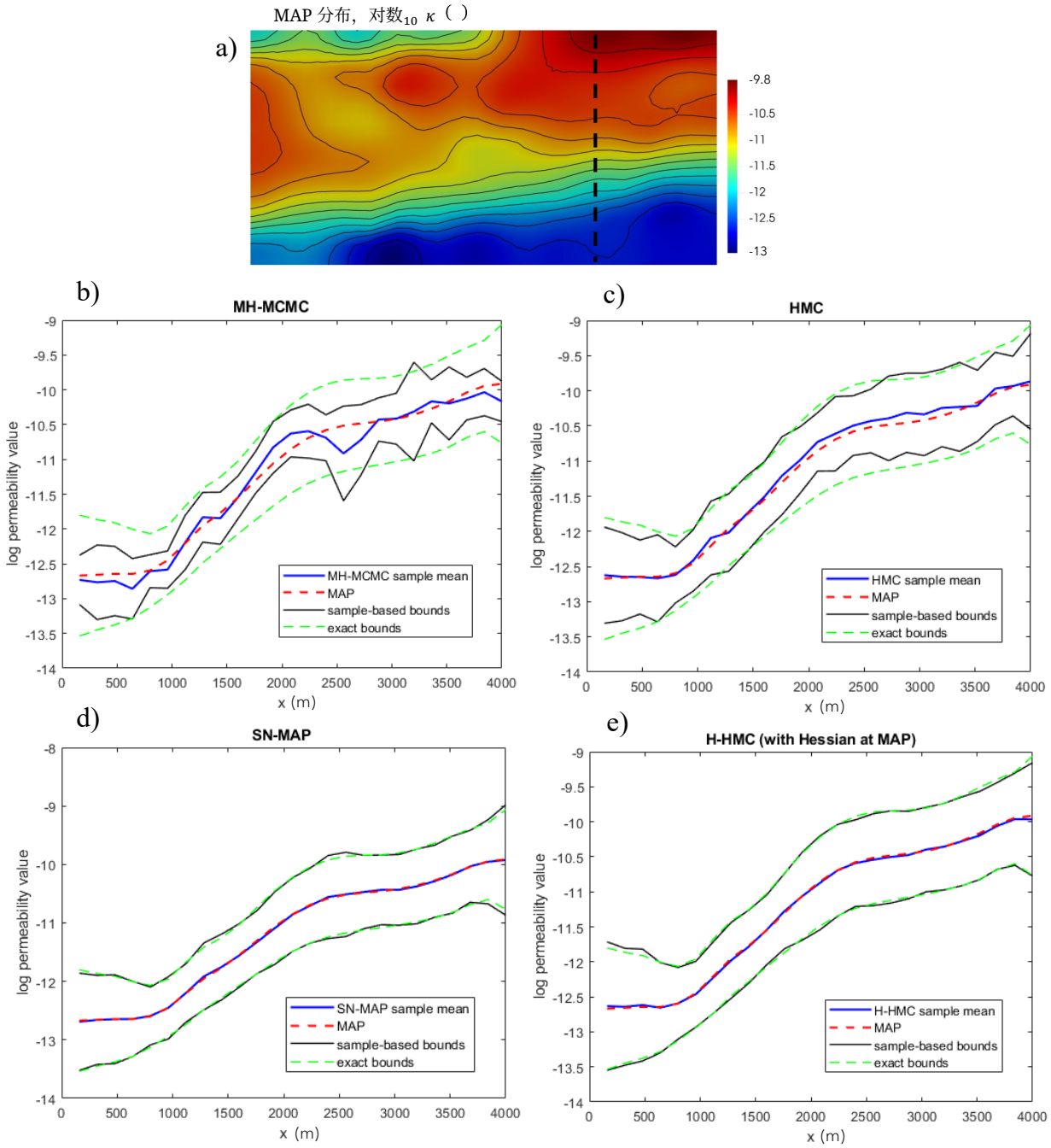

图 5：a）最大后验点，b）MH-MCMC 的区间，c）HMC 的区间，d）SN-MAP 的区间，e）H-HMC方法的区间，针对近高斯后验分布。

不同对数正态分布的接受率。此分析考察不同的对数正态分布，并计算 KLD 值，用以衡量各对数正态分布与其对应

方法	τ	N_{eff}	SE
HMC	1458.1	7	0.0062
H-HMC	70.7	141	0.0003
SN-MAP	87.7	114	0.0004

(a) 高维对数正态分布不同采样方法收敛性分析总结。

案例 #	KLD 值	接受率 %
1	2.47	8.9
2	3.15	8.3
3	5.04	7.3
4	13.55	4.3
5	69.03	1.7
6	195.02	0.3

(b) 不同对数正态分布的 SN-MAP采样方法结果。

表 2：高维对数正态示例的汇总分析。

在最大后验估计处进行正态近似。对于每个对数正态分布，我们运行 SN-MAP采样方法并获取10,000 个样本。结果表明，随着 KLD 值增加，接受率降低；这说明，对于与高斯分布不接近的分布，由于接受率较低，生成样本的成本更高。关于计算正态分布与对数正态分布之间 KLD 的详细信息见附录D。

5.3. 非高斯后验分布：根据压力数据推断渗透率。在本节中，我们根据稀疏的孔隙压力数据推断渗透率场 κ 。正向模型的偏微分方程控制方程已在第4节中描述。假定地下层的属性列于表3中。边界条件假定如下：

$$\text{top BC}: \ p = \text{const.} \ , \ \text{bottom} \quad \text{BC}: \ \boldsymbol{\delta} = \mathbf{0}$$

假设初始孔隙度为 $\phi_{0f} = 0.2$。我们考虑一个 $8,000 \times 4{,}000$ 深层地下层，其顶部边界的流体压力恒定为 $p = 500$ 兆帕 。图 6 展示了假设的（隐藏的）渗透率场 κ，用于

属性	数值
固相拉梅常数， λ	40 兆帕
固相拉梅常数， μ	40 兆帕
流体体积模量， K_f	2270 兆帕
水密度， ρ_f	1000 千克/米^3
水的黏度， γ	0.001 帕 · 秒
固体密度， ρ_s	2000 千克/米^3

表3：孔弹性模型中固相和流体相的性质，参见第 4 节。

通过运行正向模型生成人工测量值。在本例中，我们生成了52个孔隙压力的合成观测值，这些观测值受到标准差为 $\sigma_\epsilon = 1.0$ 兆帕的随机噪声影响。假设观测值位于注入开始后10天这一单一时间范围内。

如第 2 节所述，先验协方差基于拉普拉斯算子 $\mathcal{A}^{-1} = -\gamma \boldsymbol{\Psi} \boldsymbol{\Delta} + \delta \boldsymbol{I}$ 定义，其中 $\gamma = 0.5$, $\delta = 5 \times 10^{-3}$ （1/千米^2）， $a = 0.018$， $b = 0.97$，以及 $\beta = 1.017\pi$ （先验参数定义于第 2.1 节）。

5.3.1. 计算最大后验点。为求得最大后验点，我们采用拟牛顿算法（BFGS）求解非线性最小化问题。拉格朗日方法的详细表述见附录C。假设先验均值为常数（$m_\pi = 33\ \text{m}^2$）。我们从场的初始猜测开始，采用牛顿法迭代更新场值。图7 展示了最大后验点（其与实际渗透率相似）

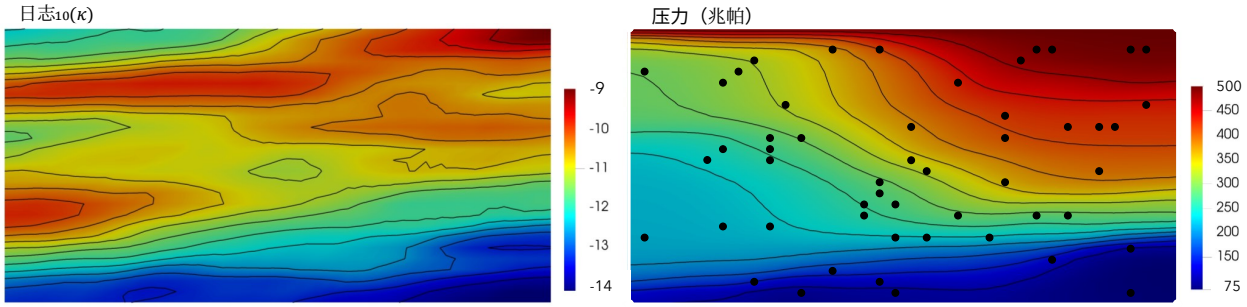

图6：（左）目标渗透率分布 $\log_{10}(\kappa)$，（右）注入开始10天后的合成逐点压力 p 观测值。

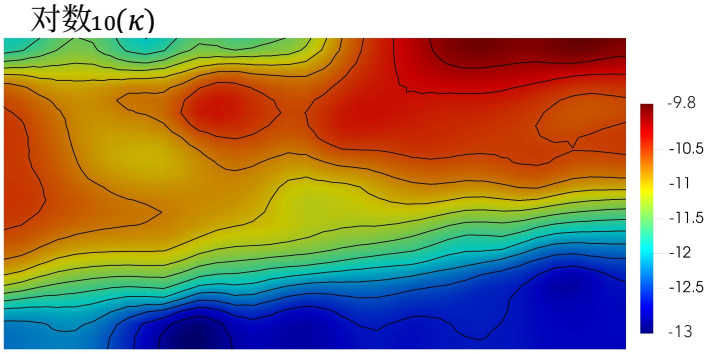

图7： $\sigma_\epsilon = 1.0$ 兆帕下的最大后验点场。

如图6所示）。图8展示了考虑不同噪声水平时识别出的最大后验点。随着噪声水平 σ_ϵ增加，先验项变得更加占主导地位，因此最大后验场更加平滑。同样，图9展示了测量值数量对计算最大后验点的影响。观测值数量减少时，最大后验点所受信息约束更少，因而更加平滑，也更偏离目标分布。

5.3.2. 探索后验分布。我们使用 HMC 和 H-HMC 方法从后验分布中生成样本。在 H-HMC 方法中，我们利用最大后验点处的海森信息来提高采样性能。为获得相近的接受率，HMC 和 H-HMC 方法的步长 Δt 分别设为0。1和0。3。HMC 采集20,000 个样本后，HMC 和 H-HMC 方法对应的接受率分别约为85% 和83 %。图10 展示了使用 HMC 和 H-HMC 方法计算的区域内三个点处渗透率值的自相关函数随滞后的变化。

在本分析中，由于 MH-MCMC方法效率低且计算成本高，我们没有采用该方法。此外，由于后验分布高度非线性且呈非高斯，SN-MAP方法的接受率很低，因此效率不高。此外，图 11 展示了使用 H-HMC方法获得 50,000 个样本后，反问题解沿垂直虚线的可信区间。图中还展示了先验样本的均值和 95% 可信区间。比较先验区间和后验区间可以看出，后验方差显著减小，表明观测值对后验分布有很大影响。此外，图 11 还展示了后验样本的均值、最大后验点和目标值。结果表明，后验样本的均值与最大后验点吻合良好。

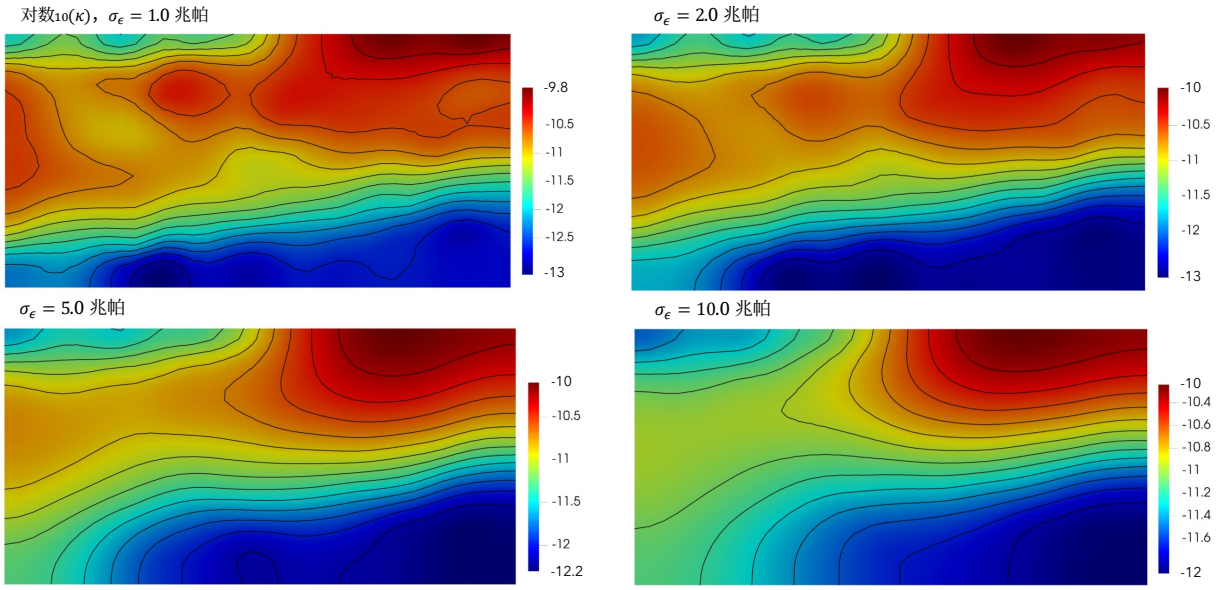

图 8：基于 52 个观测值、考虑不同噪声水平的最大后验场。

6. 结论。本文提出了一种用于由偏微分方程控制的高维反问题的贝叶斯推断框架。我们采用基于类拉普拉斯算子的连续高斯先验分布，以确保无限维反问题的适定性。这种先验函数的主要优点是可应用于连续区域，并且便于进行离散化。我们在*FEniCS* 库中实现了该反问题，并使用*dolfin-adjoint* 软件包中的拟牛顿求解器（BFGS）求解最小化问题。

我们研究了多种用于描述后验分布的采样方法，探讨了高维参数空间中采样方法的复杂性，并比较了 MH-MCMC 和 HMC 采样方法与利用海森信息和梯度信息加速的方法（如 SN-MCMC、MALA 和 H-HMC）的性能。通过考察若干具有高斯和非高斯后验分布的一维及高维问题，我们表明，采用改进的基于海森矩阵的采样方法，可以显著加快收敛速度，并提高在高维逆问题中的探索效率。然而，在非线性高维问题中使用局部海森信息的计算成本很高。为解决这一问题，我们考虑了使用在最大后验点处计算的海森信息的海森矩阵采样方法。结果表明，基于最大后验估计的海森矩阵采样方法在计算效率和探索高维分布的速度方面均表现出色。

我们还将所开发的框架应用于一个由孔隙弹性偏微分方程控制的高维逆问题，从逐点孔隙压力观测值中推断未知的渗透率分布。我们采用 HMC 和 H-HMC 方法，并利用最大后验点处的海森信息，计算了最大后验点和后验可信区间。结果表明，使用最大后验点处海森信息的 H-HMC 方法，在探索这一非高斯高维分布方面表现更佳。

软件可用性。本研究开发的代码版本可在以下网址获取：https://github.com/minakari/Bayesian。

致谢。 我们感谢美国国家科学基金会通过匹兹堡超级计算中心提供的 XSEDE 资源给予支持。MINA KARIMI 感谢 Scott Institute 提供的经费支持。KAUSHIK DAYAL 感谢美国国家科学基金会（CMMI MOMS 1635407、DMS 2108784）、美国陆军研究办公室（多学科大学研究计划 W911NF- 19-1-0245）、美国海军研究办公室（N00014-18-1-2528）、美以双边科学基金会（2018183）提供的经费支持，以及受聘于国家能源技术

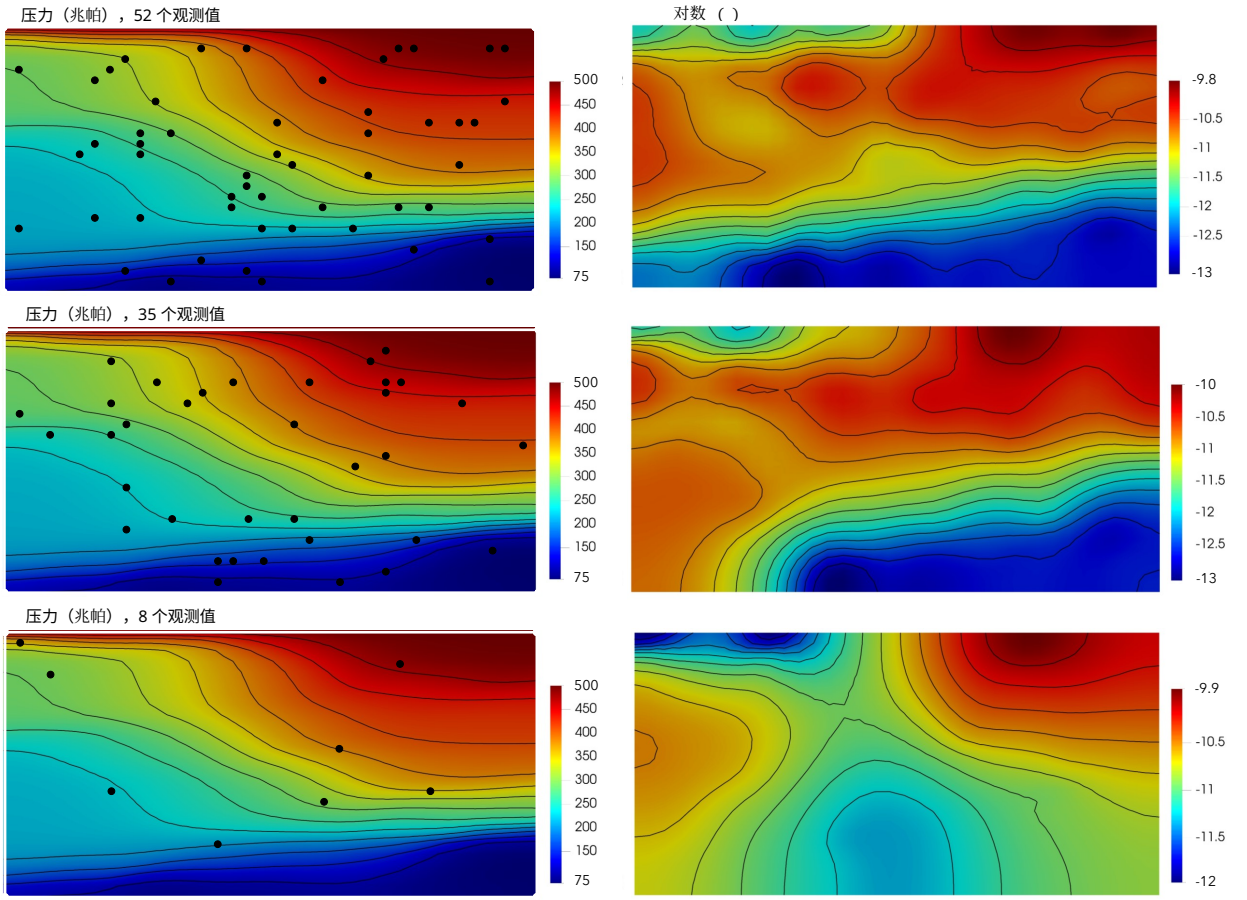

图 9：不同测量次数下的最大后验场，单位为 $\sigma_\epsilon = 1$兆帕。

实验室（由美国能源部资助）。MATTEO POZZI 感谢美国国家科学基金会（CMMI 1638327）提供的经费支持。本研究（部分）由卡内基梅隆大学工程学院的 Dowd 奖学金资助。

附录 A. 构建无限维高斯先验。

无限维区域中的高斯先验场可以通过 Karhunen-Loeve 展开定义 [11],如下：

$$\theta = m_\pi + \sum_{j=1}^{\infty} \lambda_j \xi_j \tilde{\phi}_j \tag{A.1}$$

其中 $\{\phi_j\}_{j=1}^{\infty}$ 是区域的一组正交归一基函数，$\{\lambda_j\}_{j=1}^{\infty}$ 是一个确定性序列，且 $\{\xi_j\}_{j=1}^{\infty}$是一组相互独立且同分布的正态随机变量 $\xi_j \sim \mathcal{N}(0,\ 1)$。利用方程（A.1），协方差算子

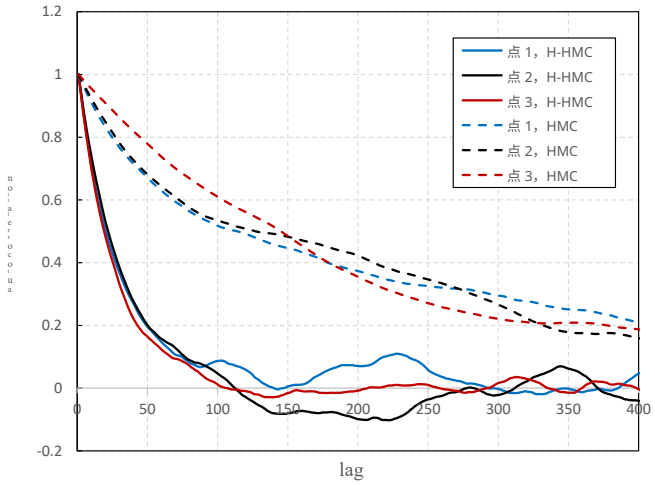

图10：3个随机点处的自相关与滞后关系：点1：(7314.3,800)，点2：(5028.8,2880)，点3：(10,3200)。

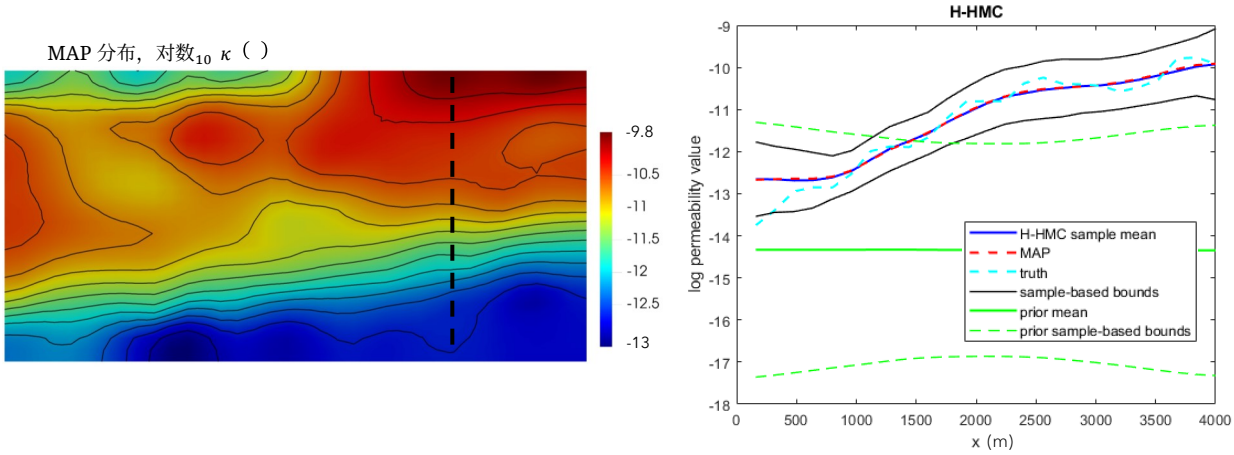

图11：基于 H-HMC方法的95%可信区间。

方程（2.8）可写为：

$$\mathcal{C}_\pi = \mathbb{E}\left(\sum_{j=1}^{\infty}\sum_{k=1}^{\infty}\lambda_j\lambda_k\xi_j\xi_k\tilde{\phi}_j\otimes\tilde{\phi}_k\right) = \left(\sum_{j=1}^{\infty}\sum_{k=1}^{\infty}\lambda_j\lambda_k\,\mathbb{E}(\xi_j\xi_k)\tilde{\phi}_j\otimes\tilde{\phi}_k\right)$$
$$= \sum_{j=1}^{\infty}\sum_{k=1}^{\infty}\lambda_j\lambda_k\delta_{jk}\tilde{\phi}_j\otimes\tilde{\phi}_k = \sum_{j=1}^{\infty}\lambda_j^2\tilde{\phi}_j\otimes\tilde{\phi}_j$$

˜根据集合 $\{\phi_j\}_{j=1}^{\infty}$的正交归一性，我们可以写出：

$$\mathcal{C}_\pi\tilde{\phi}_k = \Big(\sum_{j=1}^{\infty}\lambda_j^2\tilde{\phi}_j\otimes\tilde{\phi}_j\Big)\tilde{\phi}_k = \sum_{j=1}^{\infty}\lambda_j^2\langle\tilde{\phi}_j,\tilde{\phi}_k\rangle\tilde{\phi}_j = \sum_{j=1}^{\infty}\lambda_j^2\delta_{jk}\tilde{\phi}_k = \lambda_k^2\tilde{\phi}_k$$

因此，上述计算表明，$\big(\lambda_j^2,\ \tilde{\phi}_j\big)_{j=1}^{\infty}$是协方差算子 $\mathcal{C}_\pi$的特征对。要为随机场 θ构造方程（A.1）中的先验，我们需要指定均值函数 m_π 和协方差算子 $\mathcal{C}_\pi$。

附录 B. 低秩海森矩阵。

目标函数（方程（2.17））关于控制参数的海森矩阵（即二阶导数）可写为：

$$\boldsymbol{H} = \boldsymbol{H}_{\text{misfit}} + \boldsymbol{A}^{-1} \tag{B.1}$$

其中，$\boldsymbol{H}_{\text{misfit}}$ 是目标函数第一项的海森矩阵，该项表示观测值与预测值之间的失配；$\boldsymbol{A}$ 是先验项的海森矩阵。我们可以将先验协方差进行乔列斯基分解，表示为 $\boldsymbol{A}=\boldsymbol{L}\boldsymbol{L}^\top$。利用该分解，我们可以将方程（B.1）改写为 $\boldsymbol{H}=\boldsymbol{L}^{-\top}\big(\boldsymbol{L}^\top\boldsymbol{H}_{\text{misfit}}\boldsymbol{L}+\boldsymbol{I}\big)\boldsymbol{L}^{-1}$。

其中，第一项 $\boldsymbol{L}\boldsymbol{H}_{\text{misfit}}\boldsymbol{L}$ 对控制参数提供了更多信息，其 $r-$ 维低秩近似可写为 $\boldsymbol{L}^\top\boldsymbol{H}_{\text{misfit}}\boldsymbol{L}\approx\boldsymbol{V}_r\boldsymbol{D}_r\boldsymbol{V}_r^\top$。其中，$\boldsymbol{V}_r$ 是主导特征向量的 $n\times r$ 矩阵，$\boldsymbol{D}_r$ 是主导特征值的 $r\times r$ 对角矩阵。因此，$\boldsymbol{H}$ 的低秩近似可写为 $\tilde{\boldsymbol{H}}=\boldsymbol{L}^{-\top}\big(\boldsymbol{V}_r\boldsymbol{D}_r\boldsymbol{V}_r^\top+\boldsymbol{I}\big)\boldsymbol{L}^{-1}$。

附录 C. 拉格朗日公式。

We minimize the objective (负对数后验)进行优化以找到最大后验点。在本例中，目标函数定义为 $\mathcal{J}(\theta)=\dfrac{1}{2\sigma_\epsilon^2}\sum_{i=1}^{n}(p(\theta)-p_{obs}^i)^2+\dfrac{1}{2}||\mathcal{A}(\theta-\theta_{pr})||^2$。

其中，$p(k)$、$\{p_{obs}\}_{i=1}^{N}$和 σ_ϵ^2 分别为孔隙压力的预测值、压力测量值和噪声协方差。θ_{pr} 为先验均值。此外，拉格朗日形式可以写成目标函数与控制方程弱形式之和：

$$\begin{aligned}\mathcal{L}(\theta,p,\tilde{p},\boldsymbol{\epsilon},\tilde{\boldsymbol{u}}) = \mathcal{J}(\theta) &+ \int_0^\top\int_\Omega(\mathrm{tr}(\boldsymbol{\epsilon})_t+p_t)\tilde{p}\,\mathrm{d}\boldsymbol{x}\,\mathrm{d}t + \int_0^\top\int_\Omega\frac{e^{-\theta}}{\gamma}(\nabla p+\boldsymbol{b})\cdot\nabla\tilde{p}\,\mathrm{d}\boldsymbol{x}\,\mathrm{d}t\\ &- \int_0^\top\int_{\partial\Omega}\frac{e^{-\theta}}{\gamma}(\nabla p+\boldsymbol{b})\cdot\boldsymbol{n}\tilde{p}\,\mathrm{d}\boldsymbol{x}\,\mathrm{d}t + \int_0^\top\int_\Omega(\boldsymbol{b}\cdot\tilde{\boldsymbol{u}}-\boldsymbol{\sigma}\cdot\tilde{\boldsymbol{u}})\,\mathrm{d}\boldsymbol{x}\,\mathrm{d}t\end{aligned}$$

其中，$\tilde{p}$ 和 $\tilde{\boldsymbol{u}}$ 分别是质量守恒方程和动量守恒方程的拉格朗日乘子。$(\cdot)_t$表示对时间求导，即 $\mathrm{d}(\cdot)\,\mathrm{d}t$ 。求解最小化问题的最优性条件（Karush-Kuhn-Tucker条件）可通过计算相对于拉格朗日函数变量的变分得到。我们使用了 *dolfin-adjoint* Python库来计算最优性条件并找到最大后验点。

附录D. 正态分布与对数正态分布之间的Kullback-Leibler散度 **s.**

我们假设对数正态分布是真实分布或精确测得的分布，而正态分布是对数正态分布的近似分布。

$$p(\boldsymbol{x}) = \mathcal{N}(\boldsymbol{x}|\boldsymbol{m}_p,\boldsymbol{\Sigma}_p) = c\,|\boldsymbol{\Sigma}_p|^{-1/2}\exp\Big[-\frac{1}{2}||\boldsymbol{\Sigma}_p^{-1/2}\big(\boldsymbol{x}-\boldsymbol{m}_p\big)||^2\Big]$$

$$q(\boldsymbol{x}) = \log\mathcal{N}(\boldsymbol{x}|\boldsymbol{m}_q,\boldsymbol{\Sigma}_q) = c\,|\boldsymbol{\Sigma}_q|^{-1/2}\exp\Big[-\frac{1}{2}||\boldsymbol{\Sigma}_q^{-1/2}\big(\log(\boldsymbol{x})-\boldsymbol{m}_q\big)||^2\Big]\Pi_{i=1}^{n}x_i^{-1}$$

其中 $c=(2\pi)^{-n/2}$ ，$\boldsymbol{m}_q$ 和 $\boldsymbol{\Sigma}_q$ 分别是 $\log(\boldsymbol{x})$ 的均值和协方差。因此，KL -

散度定义为：

$$D_{\mathrm{KL}}(q(\boldsymbol{x})||p(\boldsymbol{x})) = \int q(\boldsymbol{x}) \log \frac{q(\boldsymbol{x})}{p(\boldsymbol{x})} = \mathbb{E}_q \left[\log \big(q(\boldsymbol{x})\big) - \log \big(p(\boldsymbol{x})\big) \right]$$

$$\log \big(q(\boldsymbol{x})\big) = -\frac{n}{2} \log(2\pi) - \frac{1}{2} \log |\boldsymbol{\Sigma}_q| - \frac{1}{2} ||\boldsymbol{\Sigma}_q^{-1/2} \big(\log(\boldsymbol{x}) - \boldsymbol{m}_q \big)||^2 - \sum_{i=1}^{n} \log(x_i)$$

$$\log \big(p(\boldsymbol{x})\big) = -\frac{n}{2} \log(2\pi) - \frac{1}{2} \log |\boldsymbol{\Sigma}_p| - \frac{1}{2} ||\boldsymbol{\Sigma}_p^{-1/2} (\boldsymbol{x} - \boldsymbol{m}_p)||^2$$

$$\mathbb{E}_q \left[\log \big(q(\boldsymbol{x})\big) \right] = -\frac{n}{2} \log(2\pi) - \frac{1}{2} \log |\boldsymbol{\Sigma}_q| - \frac{1}{2} \operatorname{tr}(\boldsymbol{I}_{n\times n}) - \sum_{i=1}^{n} m_q^i$$

$$\mathbb{E}_q \left[\log \big(p(\boldsymbol{x})\big) \right] = -\frac{n}{2} \log(2\pi) - \frac{1}{2} \log |\boldsymbol{\Sigma}_p| - \frac{1}{2} \mathbb{E}_q \left[||\boldsymbol{\Sigma}_p^{-1/2} (\boldsymbol{x} - \boldsymbol{m}_p)||^2 \right]$$

其中，我们可以用 $\mathrm{tr}\{(\boldsymbol{x} - \boldsymbol{m}_p)^\top \boldsymbol{\Sigma}_p^{-1} (\boldsymbol{x} - \boldsymbol{m}_p)\}$ 代替 $(\boldsymbol{x} - \boldsymbol{m}_p)^\top \boldsymbol{\Sigma}_p^{-1} (\boldsymbol{x} - \boldsymbol{m}_p)$，它可以重写为 $\mathrm{tr}\{(\boldsymbol{x} - \boldsymbol{m}_p)(\boldsymbol{x} - \boldsymbol{m}_p)^\top \boldsymbol{\Sigma}_p^{-1}\}$，并且可以证明：

$$\begin{aligned} \frac{1}{2} \mathbb{E}_q \left[||\boldsymbol{\Sigma}_p^{-1/2} (\boldsymbol{x} - \boldsymbol{m}_p)||^2 \right] &= \frac{1}{2} \mathbb{E}_q \left[\operatorname{tr} \left\{ (\boldsymbol{x} - \boldsymbol{m}_p)(\boldsymbol{x} - \boldsymbol{m}_p)^\top \boldsymbol{\Sigma}_p^{-1} \right\} \right] \\ &= \frac{1}{2} \operatorname{tr} \left\{ \mathbb{E}_q \left[(\boldsymbol{x} - \boldsymbol{m}_p)(\boldsymbol{x} - \boldsymbol{m}_p)^\top \right] \boldsymbol{\Sigma}_p^{-1} \right\} \end{aligned}$$

其中 $\mathbb{E}_q \left[(\boldsymbol{x} - \boldsymbol{m}_p)(\boldsymbol{x} - \boldsymbol{m}_p)^\top \right] = \tilde{\boldsymbol{\Sigma}}_q + (\tilde{\boldsymbol{m}}_q - \boldsymbol{m}_p)(\tilde{\boldsymbol{m}}_q - \boldsymbol{m}_p)^\top$。因此，$\mathbb{E}_q \left[\log \big(p(\boldsymbol{x})\big) \right]$ 可推导为：

$$\mathbb{E}_q \left[\log \big(p(\boldsymbol{x})\big) \right] = -\frac{n}{2} \log(2\pi) - \frac{1}{2} \log |\boldsymbol{\Sigma}_p| - \frac{1}{2} \operatorname{tr}\{(\tilde{\boldsymbol{\Sigma}}_q + (\tilde{\boldsymbol{m}}_q - \boldsymbol{m}_p)(\tilde{\boldsymbol{m}}_q - \boldsymbol{m}_p)^\top) \boldsymbol{\Sigma}_p^{-1}\}$$

KL 散度为：

$$D_{\mathrm{KL}}\big(q(\boldsymbol{x})||p(\boldsymbol{x})\big) = -\frac{1}{2} \log \frac{|\boldsymbol{\Sigma}_q|}{|\boldsymbol{\Sigma}_p|} - \frac{n}{2} - \sum_{i=1}^{n} m_q^i + \frac{1}{2} \operatorname{tr} \left\{ \big(\tilde{\boldsymbol{\Sigma}}_q + (\tilde{\boldsymbol{m}}_q - \boldsymbol{m}_p)(\tilde{\boldsymbol{m}}_q - \boldsymbol{\mu}_p)^\top\big) \boldsymbol{\Sigma}_p^{-1} \right\}$$

其中 $\tilde{\boldsymbol{\Sigma}}_q$和 $\tilde{\boldsymbol{m}}_q$ 是 $\boldsymbol{x}$ 在对数正态分布中的协方差和均值，可推导如下：

$$\tilde{\Sigma}_q^{ij} = \exp \left[m_q^i + m_q^j + 0.5 \left(\Sigma_q^{ii} + \Sigma_q^{jj} \right) \right] \big(\exp(\Sigma_q^{ij}) - 1 \big)$$

$$\tilde{\boldsymbol{m}}_q = \exp \left(\boldsymbol{m}_q + 0.5 \boldsymbol{\Sigma}_q \mathbf{1} \right)$$

附录 E. 高维对数正态分布的采样。

In this section, we consider a high-dimensional log-normal distribution, i.e., $\psi \sim \log \mathcal{N}(\boldsymbol{m}, \boldsymbol{\Sigma})$. 我们假设为第一个示例 1（第 5.1 节）中的协方差和均值。由于 ψ 值的范围较小，对数正态分布不定义负值和零值；从该分布中抽取样本较为复杂。为解决这一问题，我们通过添加常向量 $\boldsymbol{m}' = \boldsymbol{m} + c\mathbf{1} \Rightarrow \boldsymbol{\psi}' \sim \log \mathcal{N}(\boldsymbol{m} + \mathbf{c1}, \boldsymbol{\Sigma})$ 来增大均值。

我们采用 MH-MCMC、HMC 和 H-HMC（使用最大后验点处的海森矩阵）来探索对数正态分布。MH-MCMC 和 HMC方法的步长设为 $\Delta t = 1$，H-HMC方法的步长设为 $\Delta t = 0.01$，这些方法的接受率均约为 98%。

我们进行 5×10^5 次采样来探索对数正态分布；由于 $\boldsymbol{\psi}'$ 涵盖了很宽的取值范围，MH-MCMC和HMC方法的收敛速度明显低于H-HMC方法。将海森矩阵信息用作H-HMC方法中的质量矩阵，可以调整采样过程，使其更快地沿不同方向（众数）探索该分布。即使经过 5×10^5 次采样，MH-MCMC和HMC方法仍远未收敛。然而，H-HMC采样链在经过 2×10^5 次采样后便收敛。

参考文献

[1] A. Alghamdi、M. A. Hesse、J. Chen和O. Ghattas，《基于InSAR地表变形数据的贝叶斯孔弹性含水层表征。第一部分：最大后验估计》，arXiv预印本 arXiv:2002.10706，（2020）。[2] S. Arridge、J. Kaipio、V. Kolehmainen、M. Schweiger、E. Somersalo、T. Tarvainen和M. Vauhkonen，《近似误差与模型降阶及其在光学扩散层析成像中的应用》，《反问题》，22（2006），第175页。[3] M. Betancourt，《黎曼流形哈密顿蒙特卡罗的一般度量》，载于《信息几何科学国际会议》，Springer，2013年，第327–334页。[4] M. Betancourt，《哈密顿蒙特卡罗概念导论》，arXiv预印本 arXiv:1701.02434，（2017）。[5] M. Broccardo、A. Mignan、F. Grigoli、D. Karvounis、A. P. Rinaldi、L. Danciu、H. Hofmann、C. Milkereit、T. Dahm、G. Zimmermann等，《冰岛Geldinganes地热井水力刺激诱发地震风险分析》，《自然灾害与地球系统科学》，20（2020），第1573–1593页。[6] T. Bui-Thanh、O. Ghattas、J. Martin和G. Stadler，《无穷维贝叶斯逆问题的计算框架，第一部分：线性化情形及其在全球地震反演中的应用》，《工业与应用数学学会计算科学杂志》，35（2013），第A2494–A2523页。[7] T. Bui-Thanh和M. Girolami，《使用黎曼流形哈密顿蒙特卡罗求解大规模偏微分方程约束贝叶斯逆问题》，《反问题》，30（2014），第114014页。[8] T. Bui-Thanh、K. Willcox和O. Ghattas，《用于非定常空气动力学应用概率分析的参数化降阶模型》，《AIAA杂志》，46（2008），第2520–2529页。[9] N. Cressie和C. K. Wikle，《时空数据统计学》，John Wiley & Sons，2015年。[10] T. Cui、J. Martin、Y. M. Marzouk、A. Solonen和A. Spantini，《非线性逆问题的似然信息降维》，《反问题》，30（2014），第114015页。[11] M. Dashti和A. M. Stuart，《逆问题的贝叶斯方法》，arXiv预印本 arXiv:1302.6989，（2013）。[12] H. W. Engl、M. Hanke和A. Neubauer，《逆问题正则化》，第375卷，Springer Science & Business Media，1996年。[13] A. Fichtner、A. Zunino和L. Gebraad，《哈密顿蒙特卡罗求解弱非线性逆问题的教程式介绍》，（2018）。[14] S. W. Funke和P. E. Farrell，《自动化偏微分方程约束优化框架》，arXiv预印本 arXiv:1302.3894，（2013）。[15] D. Galbally、K. Fidkowski、K. Willcox和O. Ghattas，《用于大规模逆问题不确定性量化的非线性模型降阶》，《国际工程数值方法杂志》，81（2010），第1581–1608页。[16] L. Gebraad、C. Boehm和A. Fichtner，《使用哈密顿蒙特卡罗的贝叶斯弹性全波形反演》，《地球物理研究杂志：固体地球》，125（2020），第e2019JB018428页。[17] A. Gelman、J. B. Carlin、H. S. Stern、D. B. Dunson、A. Vehtari和D. B. Rubin，《贝叶斯数据分析》，CRC出版社，2013年。[18] M. Girolami和B. Calderhead，《黎曼流形朗之万与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（2011），第123–214页。[19] P. K. Jha、L. Cao和J. T. Oden，《利用多组分混合理论连续体模型对得克萨斯州COVID-19演变进行贝叶斯预测》，《计算力学》，66（2020），第1055–1068页。[20] P. K. Jha和J. T. Oden，《目标导向的后验模型误差估计对参数估计的辅助作用》，《计算物理学杂志》，（2022）。[21] P. K. Jha、C. Walker、D. Mitchell、J. T. Oden、D. Schellingerhout、J. A. Bankson和D. T. Fuentes，《基于互信息的超极化 ^{13}C-丙酮酸MRI最优实验设计》，arXiv预印本 arXiv:2206.12509，（2022）。[22] M. Karimi、M. Massoudi、N. Walkington、M. Pozzi和K. Dayal，《大变形孔弹性的能量表述》，《国际岩土力学数值与解析方法杂志》，46（2022），第910–932页。[23] K. M. Keranen、H. M. Savage、G. A. Abers和E. S. Cochran，《美国俄克拉何马州可能由人为活动诱发的地震：废水注入与2011年震级5.7地震序列之间的联系》，《地质学》，41（2013），第699–702页。[24] K. M. Keranen和M. Weingarten，《诱发地震活动》，《地球与行星科学年度评论》，46（2018），第149–174页。[25] C. Langenbruch、M. Weingarten和M. D. Zoback，《基于物理学的俄克拉何马州和堪萨斯州人为地震灾害预测》，《自然通讯》，9（2018），第1–10页。[26] Y. T. Lee和S. S. Vempala，《黎曼流形哈密顿蒙特卡罗的收敛率与更快速的多面体体积计算》，载于《第50届ACM SIGACT计算理论年会论文集》，2018年，第1115–1121页。[27] Z. Li、M. Dong、S. Li和S. Huang，《枯竭油气藏中的二氧化碳封存——盖层表征与储存容量》，《能源转换与管理》，47（2006），第1372–1382页。[28] J. Martin、L. C. Wilcox、C. Burstedde和O. Ghattas，《用于大规模统计逆问题的随机牛顿马尔可夫链蒙特卡罗方法及其在地震反演中的应用》，《工业与应用数学学会计算科学杂志》，34（2012），第A1460–A1487页。[29] K. P. Murphy，《机器学习：概率视角》，MIT出版社，2012年。[30] R. M. Neal等，《使用哈密顿动力学的马尔可夫链蒙特卡罗》，《马尔可夫链蒙特卡罗手册》，2（2011），第2页。[31] R. K. Pandit和D. Infield，《基于不同平稳协方差函数的高斯过程功率曲线模型对比分析及其在提高模型精度方面的应用》，《可再生能源》，140（2019），第190–202页。[32] D. Patsialis、A. P. Kyprioti和A. A. Taflanidis，《面向地震工程应用的滞回降阶结构模型贝叶斯校准》，《工程结构》，224（2020），第111204页。[33] N. Petra、J. Martin、G. Stadler和O. Ghattas，《无穷维贝叶斯逆问题的计算框架，第二部分：随机牛顿马尔可夫链蒙特卡罗及其在冰盖流动逆问题中的应用》，《工业与应用数学学会计算科学杂志》，36（2014），第A1525–A1555页。[34] N. Petra和G. Stadler，《由偏微分方程支配的模型变分逆问题》，技术报告，得克萨斯大学位于

奥斯汀计算工程与科学研究所，*2011*年。[35] *Y. Qi* 和 *T. P. Minka*，基于海森矩阵的马尔可夫链蒙特卡罗算法，（*2002*）。[36] *G. O. Roberts* 和 *J. S. Rosenthal*，朗之万扩散离散近似的最优尺度，皇家统计学会杂志：*B*辑（统计方法学），*60*（*1998*），第*255–268*页。[37] *G. O. Roberts and R. L. Tweedie*，朗之万分布及其离散近似的指数收敛性，*Bernoulli*，（*1996*），第*341–363*页。[38] *M. Schoenball* 和 *W. L. Ellsworth*，系统评估俄克拉何马州和堪萨斯州南部诱发地震活动导致的断层激活时空演化，地球物理研究杂志：固体地球，*122*（*2017*），第*10–189*页。[39] *P. Segall* 和 *S. Lu*，注入诱发地震活动：孔弹性与地震成核效应，地球物理研究杂志：固体地球，*120*（*2015*），第*5082–5103*页。[40] *S. A. Shapiro*，流体诱发地震活动，剑桥大学出版社，*2015*年。[41] *D. F. Sumy*、*C. J. Neighbors*、*E. S. Cochran* 和 *K. M. Keranen*，*2011*年俄克拉何马州普拉格*Mw 5.7*级地震余震的应力降较低，地球物理研究杂志：固体地球，*122*（*2017*），第*3813–3834*页。[42] *A. Tarantola*，模型参数估计的逆问题理论与方法，工业与应用数学学会，*2005*年。[43] *F. Uribe*、*I. Papaioannou*、*W. Betz* 和 *D. Straub*，利用 *Karhunen–Loève*展开表示的随机场的贝叶斯推断，计算机方法：应用力学与工程，*358*（*2020*），第*112632*页。[44] *F. Uribe*、*I. Papaioannou*、*J. Latz*、*W. Betz*、*E. Ullmann* 和 *D. Straub*，应用于 *Karhunen–Loève*展开的变化维度子集模拟贝叶斯推断，*2020*年。[45] *U. Villa*、*N. Petra* 和 *O. Ghattas*，*hippylib*：用于偏微分方程控制的大规模逆问题的可扩展软件框架；第一部分：确定性反演与线性化贝叶斯推断，*arXiv*预印本 *arXiv:1909.03948*，（*2019*）。[46] *P. Wang*、*M. Pozzi*、*M. J. Small* 和 *W. Harbert*，检测与废水注入相关的地震活动率变化的统计方法，美国地震学会公报，*105*（*2015*），第*2852–2862*页。[47] *P. Wang*、*M. J. Small*、*W. Harbert* 和 *M. Pozzi*，一种评估与废水注入相关的地震活动转变的贝叶斯方法，美国地震学会公报，*106*（*2016*），第*832–845*页。[48] *Z. Wang*、*S. Mohamed* 和 *N. Freitas*，自适应哈密顿蒙特卡罗与黎曼流形蒙特卡罗，载于《国际机器学习会议》，*2013*年，第*1462–1470*页。[49] *Y. Zhang* 和 *C. A. Sutton*，马尔可夫链蒙特卡罗的拟牛顿方法，载于《神经信息处理系统进展》，*2011*年，第*2393–2401*页。