arXiv:2301.03644v1 [math.NA] 9 Jan 2023

反问题中海森矩阵的分层非对角低秩近似及其在冰盖模型初始化中的应用

Tucker Hartland[1], **Georg Stadler**[2], **Mauro Perego**[3], **Kim Liegeois**[3], **No´emi Petra**[1]
[1] 应用数学系，加州大学默塞德分校

电子邮件：thartland@ucmerced.edu[2] 纽约大学库朗数学科学研究所[3] 桑迪亚国家实验室计算研究中心

摘要。 在偏微分方程（PDE）支配的反问题中，获得轻量且准确的海森矩阵作用近似，对于使确定性和贝叶斯统计大规模反问题在计算上变得可行至关重要。通过利用分层非对角低秩（HODLR）矩阵近似，可将稠密线性代数程序（例如从高斯提议分布中采样以及使用直接线性方法进行牛顿求解所需的程序）的计算复杂度从 $\mathcal{O}\left(N^3\right)$ 降低至对数线性复杂度。在本文中，我们证明，一类源自偏微分方程支配的反问题的海森矩阵可以用 HODLR 矩阵格式很好地近似。具体而言，我们研究了描述冰盖瞬时黏性流动的偏微分方程支配反问题。在这些问题中，我们旨在求得一个空间分布的基底滑动参数场，使冰盖模型预测的流动与冰盖表面速度观测一致。我们展示了 HODLR 近似的应用：通过高效生成海森矩阵近似，实现从高斯化后验提议分布中快速生成样本。我们开展了计算研究，说明在某些冰盖问题情形下，使用 HODLR 矩阵格式近似高斯-牛顿数据失配海森矩阵，比使用低秩（LR）格式更高效。随后，我们研究洪堡冰川和格陵兰冰盖上斯托克斯流模型支配的反问题所对应的数据失配海森矩阵，并据此证明，对于大规模问题，与全局低秩近似相比，HODLR 近似可能更具优势。

1. 引言

基于模型的复杂物理系统模拟对于理解现实世界中的现象具有重要作用。这些模型通常由偏微分方程（PDE）描述，并且往往受到未知系数场、本构定律、源项、初始条件和/或边界条件、几何形状等因素所带来的不确定性影响。当存在观测数据时，可以通过求解由底层模型（例如偏微分方程）支配的反问题来估计这些参数。众所周知，不确定性是反问题的基本特征，因此，除了推断感兴趣的参数外，我们还需要量化与此推断相关的不确定性。可以通过贝叶斯推断实现这种不确定性量化。求解由复杂偏微分方程支配的贝叶斯反问题极具挑战性，这是因为将无限维参数场离散化会产生高维参数空间，而且还需要反复求解底层偏微分方程。为克服这些计算挑战，在可能的情况下，利用问题结构至关重要。例如，底层偏微分方程解算子通常具有扩散性，而观测数据可能较为稀疏，或仅包含有关参数场的有限信息。这些特点使反问题目标函数的数据失配项（或负对数似然）的二阶导数呈现低秩结构，下文将其称为数据失配海森矩阵。在先前的工作 [1, 2] 中，我们在冰盖流动反问题中利用了这种低秩结构。然而，对于这种“低秩”实际上很大的情况——许多具有实际意义的反问题都属于此类，因为观测数据包含的信息非常丰富——低秩近似并不足够。本文利用模型预测对参数的局部敏感性，由此形成非对角低秩结构。为此，我们采用分层非对角低秩（HODLR）矩阵近似，并详细说明如何利用它们降低求解大规模基于偏微分方程的反问题的计算成本。

相关工作 反问题中的海森矩阵全局低秩近似已得到成功应用， [1, 3, 4, 5, 6], 并且已有确定性和随机化方法 [7, 5]可用于生成此类近似。然而，有些问题，尤其是观测数据包含大量信息的问题，并不适合采用全局低秩近似，因此需要采用其他利用结构特征的策略，例如基于局部平移不变性和局部化灵敏度的策略 [8, 9, 10]。本文重点研究分层低秩方法，这类方法已有便捷的随机化方法可用 [11, 12]。

研究表明，在 [13, 14] 中，层次矩阵是近似大规模高斯过程相关协方差矩阵的有效方法。在 [15],中，利用具有通用分层划分模式的层次矩阵近似，构建海森矩阵逆的显式表示。在所研究的一个示例中，作者发现，参数到偏微分方程解映射的扩散率

以及观测数据的信息量会影响数据失配海森矩阵更适合采用层次格式还是全局低秩格式进行压缩。本文在这项研究的基础上，聚焦于陆地冰建模中一个特定的反问题。

贡献 本工作的主要贡献如下。(1) 我们阐述了在偏微分方程支配的反问题中，对数据失配海森矩阵进行 HODLR压缩的动机，并针对大规模冰盖反问题（例如格陵兰冰盖）进行了详细研究。(2) 我们介绍了一种利用 HODLR矩阵快速运算的策略，以高效生成高斯后验分布的近似样本，用于不确定性量化。(3) 我们通过数值研究，考察了各种问题设置对数据失配海森矩阵非对角低秩结构的影响。结果表明，HODLR近似在不同问题规模下均有效，包括一个参数维数离散化为 3.2×10^5 的格陵兰冰盖反问题。

2. 预备知识

本节概述无限维反问题离散化求解的背景知识，并简要回顾 HODLR矩阵。具体而言，我们定义 HODLR矩阵，列举其若干性质，并概述如何计算对称算子的对称 HODLR 矩阵近似的计算复杂度，其中这些算子只能通过作用于向量来获取。关于层次矩阵的更深入讨论，请参阅 [16, 17] ；关于 HODLR矩阵的更多细节，请参阅 [12]。

2.1. 贝叶斯反问题

在参数推断中处理不确定性的一种方法，是采用贝叶斯方法解决反问题 [18, 19, 20], 。该方法以观测数据 $\boldsymbol{d}$为输入，即数据、参数的先验知识，以及一个关于给定 β时数据似然的模型。离散化参数 $\boldsymbol{\beta}$ 的先验知识通常由领域科学家的专业知识确定，并以概率密度函数 $\pi_{\text{prior}}(\boldsymbol{\beta})$的形式进行数学编码。似然 $\pi(\boldsymbol{d}|\boldsymbol{\beta})$涉及数据不确定性以及参数到观测量过程的数学模型。贝叶斯反问题的解是离散化参数$\boldsymbol{\beta}$的概率密度函数，该函数根据贝叶斯公式以观测数据为条件

$$\pi_{\text{post}}(\boldsymbol{\beta}) = \pi(\boldsymbol{\beta}|\boldsymbol{d}) \propto \pi_{\text{prior}}(\boldsymbol{\beta})\,\pi(\boldsymbol{d}|\boldsymbol{\beta}),$$

它给出了后验分布的形式表达式。这里，“$\propto$”表示在归一化常数意义下相等。对于高斯先验 $\mathcal{N}\left(\boldsymbol{\beta},\ \boldsymbol{\Gamma}_{\text{prior}}\right)$

以及由零均值高斯分布 $\mathcal{N}(\mathbf{0},\ \boldsymbol{\Gamma}_{\text{noise}})$ 描述的数据噪声 $\boldsymbol{\eta}$，$\pi_{\text{post}}(\cdot)$ 具有以下形式

$$\pi_{\text{post}}(\boldsymbol{\beta}) \propto \exp\left(-\frac{1}{2}\|\boldsymbol{\mathcal{F}}(\beta)-\boldsymbol{d}\|^2_{\boldsymbol{\Gamma}^{-1}_{\text{noise}}} - \frac{1}{2}\|\boldsymbol{\beta}-\overline{\boldsymbol{\beta}}\|^2_{\boldsymbol{\Gamma}^{-1}_{\text{prior}}}\right), \tag{1}$$

其中 $\boldsymbol{\mathcal{F}}$ 是参数到观测量映射。记号 $\|\cdot\|_{\boldsymbol{A}}$ 表示该范数由正定矩阵 $\boldsymbol{A}$加权，即 $\|\boldsymbol{v}\|_{\boldsymbol{A}}=\sqrt{\boldsymbol{v}^\top\boldsymbol{A}\boldsymbol{v}}$。参数到偏微分方程解的映射通常是非线性的，因此后验分布不是高斯分布。后验分布的一个特征是其取得最大值的点，或者等价地，使负对数后验最小的点，即所谓的最大后验（MAP）点，

$$\boldsymbol{\beta}^\star := \arg\min_{\boldsymbol{\beta}} J(\boldsymbol{\beta}) := \frac{1}{2}\|\boldsymbol{\mathcal{F}}(\beta)-\boldsymbol{d}\|^2_{\boldsymbol{\Gamma}^{-1}_{\text{noise}}} + \frac{1}{2}\|\boldsymbol{\beta}-\overline{\boldsymbol{\beta}}\|^2_{\boldsymbol{\Gamma}^{-1}_{\text{prior}}}. \tag{2}$$

计算最大后验点的一种方法是采用（高斯）牛顿法进行优化 [21], ，该方法的关键在于能否获得（高斯-牛顿）海森矩阵。由于 J 是根据参数到观测量映射隐式定义的，而该映射涉及偏微分方程解算子，因此我们利用伴随方法 [22, 23, 24] 来计算其梯度及海森矩阵作用。

为了充分探索后验分布，可以采用马尔可夫链蒙特卡罗（MCMC）技术 [25, 26] 。这类技术需要一种提议分布，理想情况下该分布能够近似后验分布，且易于采样。一种生成高斯提议分布的方法是对后验在 $\boldsymbol{\beta}_k$ （或最大后验点）处进行拉普拉斯近似

$$\tilde{\pi}_{\text{post}}(\boldsymbol{\beta},\boldsymbol{\beta}_k) \propto \exp\left(-\frac{1}{2}\langle\boldsymbol{\beta}-\boldsymbol{\mu}_k, \boldsymbol{H}_k(\boldsymbol{\beta}-\boldsymbol{\mu}_k)\rangle_{\ell^2}\right), \boldsymbol{\mu}_k = \boldsymbol{\beta}_k - \boldsymbol{H}_k^{-1}\boldsymbol{g}_k,$$

其中，$\boldsymbol{g}_k$、$\boldsymbol{H}_k$ 分别是对数后验 $J(\boldsymbol{\beta})$ 在 $\boldsymbol{\beta}_k$ 处的梯度和海森矩阵。另一种 MCMC 采样方法是广义预条件 Crank-Nicholson（gpCN）方法 [27, 28]。预条件子的一个理想选择是最大后验点处的海森矩阵， [29]。

对于这些及其他 MCMC 采样器，通常需要反复且高效地应用海森矩阵的逆 $\boldsymbol{H}_k^{-1}$ 或其平方根 $\boldsymbol{H}_k^{-1/2}$ ，这也促使了本文所述研究的开展。具体而言，在第 3.2 节中，我们将讨论如何利用 HODLR 近似快速应用海森矩阵平方根。

2.2. 对称 *HODLR* 矩阵

HODLR 矩阵 $\boldsymbol{A}\in\mathbb{R}^{N\times N}$ 是一种矩阵，它具有深度为 $L\in\mathbb{N}$ 的层次划分，将指标集 $\mathcal{I}=\{1,2,\ldots,N\}$ 划分为连续子集，并根据该划分定义低秩非对角块。相关内容详见例如 [12]。图1展示了不同层次深度下 HODLR 矩阵的块秩结构。HODLR 矩阵还必须满足另外两个性质。

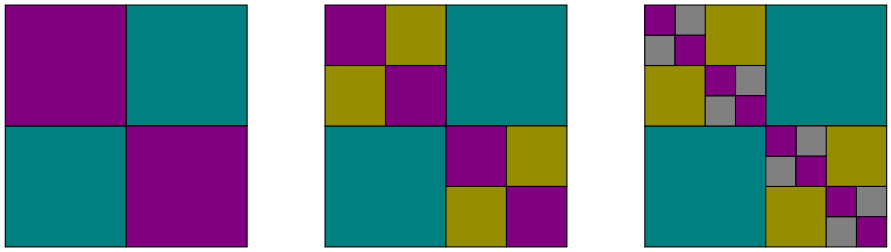


图1：具有不同层次深度的矩阵的秩结构： $\boldsymbol{A}$ $L = 1$ （左）、 $L = 2$（中）和 $L = 3$ （右）。假设非对角块为低秩。

(i) 分层划分的深度随矩阵大小的对数变化，即，

$$L = \mathcal{O}\left(\log\ N\right).$$

（ii）对于每个层次级别 ℓ ，其非对角块 r_ℓ的最大秩有一个上界 r ，该上界与问题规模 N无关 ℓ

$$\max_{1\leq\ell\leq L} r_\ell \leq r = \mathcal{O}\left(1\right).$$

由于低秩块使这类矩阵能够以少于 $\mathcal{O}$ （N^2）个浮点数进行计算表示，因此称其为数据稀疏矩阵。具体而言，存储一个 HODLR 矩阵需要 $\mathcal{O}$（Nlog N），计算一次 HODLR 矩阵向量乘积 [7], 需要 $\mathcal{O}$（Nlog N）次浮点运算；而使用直接方法计算逆 HODLR 矩阵向量乘积 [30]，以及平方根和逆平方根矩阵向量乘积 [31]，则需要 $\mathcal{O}$（N log^2 N）次浮点运算。

压缩我们的目标是生成反问题中数据失配海森矩阵的 HODLR 近似。对于大规模问题，数据失配海森矩阵只能以无矩阵算子的形式获得。为了构造对称无矩阵算子的 HODLR 近似，我们采用了先前开发的随机线性代数算法。这些算法只要求计算算子作用于有限个具有指定零元素的随机向量的结果，这类向量称为结构化随机向量。海森矩阵作用于这些结构化随机向量的结果用于采样非对角海森子矩阵的行空间和列空间，并对上述非对角子矩阵进行随机近似截断奇异值分解。更多详情请参见附录中的算法 2。

对于我们在第 5 节中展示的结果，采用了一种秩自适应的对称矩阵无关 [32, 33], 分层压缩算法，该算法基于 [12]。 [34], 中提出了一种类似算法，其中分层划分更加通用，且低秩块具有嵌套基。秩自适应性提供了一种高概率方法，可将非对角块解析到所需的精度。利用现有的矩阵向量乘积信息和瑞利商，可以实现一种秩自适应相对容差算法。

生成 *HODLR* 近似的计算成本 将对称矩阵压缩为层级为 L 、非对角秩为 $\{r_\ell\}_{\ell=1}^{L}$ 的 HODLR 矩阵时，使用 d 个过采样向量所需的矩阵向量乘积次数 ζ 为

$$\zeta = 2\left(\langle r\rangle + d\right) L + N/2^{L}, \text{ where } \langle r\rangle := \frac{1}{L}\sum_{\ell=1}^{L} r_\ell. \tag{3}$$

从附录 8.1 中的算法 2 可以理解方程 3：对于每个层级ℓ ，需要计算 $r_\ell + d$ 个海森矩阵向量积，以计算 $\boldsymbol{Y}$（算法 2 第 7 行）；还需要计算 $r_\ell + d$ 个海森矩阵向量积，以计算 $\boldsymbol{Z}$ （算法 2 第 14 行）。其余的 $N/2^{L}$ 个海森矩阵向量积用于确定对角子块，详情见 [7]。我们注意到，若采用自适应过程来确定近似基 $\boldsymbol{Q}$，例如 [33], 中用于分块矩阵列空间的方法，则计算成本可降至 $\zeta_{\text{adaptive}} = 2\left(\langle r\rangle + d/2\right) L + N/2^{L}$，但需要额外调用正交化例程，从而增加计算负担。我们注意到，要生成具有 HODLR 结构的矩阵的 HODLR近似，需要 $\zeta = \mathcal{O}(\log\ N)$ 次矩阵向量乘积。对于足够大的问题，HODLR压缩预计不会比全局低秩（LR）压缩更具计算效率，因为 $\zeta^{\text{LR}} = r + d$，单遍算法 [12] 使用 d 个过采样向量生成秩为 r 的压缩时，所需的矩阵向量乘积次数与问题规模无关。不过，对于规模相当大的问题，我们观察到 HODLR 格式确实能节省计算量（见第 6 节）。

3. 偏微分方程支配的反问题中的 HODLR 矩阵

这里，我们说明偏微分方程支配的反问题中的数据失配海森矩阵为何可能包含数值低秩的非对角块，描述如何对参数进行置换以显露这种 HODLR 结构，并展示如何利用 HODLR 近似从贝叶斯后验分布的高斯近似中抽取样本。

3.1. 动机

考虑以下数据失配代价泛函

$$J_{\text{misfit}}\left(\beta\right) := \frac{1}{2}\|\boldsymbol{\mathcal{F}}(\beta) - \boldsymbol{d}\|^2_{\boldsymbol{\Gamma}^{-1}_{\text{noise}}}, \quad \text{with} \quad \mathcal{F}(\beta) = \boldsymbol{\mathcal{B}}u,$$

其中，$\boldsymbol{\mathcal{B}}$ 将空间分布参数场 β对应的偏微分方程解 $u = u(\beta)$ 线性映射为与数据 $\boldsymbol{d}$ 相关的模型预测。此外， $\boldsymbol{\Gamma}_{\text{noise}}$是描述观测数据中高斯噪声的协方差矩阵。为便于说明，我们假设参数函数 β 定义在区域Γ_1 上，而数据 $\boldsymbol{d}$ 在区域 Γ_2上观测，这两个区域可以相同，也可以不同。这些量通过控制偏微分方程的解和测量算子 $\boldsymbol{\mathcal{B}}$相互关联。这种关系的特征取决于控制偏微分方程的性质。下文假设，空间（或时间）上的局部扰动

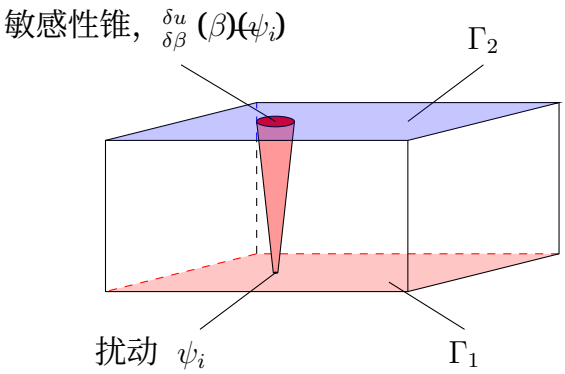


图2：示意图展示了参数 β的变化对 Γ_2 中的偏微分方程解 u 的影响集中在一个小区域内的情形。为说明这一点，我们展示了一个敏感性锥，即偏微分方程解 u 主要在局部化参数扰动支撑集周围的锥形区域内受到影响。

β 场中的扰动会对偏微分方程解 u产生主要局限于局部的影响，进而影响模型预测 $\boldsymbol{\mathcal{B}}u$。图2展示了这一性质，其中我们用敏感性锥来说明在 Γ_1上定义的 β中的局部扰动对 Γ_2中偏微分方程解 u 的影响。众所周知，对于椭圆型偏微分方程，局部扰动会对解产生全局影响；但根据区域几何和方程的不同，这种全局影响可能会在 Γ_2 的某个子集之外迅速衰减，而该子集涵盖了扰动的主要影响。例如，在图2所示的问题中，β 中的扰动对 u 的影响可能会随着Γ_1 与 Γ_2 之间距离的减小而变得更加局部化。

接下来，我们讨论上文所述偏微分方程的性质与海森矩阵（或其高斯–牛顿变体）中的非对角块之间的关系。数据失配海森矩阵，即代价泛函中数据失配部分的海森矩阵，可以通过伴随方法 [22, 23, 24]推导得到。然而，我们发现，通过研究其关于一阶和二阶敏感度 $\delta u/\delta\beta$, $\delta^2 u/\delta\beta^2$的形式表达式，最容易看出数据失配海森矩阵的 HODLR 结构。

$$\frac{\delta^2}{\delta\beta^2} J_{\text{misfit}}(\beta)(\beta_1,\beta_2) = (\boldsymbol{\mathcal{B}}u-\boldsymbol{d})^\top \boldsymbol{\Gamma}_{\text{noise}}^{-1}\left(\boldsymbol{\mathcal{B}}\frac{\delta^2 u}{\delta\beta^2}(\beta)(\beta_1,\beta_2)\right) + \left(\boldsymbol{\mathcal{B}}\frac{\delta u}{\delta\beta}(\beta)(\beta_1)\right)^\top \boldsymbol{\Gamma}_{\text{noise}}^{-1}\left(\boldsymbol{\mathcal{B}}\frac{\delta u}{\delta\beta}(\beta)(\beta_2)\right),$$

其中，$\delta u/\delta\beta\,(\beta)(\beta_1)$ 是 u 关于 β 沿 β_1 方向的 一阶变分 [35]，而 $\delta^2 u/\delta\beta^2\,(\beta)\,(\beta_1,\beta_2)$ 是 u 关于 β 沿 β_1，β_2 方向的二阶变分，也就是说，

$$\frac{\delta u}{\delta\beta}(\beta)(\beta_1) := \left[\frac{\mathrm{d}}{\mathrm{d}\epsilon}u(\beta+\epsilon\beta_1)\right]_{\epsilon=0},$$
$$\frac{\delta^2 u}{\delta\beta^2}(\beta)(\beta_1,\beta_2) := \left[\frac{\mathrm{d}}{\mathrm{d}\epsilon}\frac{\delta u}{\delta\beta}(\beta+\epsilon\beta_2)(\beta_1)\right]_{\epsilon=0}.$$

将 β 用有限元离散后，我们得到数据失配海森矩阵 $\boldsymbol{H}_{\text{misfit}}$ 和高斯-牛顿数据失配海森矩阵 $\boldsymbol{H}_{\text{misfit}}^{\text{GN}}$ 的 $(i,\ j)$ 项的如下形式表达式

$$\left(\boldsymbol{H}_{\text{misfit}}\right)_{i,j} = \frac{\delta^2}{\delta\beta^2}\Big(J_{\text{misfit}}\left(\beta\right)\Big)\left(\psi_i,\psi_j\right), \tag{4}$$

$$\left(\boldsymbol{H}_{\text{misfit}}^{\text{GN}}\right)_{i,j} = \left(\boldsymbol{\mathcal{B}}\frac{\delta u}{\delta\beta}\left(\beta\right)\left(\psi_i\right)\right)^{\top}\boldsymbol{\Gamma}_{\text{noise}}^{-1}\left(\boldsymbol{\mathcal{B}}\frac{\delta u}{\delta\beta}\left(\beta\right)\left(\psi_j\right)\right), \tag{5}$$

其中，$\{\psi_j\}_{j=1}^{N}$ 是节点有限元空间的一组基，用于近似 β。

如上所述，当灵敏度主要具有局部性，且两个有限元基函数 ψ_i、ψ_j 的支撑集相距较远时，下列项

$$\left(\boldsymbol{\mathcal{B}}\frac{\delta u}{\delta\beta}\left(\beta\right)\left(\psi_i\right)\right)^{\top}\boldsymbol{\Gamma}_{\text{noise}}^{-1}\Big(\boldsymbol{\mathcal{B}}\frac{\delta u}{\delta\beta}\left(\beta\right)\left(\psi_j\right)\Big) \quad \text{and} \quad \boldsymbol{\mathcal{B}}\Big(\frac{\delta^2 u}{\delta\beta^2}\left(\beta\right)\left(\psi_i,\psi_j\right)\Big),$$

（假设噪声协方差矩阵具有对角占优性）相对较小。例如，这是因为当$\boldsymbol{\mathcal{B}}\delta u/\delta\beta(\beta)(\psi_j)$较大时，$\boldsymbol{\mathcal{B}}\delta u/\delta\beta(\beta)(\psi_i)$的值较小。现在，设 $\mathcal{I}$、$\mathcal{J}$是 $\{1,2,\ldots,N\}$的不相交索引子集，那么只要$\cup_{i\in\mathcal{I}}$supp（ψ_i）与 $\cup_{j\in\mathcal{J}}$supp（ψ_j）相距较远，数据失配海森矩阵的矩阵块 $\{(\boldsymbol{H}_{\text{misfit}})_{i\in\mathcal{I},\ j\in\mathcal{J}}\}$ 中的元素就相对较小。此类海森矩阵块适合用低秩矩阵近似。若对有限元基函数 ψ_i 对应的自由度进行排序，使得 $\mathcal{I}$、$\mathcal{J}$ 连续排列，则（$\boldsymbol{H}_{\text{misfit}}$）$_{\mathcal{I},\mathcal{J}}$ 是 $\boldsymbol{H}_{\text{misfit}}$ 的一个非对角子块，而 $\boldsymbol{H}_{\text{misfit}}$ 往往具有第 2.2 节所定义的 HODLR 结构。出于相同原因，高斯-牛顿数据失配海森矩阵也可能具有 HODLR 结构。在这两种情况下，基函数的排列顺序因而也包括自由度的排列顺序，都会影响这种结构。理想情况下，排序应保持局部性，即连续索引对应彼此接近的基函数；因此，索引相差较大的基函数彼此距离较远，使得相应非对角块中的元素较小，并且可以用低秩矩阵近似很好地逼近。关于自由度排序方法及相关数值实验的讨论，请参见第 6.2 节。

3.2. 利用 *HODLR* 结构快速采样高斯后验近似

在 [2], 下文中，我们将给出高斯化后验协方差的表达式，

$$\begin{aligned}
\boldsymbol{\Gamma}_{\text{post}} &= \left(\boldsymbol{H}_{\text{misfit}}+\boldsymbol{\Gamma}_{\text{prior}}^{-1}\right)^{-1} = \boldsymbol{\Gamma}_{\text{prior}}^{1/2}\left(\boldsymbol{H}'_{\text{misfit}}+\boldsymbol{I}\right)^{-1}\boldsymbol{\Gamma}_{\text{prior}}^{\top/2},\\
\boldsymbol{H}'_{\text{misfit}} &:= \boldsymbol{\Gamma}_{\text{prior}}^{\top/2}\boldsymbol{H}_{\text{misfit}}\boldsymbol{\Gamma}_{\text{prior}}^{1/2},\\
\boldsymbol{\Gamma}_{\text{post}}^{1/2} &= \boldsymbol{\Gamma}_{\text{prior}}^{1/2}\left(\boldsymbol{H}'_{\text{misfit}}+\boldsymbol{I}\right)^{-1/2},
\end{aligned}$$

其中，矩阵平方根 $\boldsymbol{A}^{1/2}$ 满足 $\boldsymbol{A}=\boldsymbol{A}^{1/2}\left(\boldsymbol{A}^{1/2}\right)^{\top}$。对于参数场在有界子集上空间分布的贝叶斯反问题

$\mathbb{R}^m$, $m = 2$, 3, 一个合理的选择是采用椭圆型偏微分方程逆算子的平方作为先验协方差 [20], , 这样便可以得到 $\mathbf{\Gamma}_{\text{prior}}$的对称平方根。在以往的研究中，例如 [1, 3, 4, 5, 6], , 先验预条件化的数据失配 Hessian $\boldsymbol{H}'_{\text{misfit}}$采用全局低秩压缩进行近似。对于数据集所含信息量足够少的反问题，这种策略能有效地近似后验协方差矩阵。在此，我们提出利用 HODLR问题结构，通过对先验预条件化的数据失配~$\boldsymbol{H}'_{\text{misfit}}$进行HODLR近似，生成近似后验协方差矩阵；关于此类近似如何影响近似后验协方差的精度，请参见附录8.3

$$\tilde{\mathbf{\Gamma}}_{\text{post}} = \mathbf{\Gamma}_{\text{prior}}^{1/2} \left(\tilde{\boldsymbol{H}}'_{\text{misfit}} + \boldsymbol{I} \right)^{-1} \mathbf{\Gamma}_{\text{prior}}^{\top/2}.$$

随后，以 $\mathcal{O}\left(N \log^2 N\right)$ 次浮点运算 [31]生成 $\tilde{\boldsymbol{H}}'_{\text{misfit}} + \boldsymbol{I}$ 的对称平方根分解。该对称分解能够以 $\mathcal{O}(N \log N)$ 的方式同时计算平方根和逆平方根。

4. 贝叶斯冰盖反问题

模拟冰盖动力学（例如格陵兰或南极洲的冰盖）是耦合气候模拟的重要组成部分。这类模拟需要估计与现有观测数据相一致的冰的当前状态，这一过程有时称为模型初始化。该估计问题既可以表述为确定性反问题（即由偏微分方程约束的非线性最小二乘优化问题），也可以表述为贝叶斯反问题（即旨在刻画状态分布的统计问题）。后一种方法虽然计算成本更高，但除了确定最佳参数拟合外，还能提供不确定性估计。

冰盖动力学 [36] 通常由非线性斯托克斯方程或其简化形式（例如一阶方程，参见 [37]）描述。在冰盖模拟中，最不确定的部分通常是底部边界条件，即冰盖底部与岩石、沙土、水或其混合物之间的相互作用。因此，利用顶面的速度观测、冰盖几何形状以及描述其动力学的模型来估计冰盖的有效边界条件，是一个重要问题，可以将其数学表述为反问题 [1, 38, 39, 40, 41]。

接下来，我们概述这一反问题的表述。按照文献中的常见做法，我们采用快照优化方法，假设所有数据均在一段较短的时间内采集，在此期间冰的几何形状变化可以忽略不计。我们将被冰覆盖的有界区域记为 $\Omega \subset \mathbb{R}^m$, $m \in \{2, 3\}$，并将区域边界 $\partial \Omega$ 的底部、侧面和顶部部分分别记为 Γ_b、 Γ_l 和 Γ_t，如图3所示。

控制方程为非线性不可压缩斯托克斯方程，其解为冰流速度 $\boldsymbol{u} : \Omega \to \mathbb{R}^m$ 和压力 $p : \Omega \to \mathbb{R}$ ，如下所示:

:

$$-\nabla \cdot \boldsymbol{\sigma_u} = \rho \boldsymbol{g} \quad \text{in } \Omega, \tag{6}$$

$$\nabla \cdot \boldsymbol{u} = 0 \qquad \text{in } \Omega, \tag{7}$$

$$\boldsymbol{\sigma_u} \boldsymbol{n} = \boldsymbol{0} \qquad \text{关于 } \Gamma_t, \tag{8}$$

$$\boldsymbol{u} \cdot \boldsymbol{n} = 0 \text{ and } \boldsymbol{T}\left(\boldsymbol{\sigma_u} \boldsymbol{n} + \exp\left(\beta\right) \boldsymbol{u}\right) = \boldsymbol{0} \quad \text{ on } \Gamma_b, \tag{9}$$

以及其他侧边界条件。这里， β 是基底滑动参数场， $\rho \boldsymbol{g}$ 是体力密度，其中 ρ 是冰的质量密度， $\boldsymbol{g}$ 是重力加速度。方程 6 描述动量守恒，方程 7 描述质量守恒，方程 8 是顶面（冰-空气界面）的无应力边界条件。在法向上，方程 9 表示不可穿透条件，即冰不能流入支撑它的岩石/沙层（这里$\boldsymbol{n}$ 表示边界 $\partial\Omega$ 的外单位法向量， $\boldsymbol{T}$ 表示切向算子，$\boldsymbol{T}\boldsymbol{v} = \boldsymbol{v} - \boldsymbol{n}(\boldsymbol{n}^\top \boldsymbol{v})$）。在切向上，方程 9 指定了切向滑动条件，该条件通过（对数）基底滑动场 $\beta = \beta(x)$、 $x \in \Gamma_b$将切向滑动比例与切向应力联系起来。我们采用格伦流动定律 [42],这一冰的本构定律，它将应力张量 $\boldsymbol{\sigma_u}$ 与应变率张量 $\dot{\boldsymbol{\varepsilon}}_{\boldsymbol{u}} = \frac{1}{2}\left(\boldsymbol{\nabla u} + \boldsymbol{\nabla u}^\top\right)$联系起来，

–

$$\boldsymbol{\sigma_u} = 2\eta\left(\boldsymbol{u}\right)\dot{\boldsymbol{\varepsilon}}_{\boldsymbol{u}} - \boldsymbol{I}p, \text{ with } \eta\left(\boldsymbol{u}\right) = \frac{1}{2}A^{-1/n}\dot{\boldsymbol{\varepsilon}}_{\mathrm{II}}^{\frac{1-n}{2n}}, \tag{10}$$

其中， η为有效黏度， $\boldsymbol{I}$为单位矩阵， $\dot{\boldsymbol{\varepsilon}}_{\mathrm{II}} = \mathrm{tr}\left(\dot{\boldsymbol{\varepsilon}}_{\boldsymbol{u}}^2\right)$为应变率张量的第二不变量， A 为流动率因子， n 为格伦指数。冰通常使用 $n \approx 3$建模，这对应于剪切变稀本构关系；此处我们使用 $n = 3$。

如上所述，不确定性最大的参数是（对数）基底滑动场 $\beta = \beta(x)$。因此，通常需要从（一般为卫星）观测数据 $\boldsymbol{d}$中推断该参数，此处观测数据的形式为表面速度测量。使用适当的点观测算子$\boldsymbol{\mathcal{B}}$ 从控制方程 6-9 的解 $\boldsymbol{u}$ 中提取点数据，并假设观测误差 $\boldsymbol{\eta}$为加性误差，则模型与数据之间的关系具有如下典型形式

$$\boldsymbol{d} = \boldsymbol{\mathcal{B}}\boldsymbol{u} + \boldsymbol{\eta}. \tag{11}$$

假设观测误差 $\boldsymbol{\eta}$ 和参数场 β的先验服从高斯分布，则我们处于第2.1节所概述的贝叶斯反问题框架中。

5. 示例 I：二维 ISMIP-HOM 基准测试

我们首先研究在一个受 ISMIP-HOM 冰盖模拟基准问题集合启发的问题中，对高斯-牛顿数据失配海森矩阵进行压缩的前景 [43]。该问题集曾用于探索冰盖反问题 [40, 41]。简要介绍问题设置后，我们将展示最大后验

点估计 $\boldsymbol{\beta}^\star$ ，以及使用后验协方差的 HODLR 压缩得到的近似高斯化后验样本。随后，我们研究各种问题特征对高斯-牛顿数据失配海森矩阵适用于 HODLR 和全局低秩格式压缩的影响。

5.1. 问题设置

该问题设置包含一个位于坡面上的矩形冰块，如图3所示。这个简单示例使我们能够研究区域纵横比、观测数据数量和网格细化程度对高斯-牛顿数据失配海森矩阵性质的影响。区域宽度为 $W = 10^4\,[\mathrm{m}]$，高度为 $H = 10^2\,[\mathrm{m}]$。在侧边界上采用周期边界条件，因此该设置模拟了位于坡面上的无限冰层。控制方程和其他边界条件如方程6-9所述。

斯托克斯方程在网格上使用 Taylor-Hood 有限元进行离散，该网格由 256×10 个矩形组成，每个矩形细分为两个三角形，区域长度为 $[0, W)$，高度为 $[0, H]$。为计算最大后验估计，我们使用“真实的”对数基底滑动场生成合成表面速度数据，$\beta_{\mathrm{true}}(x) := \log\left(1\,200 + 1\,100 \sin\left(\frac{2\pi x}{W}\right)\right)$。给定该基底滑动场后，我们求解方程 6-9，提取上边界 Γ_t上100个均匀分布点处的切向速度分量，并向每个数据点添加1%的相对高斯噪声，从而得到合成数据 $\boldsymbol{d}$。

接下来需要定义参数场 β的先验分布。将 β_{true}的平均值用作常数先验均值 $\beta(x) = 6.73315 \approx \frac{1}{W}\int_0^W \beta_{\mathrm{true}}(s)\mathrm{d}s$。先验协方差矩阵 $\boldsymbol{\Gamma}_{\mathrm{prior}}$ 是协方差偏微分方程算子$\mathcal{C} := (\delta I - \gamma\Delta)^{-1}$的离散化，其中 $\gamma = 6 \times 10^2$ 和 $\delta = 2.4 \times 10^{-3}$，并采用罗宾边界条件 [44]。选择这些数值是为了获得相对较大的先验相关长度 $10^3\,[\mathrm{m}]$ [45]。接下来，我们概述最大后验点的计算，以及在最大后验点处对高斯-牛顿数据失配海森矩阵进行压缩的方法。

5.2. 最大后验点与 *HODLR* 高斯化后验

使用带回溯线搜索 [21], 的非精确高斯–牛顿最小化方法求解最大后验估计的非线性优化问题，其中线性方程组通过共轭梯度法迭代求解。所得最大后验点如图4所示。最大后验参数场 $\beta^\star$与真实参数 β_{true}非常相似，这是可用数据量大且噪声水平相对较低所致。

接下来，我们使用具有压缩后的先验预条件数据失配海森矩阵的高斯化后验分布 $\boldsymbol{H}'_{\mathrm{misfit}}$ ，从后验分布中生成近似样本。构建先验预条件数据失配海森矩阵的 HODLR 压缩后（详情及比较见下文

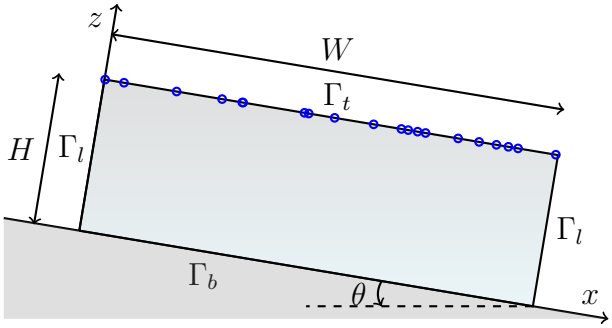


图3：第 5 节示例 I 所用二维冰层的示意图。蓝色圆圈表示具有代表性的（随机）测量位置。角度 θ 是冰层的坡度。

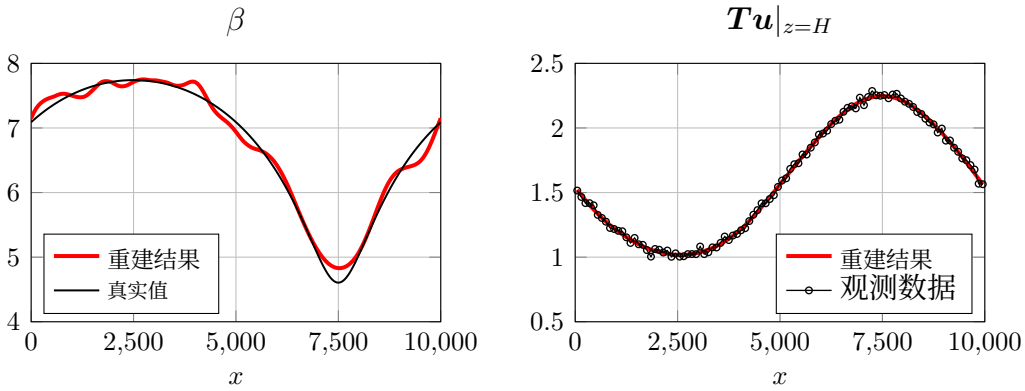


图4：示例 I 左侧展示的是用于生成上表面切向速度分量合成观测的最大后验点 $\beta^\star$（红色）和真实基底滑移参数 β_{true}（黑色） Γ_t。右侧展示的是用于计算最大后验点的含噪合成观测（黑点）及其对应的切向表面速度重构（红色）。

第5.3节），我们按照第3.2节所述，从 HODLR 高斯化后验中抽取样本。在图5中，我们比较了先验分布和后验分布的均值、逐点标准差及样本。正如预期的那样，我们发现数据更新了我们对空间分布参数场的认识，并降低了不确定性。具体而言，图中显示了高斯化后验分布和先验分布的一维点边缘分布 $\sigma(x)$ 的 2σ 界限 $\boldsymbol{\sigma}_i = [\boldsymbol{\Gamma}_{i,i}]^{-1/2}$，以验证样本大多位于各自均值的两个标准差范围内。先验预条件数据失配海森矩阵 $\boldsymbol{H}'_{\text{misfit}}$ 采用相对容差 10^{-6} 进行压缩；也就是说，高概率下满足 $\|\boldsymbol{H}'_{\text{misfit}} - \tilde{\boldsymbol{H}}'_{\text{misfit}}\|_2 / \|\boldsymbol{H}'_{\text{misfit}}\|_2 \leq 10^{-6}$。

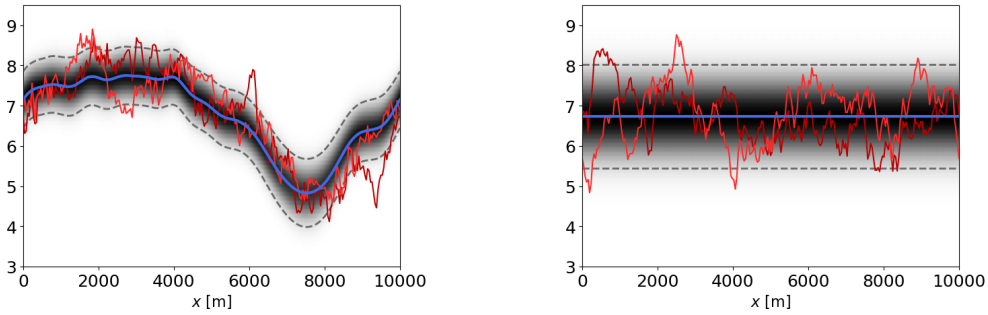


图5：示例 I 的结果：对于先验（左）和使用第3.2节所述方案得到的 HODLR 高斯化后验（右），展示了两个随机样本（红色）、均值 β（蓝色）以及满足 $0 \leq x \leq W$ 和 $\beta(x) - 2\sigma(x) \leq y \leq \beta(x) + 2\sigma(x)$ 的区域 $R = \{(x, y)$ 的边界。

5.3. 海森矩阵块谱对问题设置的依赖性

接下来，我们研究问题特征如何影响使用全局低秩和 HODLR压缩近似高斯-牛顿数据失配海森矩阵的数值适用性。在本节及后续各节中，我们以海森矩阵向量积（也称为海森矩阵作用）的次数来衡量生成矩阵压缩的成本，因为每次向量积都需要进行两次线性化偏微分方程求解，因此其计算成本占主导地位。根据附录 8.2 的结果，当每个非对角块的绝对误差不超过 ε/L 时，我们断言第 L 层 HODLR近似中的绝对误差为 ε。本节的特殊之处在于，采用自适应 单遍算法和 HODLR算法，基于绝对容差准则生成全局低秩和 HODLR近似。算法的绝对容差输入会按全局低秩奇异值中的最大值进行缩放，以报告相对近似误差。需要注意的是，报告的近似误差忽略了其他误差，例如剥离过程中产生的误差 [11, 12]以及单遍算法中的其他近似假设所带来的误差；预计这两类误差都不会很显著。

纵横比的影响在此，我们改变区域 $\phi = H/W$ 的纵横比，其中 H 和 W 分别为区域的高度和宽度，以研究其如何影响高斯-牛顿数据失配海森矩阵的块谱，并最终影响计算成本。图 6 显示，与非对角块的谱相比，全局谱对相对长度尺度 ϕ 的变化更为敏感。随着 ϕ 减小，非对角块的低秩近似计算成本降低，这是因为冰盖厚度减小时，敏感性锥会变得越来越局部化。随着 ϕ 减小，全局低秩近似的计算成本会增加，这是因为数据提供了更多信息。我们注意到，洪堡冰川和格陵兰冰盖等实际问题（将在第 6 节中研究）

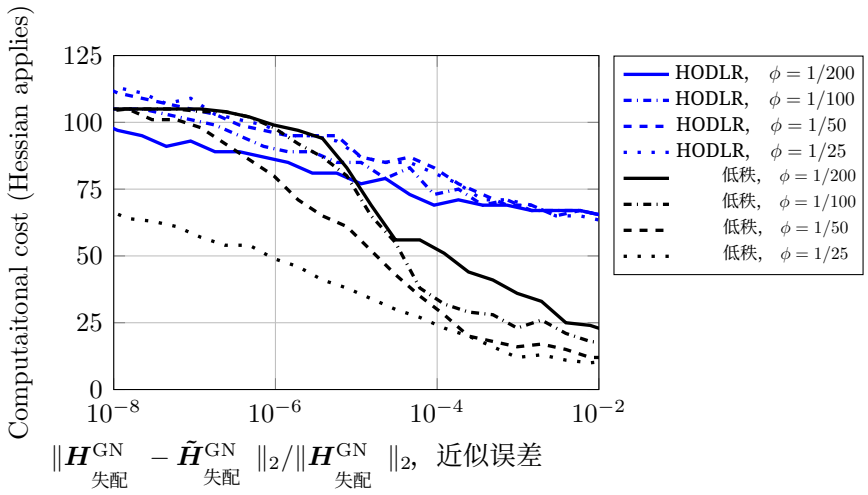


图 6：示例 I 中，冰盖纵横比为 ϕ 时，高斯-牛顿数据失配海森矩阵 $\boldsymbol{H}^{\mathrm{GN}}_{\mathrm{misfit}}$ 的 HODLR 压缩与全局低秩压缩成本对比。该图表明，在纵横比较低时，对于中等目标精度，HODLR 比全局低秩方法更高效。

纵横比较小，因此预计其数据失配海森矩阵不太适合进行全局低秩近似。

参数维数的影响 现在，我们改变网格离散细化程度，以研究数据信息量如何通过离散参数维数 $N = \dim(\boldsymbol{\beta})$ 影响生成高斯-牛顿数据失配海森矩阵的 HODLR 和全局低秩近似所需的计算成本。每当离散参数维数翻倍，层次深度 L 便递增一级，使层次深度随海森矩阵大小的对数变化；这一条件见第 2.2 节。图 7 提供了计算证据，支持第 2.2 节中的论断：对具有 HODLR 结构的算子进行层次压缩所需的应用次数为 $\mathcal{O}(\log N)$。相比之下，生成全局低秩近似所需的应用次数对网格细化程度相当不敏感。

数据维度的影响图 8 显示，全局秩会随观测点数量的增加而增大，因此，对于观测数据具有强信息性的反问题，全局低秩压缩往往效率较低。高斯-牛顿数据失配海森矩阵的谱衰减速率与未正则化逆问题的不适定程度有关。随着观测数量增加，相关模型预测对基底滑动场中的小尺度变化越来越敏感。因此，通常来说，数据越多，数据集对参数的信息量越大，而高斯-牛顿数据失配海森矩阵的谱衰减速率则越低。

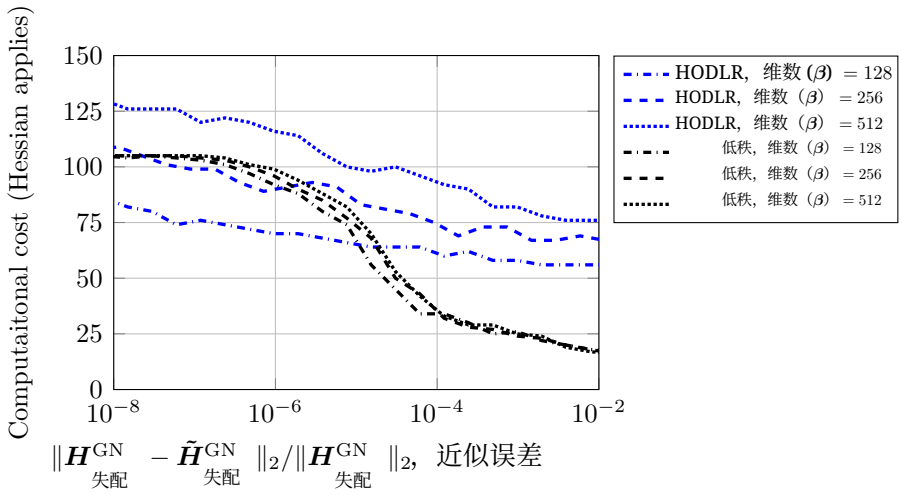


图 7：示例 I 中，高斯-牛顿数据失配海森矩阵的 HODLR 压缩和全局低秩（低秩）压缩成本与 dim(β)（离散化对数基底滑动场的维数）的关系。全局低秩压缩的成本几乎保持不变，而 HODLR 压缩的成本随着网格细化而增加。

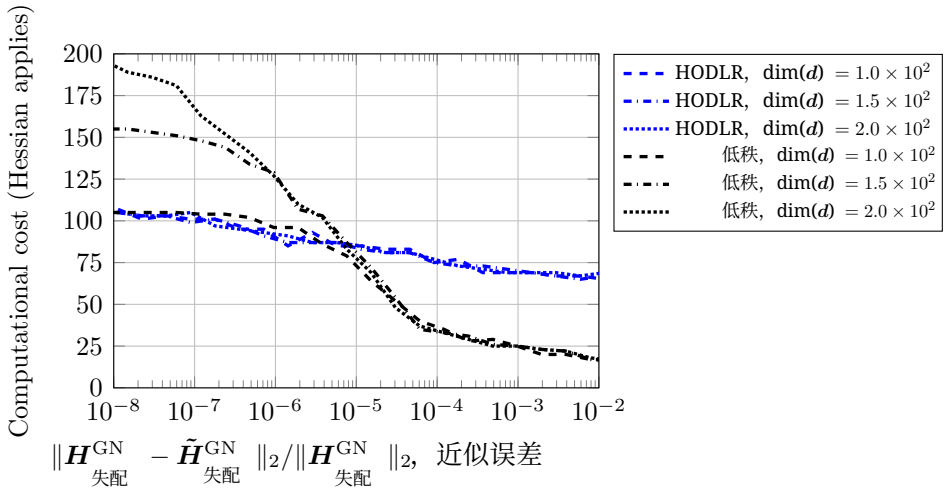


图 8：示例 I 中，高斯-牛顿数据失配海森矩阵的 HODLR 压缩和全局低秩（低秩）压缩成本对 dim(d)（即数据维度）的依赖关系。全局低秩近似的计算成本随观测数据数量的增加而增加，而 HODLR 压缩的成本则相对不敏感。

6. 示例 II：洪堡冰川与格陵兰冰盖

本文使用气候模拟中常见的大规模冰盖问题，研究所提出方法的可扩展性。具体而言，我们关注格陵兰西北部的洪堡冰川，以及整个格陵兰冰盖。对于这些模拟，我们使用冰盖模型 MALI，[46]，该模型依赖于

Albany 是一种 C++多物理场库，用于实现所谓的斯托克斯方程一阶近似。该一阶近似基于由冰盖浅薄特性所启发的尺度分析论证，并利用不可压缩条件将未知量简化为水平速度。我们使用 PyAlbany [48]，这是 Albany 软件包的便捷 Python 接口，而 Albany 又构建于 Trilinos [49]之上。Albany 旨在支持并行且可扩展的有限元离散化 PDE 求解器，以及各种分析功能。表 1 列出了计算中使用的参数、状态、数据维度，以及核心数和层次级别数的详细信息。

	洪堡	格陵兰
维度($\boldsymbol{\beta}$)	11 608	320 116
dim($\boldsymbol{u}$)	255 376	7 042 552
dim($\boldsymbol{d}$)	23 216	640 232
核心数	120	2 048
L	8	10

表 1：洪堡冰川和格陵兰冰盖问题（示例 II）的问题规格。离散化参数场的维数 dim($\boldsymbol{\beta}$)、离散化速度场的维数 dim($\boldsymbol{u}$)、观测数据的维数 dim($\boldsymbol{d}$)、计算所用处理器数量以及 HODLR 层次划分深度 L。

以下研究部分受到第 5 节研究结果的启发，具体而言，垂直方向与水平方向的纵横比（见第5.3节）会影响全局低秩压缩的适用性，并使 HODLR压缩更具优势。我们生成 HODLR 和全局低秩近似，然后根据计算得到的谱、公式 3 和 $\zeta^{\mathrm{LR}} = r + d$ 估算计算成本。此外，我们还研究自由度的排序如何影响数据失配海森矩阵非对角块的谱衰减。我们给出洪堡冰川和格陵兰冰盖的结果：洪堡冰川横向延伸约 4×10^2 [km]，格陵兰冰盖延伸约 1.8×10^3 [km]。冰层最厚处为 3.4 [km]，因此洪堡冰川和格陵兰冰盖的纵横比分别约为 $8.5{\times}10^{-3}$ 和 $1.9{\times}10^{-3}$。我们采用格陵兰冰盖的非均匀三角剖分，网格尺寸从 1 到 10 [km], 不等，然后沿垂直方向挤出，生成一个包含 10 层棱柱单元的三维网格。格陵兰冰盖顶面的速度观测值来自卫星观测 [50]。作为初始化过程的一部分，我们采用数值优化方法求得 MAP 基底滑动场和温度场，以匹配冰速度观测值；优化过程受与温度模型耦合的一阶流动模型约束 [51]。有关网格几何和数据的更多细节，尤其是洪堡冰川的相关信息，可参见 [52]。

在图中 图 9 中，我们展示了观测到的表面速度 $\boldsymbol{d}$，单位为 [米/年], ，以及最大后验估计 s

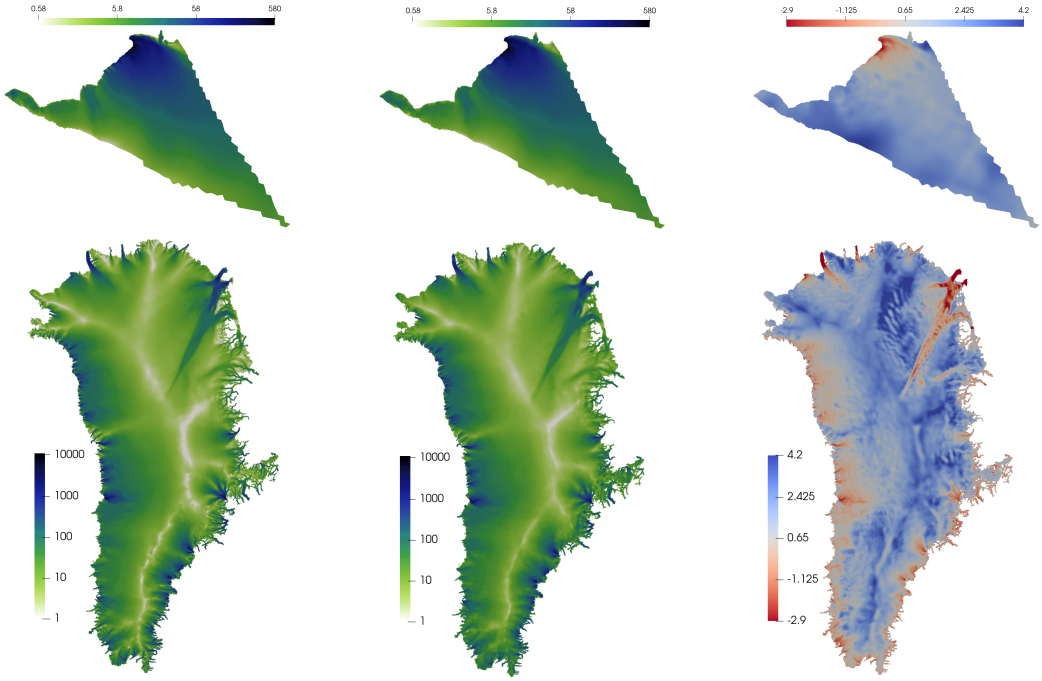


图 9：示例 II 的数据和最大后验估计。图中展示了地表速度观测数据（左）、基于对数基底滑动场最大后验估计重建的地表速度场（中）。顶行对应洪堡冰川，底行对应格陵兰冰盖。

对数基底滑动场 $\beta^\star$ （$\exp(\beta^\star)$ 的单位为 [千帕·年/米]）和模型生成的表面速度，单位为 [米/年] 。

*6.1. HODLR*可压缩性

接下来，我们分别对格陵兰和洪堡数据失配海森矩阵生成全局低秩近似，并对各种非对角块生成低秩近似。图10给出了估计奇异值的曲线。我们观察到，格陵兰冰盖的谱衰减速度明显慢于洪堡冰川。除了这两个离散化问题的规模不同之外，这也归因于它们的纵横比不同。在估计出数据失配海森矩阵及相应非对角块的奇异值后，便可以估算将它们压缩为全局低秩格式和 HODLR 矩阵格式的计算成本。图11展示了计算成本随海森矩阵近似目标精度的变化情况，其中表明，对于由复杂冰盖模型支配的大规模反问题，HODLR压缩格式是近似数据失配海森矩阵的一种有效方式。

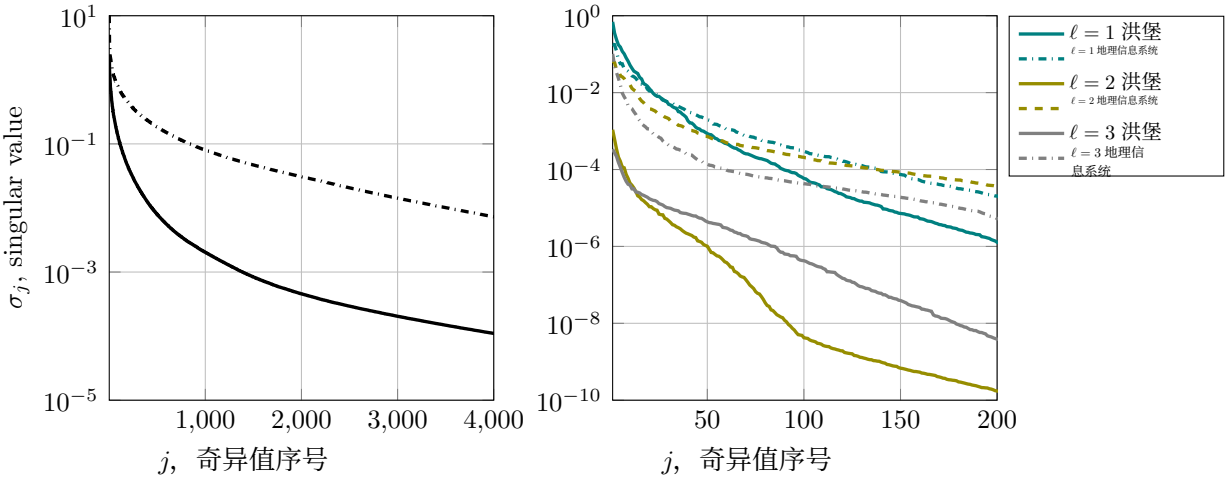


图10：示例 II 中数据失配海森矩阵（左图）及其各非对角块（右图）的奇异值。最右侧图中的配色方案与图1一致。左图中，洪堡和格陵兰数据失配海森矩阵的奇异值分别用实线和点划线表示。右图中，我们展示最上方各块的奇异值，即 $\boldsymbol{A}_{1\,2}^{(\ell)}$，如第8.2节所定义。

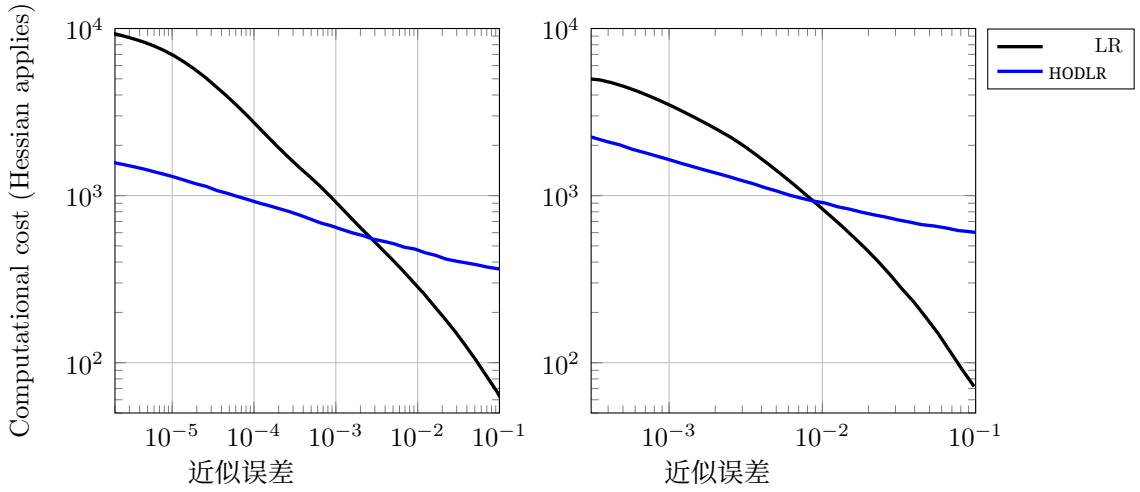


图11：将洪堡冰川（左）和格陵兰冰盖（右）的数据失配海森矩阵压缩为全局低秩（LR）和分层非对角低秩（HODLR）˜格式所需的估算计算成本（以海森矩阵作用次数衡量），随近似误差 $\|\boldsymbol{H}_{\text{misfit}} - \boldsymbol{H}_{\text{misfit}}\|_2 / \|\boldsymbol{H}_{\text{misfit}}\|_2$ 的变化。

6.2. 参数自由度排序的影响

我们旨在确保通过第 2.2 节所述的分层划分确定的数据失配海森矩阵的非对角块具有低秩性质。对于

出于这个原因，与自由度（dofs）$\{\boldsymbol{x}_i\}_i$ 相关联的节点按照 kd 树排序，即递归进行超平面划分。kd 树提供的排序具有这样的特点：当 $|i-j|$ 较小时，距离矩阵 $\boldsymbol{D}_{i,j} = \|\boldsymbol{x}_i - \boldsymbol{x}_j\|_2$ 的 $(i,\ j)$ 项通常也较小，也就是说，自由度排序保留了某种局部性（见第 3.1 节）。具体而言，确定一个稀疏置换矩阵 $\boldsymbol{B}$，其作用是将自由度从有限元离散提供的默认排序重新排列为 kd 树指定的排序。相对于 kd 树排序的数据失配海森矩阵 $\boldsymbol{H}_{\text{misfit}}^{\text{kd}} := \boldsymbol{B}\boldsymbol{H}_{\text{misfit}}\boldsymbol{B}^\top$ 随后便适合进行 HODLR压缩。之后，$\boldsymbol{B}^\top \tilde{\boldsymbol{H}}_{\text{misfit}}^{\text{kd}} \boldsymbol{B}$ 是相对于默认排序的数据失配海森矩阵的一个近似。

自由度排序不会影响矩阵的全局数值秩，但确实会影响其众多由固定划分方案定义的子矩阵的数值秩，例如 HODLR 矩阵的非对角块（见第 2.2 节）。在此，我们通过比较采用 Albany 提供的默认排序与通过 kd树递归超平面划分获得的排序时，非对角块奇异值的衰减速率，研究洪堡冰川数据失配海森矩阵的 HODLR可压缩性。如图12所示，一级非对角块奇异值的衰减速率在很大程度上取决于自由度排序。这是因为 kd树排序能够更好地保持局部性，因此根据第 3.1 节中的论述，采用 kd树排序时奇异值衰减得快得多。因此，kd树排序为生成数据失配海森矩阵的 HODLR近似提供了一种计算成本显著更低的方法。图12还展示了默认基和 kd树基对应的距离矩阵，体现了 kd树排序更好的局部性。需要注意的是，数据失配海森矩阵预计会呈现与这些距离矩阵相似的结构，这解释了为什么与自由度的默认排序相比，其非对角块在 kd树排序下能够得到更有效的压缩。

7. 结论

在本研究中，我们阐述了为什么一类偏微分方程支配的反问题所产生的数据失配海森矩阵具有 HODLR 矩阵结构。HODLR矩阵可以高效地求逆和分解，这些操作对于用牛顿法求解偏微分方程支配的反问题、构造高斯近似以及采用马尔可夫链蒙特卡罗采样方法都必不可少。我们研究了冰盖反问题。在某些条件下，与全局低秩矩阵格式相比，HODLR矩阵能够提供计算效率更高的近似格式。这类问题具有信息量很高的数据和纵横比较小的冰盖。全局低秩矩阵适用于离散化参数维数较大而数据维数较小的情况；而我们发现，对于大规模反问题，例如使用卫星观测数据、离散化参数维数超过 10^5 的格陵兰冰盖反问题，HODLR矩阵能够节省计算量。

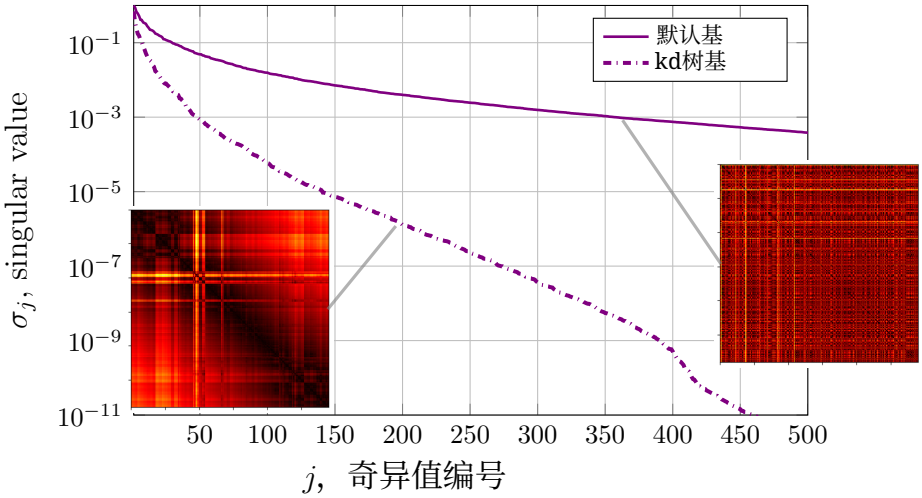


图12：洪堡冰川数据失配海森矩阵的层次第1级非对角块 $\boldsymbol{A}_{1\,2}^{(1)}$ 、，分别用kd树基和默认基表示时的奇异值。同时还展示了距离矩阵$\boldsymbol{D}_{i,\,j} = \|\boldsymbol{x}_i - \boldsymbol{x}_j\|_2$的热图，其中，与有限元自由度相关联的节点 $\{\boldsymbol{x}_i\}_i$按照默认标准和kd树进行了排序。

对于未来的工作，我们认为，通过采用满足强可容许条件 [17]，的层次矩阵划分，可以进一步降低计算成本，因为这种划分更适合利用数据失配海森矩阵的结构。然而，生成采用此类划分的层次矩阵近似，例如使用剥离法 [11, 12]，，需要大量增加海森矩阵向量积。归根结底，要进一步降低偏微分方程支配的反问题中海森矩阵近似的计算成本，关键在于进一步利用问题结构。

8. 附录

8.1. 随机压缩算法

在此，为了完整起见，我们概述随机化的无矩阵双遍全局低秩算法和 HODLR 压缩算法。随机化双遍低秩算法 [32] 的核心思想是

(i) 将一个元素为随机数的向量 $\boldsymbol{\omega}$ 作用于矩阵 $\boldsymbol{A}$，得到的向量 $\boldsymbol{y} = \boldsymbol{A}\boldsymbol{\omega}$很可能与 $\boldsymbol{A}$的主导左奇异向量对齐；(ii) 一个列向量几乎与 $\boldsymbol{A}$的主导左奇异向量对齐的矩阵 $\boldsymbol{Q}$，可用于构造精确的低秩近似

$\tilde{\boldsymbol{A}} = \boldsymbol{Q}\boldsymbol{Q}^\top \boldsymbol{A}$ $\boldsymbol{A}$的。

双遍随机奇异值分解算法见算法1，与 [32], 中的算法差异不大；具体而言，只有第 7、8 和 9 行不同。这一细微修改使我们不必对（分布式） $N \times k$ 矩阵（例如 $\boldsymbol{Z}$）进行（并行）奇异值分解（SVD）。在这里，我们只需对较小的 $k \times k$矩阵 $\boldsymbol{R}_{\boldsymbol{Z}}$进行奇异值分解。在第 6 节的分布式内存并行环境中，这一算法修改使我们只需调用串行奇异值分解程序，因为 $\boldsymbol{R}_{\boldsymbol{Z}}$通常较小，且每个处理器上都有该矩阵。

算法1 双遍随机奇异值分解。**输入：** $\boldsymbol{A} \in \mathbb{R}^{N \times N}$， $r \in \mathbb{N}$ 期望秩和过采样参数 $d \in \mathbb{N}$。~**输出：** 低秩近似 $\boldsymbol{A}$ $\boldsymbol{A}$的

1: $k = r + d$	
2: $\boldsymbol{\Omega}$ $=$ `randn`(N, k)	{生成随机矩阵}
3: $\boldsymbol{Y} = \boldsymbol{A}\boldsymbol{\Omega}$	{采样列空间}
4: $\boldsymbol{Q}_{\boldsymbol{Y}} =$ `orthog`$(\boldsymbol{Y})$	{正交化列样本}
5: $\boldsymbol{Z} = \boldsymbol{A}^\top \boldsymbol{Q}_{\boldsymbol{Y}}$	{采样行空间}
6: $\boldsymbol{Q}_{\boldsymbol{Z}} =$ `orthog`$(\boldsymbol{Z})$	{正交化行样本}
7: $\boldsymbol{R}_{\boldsymbol{Z}} = \boldsymbol{Q}_{\boldsymbol{Z}}^\top \boldsymbol{Z}$	{压缩行样本}
8: $\boldsymbol{R}_{\boldsymbol{Z}} = \hat{\boldsymbol{V}}\boldsymbol{\Sigma}\hat{\boldsymbol{U}}^\top$	{ $k \times k$ 压缩行样本矩阵}的奇异值分解
9: $\boldsymbol{V} = \boldsymbol{Q}_{\boldsymbol{Z}}\hat{\boldsymbol{V}}$	{投影行空间信息}
10: $\boldsymbol{U} = \boldsymbol{Q}_{\boldsymbol{Y}}\hat{\boldsymbol{U}}$	{投影列空间信息}
11: $\tilde{\boldsymbol{A}} = \boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}^\top$	{构造低秩近似}

随机分层非对角低秩算法通过双遍算法压缩非对角块。较大的非对角块先于较小的非对角块进行压缩，这一过程通过~剥离过程 [11]实现。这里， $\boldsymbol{A}$ 和 $\boldsymbol{A}$ 均假设为对称的，因为我们旨在压缩对称算子并计算对称近似矩阵。

算法 2 对称无矩阵随机HODLR。

输入： 对称 $\boldsymbol{A} \in \mathbb{R}^{N\times N}$，层次深度 $L \in \mathbb{N}$, $r_1, \ldots, r_L$ 各层次深度的非对角块的目标秩以及过采样参数 d。

输出： 对称 HODLR近似 $\boldsymbol{A}$ 的 $\boldsymbol{A}$

</style>对 $\ell = 1, 2, \ldots, L$ **do**

  $k_\ell = r_\ell + d$

  $\boldsymbol{\Omega} = \texttt{zeros}(N, k_\ell)$

  对 $j = 1, \ldots, 2^{\ell-1}$ **执行**

    $\boldsymbol{\Omega}(\mathcal{I}_{2j}^{(\ell)}, :) = \texttt{randn}(|\mathcal{I}_{2j}^{(\ell)}|, k_\ell)$ {初始化结构化随机矩阵}

  结束循环

  $\boldsymbol{Y} = \left(\boldsymbol{A} - \sum_{j=1}^{\ell-1} \boldsymbol{A}^{(j)}\right) \boldsymbol{\Omega}$ {示例非对角块列空间}

  对于 $j = 1$、...、$2^{\ell-1}$ **执行**

    $\boldsymbol{Y}^{(j)} =$ 零矩阵(N, k_ℓ)

    $\boldsymbol{Y}^{(j)}(\mathcal{I}_{2j-1}^{(\ell)}, :) = \boldsymbol{Y}(\mathcal{I}_{2j-1}^{(\ell)}, :)$

    $\boldsymbol{Q}_{\boldsymbol{Y}}^{(j)} = \texttt{orthog}(\boldsymbol{Y}^{(j)})$ {对第 ℓ 层的列样本进行正交化非对角块}

  结束循环

  $\boldsymbol{Q}_{\boldsymbol{Y}} = \sum_{j=1}^{2^{\ell-1}} \boldsymbol{Q}_{\boldsymbol{Y}}^{(j)}$ {行空间采样矩阵}

  $\boldsymbol{Z} = \left(\boldsymbol{A} - \sum_{j=1}^{\ell-1} \boldsymbol{A}^{(j)}\right) \boldsymbol{Q}_{\boldsymbol{Y}}$ {对非对角块行空间进行采样}

  对 $j = 1, \ldots, 2^{\ell-1}$ **do**

    $\boldsymbol{Z}^{(j)} = \boldsymbol{Z}(\mathcal{I}_{2j}^{(\ell)}, :)$

    $\boldsymbol{Q}_{\boldsymbol{Z}}^{(j)} = \texttt{orthog}(\boldsymbol{Z}^{(j)})$ {正交化第 ℓ 层非对角块}

    $\boldsymbol{R}_{\boldsymbol{Z}}^{(j)} = \left(\boldsymbol{Q}_{\boldsymbol{Z}}^{(j)}\right)^\top \boldsymbol{Z}^{(j)}$ {压缩第 ℓ 层非对角块行样本}

    $\hat{\boldsymbol{V}}_{2j-1}^{(\ell)} \boldsymbol{\Sigma}_{2j-1}^{(\ell)} \hat{\boldsymbol{U}}_{2j-1}^{(\ell)} = \boldsymbol{R}_{\boldsymbol{Z}}^{(j)}$ { $k_\ell \times k_\ell$ 压缩行样本矩阵的奇异值分解}

    $\boldsymbol{V}_{2j-1}^{(\ell)} = \boldsymbol{Q}_{\boldsymbol{Z}}^{(j)} \hat{\boldsymbol{V}}_{2j-1}^{(\ell)}$ {投影行空间信息}

    $\boldsymbol{U}_{2j-1}^{(\ell)} = \boldsymbol{Q}_{\boldsymbol{Y}}^{(j)} \hat{\boldsymbol{U}}_{2j-1}^{(\ell)}$ {投影列空间信息}

    $\boldsymbol{V}_{2j}^{(\ell)} = \boldsymbol{U}_{2j-1}^{(\ell)}$

    $\boldsymbol{U}_{2j}^{(\ell)} = \boldsymbol{V}_{2j-1}^{(\ell)}$

    $\boldsymbol{\Sigma}_{2j}^{(\ell)} = \boldsymbol{\Sigma}_{2j-1}^{(\ell)}$

  结束循环

  $\boldsymbol{A}^{(\ell)} = \sum_{j=1}^{2^\ell} \boldsymbol{U}_j^{(\ell)} \boldsymbol{\Sigma}_j^{(\ell)} \left(\boldsymbol{V}_j^{(\ell)}\right)^\top$

结束循环

通过采样 $\boldsymbol{A} - \sum_{j=1}^{L} \boldsymbol{A}^{(j)}$，获取 $\boldsymbol{A}$ 的块对角矩阵 $\boldsymbol{D}$

$\tilde{\boldsymbol{A}} = \boldsymbol{D} + \sum_{\ell=1}^{L} \boldsymbol{A}^{(\ell)}$

8.2. 块低秩非对角近似误差累积产生的全局 *HODLR* 近似误差

设 $\boldsymbol{A}$ 为一个 $N \times N$ 矩阵，并考虑以下分块

$$\begin{aligned}
\boldsymbol{A}^{(1)} &= \begin{pmatrix} \mathbf{0} & \boldsymbol{A}_{1,2}^{(1)} \\ \boldsymbol{A}_{2,1}^{(1)} & \mathbf{0} \end{pmatrix}, \\
\boldsymbol{A}^{(2)} &= \begin{pmatrix} \mathbf{0} & \boldsymbol{A}_{1,2}^{(2)} & \mathbf{0} & \mathbf{0} \\ \boldsymbol{A}_{2,1}^{(2)} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \boldsymbol{A}_{3,4}^{(2)} \\ \mathbf{0} & \mathbf{0} & \boldsymbol{A}_{4,3}^{(2)} & \mathbf{0} \end{pmatrix}, \\
\boldsymbol{D} &= \begin{pmatrix} \boldsymbol{A}_{1,1}^{(2)} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{A}_{2,2}^{(2)} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \boldsymbol{A}_{3,3}^{(2)} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \boldsymbol{A}_{4,4}^{(2)} \end{pmatrix},
\end{aligned}$$

其中 $\boldsymbol{A}_{i}^{(\ell)}{}_{,j}$ 是 $2^{\ell} \times 2^{\ell}$ 的i,j块分块中的一个块，其中 $1 \le \ell \le L$。 $\boldsymbol{A}^{(\ell)}$包含满足 $|i-j|=1$ 的所有块 $\boldsymbol{A}_{i}^{(\ell)}{}_{,j}$，而 $\boldsymbol{D}$ 包含对角块 $\boldsymbol{A}_{i\,i}^{(L)}{}_{,}$。上文展示了 $L=2$ 层次深度下的分解 $\boldsymbol{A} = \sum_{\ell=1}^{L} \boldsymbol{A}^{(\ell)} + \boldsymbol{D}$，但在以下分析中， L 是任意的。设 $\boldsymbol{x} \in \mathbb{R}^{N}$，则

$$\begin{aligned}
\boldsymbol{A}\boldsymbol{x} &= \sum_{j=1}^{L} \boldsymbol{A}^{(j)}\boldsymbol{x} + \boldsymbol{D}\boldsymbol{x}, \\
\boldsymbol{A}^{(1)}\boldsymbol{x} &= \begin{pmatrix} \boldsymbol{A}_{1,2}^{(1)}\boldsymbol{x}_{2}^{(1)} \\ \boldsymbol{A}_{2,1}^{(1)}\boldsymbol{x}_{1}^{(1)} \end{pmatrix}, \quad \boldsymbol{x} = \begin{pmatrix} \boldsymbol{x}_{1}^{(1)} \\ \boldsymbol{x}_{2}^{(1)} \end{pmatrix}, \\
\boldsymbol{A}^{(j)}\boldsymbol{x} &= \begin{pmatrix} \boldsymbol{A}_{1,2}^{(j)}\boldsymbol{x}_{2}^{(j)} \\ \boldsymbol{A}_{2,1}^{(j)}\boldsymbol{x}_{1}^{(j)} \\ \vdots \\ \boldsymbol{A}_{2^{j}-1,2^{j}}^{(j)}\boldsymbol{x}_{2^{j}}^{(j)} \\ \boldsymbol{A}_{2^{j},2^{j}-1}^{(j)}\boldsymbol{x}_{2^{j}-1} \end{pmatrix}, \quad \boldsymbol{x} = \begin{pmatrix} \boldsymbol{x}_{1}^{(j)} \\ \boldsymbol{x}_{2}^{(j)} \\ \vdots \\ \boldsymbol{x}_{2^{j}-1}^{(j)} \\ \boldsymbol{x}_{2^{j}}^{(j)} \end{pmatrix},
\end{aligned}$$

由此可得以下表达式

$$\|\boldsymbol{A}^{(j)}\boldsymbol{x}\|_{2}^{2} = \sum_{k=1}^{2^{j-1}} \left(\|\boldsymbol{A}_{2\,k-1,2\,k}^{(j)}\boldsymbol{x}_{2\,k}^{(j)}\|_{2}^{2} + \|\boldsymbol{A}_{2\,k,2\,k-1}^{(j)}\boldsymbol{x}_{2\,k-1}^{(j)}\|_{2}^{2} \right).$$

˜现在假设 $\boldsymbol{A}$ 是 $\boldsymbol{A}$的 HODLR 近似，其对角 $\boldsymbol{D}$ 与 $\boldsymbol{A}$ 的对角相等，因此

$$\begin{aligned}
\left(\boldsymbol{A} - \tilde{\boldsymbol{A}}\right) &= \sum_{j=1}^{L} \Delta\boldsymbol{A}^{(j)}, \\
\Delta\boldsymbol{A}^{(j)} &:= \left(\boldsymbol{A}^{(j)} - \tilde{\boldsymbol{A}}^{(j)}\right).
\end{aligned}$$

这里假设每个非对角块都已近似到某个绝对容差 $\varepsilon > 0$，因此 $\|\Delta\boldsymbol{A}_{2\,k-1,\ 2\,k}^{(j)}\|_{2}$, $\|\Delta\boldsymbol{A}_{2\,k,\ 2\,k-1}^{(j)}\| \le \varepsilon$对于每个 $j = 1,\ 2,\ \ldots,\ L$和

$k = 1,\ 2,\ \ldots,\ 2^{j-1}$。对于 $\boldsymbol{x} \in \mathbb{R}^N$，我们有

$$
\begin{aligned}
&\| \left(\boldsymbol{A} - \tilde{\boldsymbol{A}} \right) \boldsymbol{x} \|_2 \leq \sum_{j=1}^{L} \| \Delta \boldsymbol{A}^{(j)}\, \boldsymbol{x} \|_2, \\
&\| \Delta \boldsymbol{A}^{(j)}\, \boldsymbol{x} \|_2 \quad = \sqrt{\sum_{k=1}^{2^{j-1}} \left(\| \Delta \boldsymbol{A}^{(j)}_{2\,k-1,2\,k}\, \boldsymbol{x}^{(j)}_{2\,k} \|_2^2 + \| \Delta \boldsymbol{A}^{(j)}_{2\,k,2\,k-1}\, \boldsymbol{x}^{(j)}_{2\,k-1} \|_2^2 \right)} \\
&\qquad\qquad \leq \sqrt{\sum_{k=1}^{2^{j-1}} \left(\varepsilon^2 \| \boldsymbol{x}^{(j)}_{2\,k} \|_2^2 + \varepsilon^2 \| \boldsymbol{x}^{(j)}_{2\,k-1} \|_2^2 \right)}, \\
&\| \Delta \boldsymbol{A}^{(j)}\, \boldsymbol{x} \|_2 \quad \leq \varepsilon \sqrt{\sum_{k=1}^{2^{j-1}} \left(\| \boldsymbol{x}^{(j)}_{2\,k} \|_2^2 + \| \boldsymbol{x}^{(j)}_{2\,k-1} \|_2^2 \right)} = \varepsilon \| \boldsymbol{x} \|_2, \\
&\| \left(\boldsymbol{A} - \tilde{\boldsymbol{A}} \right) \boldsymbol{x} \|_2 \leq \varepsilon\, L\, \| \boldsymbol{x} \|_2, \\
&\| \boldsymbol{A} - \tilde{\boldsymbol{A}} \|_2 \quad := \sup_{\boldsymbol{x} \neq \boldsymbol{0}} \left(\frac{\| \left(\boldsymbol{A} - \tilde{\boldsymbol{A}} \right) \boldsymbol{x} \|_2}{\| \boldsymbol{x} \|_2} \right) \leq \varepsilon\, L.
\end{aligned}
$$

8.3. 后验协方差的误差分析

考虑一个对称矩阵 $\boldsymbol{A} \in \mathbb{R}^{N \times N}$，其特征值下界大于 -1 ~，以及一个对称近似矩阵 $\boldsymbol{A}$，二者的差异为$\Delta \boldsymbol{A} = \boldsymbol{A} - \boldsymbol{A}$。我们用 $\lambda(\boldsymbol{S})$ 表示 $\boldsymbol{S}$ 的一般特征值，因此 $s_1 \leq \lambda(\boldsymbol{S}) \leq s_2$表示 $\boldsymbol{S}$ 的所有特征值均有下界 s_1 、上界 s_2。现在，在已知 $\| \Delta \boldsymbol{A} \|_2 = \varepsilon$ 的情况下，我们给出 $(\boldsymbol{I} + \boldsymbol{A})^{-1} - \left(\boldsymbol{I} + \tilde{\boldsymbol{A}} \right)^{-1}$ 的误差界，以便评估 HODLR 高斯化后验协方差的精度。当 $\boldsymbol{A}$ 如第3.2节所述为先验预条件化海森失配时，$\| (\boldsymbol{I} + \boldsymbol{A})^{-1} - \left(\boldsymbol{I} + \tilde{\boldsymbol{A}} \right)^{-1} \|_2$量化了 HODLR近似高斯化后验协方差与真实高斯化后验协方差之间的差异。

$$
\begin{aligned}
&(\boldsymbol{I} + \boldsymbol{A})^{-1} - \left(\boldsymbol{I} + \tilde{\boldsymbol{A}} \right)^{-1} = (\boldsymbol{I} + \boldsymbol{A})^{-1} - (\boldsymbol{I} + \boldsymbol{A} - \Delta \boldsymbol{A})^{-1} = \\
&(\boldsymbol{I} + \boldsymbol{A})^{-1} - \left((\boldsymbol{I} + \boldsymbol{A}) \left(\boldsymbol{I} - (\boldsymbol{I} + \boldsymbol{A})^{-1} \Delta \boldsymbol{A} \right) \right)^{-1} = \\
&(\boldsymbol{I} + \boldsymbol{A})^{-1} - \left(\boldsymbol{I} - (\boldsymbol{I} + \boldsymbol{A})^{-1} \Delta \boldsymbol{A} \right)^{-1} (\boldsymbol{I} + \boldsymbol{A})^{-1} = \\
&\left(\boldsymbol{I} - \left(\boldsymbol{I} - (\boldsymbol{I} + \boldsymbol{A})^{-1} \Delta \boldsymbol{A} \right)^{-1} \right) (\boldsymbol{I} + \boldsymbol{A})^{-1}.
\end{aligned}
$$

鉴于 $\| \Delta \boldsymbol{A} \|_2 = \varepsilon$，我们有

$$
\begin{aligned}
&-\varepsilon \leq \lambda \left(\Delta \boldsymbol{A} \right) \leq \varepsilon, \\
&-\varepsilon^* \leq \lambda \left((\boldsymbol{I} + \boldsymbol{A})^{-1} \Delta \boldsymbol{A} \right) \leq \varepsilon^*, \\
&\varepsilon^* := \varepsilon (1 + \lambda_{\min}(\boldsymbol{A}))^{-1}, \\
&1 + \varepsilon^* \geq \lambda \left(\boldsymbol{I} - (\boldsymbol{I} + \boldsymbol{A})^{-1} \Delta \boldsymbol{A} \right) \geq 1 - \varepsilon^*,
\end{aligned}
$$

接下来我们假设 $\varepsilon^* < 1$，因此 $\boldsymbol{I} - (\boldsymbol{I} + \boldsymbol{A})^{-1} \Delta \boldsymbol{A}$ 的特征值必然为正，且

$$(1 + \varepsilon^*)^{-1} \leq \lambda \left(\left(\boldsymbol{I} - (\boldsymbol{I} + \boldsymbol{A})^{-1} \Delta \boldsymbol{A} \right)^{-1} \right) \leq (1 - \varepsilon^*)^{-1} .$$

由此可得

$$\| (\boldsymbol{I} + \boldsymbol{A})^{-1} - \left(\boldsymbol{I} + \tilde{\boldsymbol{A}} \right)^{-1} \|_2 / \| (\boldsymbol{I} + \boldsymbol{A})^{-1} \|_2 \leq \left(1 - (1 + \varepsilon^*)^{-1} \right)$$

$$\| (\boldsymbol{I} + \boldsymbol{A})^{-1} - \left(\boldsymbol{I} + \tilde{\boldsymbol{A}} \right)^{-1} \|_2 / \| (\boldsymbol{I} + \boldsymbol{A})^{-1} \|_2 \leq \frac{\varepsilon^*}{1 + \varepsilon^*},$$

其中，与之前一样，$\varepsilon^* = \|\Delta \boldsymbol{A}\|_2 / (1 + \lambda_{\min}(\boldsymbol{A}))$.

致谢

作者感谢洛斯阿拉莫斯国家实验室的 Trevor Hillebrand 在搭建洪堡和格陵兰冰盖网格及数据集方面提供的帮助。本研究获得美国国家科学基金会 DMS-1840265 和 CAREER-1654311 项目资助，并通过 ProSPect SciDAC 项目获得支持。该项目由美国能源部（DOE）科学办公室的先进科学计算研究和生物与环境研究项目资助。本研究使用了国家能源研究科学计算中心（NERSC）的资源。该中心是美国能源部科学办公室的用户设施，依据合同编号 DE-AC02-05CH11231 运营；本研究使用的 NERSC 奖项编号为 ERCAP0020130。

免责声明

本文介绍客观的技术成果与分析。文中可能表达的任何主观观点或意见，并不一定代表美国能源部或美国政府的观点。桑迪亚国家实验室是一家多任务实验室，由桑迪亚国家技术与工程解决方案有限责任公司管理和运营；该公司是霍尼韦尔国际公司的全资子公司，负责为美国能源部国家核安全管理局提供服务，合同编号为 DE-NA-0003525。

参考文献

[1]Isaac T、Petra N、Stadler G 和 Ghattas O，2015年，《用于将不确定性从数据经由推断传播至预测的可扩展高效算法：及其在南极冰盖流动中的应用》《计算物理学杂志》**296** 348–368[2]Petra N、Martin J、Stadler G 和 Ghattas O，2014年，《无限维贝叶斯反问题的计算框架：第二部分：随机牛顿马尔可夫链蒙特卡罗及其在冰盖流动反问题中的应用》《*SIAM*科学计算杂志》**36** A1525–A1555[3]Spantini A、Solonen A、Cui T、Martin J、Tenorio L 和 Marzouk Y，2015年，《贝叶斯线性反问题的最优低秩近似》《*SIAM*科学计算杂志》**37**A2451–A2487

[4] Flath H P、Wilcox L C、Akçelik V、Hill J、van Bloemen Waanders B 和 Ghattas O，2011年，《基于低秩部分海森矩阵近似的大规模线性反问题贝叶斯不确定性量化快速算法》，SIAM Journal on Scientific Computing 33 407–432[5] Bui-Thanh T、Ghattas O、Martin J 和 Stadler G，2013年，《无限维贝叶斯反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》，SIAM Journal on Scientific Computing 35 A2494–A2523[6] Saibaba A K 和 Kitanidis P K，2015年，《地统计方法求解反问题中不确定性量化指标的快速计算》，《水资源进展》82 124–138[7] Martinsson P G，2011年，《一种快速随机化算法：计算矩阵的层次半可分表示》，SIAM Journal on Matrix Analysis and Applications 32 1251–1274[8] Alger N、Rao V、Meyers A、Bui-Thanh T 和 Ghattas O，2019年，《局部平移不变算子的可扩展无矩阵自适应乘积卷积近似》，SIAM Journal on Scientific Computing 41 A2296–A2328[9]Alger N、Hartland T、Petra N 和 Ghattas O，《具有局部支撑的非负积分核的算子的高效无矩阵点扩散函数近似及其在偏微分方程约束反问题海森矩阵中的应用》，待投稿[10] Zhu H、Li S、Fomel S、Stadler G 和 Ghattas O，2016年，《利用深度偏移先验信息，通过贝叶斯方法估计全波形反演的不确定性》，Geophysics 81 R307–R323[11] Lin L、Lu J 和 Ying L，2011年，《通过矩阵-向量乘法快速构造层次矩阵表示》，Journal of Computational Physics 230 4071–4087[12]Martinsson P G，2016年，《通过随机采样压缩秩结构矩阵》，SIAM Journal on Scientific Computing 38 A1959–A1986[13]Geoga C J、Anitescu M 和 Stein M L，2020年，《利用层次矩阵进行可扩展的高斯过程计算》，Journal of Computational and Graphical Statistics 29 227–237[14] Litvinenko A、Sun Y、Genton M G 和 Keyes D E，2019年，《利用层次矩阵近似大型空间数据集的似然》，Computational Statistics & Data Analysis 137 115–132[15] Ambartsumyan I、Boukaram W、Bui-Thanh T、Ghattas O、Keyes D、Stadler G、Turkiyyah G 和 Zampini S，2020年，《偏微分方程支配的反问题中海森矩阵的层次矩阵近似》，SIAM Journal on Scientific Computing 42 A3397–A3426[16] Hackbusch W，1999年，《基于 $\mathcal{H}$-矩阵的稀疏矩阵运算。第一部分：介绍》，Computing 62 89–108[17] Hackbusch W 和 B¨orm S，2002年，《通过自适应 $\mathcal{H}^2$-矩阵进行数据稀疏近似》，Computing 69 1–35[18] Tarantola A，2005年，《模型参数估计的反问题理论与方法》（SIAM）[19] Kaipio J 和 Somersalo E，2006年，《统计与计算反问题》第160卷（施普林格科学与商业媒体）[20] Stuart A M，2010年，《反问题：贝叶斯视角》，数值学报 19 451–559[21] Nocedal J 和 Wright S J，2006年，《数值优化》第2版（柏林、海德堡、纽约：施普林格出版社）[22] Borz`ı A 和 Schulz V，2011年，《偏微分方程支配系统的计算优化》（SIAM）[23] Gunzburger M D，2002年，《流动控制与优化展望》（SIAM）[24] Petra N 和 Sachs E W，2021年，《优化中的二阶伴随》，载于《数值分析与优化》，编者：Al-Baali M、Purnama A 和 Grandinetti L（尚：施普林格国际出版），第209–230页[25] Hastings W K，1970年，《使用马尔可夫链的蒙特卡罗抽样方法及其应用》，生物统计学 57 97–109[26] Robert C P、Casella G 和 Casella G，1999年，《蒙特卡罗统计方法》第2卷（施普林格）[27]Rudolf D 和 Sprungk B，2018年，《预条件Crank–Nicolson Metropolis算法的一种推广》，计算数学基础 18 309–343

[28]Pinski F J, Simpson G, Stuart A M and Weber H 2015年 无限维概率测度的 Kullback–Leibler 近似算法 《SIAM科学计算杂志》37 A2733–A2757[29] Kim K T, Villa U, Parno M, Marzouk Y, Ghattas O and Petra N 2021 hIPPYlib-MUQ：一种贝叶斯推断软件框架，用于在不确定性条件下将数据与复杂预测模型相结合 arXiv 预印本 arXiv:2112.00713[30] Ambikasaran S and Darve E 2013 一种 $\mathcal{O}(n \log n)$快速直接求解器，用于部分层次半可分矩阵 《科学计算杂志》57 477–501[31] Ambikasaran S, O'Neil M and Singh K R 2014 层次矩阵的快速对称分解及其应用 arXiv 预印本 arXiv:1405.0223[32] Halko N, Martinsson P G and Tropp J A 2011 利用随机性发现结构：用于构造近似矩阵分解的概率算法 《SIAM评论》53 217–288[33] Xi Y, Xia J and Chan R 2014 一种带结构化 LDL 分解更新的快速随机化特征值求解器 SIAM矩阵分析与应用期刊35 974–996[34] Boukaram W, Turkiyyah G and Keyes D 2019 用于通过矩阵-向量运算构造层次矩阵的随机化 GPU算法 《SIAM科学计算杂志》41 C339–C366[35] Gelfand I M and Fomin S 1963 变分法 (Dover)[36] Cuffey K M and Paterson W S B 2010 冰川物理学 (Academic Press)[37] Dukowicz J K, Price S F and Lipscomb W H 2010 基于最小作用量原理的冰盖动力学一致近似与边界条件 《冰川学杂志》56 480–496[38] Larour E, Seroussi H, Morlighem M and Rignot E 2012 使用冰盖系统模型（ISSM）进行大陆尺度、高阶、高空间分辨率冰盖建模 《地球物理研究杂志：地表》117[39] Morlighem M, Rignot E, Seroussi H, Larour E, Ben Dhia H and Aubry D 2010 利用全斯托克斯模型和较简单模型，通过控制方法推断的西南极洲派恩岛冰川基底阻力空间分布 《地球物理研究快报》37 L14502[40] Perego M, Price S and Stadler G 2014 耦合冰盖模型与地球系统模型的最优初始条件 《地球物理研究杂志：地表》119 1894–1917[41] Petra N, Zhu H, Stadler G, Hughes T J R and Ghattas O 2012 一种用于反演非线性斯托克斯冰盖模型中基底滑动和流变参数的非精确高斯–牛顿法 《冰川学杂志》58 889–903[42] Glen J W 1955 多晶冰的蠕变 《伦敦皇家学会会刊 A 辑：数学与物理科学》228 519–538[43] Pattyn F, Perichon L, Aschwanden A, Breuer B, de Smedt B, Gagliardini O, Gudmundsson G H, Hindmarsh R C A, Hubbard A, Johnson J V, Kleiner T, Konovalov Y, Martin C, Payne A J, Pollard D, Price S, Ruckamp M, Saito F, Soucek O, Sugiyama S and Zwinger T 2008 高阶和全斯托克斯冰盖模型的基准实验（ISMIP-HOM）《冰冻圈》2 95–108[44] Daon Y and Stadler G 2018 减轻边界条件对由椭圆型偏微分方程导出的协方差算子的影响 《反问题与成像》12 1083–1102[45] Lindgren F, Rue H and Lindstr¨om J 2011 高斯场与高斯马尔可夫随机场之间的显式联系：随机偏微分方程方法 《英国皇家统计学会杂志：B辑（统计方法学）》73 423–498[46] Hoffman M J, Perego M, Price S F, Lipscomb W H, Zhang T, Jacobsen D, Tezaur I, Salinger A G, Tuminaro R and Bertagna L 2018 MPAS-Albany陆冰模型（MALI）：一种使用沃罗诺伊网格进行地球系统建模的可变分辨率冰盖模型 《地球科学模型发展》11 3747–3780[47] Tezaur I K, Perego M, Salinger A G, Tuminaro R S and Price S F 2015 Albany/FELIX：一种并行、可扩展且稳健的有限元一阶斯托克斯近似冰盖求解器，专为

高级分析《地球科学模型发展》**8** 1197–1220[48] Liegeois K, Perego M and Hartland T 2022 PyAlbany：一个面向 C++ 多物理场求解器 Albany 的 Python 接口《计算与应用数学杂志》115037[49] Trilinos 项目团队 T 2020 *Trilinos* 项目网站网址 https://trilinos.github.io[50] Joughin I, Smith B, Howat I and Scambos T 2015 MEaSUREs 基于 InSAR 数据的格陵兰冰盖速度图，版本 2 网址 https://nsidc.org/data/NSIDC-0478/versions/2[51] Perego M 2022 面向冰盖模型初始化的大规模偏微分方程约束优化 *SIAM* 在线新闻[52] Hillebrand T R, Hoffman M J, Perego M, Price S F and Howat I M 2022 在近期加速和退缩观测的约束下，洪堡冰川对格陵兰北部至 2100 年海平面上升的贡献 《冰冻圈讨论》**2022年** 1–33