arXiv:2301.12019v1 [math.NA] 27 Jan 2023

超参数化线性贝叶斯逆问题的贪心传感器选择算法

Nicole Aretz[a]、Peng Chen[b]、Denise D. Degen[c]、Karen Veroy[d]

[a]奥登计算工程与科学研究所，德克萨斯大学奥斯汀分校，*201 E 24th St, Austin, TX 78712, USA*[b]佐治亚理工学院计算科学与工程学院，*756 W Peachtree St NW, Atlanta, GA30308, USA*[c]计算地球科学、地热学与储层地球物理学，亚琛工业大学，*Mathieustr. 30, 52074 Aachen, Germany*[d]埃因霍温理工大学数学与计算机科学系分析、科学计算与应用中心，*5612 AZ Eindhoven, The Netherlands*

摘要

我们研究一类由确定性超参数表征的线性贝叶斯逆问题的最优传感器布置。该超参数描述了可对所观测物理系统进行测量的不同配置。为通过一组传感器最大限度地降低系统模型的不确定性，初始传感器布置需考虑所有允许配置的非线性状态变化。我们通过一个可观测性系数来满足这一要求，该系数将后验分布的不确定性与传感器的选择直接关联起来。我们提出一种贪心传感器选择算法，通过正交匹配追踪迭代改进所有配置的可观测性系数。该算法支持显式相关噪声模型，即使候选传感器集合很大也不例外；同时，通过模型降阶，即便面对高维正向模型，算法仍能保持计算效率。我们在珀斯盆地的大规模地球物理模型上展示了该方法，并通过数值研究评估其相对于经典最优实验设计效用函数的最优性和可扩展性。

1. 引言

在反问题的贝叶斯方法中（参见 [1]），参数的不确定性通过概率分布进行描述。根据贝叶斯定理，当获得新信息时，对参数的先验信念会得到更新，从而使后验分布以更高的确定性描述该参数。从以下意义上说，贝叶斯后验分布是最佳的：它是后验分布与先验之间的相对熵，以及模型预测与实验数据之间的均方误差之和的唯一极小化点。噪声模型与测量值共同决定了后验分布相较于先验不确定性的降低程度。一个关键方面——尤其是对成本高昂的实验数据而言——

是如何选择测量值，以最佳地提高后验分布的可信度。因此，选择满足具体应用需求的适当传感器，是最优实验设计（OED）研究领域及其相关社群的一项重要目标。有关介绍，请参阅相关文献（例如 [3, 4, 5]）。

本文提出的分析与算法显著扩展了我们在 [6] 中提出的初步构想；在该文中，我们试图将 3D-VAR 的稳定性结果推广到概率贝叶斯设置。我们提出的算法与参数化背景数据弱（PBDW）方法的正交匹配追踪（OMP）算法 [8, 9] 以及经验插值方法（EIM）（[10, 11]）直接相关。针对偏微分方程（PDEs）上的线性贝叶斯反问题，与之密切相关的最优实验设计（OED）方法包括 [12, 13, 14, 15, 16, 17], ，主要涉及 A-OED 和 D-OED，且假设噪声不相关。近年来，这些方法也已扩展到非线性贝叶斯反问题，例如 [18, 19, 20, 21, 22], ；而 [23] 则实现了向相关噪声情形的推进。具体而言， [21, 22] 采用了与本文相似的算法思路，通过应用贪心算法来最大化期望信息增益。为应对偏微分方程模型带来的高维问题，常见策略是采用 [24] 中的框架进行离散化，并结合参数降维方法（例如 [25, 26, 27, 28, 29, 30, 31]）和用于不确定性量化（UQ）问题的模型降阶（MOR）方法（例如 [32, 33, 34, 35, 36]）。

本文考虑这样的逆问题设置：一个确定性的超参数描述预期的系统配置，例如材料属性或加载条件。每种配置都会以非线性方式改变模型，因此，对于任意测量数据，我们都会得到一族可能的后验分布。假设无论系统配置如何，数据都只能通过一组传感器获取，那么最优实验设计任务就变成了在所有超参数范围内，均匀降低每个后验分布的不确定性。对于高维模型而言，这项任务颇具挑战，因为：1）每种配置都需要单独求解模型，计算成本高昂；2）对于大量可用测量值，比较不同传感器时必须求逆相关的噪声协方差矩阵，而该矩阵可能是稠密的。本文基于 [6], ，应对了这两项挑战，并详细提出了一种即使在相关噪声模型下仍然高效的传感器选择算法。

主要贡献如下：首先，我们确定了一个可观测性系数，用以关联传感器选择与各后验分布的最大特征值。我们还分析了该系数对模型近似的敏感性。其次，我们对任意观测算子的噪声协方差矩阵进行分解，以便在通过添加附加传感器扩展观测算子时快速计算可观测性增益。第三，我们提出了一种传感器选择算法，通过从大量传感器中迭代构建观测算子，提高所有超参数下的可观测性系数。该算法适用于相关噪声模型，并通过高效利用 MOR 技术，只需对每个选定的传感器进行一次单次全阶模型

[1]例如，对于开发地热能的项目，开发成本（如钻井、增产和测试）占总预算的 50 – 70%（[2]）。由于每个钻孔的成本可能高达数百万美元，因此谨慎规划其位置至关重要。

评估。

尽管可观测性系数的主要思想和推导与[6], 类似，但本研究还包括：1）分析模型近似对可观测性系数的影响；2）详细说明处理相关噪声模型的计算方法；以及 3）全面讨论传感器选择算法中的各个步骤。此外，本文还使用珀斯盆地的大规模地球物理模型对所提出的方法进行了测试。

本文结构如下：第 2 节介绍超参数化反问题的设定，包括先验分布、噪声模型和正向模型的所有假设。第 3 节建立并分析可观测性系数与后验不确定性之间的联系。最后，我们在第 4 节提出传感器选择算法，该算法利用上述分析，即使在超参数化设定下，也能够选择出可提高可观测性系数的传感器。在第 5 节中，我们通过一个高维地球物理模型展示所提方法的适用性和可扩展性，并在第 6 节中进行总结。

2. 问题设定

令 $\mathcal{X}$ 为一个内积为 $\langle\cdot,\cdot\rangle_{\mathcal{X}}$ 、诱导范数为 $\|x\|_{\mathcal{X}}^2 := \langle x,x\rangle_{\mathcal{X}}$ 的希尔伯特空间。我们考虑在可变配置 θ 下，根据带噪测量值识别单个物理系统的未知状态 $x_{\text{true}}(\theta) \in \mathcal{X}$ 的问题。

$$\mathbf{d}(\theta) \approx [\ell_1(x_{\text{true}}(\theta)), \dots, \ell_K(x_{\text{true}}(\theta))]^T \in \mathbb{R}^K.$$

测量值由一组 K 个互不相同的传感器（或实验） $\ell_1, \dots, \ell_K \in \mathcal{X}'$ 获得。我们的目标是从大量可选项组成的传感器库 $\mathcal{L} \subset \mathcal{X}'$ 中选择这些传感器，以便优化在任意配置 θ 下通过其测量所获得的信息量。

超参数化正向模型

我们认为，未知状态x_{true} 由两类信息唯一确定：

- 描述支配物理规律不确定性的一个未知参数 $\mathbf{u}_{\text{true}} \in \mathbb{R}_M$ ，以及
- 一个超参数（或配置[2]） $\theta \in \mathcal{P} \subset \mathbb{R}_p$ ，用于描述系统在可控配置（例如材料属性或加载条件）下的依赖关系，其中 $\mathcal{P}$ 是一个给定的紧集，包含所有可能的配置。

对于任意给定的$\mathbf{u} \in \mathbb{R}^M$ 和 $\theta \in \mathcal{P}$，令$x_\theta(\mathbf{u}) \in \mathcal{X}$为抽象模型方程 $\mathcal{M}_\theta(x_\theta(\mathbf{u});\mathbf{u}) = 0$ 的解，并假设映射$\mathbf{u} \to x_\theta(\mathbf{u})$ 是定义良好的、线性的，

[2]我们将 θ 交替称为超参数或配置，以强调其在数学模型中的作用或物理解释。

且关于$\mathbf{u}$一致连续，即

$$\exists\, \bar{\eta} > 0: \qquad \bar{\eta}(\theta) := \sup_{\mathbf{u}\in\mathbb{R}^M} \frac{\|x_\theta(\mathbf{u})\|_{\mathcal{X}}}{\|\mathbf{u}\|_{\Sigma_{\text{pr}}^{-1}}} < \bar{\eta} \qquad \forall\, \theta \in \mathcal{P}. \tag{1}$$

备注 1. 虽然我们假设$\mathbf{u}_{\text{true}}$属于欧几里得空间 $\mathbb{R}^M$，但也可以通过将其仿射变换到适当的基上来考虑其他线性空间（见 *[12, 37]*）。对于无限维空间，我们首先进行离散化，并对伴随算子作适当处理（参见 *[24]*）。

备注 2. 通过保持模型方程的一般性，我们强调了本方法适用于广泛的问题。例如，可以通过将 $\mathcal{X}$ 选为博赫纳空间或其离散化来处理时变状态（参见 *[38]*）。我们也不对 $\mathcal{X}$的维数作正式限制，不过任何实现都依赖于能否以足够的精度计算 $x_\theta(\mathbf{u})$。为此，我们指出，第 *3.2* 节中的分析可用于确定离散化误差如何影响传感器选择中的可观测性准则。

遵循逆问题的概率方法，我们通过一个具有高斯先验 $\mu_{\text{pr}} = \mathcal{N}\left(\mathbf{u}^{\text{pr}}, \Sigma_{\text{pr}}\right)$的随机变量$\mathbf{u}$，来表达配置 θ 中任意$x_{\text{true}} = x_\theta(\mathbf{u}_{\text{true}})$的初始不确定性$\mathbf{u}_{\text{true}} = \mathbf{u}_{\text{true}}(\theta)$，其中$\mathbf{u}_{\text{pr}} \in \mathbb{R}^M$ 是先验均值，$\Sigma_{\text{pr}} \in \mathbb{R}^{M\times M}$ 是对称正定（s.p.d.）协方差矩阵。后者定义了内积 $\langle\cdot,\cdot\rangle_{\Sigma_{\text{pr}}^{-1}}$及其诱导范数 $\|\cdot\|_{\Sigma_{\text{pr}}^{-1}}$，如下所示

$$\langle \mathbf{u}, \mathbf{v}\rangle_{\Sigma_{\text{pr}}^{-1}} := \mathbf{u}^T \Sigma_{\text{pr}}^{-1} \tilde{\mathbf{u}}, \qquad \|\mathbf{u}\|^2_{\Sigma_{\text{pr}}^{-1}} := \langle \mathbf{u}, \mathbf{u}\rangle_{\Sigma_{\text{pr}}^{-1}}, \qquad \forall\, \mathbf{u}, \mathbf{v} \in \mathbb{R}^M. \tag{2}$$

根据这些定义，μ^{pr} 的概率密度函数（pdf）为

$$\pi_{\text{pr}}(\mathbf{u}) = \frac{1}{\sqrt{(2\pi)^M \det \Sigma_{\text{pr}}}} \exp\left(-\frac{1}{2}\|\mathbf{u} - \mathbf{u}_{\text{pr}}\|^2_{\Sigma_{\text{pr}}^{-1}}\right).$$

为简单起见，我们假设 $\{\mathbf{u}_{\text{true}}(\theta)\}_{\theta\in\mathcal{P}}$ 是$\mathbf{u}$ 的独立实现，因此可以对所有 θ 采用相同的先验，而无需考虑不同配置下可能存在的测量历史。

传感器库和噪声模型

为了测量未知状态 $\{x_{\text{true}}(\theta)\}_\theta$，我们将任意线性泛函 $\ell \in \mathcal{X}'$ 称为传感器，而将其作用于状态$x \in \mathcal{X}$ 所得的结果称为其测量值 $\ell(x) \in \mathbb{R}$。我们将实际物理状态x_{true}的实验测量$d_\ell \in \mathbb{R}$ 建模为$d_\ell = \ell(x_{\text{true}}) + \varepsilon_\ell$，其中 $\varepsilon_\ell \sim \mathcal{N}(0,$**协方差**$(\varepsilon_\ell, \varepsilon_\ell))$ 是一个高斯随机变量。我们允许不同传感器测量中的噪声彼此相关，并以已知的协方差函数**协方差**描述。略微滥用符号时，我们将**协方差**$: \mathcal{L}\times\mathcal{L} \to \mathbb{R}$,**协方差**$(\ell_i, \ell_j) :=$ **协方差**$(\varepsilon_{\ell_i}, \varepsilon_{\ell_j})$ 视为传感器库上的对称双线性型。于是，传感器的任意有序子集 $\mathcal{S} = \{\ell_1, \ldots, \ell_K\} \subset \mathcal{L}$都可以构成一个（线性且连续的）观测算子，其形式为

$$L := [\ell_1, \ldots, \ell_K]^T : \mathcal{X} \to \mathbb{R}^K, \qquad Lx := [\ell_1(x), \ldots, \ell_K(x)]^T.$$

L 的实验测量具有如下形式

$$\mathbf{d} = [\ell_1(x_{\text{true}}) + \varepsilon_{\ell_1}, \ldots, \ell_K(x_{\text{true}}) + \varepsilon_{\ell_K}]^T = Lx_{\text{true}} + \varepsilon \qquad \text{with} \qquad \varepsilon = [\varepsilon_{\ell_1}, \ldots, \varepsilon_{\ell_K}]^T \sim \mathcal{N}(\mathbf{0}, \sigma^2\Sigma_L), \tag{3}$$

其中， $\sigma^2\Sigma_L$ 是通过以下方式定义的噪声协方差矩阵

$$\Sigma_L \in \mathbb{R}^{K\times K}, \qquad \text{such that} \qquad \left[\sigma^2\Sigma_L\right]_{i,j} := \mathbf{cov}(\ell_j, \ell_i) = \mathbf{cov}(\varepsilon_{\ell_j}, \varepsilon_{\ell_i}) \tag{4}$$

其中包含辅助缩放参数[3] $\sigma^2 > 0$。我们假设，库 $\mathcal{L}$ 和噪声协方差函数**cov** 经过选择，使得对于 $\mathcal{L}$中的任意传感器组合， Σ_L 都是对称正定的。由此可得依赖于 L的内积及其诱导范数

$$\left\langle \mathbf{d}, \tilde{\mathbf{d}} \right\rangle_{\Sigma_L^{-1}} := \mathbf{d}^T\Sigma_L^{-1}\tilde{\mathbf{d}}, \qquad \|\mathbf{d}\|^2_{\Sigma_L^{-1}} := \langle \mathbf{d}, \mathbf{d}\rangle_{\Sigma_L^{-1}}, \qquad \forall \mathbf{d}<\text{style id="2"}>, \tilde{\mathbf{d}} \in \mathbb{R}^K<\text{style id="6"}>。 \tag{5}$$

相对于此范数进行度量时，任意（归一化）状态的最大观测因此为

$$\gamma_L := \sup_{\|x\|_{\mathcal{X}}=1} \|Lx\|_{\Sigma_L^{-1}} = \sup_{x\in\mathcal{X}} \frac{\|Lx\|_{\Sigma_L^{-1}}}{\|x\|_{\mathcal{X}}}. \tag{6}$$

我们在第4.1节证明，尽管范数发生变化，随着附加传感器的加入， γ_L 会增大<style id="2">L 的欠展开，因此它有界于 $\gamma_L \le \gamma_{\mathcal{L}}$。

我们还定义参数到可观测量映射

$$G_{L,\theta} : \mathbb{R}^M \to \mathbb{R}^K, \quad \text{such that} \quad G_{L,\theta}(\mathbf{u}) := Lx_\theta(\mathbf{u}). \tag{7}$$

根据上述假设——特别是x 关于**u** 的线性性和一致连续性（1）——映射$G_{L,\theta}$ 关于**u**是线性的且一致有界。我们记$\mathbf{G}_{L,\theta} \in \mathbb{R}^{K\times M}$为其相对于单位基 $\{\mathbf{e}_m\}_{m=1}^M$的矩阵表示。对于参数$\mathbf{u} \in \mathbb{R}^M$ 和系统配置 θ，通过观测算子L 得到$\mathbf{d} \in \mathbb{R}^K$ 的似然为

$$\Phi_L\left(\mathbf{d} \,\middle|\, \mathbf{u}, \theta\right) := \frac{1}{\sqrt{2^K \det\Sigma_L}} \exp\left(-\frac{1}{2\sigma^2} \left\|\mathbf{d} - G_{L,\theta}(\mathbf{u})\right\|^2_{\Sigma_L^{-1}}\right).$$

请注意，$G_{L,\theta}$ 和$\mathbf{G}_{L,\theta}$ 可能非线性地依赖于 θ。

后验分布

一旦获得含噪测量数据 $\mathbf{d} \approx Lx_{\text{true}}(\theta)$，贝叶斯定理便给出后验概率密度函数：

$$\pi^{L,\theta}_{\text{post}}(\mathbf{u} \,|\, \mathbf{d}) = \frac{1}{Z(\theta)} \exp\left(-\frac{1}{2\sigma^2} \left\|G_{L,\theta}(\mathbf{u}) - \mathbf{d}\right\|^2_{\Sigma_L^{-1}} - \frac{1}{2}\|\mathbf{u} - \mathbf{u}_{\text{pr}}\|^2_{\Sigma_{\text{pr}}^{-1}}\right) \propto \pi_{\text{pr}}(\mathbf{u}) \cdot \Phi_L\left(\mathbf{d} \,\middle|\, \mathbf{u}, \theta\right), \tag{8}$$

[3]我们在此引入 σ^2 作为附加变量，以便于讨论第13节中的缩放问题。然而，不失一般性（w.l.o.g.），我们可以令 $\sigma^2 = 1$。

其归一化常数为

$$Z(\theta) := \int_{\mathbb{R}^p} \exp\left(-\frac{1}{2\sigma^2}\left\|G_{L,\theta}(\mathbf{u}) - \mathbf{d}\right\|^2_{\Sigma_L^{-1}}\right) d\mu_{\text{pr}}.$$

由于参数到可观测量映射具有线性性，后验测度 $\mu^{L,\,\theta}{}_{\text{post}}$ 服从高斯分布

$$\mu^{L,\theta}_{\text{post}} = \mathcal{N}(\mathbf{u}^{L,\theta}_{\text{post}}(\mathbf{d}), \Sigma^{L,\theta}_{\text{post}})$$

其均值和协方差矩阵已知（参见 [1]）

$$\mathbf{u}^{L,\theta}_{\text{post}}(\mathbf{d}) = \Sigma^{L,\theta}_{\text{post}}\left(\tfrac{1}{\sigma^2}\mathbf{G}^T_{L,\theta}\Sigma_L^{-1}\mathbf{d} + \Sigma_{\text{pr}}^{-1}\mathbf{u}_{\text{pr}}\right) \qquad \in \mathbb{R}^M, \tag{9}$$

$$\Sigma^{L,\theta}_{\text{post}} = \left(\tfrac{1}{\sigma^2}\mathbf{G}^T_{L,\theta}\Sigma_L^{-1}\mathbf{G}_{L,\theta} + \Sigma_{\text{pr}}^{-1}\right)^{-1} \qquad \in \mathbb{R}^{M\times M}. \tag{10}$$

因此，后验分布 $\mu^{L,\,\theta}{}_{\text{post}}$ 不仅取决于传感器的选择，还取决于获取测量值时所采用的配置 θ。因此，为了通过单个与 θ 无关的观测算子L降低所有可能后验分布中的不确定性，构造L 时应考虑所有可能观测x_{true} 的允许的配置 $\theta \in \mathcal{P}$。

备注 3。 $x_\theta(\mathbf{u})$ 关于$\mathbf{u}$ 的线性性是一项强假设，它决定了后验分布为高斯分布。然而，结合超参数 θ，这里的设定也可以重新解释为针对非线性状态映射 $\theta \mapsto x(\theta)$的拉普拉斯近似（参见 *[39, 21, 40]*）。因此，这里介绍的传感器选择是针对非线性前向模型进行最优实验设计的中间步骤。

3. 可观测性系数

本节将说明观测算子 L 中传感器的选择及其相关噪声协方差矩阵 Σ_L 如何影响后验分布 $\mu^{L,\theta}{}_{\text{post}}$, $\theta \in \mathcal{P}$中的不确定性。我们将确定一个可观测性系数，该系数能够界定相对于 L的后验协方差矩阵 $\Sigma^{L,\theta}{}_{\text{post}}$, $\theta \in \mathcal{P}$的特征值，并有助于实现第 4 节提出的传感器选择算法。

3.1. 后验协方差矩阵的特征值

对于任意配置 $\theta \in \mathcal{P}$，后验分布 $\pi^{L,\theta}{}_{\text{post}}$ 的不确定性可由后验协方差矩阵 $\Sigma^{L,\theta}{}_{\text{post}}$唯一表征；而后验协方差矩阵又通过参数到可观测量映射$G_{L,\theta}$ 与噪声协方差矩阵 Σ_L，和观测算子 L 相关联。为衡量 $\Sigma^{L,\theta}{}_{\text{post}}$中的不确定性，最优实验设计文献提出了多种不同的效用函数，并建议通过最小化这些效用函数来优化传感器选择。这些效用函数中有许多都可以用特征值表示

$\lambda^{\theta,1}{}_L \geq \cdots \geq \lambda^{\theta,M}{}_L > 0$ 的 $\Sigma^{L,\theta}{}_{\text{post}}$，例如，

$$\begin{aligned}
&\text{A-OED:} && \operatorname{trace}(\Sigma^{L,\theta}_{\text{post}}) = \sum_{m=1}^{M} \lambda_L^{\theta,m} && \text{(mean variance)}\\
&\text{D-OED:} && \det(\Sigma^{L,\theta}_{\text{post}}) = \prod_{m=1}^{M} \lambda_L^{\theta,m} && \text{(volume)}\\
&\text{E-OED:} && \lambda_{\max}(\Sigma^{L,\theta}_{\text{post}}) = \lambda_L^{\theta,1} && \text{(spectral radius).}
\end{aligned}$$

在实践中，效用函数的选择取决于具体应用。例如，在 E-最优实验设计（E-OED）中，会避免后验分布的不确定性椭球沿任一方向过度延伸；而 D-OED 则最小化不确定性椭球的总体体积，而不考虑任一参数方向上的不确定性。有关详细介绍及其他最优实验设计准则，请参阅 [3]。

考虑这样一种超参数化设定：每种配置 θ 都会影响后验不确定性。我们希望选择单个观测算子 L，使所选效用函数对于所有配置 $\theta \in \mathcal{P}$ 都保持较小。例如，对于 E-OED，最小化

$$\min_{\ell_1,\ldots,\ell_K \in \mathcal{L}} \ \max_{\theta \in \mathcal{P}} \ \lambda_{\max}(\Sigma^{L,\theta}_{\text{post}}) \quad \text{such that} \quad L = [\ell_1, \ldots, \ell_K]^T$$

可保证对于任意 $\theta \in \mathcal{P}$，每个后验协方差矩阵 $\Sigma^{L,\ \theta}$ 后验的最长轴都有相同的上界。这里的难点在于，对 $\mathcal{P}$ 进行最小化需要反复进行耗费大量计算的模型评估，以计算许多不同配置 θ 下的效用函数。因此，接下来我们引入后验特征值的一个上界，该上界可通过可观测性准则进行优化，所需的模型求解次数少得多。对该上界进行优化，可通过后验特征值间接降低不同的效用函数。

回忆一下，$\Sigma^{L,\theta}$ 后验 是对称正定的，令 $\{\psi_m\}_{m=1}^{M}$ 为 $\Sigma^{L,\ \theta}$ 后验的标准正交特征向量基，即 $\psi_m^T \psi_n = \delta_{m,n}$ 并且

$$\Sigma^{L,\theta}_{\text{post}} \psi_m = \lambda_L^{\theta,m} \psi_m \qquad\qquad m = 1, \ldots, M. \tag{11}$$

利用表示式 (10)，任意特征值 $\lambda^{\theta,m}{}_L$ 都可以写成如下形式

$$\frac{1}{\lambda_L^{\theta,m}} = \psi_m^T \left[\Sigma^{L,\theta}_{\text{post}}\right]^{-1} \psi_m = \psi_m^T \left[\tfrac{1}{\sigma^2} \mathbf{G}_{L,\theta}^T \Sigma_L^{-1} \mathbf{G}_{L,\theta} + \Sigma_{\text{pr}}^{-1}\right] \psi_m = \frac{1}{\sigma^2} \left\| G_{L,\theta}(\psi_m) \right\|^2_{\Sigma_L^{-1}} + \|\psi_m\|^2_{\Sigma_{\text{pr}}^{-1}}. \tag{12}$$

由于 ψ_m 通过 (11) 隐式依赖于 L 和 θ，因此我们无法直接使用这一表示来优化 L。为了消除对 ψ_m 的依赖，我们用先验协方差矩阵 Σ_{pr} 的最大特征值来界定——$\|\psi_m\|^2_{\Sigma_{\text{pr}}^{-1}} \geq \frac{1}{\lambda_{\text{pr}}^{\max}}$。同样，我们定义

$$\beta_G(\theta) := \inf_{\mathbf{u} \in \mathbb{R}^M} \frac{\left\| G_{L,\theta}(\mathbf{u}) \right\|_{\Sigma_L^{-1}}}{\|\mathbf{u}\|_{\Sigma_{\text{pr}}^{-1}}} = \inf_{\mathbf{u} \in \mathbb{R}^M} \frac{\|L x_\theta(\mathbf{u})\|_{\Sigma_L^{-1}}}{\|\mathbf{u}\|_{\Sigma_{\text{pr}}^{-1}}}, \tag{13}$$

为相对于先验协方差范数的参数 $\mathbf{u}$ 观测的最小比值。由 (12) 和 (13) 可得上界

$$\lambda_L^{\theta,m} = \left(\frac{1}{\sigma^2} \frac{\left\| G_{L,\theta}(\psi_m) \right\|_{\Sigma_L^{-1}}^2}{\|\psi_m\|_{\Sigma_{\text{pr}}^{-1}}^2} + 1 \right)^{-1} \|\psi_m\|_{\Sigma_{\text{pr}}^{-1}}^{-2} \leq \left(\frac{1}{\sigma^2} \beta_G(\theta)^2 + 1 \right)^{-1} \lambda_{\text{pr}}^{\max}.$$

从几何上看，这个界意味着，后验不确定性椭球周围外球的半径 $\lambda^{\theta,1}{}_L$ 至少比先验不确定性椭球的半径小因子 $\left(\frac{1}{\sigma^2} \beta_G(\theta)^2 + 1 \right)^{-1}$。因此，选择使 $\min_\theta \beta_G(\theta)$ 最大的 L，就能使包含所有不确定性椭球的外球最小（即对任意 $\theta \in \mathcal{P}$）。正如预期，当测量噪声较小时，数据更可信（$\sigma^2 \ll 1$），L 的影响最显著；随着噪声水平增大（$\sigma^2 \gg 1$），其影响会减弱。

3.2. 参数限制

$\beta_G(\theta)$ 的一个重要性质是：当 $K < M$ 时，$\beta_G(\theta) = 0$；也就是说，L 中的传感器数量少于参数维度的数量。在这种情况下，在迭代算法的前 $M - 1$ 步中，$\beta_G(\theta)$ 无法区分不同传感器；一般而言，当预期选择的传感器总数少于 M 个时，也会出现这种情况。对于中等维度的参数空间（$M \in O(10)$），我们通过将 $\mathbf{u}$ 限制在由前 $\min\{K, M\}$ 个特征向量张成的子空间 $\text{span}\{\varphi_1, \ldots \varphi_{\min\{K,M\}}\} \subset \mathbb{R}^M$ 中来缓解这一问题。这些特征向量对应于 Σ_{pr} 最大的特征值，即先验不确定性最大的子空间。对于高维参数空间，或模型 $\mathcal{M}_\theta$ 具有非平凡零空间的情况，我们进一步对 $\beta_G(\theta)$ 进行界定

$$\beta_G(\theta) = \inf_{\mathbf{u} \in \mathbb{R}^M} \frac{\|Lx_\theta(\mathbf{u})\|_{\Sigma_L^{-1}}}{\|x_\theta(\mathbf{u})\|_{\mathcal{X}}} \frac{\|x_\theta(\mathbf{u})\|_{\mathcal{X}}}{\|\mathbf{u}\|_{\Sigma_{\text{pr}}^{-1}}} \geq \inf_{x \in \mathcal{W}_\theta} \frac{\|Lx\|_{\Sigma_L^{-1}}}{\|x\|_{\mathcal{X}}} \inf_{\mathbf{u} \in \mathbb{R}^M} \frac{\|x_\theta(\mathbf{u})\|_{\mathcal{X}}}{\|\mathbf{u}\|_{\Sigma_{\text{pr}}^{-1}}} = \beta_{L|\mathcal{W}}(\theta)\, \underline{\eta}(\theta) \tag{14}$$

其中，我们将所有可达状态构成的线性空间定义为 $\mathcal{W}_\theta$

$$\mathcal{W}_\theta := \{x_\theta(\mathbf{u}) \in \mathcal{X} : \ \mathbf{u} \in \mathbb{R}^M\}$$

以及系数

$$\beta_{L|\mathcal{W}}(\theta) := \inf_{x \in \mathcal{W}_\theta} \frac{\|Lx\|_{\Sigma_L^{-1}}}{\|x\|_{\mathcal{X}}}, \qquad \underline{\eta}(\theta) := \inf_{\mathbf{u} \in \mathbb{R}^M} \frac{\|x_\theta(\mathbf{u})\|_{\mathcal{X}}}{\|\mathbf{u}\|_{\Sigma_{\text{pr}}^{-1}}}. \tag{15}$$

$\underline{\eta}(\theta)$ 的值描述了参数 $\mathbf{u}$ 相对于其先验诱导范数 $\|\mathbf{u}\|_{\Sigma_{\text{pr}}^{-1}}$ 所能实现的最小状态变化。它可以筛除那些对状态 $x_\theta(\mathbf{u})$ 影响很小的参数方向。相比之下，可观测性系数 $\beta_{L|\mathcal{W}}(\theta)$ 仅通过 $\mathcal{W}_\theta$ 间接依赖于先验；它量化了相对于其范数，在 $\mathcal{W}_\theta$ 中任意状态上可获得的最少信息量（根据噪声模型度量）。未来的工作将研究如何在选择使 $\beta_{L|\mathcal{W}}(\theta)$ 最大化的传感器之前，基于 $\underline{\eta}(\theta)$ 对参数空间进行最优限制。在类似情境下，现有的参数降维方法包括 [28, 41, 42, 27]。然而，在本文中，我们仅关注最大化 $\beta_G(\theta)$，进而最大化 $\beta_{L|\mathcal{W}}(\theta)$，并在下文中假设 M 足够小，且 $\underline{\eta} := \inf_{\theta \in \mathcal{P}} \underline{\eta}(\theta) > 0$ 有正下界。

3.3 模型近似下的可观测性

要优化可观测性系数 $\beta_G(\theta)$ 或 $\beta_{L|\mathcal{W}}(\theta)$，必须针对许多不同的配置 $\theta \in \mathcal{P}$ 进行计算。不断累积的计算成本促使人们采用 降阶 替代模型；与原始 全阶 模型相比，这类模型通常能显著节省计算量。然而，这会导致状态近似误差。因此，接下来我们将量化状态近似误差对可观测性系数 $\beta_G(\theta)$ 和 $\beta_{L|\mathcal{W}}(\theta)$ 的影响。关于在反问题中以替代模型 $\mathcal{M}_\theta$ 整个模型时后验分布变化的分析，可参见 [1]。

˜假设存在一个降阶替代模型 $\mathcal{M}_\theta(\tilde{x}_\theta(\mathbf{u}); \mathbf{u}) = 0$，对于任意配置 $\theta \in \mathcal{P}$ 和参数 $\mathbf{u} \in \mathbb{R}^M$，该模型都能给出唯一的解 $\tilde{x}_\theta(\mathbf{u}) \in \mathcal{X}$，使得

$$\|x_\theta(\mathbf{u}) - \tilde{x}_\theta(\mathbf{u})\|_{\mathcal{X}} \le \varepsilon_\theta \|x_\theta(\mathbf{u})\|_{\mathcal{X}} \quad \text{with accuracy} \quad 0 \le \varepsilon_\theta \le \varepsilon < 1. \tag{16}$$

与式（13）和（15）类似，我们定义降阶可观测性系数

$$\tilde{\beta}_G(\theta) := \inf_{\mathbf{u}\in\mathbb{R}^M} \frac{\|L\tilde{x}_\theta(\mathbf{u})\|_{\Sigma_L^{-1}}}{\|\mathbf{u}\|_{\Sigma_{\text{pr}}^{-1}}}, \qquad \tilde{\beta}_{L|\mathcal{W}}(\theta) := \inf_{\mathbf{u}\in\mathbb{R}^M} \frac{\|L\tilde{x}_\theta(\mathbf{u})\|_{\Sigma_L^{-1}}}{\|\tilde{x}_\theta(\mathbf{u})\|_{\mathcal{X}}} \tag{17}$$

以量化代理状态的最小观测值。对于许多应用，可以选择一种降阶模型，其解的计算成本显著降低˜ ˜，使得计算 $\beta_G(\theta)$ 和 $\beta_{L|\mathcal{W}}(\theta)$ 的成本远低于其全阶对应量 $\beta_G(\theta)$ 和 $\beta_{L|\mathcal{W}}(\theta)$。由于此类替代模型的构建很大程度上取决于具体应用，我们请读者参阅相关文献（例如 [43, 44, 45, 46, 47]）了解具体方法。

回顾式（6）中 γ_L 的定义，我们首先对替代模型˜的可观测性系数 $\beta_{L|\mathcal{W}}(\theta)$ 对全阶可观测性系数 $\beta_{L|\mathcal{W}}(\theta)$ 的近似精度进行界定。

命题 1. 设 $\eta(\theta) > 0$ 成立，并设 $\tilde{x}_\theta(\mathbf{u}) \in \mathcal{X}$ 是 $x_\theta(\mathbf{u})$ 的一个近似，使得对所有 $\theta \in \mathcal{P}$,$\mathbf{u} \in \mathbb{R}^M$，(16) 均成立。于是

$$(1 - \varepsilon_\theta)\,\tilde{\beta}_{L|\mathcal{W}}(\theta) - \gamma_L\varepsilon_\theta \;\le\; \beta_{L|\mathcal{W}}(\theta) \;\le\; (1 + \varepsilon_\theta)\,\tilde{\beta}_{L|\mathcal{W}}(\theta) + \gamma_L\varepsilon_\theta. \tag{18}$$

证明. 任取$\mathbf{u} \in \mathbb{R}^M \setminus \{\mathbf{0}\}$。利用 (16) 和（反向）三角不等式，可得如下界

$$\frac{\|\tilde{x}_\theta(\mathbf{u})\|_{\mathcal{X}}}{\|x_\theta(\mathbf{u})\|_{\mathcal{X}}} \ge \frac{\|x_\theta(\mathbf{u})\|_{\mathcal{X}} - \|x_\theta(\mathbf{u}) - \tilde{x}_\theta(\mathbf{u})\|_{\mathcal{X}}}{\|x_\theta(\mathbf{u})\|_{\mathcal{X}}} \ge 1 - \varepsilon_\theta. \tag{19}$$

注意，此处$\eta(\theta) > 0$ 蕴含 $\|x_\theta(\mathbf{u})\|_{\mathcal{X}} > 0$，因此商确实有定义。现在可以将观测与状态之比从下方界定为

$$\begin{aligned}
\frac{\|Lx_\theta(\mathbf{u})\|_{\Sigma_L^{-1}}}{\|x_\theta(\mathbf{u})\|_{\mathcal{X}}} &\ge \frac{\|L\tilde{x}_\theta(\mathbf{u})\|_{\Sigma_L^{-1}}}{\|x_\theta(\mathbf{u})\|_{\mathcal{X}}} - \frac{\|L(x_\theta(\mathbf{u}) - \tilde{x}_\theta(\mathbf{u}))\|_{\Sigma_L^{-1}}}{\|x_\theta(\mathbf{u})\|_{\mathcal{X}}} \\
&\ge \frac{\|\tilde{x}_\theta(\mathbf{u})\|_{\mathcal{X}}}{\|x_\theta(\mathbf{u})\|_{\mathcal{X}}} \frac{\|L\tilde{x}_\theta(\mathbf{u})\|_{\Sigma_L^{-1}}}{\|\tilde{x}_\theta(\mathbf{u})\|_{\mathcal{X}}} - \gamma_L \frac{\|x_\theta(\mathbf{u}) - \tilde{x}_\theta(\mathbf{u})\|_{\mathcal{X}}}{\|x_\theta(\mathbf{u})\|_{\mathcal{X}}} \\
&\ge (1 - \varepsilon_\theta) \frac{\|L\tilde{x}_\theta(\mathbf{u})\|_{\Sigma_L^{-1}}}{\|\tilde{x}_\theta(\mathbf{u})\|_{\mathcal{X}}} - \gamma_L\varepsilon_\theta \\
&\ge (1 - \varepsilon_\theta)\tilde{\beta}_{L|\mathcal{W}}(\theta) - \gamma_L\varepsilon_\theta,
\end{aligned}$$

其中应用了反三角不等式、定义（6）、界（16）、˜（19），以及 $\beta_{L|\mathcal{W}}(\theta)$ 的定义（17）。由于 $\mathbf{u}$ 是任意的，由 $\beta_{L|\mathcal{W}}(\theta)$ 的定义（13）可得（18）中的下界。（18）中的上界可类似推出。 □

对于参数到可观测量映射$G_{L,\theta}$ 及其近似$\mathbf{u} \mapsto L\tilde{x}_\theta(\mathbf{u})$的可观测性，我们可以得到类似的界。该界使用$x_\theta : \mathbf{u} \mapsto x_\theta(\mathbf{u})$的范数 $\eta(\theta)$，其中它是从参数到状态空间的映射，见式（1）。

命题 2. 设 $\tilde{x}_\theta(\mathbf{u}) \in \mathcal{X}$ 是 $x_\theta(\mathbf{u})$ 的一个近似，且对所有$\theta \in \mathcal{P}\mathbf{u} \in \mathbb{R}^M$，式(16) 均成立。那么

$$\tilde{\beta}_G(\theta) - \gamma_L\overline{\eta}(\theta)\varepsilon_\theta \le \beta_G(\theta) \le \tilde{\beta}_G(\theta) + \gamma_L\overline{\eta}(\theta)\varepsilon_\theta. \tag{20}$$

证明。 任取$\mathbf{u} \in \mathbb{R}^M \setminus \{\mathbf{0}\}$，则

$$\begin{aligned}
\|Lx_\theta(\mathbf{u})\|_{\Sigma_L^{-1}} &\ge \|L\tilde{x}_\theta(\mathbf{u})\|_{\Sigma_L^{-1}} - \|L(x_\theta(\mathbf{u}) - \tilde{x}_\theta(\mathbf{u}))\|_{\Sigma_L^{-1}} \\
&\ge \|L\tilde{x}_\theta(\mathbf{u})\|_{\Sigma_L^{-1}} - \gamma_L \|x_\theta(\mathbf{u}) - \tilde{x}_\theta(\mathbf{u})\|_{\mathcal{X}} \\
&\ge \|L\tilde{x}_\theta(\mathbf{u})\|_{\Sigma_L^{-1}} - \gamma_L\varepsilon_\theta \|x_\theta(\mathbf{u})\|_{\mathcal{X}} \\
&\ge \|L\tilde{x}_\theta(\mathbf{u})\|_{\Sigma_L^{-1}} - \gamma_L\varepsilon_\theta\overline{\eta}(\theta)\|\mathbf{u}\|_{\Sigma_{\mathrm{pr}}^{-1}},
\end{aligned}$$

这里我们先使用反三角不等式，再应用式（6）、（16）和（1）。两边除以 $\|\mathbf{u}\|_{\Sigma_{\mathrm{pr}}^{-1}}$，并对$\mathbf{u}$取下确界，得到

$$\beta_G(\theta) = \inf_{\mathbf{u}\in\mathbb{R}^M} \frac{\|Lx_\theta(\mathbf{u})\|_{\Sigma_L^{-1}}}{\|\mathbf{u}\|_{\Sigma_{\mathrm{pr}}^{-1}}} \ge \inf_{\mathbf{u}\in\mathbb{R}^M} \frac{\|L\tilde{x}_\theta(\mathbf{u})\|_{\Sigma_L^{-1}}}{\|\mathbf{u}\|_{\Sigma_{\mathrm{pr}}^{-1}}} - \gamma_L\,\overline{\eta}(\theta)\,\varepsilon_\theta = \tilde{\beta}_G(\theta) - \gamma_L\,\overline{\eta}(\theta)\,\varepsilon_\theta.$$

式（20）中的上界可用类似方法得到。 □

如果 ε_θ 足够小，命题 1 和命题 2 表明，可以使用代理模型˜ ˜ $\beta_{L|\mathcal{W}}(\theta)$和 $\beta_G(\theta)$，代替原始的全阶可观测性系数 $\beta_{L|\mathcal{W}}(\theta)$和 $\beta_G(\theta)$。当计算$x_\theta(\mathbf{u})$的成本过高，无法针对不同配置 θ反复计算 $\beta_{L|\mathcal{W}}(\theta)$或 $\beta_G(\theta)$时，这种替代尤其必要。

我们的传感器选择算法中的另一个近似步骤依赖于识别出一个可观测性相对较低的参数方向$\mathbf{v} \in \mathbb{R}^M$，即

$$\frac{\|Lx_\theta(\mathbf{v})\|_{\Sigma_L^{-1}}}{\|\mathbf{v}\|_{\Sigma_{\mathrm{pr}}^{-1}}} \approx \inf_{\mathbf{u}\in\mathbb{R}^M} \frac{\|Lx_\theta(\mathbf{u})\|_{\Sigma_L^{-1}}}{\|\mathbf{u}\|_{\Sigma_{\mathrm{pr}}^{-1}}} = \beta_G(\theta) \qquad \text{or} \qquad \frac{\|Lx_\theta(\mathbf{v})\|_{\Sigma_L^{-1}}}{\|x_\theta(\mathbf{v})\|_{\mathcal{X}}} \approx \inf_{x\in\mathcal{W}_\theta} \frac{\|Lx\|_{\Sigma_L^{-1}}}{\|x\|_{\mathcal{X}}} = \beta_{L|\mathcal{W}}(\theta).$$

理想的选择应分别为 $\beta_G(\theta)$ 或 $\beta_{L|\mathcal{W}}(\theta)$ 的下确界取值参数，但其计算需要进行 M 次全阶模型评估（参见第4.2节）。为避免这些代价高昂的计算，我们改为选择 $\mathbf{v}$ 作为相应降阶可观测性系数的下确界取值参数。以下命题说明，当 $\varepsilon_\theta < 1$ 较小时，这种选择是合理的：

命题 3. 设 $\eta(\theta) > 0$ 成立，且 $\tilde{x}_\theta(\mathbf{u}) \in \mathcal{X}$ 是 $x_\theta(\mathbf{u})$ 的一个近似，并且对所有 $\theta \in \mathcal{P}, \mathbf{u} \in \mathbb{R}^M$，式 (16) 均成立。假设 $\mathbf{v} \in \arg\inf_{\mathbf{u}\in\mathbb{R}^M} \|\mathbf{u}\|_{\Sigma_{\text{pr}}^{-1}}^{-1} \|L\tilde{x}_\theta(\mathbf{u})\|_{\Sigma_L^{-1}}$，则

$$\beta_G(\theta) \le \frac{\|Lx_\theta(\mathbf{v})\|_{\Sigma_L^{-1}}}{\|\mathbf{v}\|_{\Sigma_{\text{pr}}^{-1}}} \le \beta_G(\theta) + 2\gamma_L\bar{\eta}(\theta)\varepsilon_\theta. \tag{21}$$

同样，若 $\mathbf{v} \in \arg\inf_{\mathbf{u}\in\mathbb{R}^M} \|\tilde{x}_\theta(\mathbf{u})\|_{\mathcal{X}}^{-1} \|L\tilde{x}_\theta(\mathbf{u})\|_{\Sigma_L^{-1}}$，则

$$\beta_{L|\mathcal{W}}(\theta) \le \frac{\|Lx_\theta(\mathbf{v})\|_{\Sigma_L^{-1}}}{\|x_\theta(\mathbf{v})\|_{\mathcal{X}}} \le \frac{1+\varepsilon_\theta}{1-\varepsilon_\theta}\left(\beta_{L|\mathcal{W}}(\theta) + \gamma_L\varepsilon_\theta\right) + \gamma_L\varepsilon_\theta. \tag{22}$$

证明。对于 (21) 和 (22)，下界均可直接由定义 (13) 和 (15) 得出。为证明 (21) 中的上界，令 $\mathbf{v} \in \arg\inf_{\mathbf{u}\in\mathbb{R}^M} \|\mathbf{u}\|_{\Sigma_{\text{pr}}^{-1}}^{-1} \|L\tilde{x}_\theta(\mathbf{u})\|_{\Sigma_L^{-1}}$。按照命题 2 的证明中的相同步骤，我们可以得到如下界

$$\frac{\|Lx_\theta(\mathbf{v})\|_{\Sigma_L^{-1}}}{\|\mathbf{v}\|_{\Sigma_{\text{pr}}^{-1}}} \le \frac{\|L\tilde{x}_\theta(\mathbf{v})\|_{\Sigma_L^{-1}}}{\|\mathbf{v}\|_{\Sigma_{\text{pr}}^{-1}}} + \frac{\|L(x_\theta(\mathbf{v}) - \tilde{x}_\theta(\mathbf{v}))\|_{\Sigma_L^{-1}}}{\|\mathbf{v}\|_{\Sigma_{\text{pr}}^{-1}}} \le \tilde{\beta}_G(\theta) + \gamma_L\bar{\eta}(\theta)\varepsilon_\theta.$$

由命题 2 即可得出 (21) 中的上界。

为证明 (22) 中的上界，令 $\mathbf{v} \in \arg\inf_{\mathbf{u}\in\mathbb{R}^M} \|\tilde{x}_\theta(\mathbf{u})\|_{\mathcal{X}}^{-1} \|L\tilde{x}_\theta(\mathbf{u})\|_{\Sigma_L^{-1}}$。于是

$$\frac{\|Lx_\theta(\mathbf{v})\|_{\Sigma_L^{-1}}}{\|x_\theta(\mathbf{v})\|_{\mathcal{X}}} \le \frac{\|L\tilde{x}_\theta(\mathbf{v})\|_{\Sigma_L^{-1}}}{\|\tilde{x}_\theta(\mathbf{v})\|_{\mathcal{X}}} \frac{\|\tilde{x}_\theta(\mathbf{v})\|_{\mathcal{X}}}{\|x_\theta(\mathbf{v})\|_{\mathcal{X}}} + \frac{\|L(x_\theta(\mathbf{v}) - \tilde{x}_\theta(\mathbf{v}))\|_{\Sigma_L^{-1}}}{\|x_\theta(\mathbf{v})\|_{\mathcal{X}}} \le (1+\varepsilon)\,\tilde{\beta}_{L|\mathcal{W}}(\theta) + \gamma_L\varepsilon_\theta.$$

由命题 1 即可得出该结果。 □

4. 传感器选择

下文将介绍一种传感器选择算法，该算法通过迭代提高最小可观测性系数 $\min_{\theta\in\mathcal{P}} \beta_G(\theta)$，从而降低所有可行系统配置 $\theta \in \mathcal{P}$ 下后验协方差矩阵特征值的上界。这种迭代方法相对容易实现，能够以简单的方式处理组合限制，并且适用于大型传感器库[4]。

4.1. Cholesky 分解

协方差函数 **协方差**通过噪声协方差矩阵 Σ_L，将观测算子 L 与其可观测性系数 $\beta_G(\theta)$ 和 $\beta_{L|\mathcal{W}}(\theta)$ 联系起来。其逆矩阵会出现在范数 $\|\cdot\|_{\Sigma_L^{-1}}$和后验协方差矩阵 $\Sigma^{L,\,\theta}$ 后验分布中。当噪声相关时，即 Σ_L 不是对角矩阵时，反演会带来挑战，因为即使只用一个传感器 $\ell \in \mathcal{L}$ 对 L 进行展开，也会改变 Σ_L^{-1}的每个元素。在对可观测性系数和后验协方差矩阵进行朴素计算时，这会导致

[4]例如，在第 5.3 节中，我们将所介绍的算法应用于一个包含 $K_{\mathcal{L}} = 11$ 个可用传感器位置的库。

算法 1： Cholesky 扩展

输入： 观测算子 $L = [\ell_1, \ldots, \ell_K]^T$，噪声协方差矩阵 Σ_L，Cholesky 矩阵 $\mathbf{C}_L$，新传感器 $\ell \in \mathcal{X}'$

$L \leftarrow [\ell_1, \ldots, \ell_K, \ell]^T$ // 算子扩展

如果 $K = 0$**则**

$\Sigma_L \leftarrow (\mathbf{协方差}(\ell, \ell))$，$\mathbf{C}_L \leftarrow \left(\sqrt{\mathbf{协方差}(\ell, \ell)}\right) \in \mathbb{R}^{1\times 1}$ // 第一个传感器

else

$\mathbf{v} \leftarrow [\mathbf{协方差}(\ell_1, \ell), \ldots, \mathbf{协方差}(\ell_K, \ell)]^T \in \mathbb{R}^K$ // 矩阵扩展

$\mathbf{w} \leftarrow \mathbf{C}_L^{-1}\mathbf{v} \in \mathbb{R}^K$，$s \leftarrow \mathbf{协方差}(\ell, \ell)$，$c \leftarrow s - \mathbf{w}^T\mathbf{w} \in \mathbb{R}$

$$\Sigma_L \leftarrow \begin{pmatrix} \Sigma_L & \mathbf{v} \\ \mathbf{v}^T & s \end{pmatrix}, \mathbf{C}_L \leftarrow \begin{pmatrix} \mathbf{C}_L & \mathbf{0} \\ \mathbf{w}^T & c \end{pmatrix} \in \mathbb{R}^{(K+1)\times(K+1)}$$

返回 L，Σ_L，$\mathbf{C}_L$

每次展开观测算子时，都需要求解 $O((K+1)^3)$ 个阶为 O 的稠密线性方程组。因此，下面我们将阐述 Σ_L^{-1} 在扩展 L 时如何变化，以便在比较潜在传感器选择时利用其结构。

假设 $L = [\ell_1, \ldots, \ell_K]^T$ 已通过传感器 $\ell_k \in \mathcal{X}'$ 选定，但还要再加入一个传感器 ℓ，扩展为

$$[L, \ell] := [\ell_1, \ldots, \ell_K, \ell]^T : \mathcal{X} \to \mathbb{R}^{K+1}.$$

根据定义（4），扩展后算子$[L, \ell]$ 的噪声协方差矩阵 $\Sigma_{[L,\ell]}$ 具有如下形式

$$\Sigma_{[L,\ell]} = \begin{pmatrix} \Sigma_L & \mathbf{v}_{L,\ell} \\ \mathbf{v}_{L,\ell}^T & v_{\ell,\ell} \end{pmatrix} = \begin{pmatrix} \mathbf{C}_L & \mathbf{0} \\ \mathbf{c}_{L,\ell}^T & c_{\ell,\ell} \end{pmatrix} \begin{pmatrix} \mathbf{C}_L^T & \mathbf{c}_{L,\ell} \\ \mathbf{0} & c_{\ell,\ell} \end{pmatrix},$$

其中，$\mathbf{C}_L\mathbf{C}_L^T = \Sigma_L \in \mathbb{R}^{K\times K}$ 是原始观测算子 L 对应的对称正定噪声协方差矩阵 Σ_L 的 Cholesky 分解，而 $\mathbf{v}_{L,\ell}$,$\mathbf{c}_{L,\ell} \in \mathbb{R}^K$, $v_{\ell,\ell}$,$c_{\ell,\ell} \in \mathbb{R}$ 定义如下

$$[\mathbf{v}_{L,\ell}]_i := \mathbf{cov}(\ell_i, \ell), \qquad \mathbf{c}_{L,\ell} := \mathbf{C}_L^{-1}\mathbf{v}_{L,\ell},$$
$$v_{\ell,\ell} := \mathbf{cov}(\ell, \ell), \qquad c_{\ell,\ell} := \sqrt{v_{\ell,\ell} - \mathbf{c}_{L,\ell}^T\mathbf{c}_{L,\ell}}.$$

注意，根据第 2 节对**协方差** 所作的假设，$\Sigma_{[L,\ell]}$ 是对称正定的；因此，$c_{\ell,\ell}$ 定义良好且严格为正。利用这一分解，可在 $O(K^2)$ 内计算扩展的 Cholesky 矩阵 $\mathbf{C}_{[L,\ell]}$，其中$\mathbf{C}_{[L,\ell]}\mathbf{C}_{[L,\ell]}^T = \Sigma_{[L,\ell]}$；计算开销主要来自利用三角矩阵$\mathbf{C}_L$ 进行线性系统求解，以获得$\mathbf{c}_{L,\ell}$。算法 1 对此进行了总结，以供后续传感器选择算法使用。

使用 Cholesky 分解，$\Sigma_{[L,\ell]}$ 的逆可分解为

$$\Sigma_{[L,\ell]}^{-1} = \begin{pmatrix} \mathbf{C}_L^T & \mathbf{c}_{L,\ell} \\ \mathbf{0} & c_{\ell,\ell} \end{pmatrix}^{-1} \begin{pmatrix} \mathbf{C}_L & \mathbf{0} \\ \mathbf{c}_{L,\ell}^T & c_{\ell,\ell} \end{pmatrix}^{-1} = \begin{pmatrix} \mathbf{C}_L^{-T} & \mathbf{r}_{L,\ell} \\ \mathbf{0} & 1/c_{\ell,\ell} \end{pmatrix} \begin{pmatrix} \mathbf{C}_L^{-1} & \mathbf{0} \\ \mathbf{r}_{L,\ell}^T & 1/c_{\ell,\ell} \end{pmatrix},$$

算法 2： 可观测性增益

输入： 观测算子 $L = [\ell_1, \ldots, \ell_K]^T$, Cholesky 矩阵 $\mathbf{C}_L$, 传感器候选项 $\ell \in \mathcal{X}'$, 状态 $x \in \mathcal{X}$

$\mathbf{d} \leftarrow Lx, \mathbf{z} \leftarrow \mathbf{C}_L^{-1}\mathbf{d}$ // 准备

如果 $K = 0$ **那么**

  返回 $\ell(x_K)^2 / \mathbf{协方差}(\ell, \ell)$ // 仅一个传感器

else

  $\mathbf{v} \leftarrow [\mathbf{协方差}(\ell_1, \ell), \ldots, \mathbf{协方差}(\ell_K, \ell)]^T \in \mathbb{R}^K$ // 一般情况

  $\mathbf{w} \leftarrow \mathbf{C}_L^{-1}\mathbf{v} \in \mathbb{R}^K$

  返回 $\frac{\left(\ell(x_K) - \mathbf{w}^T\mathbf{z}\right)^2}{\mathbf{协方差}(\ell,\ell) - \mathbf{w}^T\mathbf{w}}$

其中

$$\mathbf{r}_{L,\ell} := -\frac{1}{c_{\ell,\ell}}\mathbf{C}_L^{-T}\mathbf{c}_{L,\ell} = -\frac{1}{c_{\ell,\ell}}\mathbf{C}_L^{-T}\mathbf{C}_L^{-1}\mathbf{v}_{L,\ell} = -\frac{1}{c_{\ell,\ell}}\Sigma_L^{-1}\mathbf{v}_{L,\ell}.$$

对于任意状态$x \in \mathcal{X}$，扩展观测 $[L, \ell](x) = \left[Lx^T, \ \ell(x)\right]^T \in \mathbb{R}^{K+1}$ 在相应范数 $\|\cdot\|_{\Sigma^{-1}_{[L,\ell]}}$ 下的范数，因此与原始观测$Lx \in \mathbb{R}^K$ 在原范数 $\|\cdot\|_{\Sigma_L^{-1}}$下通过以下关系相连：

$$\begin{aligned}
\| [L, \ell](x) \|^2_{\Sigma^{-1}_{[L,\ell]}} &= \begin{pmatrix} Lx \\ \ell(x) \end{pmatrix}^T \begin{pmatrix} \Sigma_L & \mathbf{v}_{L,\ell} \\ \mathbf{v}_{L,\ell}^T & v_{\ell,\ell} \end{pmatrix}^{-1} \begin{pmatrix} Lx \\ \ell(x) \end{pmatrix} \\
&= \begin{pmatrix} Lx \\ \ell(x) \end{pmatrix}^T \begin{pmatrix} \mathbf{C}_L^{-T} & \mathbf{r}_{L,\ell} \\ \mathbf{0} & 1/c_{\ell,\ell} \end{pmatrix} \begin{pmatrix} \mathbf{C}_L^{-1} & \mathbf{0} \\ \mathbf{r}_{L,\ell}^T & 1/c_{\ell,\ell} \end{pmatrix} \begin{pmatrix} Lx \\ \ell(x) \end{pmatrix} \\
&= \begin{pmatrix} \mathbf{C}_L^{-1}Lx \\ \mathbf{r}_{L,\ell}^T Lx + \ell(x)/c_{\ell,\ell} \end{pmatrix}^T \begin{pmatrix} \mathbf{C}_L^{-1}Lx \\ \mathbf{r}_{L,\ell}^T Lx + \ell(x)/c_{\ell,\ell} \end{pmatrix} \\
&= (Lx)^T\mathbf{C}_L^{-T}\mathbf{C}_L^{-1}Lx + (\mathbf{r}_{L,\ell}^T Lx + \ell(x)/c_{\ell,\ell})^2 \\
&= \|Lx\|^2_{\Sigma_L^{-1}} + (\mathbf{r}_{L,\ell}^T Lx + \ell_{K+1}(x)/c_{\ell,\ell})^2 \\
&\geq \|Lx\|^2_{\Sigma_L^{-1}}.
\end{aligned} \tag{23}$$

由此结果可知，任意观测的范数 $\|Lx\|_{\Sigma_L^{-1}}$，因此还有式 (6) 中定义的连续性系数 γ_L ，尽管范数发生变化，仍会随L的扩展而增大。对于任意配置 θ，当迭代选择传感器时，可观测性系数$\beta_G(\theta)$和 $\beta_{L|W}(\theta)$因此均单调不减。

给定状态 $x \in \mathcal{X}$ 和观测算子 L，我们可以通过比较所有 $\ell \in \mathcal{L}$ 的观测增量（$\mathbf{r}_{L,\ell}^T Lx + \ell(x)/c_{\ell,\ell}$），确定哪个传感器 $\ell_{K+1} \in \mathcal{L}$能最大程度地增加对 x 的观测。算法 2 总结了该可观测性增益的计算方法，以供传感器选择算法使用（见第 4.3 节）。其一般运行时间由 $K + 1$ 次传感器评估以及对三角 Cholesky 矩阵 $\mathbf{C}_L$ 进行的两次线性求解组成，复杂度为 $O(K^2)$。当使用相同的 L 和相同的

状态 x 针对不同的候选传感器 ℓ时，准备步骤只需执行一次，这使得后续每次调用的运行时间减少为一次传感器评估和一次线性方程组求解。与针对库 $\mathcal{L}$中的所有 $K_{\mathcal{L}}$ 个候选传感器计算 $\| [L, \ell](x) \|^2_{\Sigma^{-1}_{[L,\ell]}}$ 相比，我们节省了 $O(K_{\mathcal{L}}K^2)$。

4.2. 可观测性系数的计算

接下来，我们讨论如何计算给定配置 θ 和观测算子 L 下的可观测性系数 $\beta_G(\theta)$。

设 $\Sigma_{\text{pr}} = \mathbf{U}^T\mathbf{D}_{\text{pr}}\mathbf{U}$ 为对称正定先验协方差矩阵的特征值分解，其中 $\mathbf{U} = [\varphi_1, 、 \ldots\ldots 、 \varphi_M] \in \mathbb{R}^{M\times M}$、 $\varphi_j \in \mathbb{R}^M$ 在欧几里得内积下正交归一，且 $\mathbf{D}_{\text{pr}} = \text{diag}(\lambda^1_{\text{pr}}, \ldots\ldots, \lambda^M_{\text{pr}})$ 为包含特征值$\lambda^1_{\text{pr}} \geq \cdots \geq \lambda^M_{\text{pr}} > 0$的对角矩阵，特征值按降序排列。利用特征向量基 $\{\varphi_m\}_{m=1}^M$，我们定义矩阵

$$\mathbf{M}(\theta) := [Lx_\theta(\varphi_1), \ldots, Lx_\theta(\varphi_M)] \in \mathbb{R}^{K\times M} \tag{24}$$

该矩阵包含配置 θ 下相关状态 $x_\theta(\varphi_j)$ 的所有观测。随后，可将可观测性系数$\beta_G(\theta)$计算为广义特征值问题中最小特征值 $\lambda^{\min}$ 的平方根

$$\mathbf{M}(\theta)^T\mathbf{C}_L^{-T}\mathbf{C}_L^{-1}\mathbf{M}(\theta)\mathbf{u}_{\min} = \lambda^{\min}\mathbf{D}_{\text{pr}}^{-1}\mathbf{u}_{\min}. \tag{25}$$

请注意，由于左侧矩阵是对称半正定的，且M 具有$\mathbf{D}_{\text{pr}}$ 是对称正定（参见 [48]），因此（25）有M 个实的非负特征值。特征向量$\mathbf{u}_{\min}$包含“最坏情况”参数 $\{\varphi_m\}_{m=1}^M$ 在特征向量基 $\{\varphi_m\}_{m=1}^M$ 中的基系数，即$\beta_G(\theta)$的下确界取值参数。

注 4. 为了计算 $\beta_{L|\mathcal{W}}(\theta)$，我们将右端矩阵$\mathbf{D}_{\text{pr}}^{-1}$ （位于(25) 中） 替换为状态 $x_\theta(\varphi_1), \ldots, x_\theta(\varphi_M)$的 $\mathcal{X}$内积矩阵。

特征值问题的解可在 $O(M^3)$内计算，此外，计算（25）中的左侧矩阵还需要 $O(MK^2 + M^2K)$。主要计算成本隐藏在$\mathbf{M}(\theta)$中，因为这需要KM 次传感器观测和K 次全阶模型求解。因此，为降低计算成本，我们将全阶状态$x_\theta(\varphi_j)$替换为其降阶近似 $\tilde{x}_\theta(\varphi_j)$，从而用 $\beta_G(\theta)$近似$\beta_G(\theta)\tilde{}$。该过程总结于算法 3 中。

注释 5. 如果 $K < M$，算法 *3* 会按照第 *3.2* 节的讨论，将参数空间限制在前 K 个特征向量 $\varphi_1, \ldots\ldots, \varphi_K$所张成的空间内，这些特征向量对应先验中最不确定的方向。文献 *[8]* 在 *PBDW* 方法的背景下简要讨论了一种变体，旨在进一步优先考虑最不确定的参数：只有当子空间上的可观测性系数超过预定阈值后，才扩展参数空间。

4.3. 传感器选择

在我们的传感器选择算法中，我们通过迭代扩展观测算子L ，从而提高所有 $\theta \in \mathcal{P}$ 的可观测性系数L 。尽管

算法 3： 代理可观测性

输入： 配置 $\theta \in \mathcal{P}$，观测算子 $L = [\ell_1, \ldots, \ell_K]^T$ with $K > 0$，Cholesky 矩阵 $\mathbf{C}_L$

$N \leftarrow$ 最小$\{M, K\}$ // 参数限制

$\mathbf{M} \leftarrow [L\,\tilde{x}_\theta(\varphi_1), \ldots, L\,\tilde{x}_\theta(\varphi_N)]$, $\mathbf{S} \leftarrow \left[\left\langle x_\theta(\varphi_i), x_\theta(\varphi_j)\right\rangle_{\mathcal{X}}\right]_{i,\,j=1}^{N}$ // 矩阵设置

求解 $(\lambda^{\min}, \mathbf{u}_{\min})$，使 $\left[\mathbf{C}_L^{-1}\mathbf{M}\right]^T \left[\mathbf{C}_L^{-1}\mathbf{M}\right]\mathbf{u}_{\min} = \lambda^{\min}\mathbf{S}\mathbf{u}_{\min}$ // 特征值问题

返回 $\sqrt{\lambda^{\min}}, \mathbf{u}_{\min}$

该过程无法保证在所有传感器组合中找到最大可观测性，但相关的贪心搜索在实践中已得到充分验证；在密切相关的情形下，可以证明其误差率呈指数级下降，参见 [49, 8, 50, 51, 52]。在每次迭代中，算法执行两个主要步骤：

- 在训练集 $\Xi_{\text{train}} \subset \mathcal{P}$ 上进行**贪婪搜索**，以确定配置$\theta \in \Xi_{\text{train}}$，使得对于该配置，可观测性系数 $\beta_G(\theta)$ 最小；
- 执行**数据匹配步骤**，以确定库中在所选配置 θ下对“最坏情况”参数的观测最大化的传感器。

该过程总结于算法4中。当选定$K_{\max} \leq K_{\mathcal{L}}$ 个传感器后，过程终止。[5] 下面将介绍其计算细节。

准备工作

为了在超参数域 $\mathcal{P}$上均匀提高 $\beta_G(\theta)$，我们考虑一个有限训练集 $\Xi_{\text{train}} \subset \mathcal{P}$，并将其选择得足够精细，以捕捉依赖于 θ 的变化，即 $x_\theta(\mathbf{u})$ 中的变化。我们假设已有一个降阶模型，因此可以在可接受的计算时间内，为每个 $\theta \in \Xi_{\text{train}}$ 以及 $1 \leq m \leq M$ 计算近似值 $\tilde{x}_\theta(\varphi_m) \approx x_\theta(\varphi_m)$，同时保证精度要求（16）。如有必要，可以通过自适应训练域（例如，[53, 54]）来平衡这两个标准。

注6. 如果存储空间允许（例如，使用基于投影的替代模型），我们只需计算一次代理状态，并避免在每次迭代更新替代模型可观测性系数时进行不必要的重新计算$\tilde{\beta}_G(\theta)$。

作为第一个“最坏情况”参数方向，$\mathbf{u}_0$, 我们选择先验不确定性最大的向量 φ_1。同样，我们将对应状态 $\tilde{x}_\theta(\varphi_1)$ 最大的配置 $\theta_K \in \mathcal{P}$选为“最坏情况”配置。

[5]这个终止准则可以轻松调整，以指定可观测性系数的最小值。该值应根据使用整个传感器库所达到的可观测性来选取 $\beta_G(\mathcal{L})$。

算法4： 传感器选择

输入： 传感器库 $\mathcal{L} \subset \mathcal{X}'$、训练集 $\Xi_{\text{train}} \subset \mathcal{P}$、最大传感器数量 $K_{\max} \leq |K_{\mathcal{L}}|$、替代模型 $\tilde{\mathcal{M}}_\theta$、协方差函数 **协方差**： $\mathcal{L} \times \mathcal{L} \to \mathbb{R}$

计算 $\Sigma_{\text{pr}} = [\varphi_1, \ldots, \varphi_M]^T \mathbf{D}_{\text{pr}} [\varphi_1, \ldots, \varphi_M]$ // 特征值分解

对于所有 $\theta \in \Xi_{\text{train}}, 1 \leq m \leq M$，计算 $\tilde{x}_\theta(\varphi_m)$ // 准备

$K \leftarrow 0, \theta_0 \leftarrow \arg\max_{\theta \in \Xi_{\text{train}}} \|\tilde{x}_\theta(\varphi_1)\|_{\mathcal{X}}, \mathbf{u}_0 \leftarrow \varphi_1$ // 初始化

当 $K < K_{\max}$ **do**

求解全阶方程 $\mathcal{M}_{\theta_K}(x_K, \mathbf{u}_K)$，得到 x_K // “最坏情况”状态

$\ell_{K+1} \leftarrow \arg\max_{\ell \in \mathcal{L}}$ 可观测性增益$(L, \mathbf{C}_L, \ell)$ // 传感器选择

$L, \Sigma_L, \mathbf{C}_L \leftarrow$ Cholesky 扩展$(L, \Sigma_L, \mathbf{C}_L, \ell_{K+1})$ // 展开

$K \leftarrow K + 1$

对于 $\theta \in \Xi_{\text{train}}$ **执行**

$\tilde{\beta}_{L|\mathcal{W}}(\theta), \mathbf{u}_{\min}(\theta) \leftarrow$ 代理可观测性$(\theta, L, \mathbf{C}_L)$ // 更新系数

$\theta_K \leftarrow \arg\min_{\theta \in \Xi_{\text{train}}} \tilde{\beta}_{L|\mathcal{W}}(\theta)$ // 贪心步骤

$\mathbf{u}_K \leftarrow \sum_{m=1}^{\min\{M,K\}} [\mathbf{u}_{\min}(\theta_K)]_m \varphi_m$

返回 L , $\mathbf{C}_L$

数据匹配步骤

在每次迭代中，我们首先计算“最坏情况”参数 $\mathbf{u}_K$ 和配置 θ_K下的全阶状态 $x_K = x_{\theta_K}(\mathbf{u}_K)$。然后，我们选择传感器 ℓ_{K+1}，使其在扩展观测算子 $[L^T$及其关联范数下，最大程度地改善对“最坏情况”状态 x_K 的观测。由此，我们通过迭代近似获得使用库 $\mathcal{L}$中的所有传感器进行测量所能获取的信息。对于固定的 θ_K，并结合选择在 $\mathcal{W}_\theta$中可观测性最小的 x，我们得到一种类似于最坏情况正交匹配追踪的算法（参见 [8, 9]），但该算法经过推广，可处理噪声模型 (3) 中的协方差函数 **cov**。

注释7。 为了避免在降阶模型中针对局部近似误差进行训练，我们使用全阶状态 $x_{\theta_K}(\mathbf{u}_K)$，而非其降阶近似。在这里，通过使用“最坏情况”参数方向 $\mathbf{u}_K$，每次迭代只需进行一次全阶求解，而无需进行 M 次求解来构建整个后验协方差矩阵 $\Sigma^{L\theta}_{\text{post}}$。

贪心步骤

我们通过改变 θ（以计算“最坏情况”状态），在所有配置 $\theta \in \Xi_{\text{train}}$ 上训练观测算子 L。具体而言，我们采用贪心方法

其中，在第 K 次迭代中，我们选择训练域 Ξ_{train} 上使 $\beta_G(\theta)$ 取得极小值的配置 θ_K，也就是当前观测算子 L 最不利的配置。对应的“最坏情况”参数 $\mathbf{u}_K$ 是能够实现最不显著观测的参数方向。通过迭代提高“最坏情况”参数和超参数下的可观测性，我们提高了整个训练域内 $\beta_G(\theta)$ 的最小值。

注 8. 由于计算 $\beta_G(\theta)$ 需要针对替代模型上的后验协方差矩阵进行所需次数的降阶模型求解，因此可以在贪心步骤中直接以（近似的）最优实验设计效用函数为目标，而非 $\beta_{L|\mathcal{W}}(\theta)$，且不会显著牺牲计算效率。随后仍可仅通过一次全阶模型求解，针对“最坏情况”参数执行正交匹配追踪步骤，但应仔细评估该步骤对效用函数的益处。

运行时间

假设主要的计算限制来自模型评估，即求解 $x_\theta(\mathbf{u})$——偏微分方程模型通常正是如此——那么算法4中每次迭代的运行时间由一次全阶模型评估以及对全阶状态进行 $K_{\mathcal{L}}$ 次传感器测量决定。与为所选配置计算后验协方差矩阵相比，正交匹配追踪步骤可节省 $N - 1$ 次全阶模型求解。

算法4运行时间的另一个主要影响因素是每次迭代中需要进行的 $|\Xi_{\text{train}}|M$ 次降阶模型评估，每次评估还需要 $K_{\mathcal{L}}$ 次传感器评估（除非这些评估可以预先计算）。参数维度 M 不仅是一个缩放因子，还会影响降阶模型本身的成本，因为要达到所需精度，较大的 M 通常需要更大或更复杂的降阶模型 (16)。反过来，降阶模型的计算成本也决定了在给定计算预算下 Ξ_{train} 可以取多大。尽管通过自适应训练集和模型可以节省部分成本，但总体而言，这种与 M 的关联凸显了进行适当初始参数降维的必要性，正如第 3.2 节所讨论的那样。

5. 数值结果

我们使用西澳大利亚珀斯盆地某一部分的地球物理模型，对传感器选择方法的有效性进行了数值验证。由于该盆地具有很大的地热能开发潜力，因此引起了地球物理学界的关注，例如 [55, 56, 57, 58, 59]。我们重点研究其中一个子区域，该区域面积为 63 km $\times$ 70 km，向地下延伸至 19 km 深处。该模型在[60] 中提出，模型中所展示的区域也在 [61, 62] 中结合 MOR 进行了深入讨论。具体而言，地下温度分布通过一个稳态热传导问题描述，其中不同子域对应不同地质层，并可通过钻孔获取局部测量值。由于钻孔成本高昂（通常为数百万美元， [63]），其位置必须谨慎选择，这也促使我们应用算法4。为进行演示，我们对测试模型作出以下简化：1)

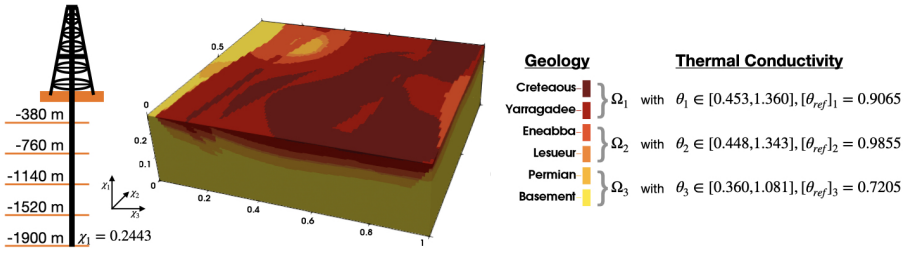


图 1：珀斯盆地剖面的示意图，包含（合并后的）地质层、潜在测量位置的深度，以及各子域热导率的配置范围 θ。范围根据参考值（参见 [60, 61]）并留出 ±50% 的裕度确定。改编自 [61]。

我们忽略放射性生热；2）将导热特性相似的地质层合并；3）对先验进行缩放，以突出不同传感器测量对后验分布的影响。所有计算均在一台配备 2.3 GHz 四核 Intel Core i5 处理器和 16 GB RAM 的计算机上，使用 Python 3.7 完成。相关代码将发布在一个公开的 GitHub 代码库中，该代码库还包含另一个地球物理测试问题。[6]

5.1. 模型描述

我们使用稳态偏微分方程对温度分布x_θ 建模

$$-\nabla\,(\theta\nabla x_\theta) = 0 \qquad \text{in } \Omega := (0, 0.2714) \times (0, 0.9) \times (0, 1) \subset \mathbb{R}^3, \tag{26}$$

其中，区域 Ω 是盆地的无量纲化表示，而 $\theta : \Omega \to \mathbb{R}_{>0}$ 表示局部热导率。该剖面包含三个主要地质层 $\Omega = \bigcup_{i=1,\ 2,\ 3}\Omega_i$，各层具有不同的岩石性质，即图 1 $\theta|_{\Omega_i} \equiv \theta_i$ 所示的热导率。我们认为地质层的位置是固定的，因为通常会事先通过地质调查和地球物理调查确定这些位置，但允许热导率变化。稍微滥用一下记号，我们可以将场 θ 与向量

$$\theta = (\theta_1, \theta_2, \theta_3) \in \mathcal{P} := [0.453, 1.360] \times [0.448, 1.343] \times [0.360, 1.081].$$

在超参数域 $\mathcal{P}$ 中。

我们在地表施加零狄利克雷边界条件[7]，，并在区域的侧面施加零诺伊曼（“无流动”）边界条件。剩余边界 Γ_{In} 对应珀斯盆地地表以下 19 km 处一片面积为 63 km $\times$ 70 km 的区域。在这一深度，热通量的局部变化大多已趋于稳定，因此可以进行建模；但由于大多数钻孔——通常源自油气勘探——都位于最上方 2 km 范围内，我们将其视为

[6]珀斯盆地模型可向第三作者索取。[7]通过提升函数和测量数据的仿射变换，可以考虑从卫星数据中获得的非零狄利克雷边界条件（见 [62]）。

不确定。具体而言，我们将其建模为诺伊曼边界条件

$$\mathbf{n} \cdot \nabla x_\theta = \mathbf{u} \cdot \mathbf{p} \qquad \text{a.e. on } \Gamma_{\text{In}} := \{0\} \times [0, 0.9] \times [0, 1]$$

其中 $\mathbf{n}: \Gamma_{\text{In}} \to \mathbb{R}^3$ 是 Ω 上的外指单位法向量，$\mathbf{p}: \Gamma_{\text{In}} \to \mathbb{R}^5$ 是由底部边界上的二次 $L^2(\Gamma_{\text{In}})$-正交归一多项式构成的向量，这些多项式沿南北方向或东西方向变化；$\mathbf{u} \sim \pi_{\text{pr}} = \mathcal{N}(\mathbf{u}^{\text{pr}}, \Sigma_{\text{pr}})$ 是一个随机变量。先验分布的选择使得最大的不确定性归于$\mathbf{p}$中的常数项，而二次项则被视为最确定的项，其先验值为零。这一设置反映了典型的地球物理边界条件：通常假设诺伊曼热通量为常数（例如 [61]），有时也假设其为线性项（例如 [60]）。通过引入二次函数，我们比通常考虑的情况多增加了一个自由度。

该问题采用维数为 132,651 的线性有限元（FE）基进行离散化。底层网格由 GemPy（[64]）和 MOOSE（[65]）创建。由于 FE 矩阵在 θ 中解耦，我们使用 DwarfElephant（[61]）预先计算并存储仿射分解。给定一个配置 θ 以及热通量在 Γ_{In} 处的系数向量$\mathbf{u}$ 后，计算全阶解$x_\theta(\mathbf{u}) \in \mathcal{X}$ 平均需要 2.96 秒。随后，我们进一步利用仿射分解，通过贪心算法构造降阶基（RB）替代模型（参见 [49, 66]）。利用内积[8]$\langle x, \phi \rangle_{\mathcal{X}} := \int_\Omega \nabla x \cdot \nabla \phi d\Omega$ 和后验 误差界 $\Delta(\theta)$，我们规定相对目标精度

$$\max_{\mathbf{u} \in \mathbb{R}^M} \frac{\|x_\theta(\mathbf{u}) - \tilde{x}_\theta(\mathbf{u})\|_{\mathcal{X}}}{\|\tilde{x}_\theta(\mathbf{u})\|_{\mathcal{X}}} \leq \max_{\mathbf{u} \in \mathbb{R}^M} \frac{\Delta(\theta)}{\|\tilde{x}_\theta(\mathbf{u})\|_{\mathcal{X}}} < \varepsilon := 1\text{e} - 4 \tag{27}$$

对于 θ 连续抽取的 511,000 个均匀分布样本，均可达到这一目标。图2总结了RB代理模型的训练阶段和最终计算性能。替代模型的加速比（不考虑误差界时约为 3,000 倍）足以抵消其离线训练时间；预计只需对 $\beta_G(\theta)$ 进行 152 次近似计算即可开始节省计算开销。

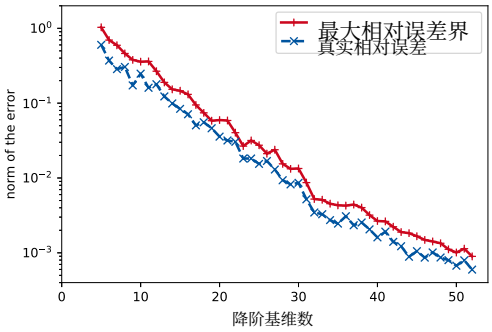
为进行测量，我们在地表上设置一个 47×47 网格，以表示可能的钻探位置。每个位置均可在图1所示的五个可能深度中的任一深度，对盆地的温度分布进行一次单点求值[9]。由此，我们总共得到一组包含 $11,045$ 个可用于测量的可行点 Ω $\mathcal{L} \subset$。我们将位于点 χ、$\tilde{\chi} \in \Omega$ 的传感器 ℓ_χ、$\ell_{\tilde{\chi}} \in \mathcal{L}$ 之间的噪声协方差建模为

$$\mathbf{cov}(\ell_\chi, \ell_{\tilde{\chi}}) := a + b - y(h)$$

采用指数变差函数模型

$$y(h) := a + (b - a)\left(\frac{3}{2}\max\{\frac{h}{c}, 1\} - \frac{1}{2}\max\{\frac{h}{c}, 1\}^3\right)$$

[8]请注意，由于狄利克雷边界条件，$\langle \cdot, \cdot \rangle_{\mathcal{X}}$ 确实是一个内积。[9]点求值是地球物理模型中的常见做法，因为与模型的尺寸相比，钻孔（直径约为 1 米）非常小。



降阶模型	
降阶基维数	83
训练时间	37.58 分钟
训练精度	1e−4
降阶基求解	0.97 毫秒
↪ 加速比	3,058
RB误差界	4.78 毫秒
↪ 加速比	515

图2：珀斯盆地剖面的 RB 代理模型训练。左图：贪心算法执行过程中，根据训练集计算得到的最大相对误差界 (27) Ξ_{train} 以及对应配置下的真实相对误差 θ。右图：达到 (27) 后所得 RB 模型的性能指标；在线计算时间和加速比为对 1000 个随机抽取的配置取平均值 θ。

其中 $h^2 := (\chi_2 - \tilde{\chi}_2)^2 + (\chi_3 - \tilde{\chi}_3)^2$ 表示点之间的水平距离，并且

$a := 2.2054073480730403$ (基台)

$b := 1.6850672040263555$ (块金)

$c := 20.606782733391228$ (范围)

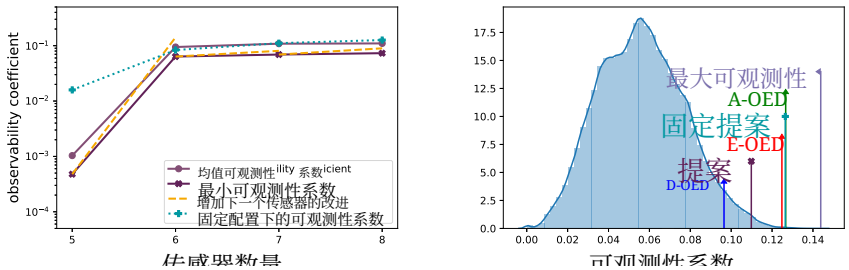
协方差函数通过克里金法（参见 [67]），利用现有测量值 [68] 计算得出。使用该协方差函数，水平面上任意两个传感器位置之间的测量噪声相关性会随着它们距离的缩短而增强。请注意，对于任意传感器位置子集，只要每个传感器都位于不同的钻井位置，相应的噪声协方差矩阵就始终是正则的。我们选择这种实验设置，是因为在典型的地热数据集中，测量值通常是在钻孔底部（“井底温度测量”）获取的，位置在地表以下 2 千米以内。

5.2. 受限库

为测试可观测性系数用于传感器选择的可行性，我们首先考虑一个较小的传感器库（记为下文的 $\mathcal{L}_{5\times5}$），其中包含位于 5×5 网格上的 25 个钻井位置。我们考虑从给定的 25 个位置中选出 8 个两两不同且无序的传感器位置；这是一个组合问题，共有 1,081,575 种可能组合。

传感器选择

我们运行算法4，使用 RB代理模型和包含 512,000 个配置的训练集 $\Xi_{\text{train}} \subset \mathcal{P}$，在 $\mathcal{P}$ 上的 $80 \times 80 \times 80$ 规则网格上进行计算。随着新传感器 ˜ 被选出，代理可观测性系数 $\beta_G(\theta)$ 单调递增；在最初的 $M = 5$ 个传感器之后，系数迅速上升，随后明显趋于停滞（见图 3a），这与类似的基于 OMP 的传感器选择算法中常见的现象一致



(a) 传感器选择过程中的可观测性　　(b) $\beta_G(\theta_{\text{ref}})$ 的直方图

图 3：从 25 个传感器位置中选择 8 个时，不同方法的可观测性系数。左：对 512,000 个配置运行一次算法4，以及对训练集 $\{\theta_{\text{ref}}\}$ 运行一次算法4 的过程中，θ 对 $\tilde{\beta}_G(\theta)$ 的最小值和均值，以及获得的 $\beta_G(\theta_{\text{ref}})$。右：所有可能传感器组合中 $\beta_G(\theta_{\text{ref}})$ 的分布，并标示 A-、D- 和 E-最优选择、具有最大可观测性的组合，以及算法4通过 Ξ_{train}-训练（“提案”，紫色，标记为“x”）和 θ_{ref}-训练（“固定提案”，青绿色，标记为“+”）所选择的传感器。请注意，指示线的高度仅为便于阅读而设定。

（例如，$[8, 69, 70, 7]$）。算法4在 7.93 分钟后终止，最小降阶可观测性为 $\beta_G(\theta)$ = 7.3227e-2，平均值为 1.0995e-1。在参考配置 θ_{ref} 下，全阶可观测性系数为 $\beta_G(\theta_{\text{ref}})$ = 1.0985，略低于降阶平均值。下文将此训练过程称为“Ξ_{train}-训练”，并将所选传感器称为“Ξ_{train}-训练传感器集”；图中则称其为“提案”。

为了准确了解替代模型 $\tilde{x}_\theta(\mathbf{u})$ 和大型配置训练集 Ξ_{train} 如何影响传感器选择，我们再次运行算法4；这次将其限制在全阶有限元模型 $x^{\theta_{\text{ref}}}(\mathbf{u})$ 上，仅使用参考配置 θ_{ref}。算法运行过程中 $\beta_G(\theta_{\text{ref}})$的增加情况如图 3a 所示。对于 Ξ_{train}-训练，曲线起始时显著高于平均值，这可能是因为不会出现相互冲突的配置，例如，在某一配置下一个传感器会显著提高可观测性，但在另一配置下却几乎没有变化。然而，在停滞阶段，该曲线逐渐接近通过Ξ_{train}-训练达到的平均值。计算在 12.53 秒内完成，表明此前较长的运行时间可归因于 Ξ_{train} 的规模。最终，使用 8 个传感器时的可观测性系数为 $\beta_G(\theta_{\text{ref}})$ = 1.2647e-1，高于在Ξ_{train}上进行训练所达到的 $\beta_G(\theta)$ 平均值。下文将这一训练过程称为“θ_{ref}-训练”，并在正文中将传感器配置称为“θ_{ref}-训练的”，或在图中称为“固定配置下的提案”。

参考配置下的比较

为了比较经过 Ξ_{train}和 θ_{ref}训练的传感器组合的性能，我们在参考配置 θ_{ref} 下计算传感器库 $\mathcal{L}_{5\times5}$中所有由 8 个不同传感器组成的无序组合对应的 1,081,575 个后验协方差矩阵 $\Sigma^{\theta_{\text{ref}},\ L\ \text{post}}$。对于每个矩阵，我们计算其迹（A-OED准则）、行列式（D-OED准则）、最大特征值（E-OED准则）以及可观测性系数 $\beta_G(\theta_{\text{ref}})$。这样便可确定 A-、D- 和 E-最优的传感器组合——

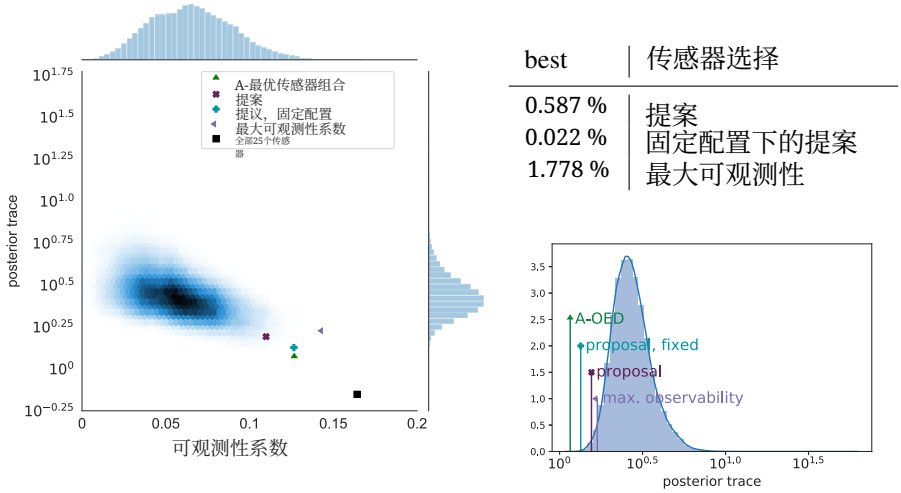


图 4：对于 $\theta = \theta_{\text{ref}}$，在从25个传感器位置中选取8个的所有1,081,575种组合中，trace($\Sigma^{L,\theta}$ post)的分布。左图：trace($\Sigma^{L,\theta}$ post)与可观测性系数 $\beta_G(\theta_{\text{ref}})$的关系分布。请注意，横轴的边缘分布见图 3b。右图：trace($\Sigma^{L,\theta}$ post)的直方图（左图的边缘分布），展示不同传感器组合所占的百分比（在1,081,575种组合中）。图中标出了A-最优传感器选择、算法4使用 Ξ_{train}训练选出的传感器（“提案”）和使用 $\{\theta_{\text{ref}}\}$训练选出的传感器（“提议，固定配置”）、具有最大可观测性 $\beta_G(\theta_{\text{ref}})$的传感器组合，以及包含全部25个传感器时的结果。

合。完成这些计算的总运行时间为 4 分钟，远高于θ_{ref}训练所需的 12.53 秒。 Ξ_{train}训练耗时（近）8 分钟，仍在可接受范围内，因为其训练使用了 $|\Xi_{\text{train}}| = 512{,}000$ 个配置，而不仅仅是 θ_{ref}。

图 3b 给出了 $\beta_G(\theta_{\text{ref}})$分布的直方图，并标出了A-、D-和E-最优选择以及经过 Ξ_{train}和 θ_{ref}训练的观测算子的取值。在这五者中，D-最优选择的值最小，因为后验行列式受最大后验特征值及可观测性系数的影响较小。相比之下，A-和E-最优传感器选择均位于 $\beta_G(\theta_{\text{ref}})$最大的700种组合之中（相当于排名前0.065%）。经过 θ_{ref}训练的传感器具有相近的可观测性，甚至位列前500种组合之中。对于经过 Ξ_{train}训练的传感器，可观测性系数较小，这可能是因为 Ξ_{train}训练并未针对 θ_{ref}进行充分优化。尽管如此，其可观测性仍位列最大的传感器组合的前0.705%。

为了直观展示可观测性系数 $\beta_G(\theta_{\text{ref}})$ 与经典 A、D 和 E 最优实验设计准则之间的联系，我们绘制了所有传感器组合下后验协方差矩阵的迹、行列式和最大特征值相对于 $\beta_G(\theta)$ 的分布，如图 4、5、6 所示。总体而言，我们观察到相应的最优实验设计准则与 $\beta_G(\theta_{\text{ref}})$ 之间存在很强的相关性：在图 6 的 E-最优性中最为明显，而在图 5 的 D-最优性中最不明显。对于所有最优实验设计

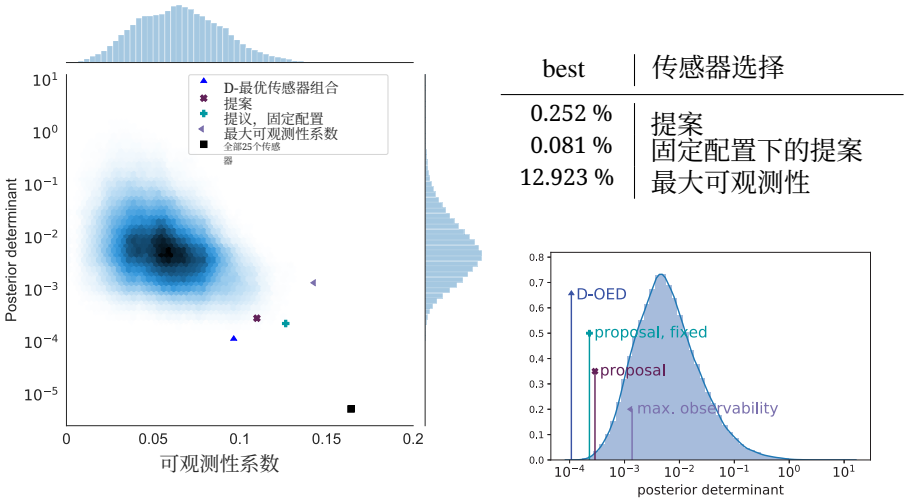


图 5： $\theta = \theta_{\text{ref}}$ 的后验行列式 $\det(\Sigma^{L,\theta}\text{post})$ 的分布。图的结构详情请参见图 4。

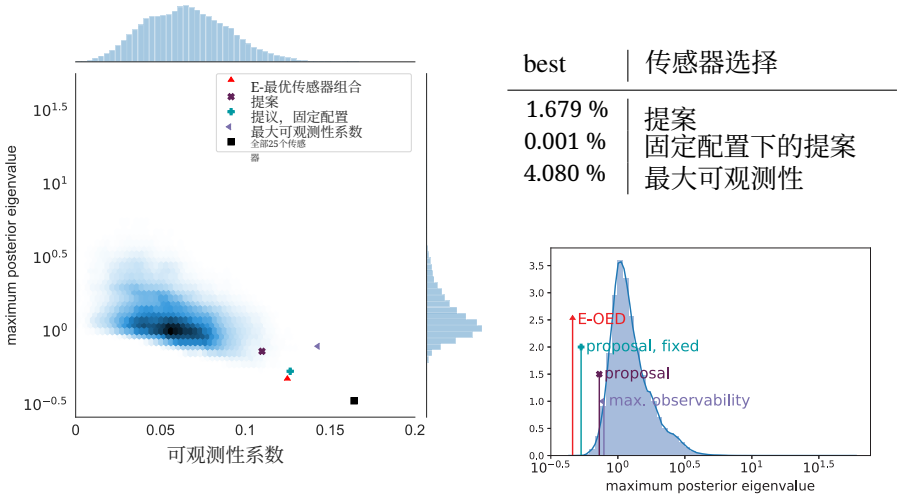
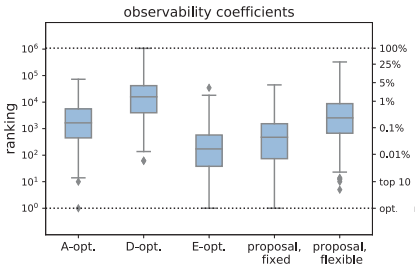


图 6：后验协方差矩阵 $\Sigma^{L,\theta}\text{post}$ 的最大特征值分布，适用于 $\theta = \theta_{\text{ref}}$。有关图表结构的详细信息，请参见图 4。请注意，经过 θ_{ref} 训练的传感器组合在所有 1,081,575 种可能组合中，最大后验特征值排名第 101 小。



pctl	设计准则 A-OED	D-OED	E-OED	训练 θ_{ref}	Ξ_{train}
第 99	3.5835	81.8508	1.1724	2.2223	9.2512
第95百分位	2.2747	26.8430	0.3601	0.7846	4.0374
第75百分位	0.5141	3.8600	0.0532	0.1419	0.8106
第50百分位	0.1527	1.4641	0.0159	0.0438	0.2354
第25百分位	0.0414	0.3669	0.0035	0.0068	0.0621

图7：在 $\beta_G(\theta_{\text{ref}})$中的排名：针对库中所有选择8个无序传感器的可能组合，A、D、E最优传感器选择以及经 θ_{ref}和 Ξ_{train}训练的传感器选择的排名。左：基于200个随机传感器库得到的箱线图。右：相应百分位数（“百分位数”）中最差情况的排名（百分比）。

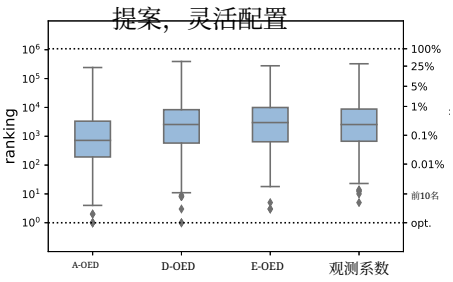
准则下，当先验受到重视时，随着缩放因子 σ^2 减小，相关性增强；当 σ^2 较大时，相关性减弱（未展示图表）。这一表现与第3.1节中的讨论一致：$\beta_G(\theta)$主要针对最大的后验特征值，对于不确定性较高的先验最具决定性。

不同库的比较

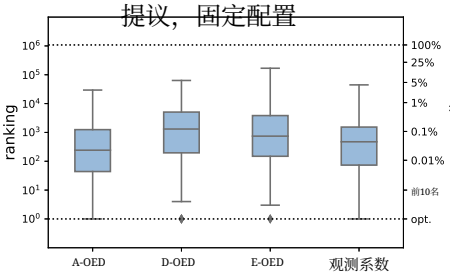
最后，我们评估库 $\mathcal{L}_{5\times5}$ 对结果的影响。为此，我们随机选择了 200 组新的测量位置，每组包含 25 个钻井位置及其对应的钻孔深度。对于每个库，我们运行算法4来选择 8 个传感器：一次使用 Ξ_{train} 训练替代模型，另一次仅使用 θ_{ref} 下的全阶模型。作为比较，随后我们针对每个库，考虑所有从 8 个无序传感器集合中进行选择的可能组合，并在参考配置 θ_{ref} 下计算相应后验协方差矩阵的迹、行列式和最大特征值，以及其可观测性系数。由此，我们可以确定 A-、D- 和 E-最优传感器组合。

图7展示了 $\beta_G(\theta_{\text{ref}})$ 在 200 个库中的分布情况，旁侧表格给出了相应的百分位数。对于 75% 的库而言，A- 和 E-最优传感器选择，以及经过 Ξ_{train} 和 θ_{ref} 训练的传感器选择，在可观测性最大的组合中均位列前 1%。由于针对 θ_{ref} 的训练未经过优化，经过 Ξ_{train} 训练的传感器集表现略逊于采用 θ_{ref} 训练所取得的结果，但其 $\beta_G(\theta_{\text{ref}})$ 值仍然相对较大。相比之下，总体而言，D-最优传感器选择的可观测性系数较小，这可能是因为后验行列式的最小化受最大后验特征值的影响较小。

图8展示了在200个库中，按后验协方差矩阵的迹、行列式和最大特征值对经过 Ξ_{train} 和 θ_{ref} 训练的传感器配置进行的排名。两者表现均良好，在75%的库中都位于组合排名前1%之内。由于排名是针对配置参数 θ_{ref} 进行的，因此经过 θ_{ref} 训练的传感器组合表现更好，在95%的库中都保持在传感器组合排名前5%之内。

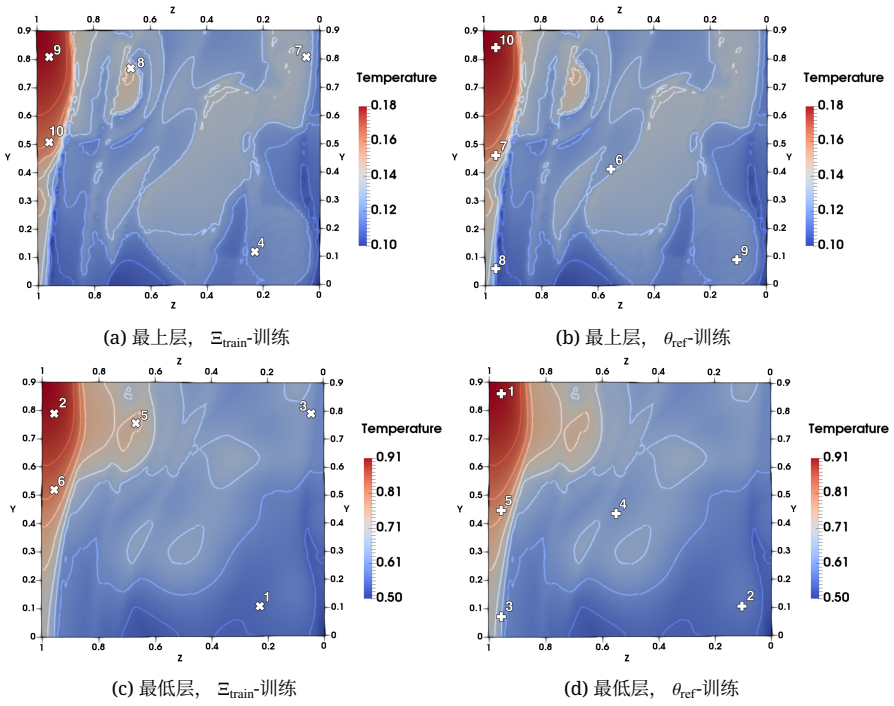


pctl	A-OED	D-OED	E-OED	$\beta_G(\theta_{\text{ref}})$
第 99 个	3.9240	6.2372	10.8391	9.2512
第 95 个	1.9093	3.1544	4.5583	4.0374
第75个	0.3083	0.7718	0.9185	0.8106
第50个	0.0664	0.2361	0.2763	0.2354
第25个	0.0177	0.0536	0.0596	0.0621



pctl	A-OED	D-OED	E-OED	β_G （θ_{ref}）
第99	2.5261	2.9752	11.1534	2.2223
第95	1.0134	1.8324	2.8458	0.7846
第75	0.1155	0.4698	0.3549	0.1419
第50	0.0224	0.1212	0.0687	0.0438
第25	0.0041	0.0181	0.0138	0.0068

图8：当使用算法4和Ξ_{train}-训练（上）或 θ_{ref}-训练（下）选择观测算子 $G_{L,\theta}$ 时，后验协方差矩阵 $\Sigma^{\theta_{\text{ref}},}$ $_L$ post 在 A-OED、D-OED、E-OED 准则和可观测性系数 $\beta_G(\theta_{\text{ref}})$ 下的排名。通过比较每个传感器库中所有可能的8个传感器无序组合，得出排名。左侧：200个传感器库排名的箱线图；右侧：不同百分位数下的排名（百分比）。



(a) 最上层， Ξ_{train}-训练　(b) 最上层， θ_{ref}-训练

(c) 最低层， Ξ_{train}-训练　(d) 最低层， θ_{ref}-训练

图 9：算法4从包含 47×47 个可用水平位置且每个位置有5个可用深度的网格中选出的传感器位置，但最终仅选择了最低层（底层）和最上层（顶层）。底层图显示了 $\theta = \theta_{\text{ref}}$处全阶解 $x_\theta(\mathbf{u})$ 的剖面。左： Ξ_{train}-训练，使用RB代理模型，训练集 $\Xi_{\text{train}} \subset \mathcal{P}$ 包含10,000个随机配置；使用10个传感器时，运行时间为14.19秒。右： θ_{ref}-训练，使用参考参数处的全阶解；使用10个传感器时，运行时间为15.85秒。

5.3. 不受限库

接下来，我们验证算法4在大型传感器库中的可扩展性：允许使用全部 2,209 个钻井位置，每个位置最多可在 5 个可用测量深度中的任意一个深度进行一次测量。选择 10 个无序传感器约有 7.29e+33 种可能组合。使用之前的RB代理模型，我们在由 10,000 个随机选取的配置组成的训练网格 $\Xi_{\text{train}} \subset \mathcal{P}$ 上仅使用替代模型运行一次算法4（运行时间为 14.19 秒），并在参考配置 θ_{ref} 上使用全阶模型运行一次（运行时间为 15.85 秒）以供比较。算法每选出 10 个传感器后即终止。与之前的训练时间 $\mathcal{L}_{5\times5}$ 相比，结果证实，库本身的规模对总体运行时间影响甚微，而全阶计算以及 Ξ_{train} 相对于替代模型计算的规模则是主要影响因素。

算法4两次运行所选出的传感器如图 9所示。它们具有许多结构相似性：

- **深度：** 尽管有 5 个测量深度可供选择，传感器却只布置在最底层和最上层，每层各 5 个。首先选择的是较低层的传感器（有一个例外，即 θ_{ref}-训练中的传感器 3），这大概是因为

较低层更靠近不确定的诺伊曼边界条件，因此测量值更大。

- **配对**最低层上的每个传感器，在最上层都有一个对应传感器，它们在水平面上的位置几乎相同。这种配对旨在降低噪声敏感性：根据给定的误差协方差函数，两次测量在水平方向上的位置越接近，其噪声相关性就越强，而与深度坐标无关。选择靠近表面零狄利克雷边界的参考测量值，有助于滤除较低测量值中的噪声项。

- **布局**每一层上的传感器均匀分布，近似排列成3行3列。这种排列有助于区分南北方向和东西方向上不确定诺伊曼通量函数的常数项、线性项和二次项。

˜图10（左侧）展示了可观测性系数 $\beta_G(\theta)$（用于Ξ_{train}训练）和 $\beta_G(\theta_{\text{ref}})$（用于 θ_{ref}训练）随所选传感器数量的增加情况。我们再次观察到， Ξ_{train}训练的传感器对应的曲线起初陡升，随后趋于停滞；而 θ_{ref}训练的曲线一开始就处于较高值，之后几乎保持不变。后者可以由图 9（右侧）中前5个传感器的位置来解释：它们在两个方向上已经彼此间隔开，足以识别二次多项式。相比之下，对于 Ξ_{train}训练，只有在选取第六个传感器后才形成“3行3列”的结构（参见图 9 左侧）。在传感器数量为6时，两种训练方案的可观测性系数均已超过此前在较小库 $\mathcal{L}_{5\times5}$ 上训练、使用8个传感器时的最终可观测性系数。参考参数 θ_{ref} 处的最终可观测性系数为： θ_{ref}训练时 $\beta_G(\theta_{\text{ref}}) = 0.4042$； Ξ_{train}训练时 $\beta_G(\theta_{\text{ref}}) = 0.3595$。

作为最后一项实验，我们将 $\Sigma^{L,\ \theta_{\text{ref}}\text{后验分布}}$协方差矩阵的特征值与 50,000 组随机传感器进行比较，每组包含 10 个传感器，分别用于对比通过 Ξ_{train} 和 θ_{ref} 训练的传感器。我们确认，全部 50,000 个传感器组合均符合组合限制。图10（右侧）给出了这些特征值的箱线图。通过算法4选出的传感器所对应的后验协方差矩阵的特征值，小于随机传感器组合的所有后验特征值[10] 。

6. 结论

在这项工作中，我们分析了在线性超参数化正向模型的贝叶斯反演中，观测算子与后验协方差矩阵特征值之间的联系。我们确定了一种可观测性系数，最大化该系数能够降低所有超参数对应的后验概率分布中的不确定性。为此，我们提出了

[10]这里，我们比较一个矩阵的最大特征值与另一个矩阵的最大特征值、次大特征值与次大特征值，依此类推。

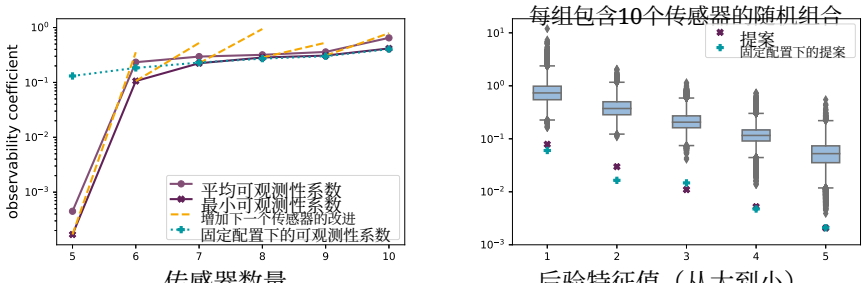


图10：左：对于具有11,045个测量位置和组合限制的库，使用 Ξ_{train}训练和 θ_{ref}训练进行传感器选择期间的可观测性系数。图中显示：1）在包含10,000个随机配置的训练集上，代理可观测性系数的最小值和均值 $\tilde{\beta}_G(\theta)$分别为最终值$\min_\theta \tilde{\beta}_G(\theta) = 0.4160$和$\text{mean}_\theta \tilde{\beta}_G(\theta) = 0.6488$；2）仅使用参考参数 θ_{ref} 进行训练时的全阶可观测性系数 $\beta_G(\theta_{\text{ref}})$（最终值为 $\beta_G(\theta_{\text{ref}}) = 0.4042$）。右：从施加组合限制的 $5 \times 47 \times 47$ 网格中均匀选取的50,000组、每组包含10个传感器的后验协方差矩阵 $\Sigma^{L,\theta}$ post的5个特征值的箱线图。根据特征值从大到小的顺序对其进行比较。同时还标出了图 9中 Ξ_{train}训练传感器（紫色，“x”标记）和 θ_{ref}训练传感器（青绿色，“+”标记）的特征值。

一种传感器选择算法，通过迭代扩展观测算子，确保所有超参数对应的可观测性系数都足够大。在贪心步骤中使用降阶模型，并通过正交匹配追踪搜索下一个传感器；该搜索只需进行一次全阶模型评估，从而保证计算上的可行性。我们在西澳大利亚珀斯盆地一处区域的大规模热传导问题上验证了该方法的有效性。未来计划从以下方面扩展这项工作：1）通过参数降维技术处理高维参数空间；2）结合 PBDW *inf-sup* 判据，除噪声协方差外，还从泛函分析角度为传感器提供依据；3）通过拉普拉斯近似将方法扩展到非线性模型。

致谢

我们谨向 Tan Bui-Thanh、Youssef Marzouk、Francesco Silva、Andrew Stuart、Dariusz Ucinski 和 Keyi Wu 致谢，感谢他们提供的宝贵讨论；同时感谢亚琛工业大学计算地球科学、地热学与储层地球物理研究所的 Florian Wellmann 提供珀斯盆地模型。本研究得到德国联邦政府和州政府卓越计划以及德国研究联合会的支持，资助项目为 GSC 111 和 33849990/GRK2379（IRTG 现代逆问题）。本项目还获得欧洲研究委员会（ERC）通过欧盟地平线2020研究与创新计划提供的资助（资助协议编号 818473）、美国能源部资助（DE-SC0021239）以及美国空军科学研究办公室资助（FA9550-21-1-0084）。Peng Chen 部分获得美国国家科学基金会 DMS #2245674 项目资助。

参考文献

[1] A. M. Stuart, 《逆问题：贝叶斯视角》, Acta Numerica 19（2010年）451–559。[2] I. Stober、K. Bucher, 《地热能》, 施普林格, 2012年。[3] D. Ucinski, 《分布参数系统辨识的最优测量方法》, CRC出版社, 2004年。[4] V. B. Melas, 《最优实验设计的泛函方法》, 第184卷, Springer Science & Business Media, 2006年。[5] L. Pronzato, 《最优实验设计及相关控制问题》, Automatica 44（2008年）303–325。[6] N. Aretz-Nellesen、P. Chen、M. A. Grepl、K. Veroy, 《超参数化线性贝叶斯反问题的序贯传感器选择策略》, 载于：数值数学与高级应用 ENUMATH 2019, 施普林格, 2021年, 第489–497页。[7] N. Aretz-Nellesen、M. A. M. Grepl、K. Veroy, 《参数化偏微分方程的 3D-VAR：一种认证降阶基方法》, 《计算数学进展》45（2019年）2369–2400。[8] P. Binev、A. Cohen、O. Mula、J. Nichols, 《利用降阶模型进行状态估计的最优测量选择贪心算法》, SIAM/ASA不确定性量化期刊 6（2018年）1101–1126。[9] Y. Maday、A. T. Patera、J. D. Penn、M. Yano, 《一种用于变分数据同化的参数化背景数据弱形式方法：公式、分析及其在声学中的应用》, 《国际工程数值方法杂志》102（2015年）933–965。[10] M. Barrault、Y. Maday、N. C. Nguyen、A. T. Patera, 《一种“经验插值”方法：在偏微分方程高效降阶基离散化中的应用》, Comptes Rendus Mathematique 339（2004年）667–672。[11] Y. Maday、O. Mula, 《广义经验插值方法：将降阶基技术应用于数据同化》, 载于：偏微分方程的分析与数值计算, 施普林格, 2013年, 第221–235页。[12] A. Alexanderian、P. J. Gloor、O. Ghattas, 《无限维空间中的贝叶斯 A-最优与 D- 最优实验设计》, 贝叶斯分析 11 （ 2016 年） 671–695。[13] A. Alexanderian、N. Petra、G. Stadler、O. Ghattas, 《采用正则化\ell 0-稀疏化的无限维贝叶斯线性反问题 A-最优实验设计》, SIAM科学计算杂志 36（2014年）A2122–A2148。

[14] A. Attia, A. Alexanderian, A. K. Saibaba，面向大规模贝叶斯线性反问题的目标导向最优实验设计，《逆问题》34（2018）095009。[15] A. Alexanderian, A. K. Saibaba，面向无限维贝叶斯线性反问题的高效 D-最优实验设计，SIAM科学计算杂志 40（2018）A2956–A2985。[16] A. Alexanderian, N. Petra, G. Stadler, I. Sunseri，在可约模型不确定性下的大规模贝叶斯线性反问题最优设计：了解自己所不知道的内容，SIAM/ASA不确定性量化期刊 9（2021）163–184。[17] K. Wu, P. Chen, O. Ghattas，一种面向目标的线性贝叶斯最优实验设计高效方法：最优传感器布置应用，arXiv预印本 arXiv:2102.06627，即将发表于《SIAM/AMS不确定性量化杂志》（2022）。[18] A. Alexanderian，由偏微分方程支配的无限维贝叶斯反问题的最优实验设计：综述，《逆问题》（2021）。[19] A. Alexanderian, N. Petra, G. Stadler, O. Ghattas，一种面向无限维贝叶斯非线性反问题的快速且可扩展的 A-最优实验设计方法，SIAM科学计算杂志 38（2016）A243–A272。[20] X. Huan, Y. M. Marzouk，面向非线性系统的基于模拟的贝叶斯最优实验设计，《计算物理学杂志》232（2013）288–317。[21] K. Wu, P. Chen, O. Ghattas，一种面向大规模、高维贝叶斯最优实验设计的快速且可扩展计算框架，arXiv预印本 arXiv:2010.15196，即将发表于SIAM科学计算杂志（2022）。[22] K. Wu, T. O’Leary-Roseberry, P. Chen, O. Ghattas，面向大规模贝叶斯最优实验设计的导数信息投影神经网络，arXiv预印本 arXiv:2201.07925，即将发表于《科学计算杂志》（2022）。[23] A. Attia, E. Constantinescu，存在观测相关性时反问题的最优实验设计，arXiv预印本 arXiv:2007.14476（2020）。[24] T. Bui-Thanh, O. Ghattas, J. Martin, G. Stadler，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科学计算杂志 35（2013）A2494–A2523。

[25] T. Cui, Y. Marzouk, K. Willcox，通过似然信息驱动的参数与状态降阶获得大规模贝叶斯反问题的可扩展后验近似，《计算物理学杂志》315（2016）363–387。[26] M. T. Parente, J. Wallin, B. Wohlmuth，活跃子空间的广义界，《电子统计学杂志》14（2020）917–943。[27] C. Lieberman, K. Willcox, O. Ghattas，大规模统计反问题的参数与状态模型降阶，SIAM科学计算杂志32（2010）2523–2542。[28] P. Chen, O. Ghattas，面向高维模型降阶的基于海森矩阵的采样，《不确定性量化国际期刊》9（2019）。[29] P. Chen, K. Wu, J. Chen, T. O'Leary-Roseberry, O. Ghattas，投影斯坦变分牛顿法：一种快速且可扩展的高维贝叶斯推断方法，NeurIPS（2019）。https://arxiv.org/abs/1901.08659。[30] P. Chen, O. Ghattas，投影斯坦变分梯度下降，载于：《神经信息处理系统进展》，2020。[31] O. Zahm, T. Cui, K. Law, A. Spantini, Y. Marzouk，非线性贝叶斯反问题中的认证降维，计算数学 91（2022）1789–1835。[32] E. Qian, M. Grepl, K. Veroy, K. Willcox，一种用于偏微分方程约束优化的认证信赖域降阶基方法，SIAM科学计算杂志39（2017）S434–S460。[33] P. Chen，不确定性量化问题的模型降阶技术，技术报告，2014。[34] P. Chen, A. Quarteroni, G. Rozza，不确定性量化的降阶基方法，SIAM/ASA不确定性量化期刊 5（2017）813–869。[35] T. O' Leary-Roseberry, U. Villa, P. Chen, O. Ghattas，由偏微分方程支配的高维参数映射的导数信息投影神经网络，《计算机方法：应用力学与工程》388（2022）114199。[36] T. O'Leary-Roseberry, P. Chen, U. Villa, O. Ghattas，导数信息神经算子：一种用于高维参数导数学习的高效框架，arXiv:2206.10745（2022）。[37] G. Da Prato，无限维分析导论，Springer Science & Business Media，2006。[38] C. Schwab, R. Stevenson，抛物型演化问题的时空自适应小波方法，计算数学 78（2009）1293–1318。

[39] Q. Long、M. Scavino、R. Tempone、S. Wang，基于拉普拉斯近似的贝叶斯实验设计中期望信息增益的快速估计，《计算机方法：应用力学与工程》259（2013）24–39。[40] N. Aretz，《可配置正向模型的数据同化与传感器选择：模型降阶方法的挑战与机遇》，博士论文，亚琛工业大学，2022年。[41] T. Cui、Y. M. Marzouk、K. E. Willcox，面向逆问题贝叶斯求解的数据驱动模型降阶，《工程数值方法国际期刊》102（2015）966–990。[42] T. Bui-Thanh、K. Willcox、O. Ghattas，面向高维参数输入空间大规模系统的模型降阶，SIAM科学计算杂志 30（2008）3270–3288。[43] P. Benner、S. Gugercin、K. Willcox，参数化动力系统基于投影的模型降阶方法综述，《SIAM评论》57（2015）483–531。[44] W. H. A. Schilders、H. A. Van der Vorst、J. Rommes，《模型降阶：理论、研究方面与应用》，第13卷，施普林格，2008年。[45] J. S. Hesthaven、G. Rozza、B. Stamm，《参数化偏微分方程的经认证约化基方法》，第590卷，施普林格，2016年。[46] A. Quarteroni、A. Manzoni、F. Negri，《偏微分方程的约化基方法：入门》，第92卷，施普林格，2015年。[47] B. Haasdonk，参数化偏微分方程的约化基方法——稳态与非稳态问题入门教程，《模型降阶与近似：理论与算法》15（2017）65。[48] G. H. Golub、C. F. Van Loan，《矩阵计算》，第3卷，约翰斯·霍普金斯大学出版社，2013年。[49] P. Binev、A. Cohen、W. Dahmen、R. DeVore、G. Petrova、P. Wojtaszczyk，约化基方法中贪心算法的收敛速率，《SIAM数学分析杂志》43（2011）1457–1472。[50] A. Cohen、W. Dahmen、R. DeVore、J. Fadili、O. Mula、J. Nichols，基于数据的状态估计最优约化模型算法，《SIAM数值分析杂志》58（2020）3355–3381。[51]A. Buffa、Y. Maday、A. T. Patera、C. Prud’homme、G. Turinici，参数化约化基方法中贪心算法的先验收敛性，ESAIM：数学建模与数值分析-Mod´elisation Math´ematique et Analyse Num´erique 46（2012）595–603。[52]J. Jagalur-Mohan、Y. M. Marzouk，非次模函数的批量贪心最大化：保证与实验设计应用，《机器学习研究杂志》22（2021）251–252。

[53] J. L. Eftang、A. T. Patera、E. M. Rønquist，参数化椭圆偏微分方程的一种“hp”经认证约化基方法，SIAM科学计算杂志 32（2010）3170–3200。[54] J. L. Eftang、D. J. Knezevic、A. T. Patera，参数化抛物型偏微分方程的一种 hp 经认证约化基方法，《数学与计算机动力系统建模》17（2011）395–422。[55] K. Regenauer-Lieb、F. Horowitz，珀斯盆地的地热开发机遇，《西澳大利亚石油》3（2007）。[56] S. Corbel、O. Schilling、F. G. Horowitz、L. B. Reid、H. A. Sheldon、N. E. Timms、P. Wilkes，澳大利亚珀斯大都市区断层的识别及其地热影响，载于：第三十七届地热储层工程研讨会，加利福尼亚州斯坦福，2012年。[57] H. A. Sheldon、B. Florio、M. G. Trefry、L. B. Reid、L. P. Ricard、K. A. R. Ghori，西澳大利亚珀斯盆地的对流潜力及其对地热能的影响，《水文地质学杂志》20（2012）1251–1268。[58] O. Schilling、H. A. Sheldon、L. B. Reid、S. Corbel，西澳大利亚珀斯大都市区的热液模型：对地热能的启示，《水文地质学杂志》21（2013）605–621。[59] M. Pujol、L. P. Ricard、G. Bolton，西澳大利亚珀斯盆地亚拉加迪含水层直接利用地热的20年开发历程，《地热学》57（2015）39–55。[60] J. F. Wellmann、L. B. Reid，盆地尺度地热模型校准：来自澳大利亚珀斯盆地的经验，《能源过程》59（2014）382–389。[61] D. Degen、K. Veroy、F. Wellmann，地球科学中的经认证约化基方法，《计算地球科学》24（2020）241–259。[62] D. M. Degen，《约化基方法在地球物理模拟中的应用：概念、实现、优势与局限》，博士论文，亚琛工业大学，2020年。doi:10.18154/RWTH-2020-12042。[63] M. Bauer、W. Freeden、H. Jacobi、T. Neu，《深层地热手册》，施普林格，2014年。[64] M. de la Varga、A. Schaaf、F. Wellmann，GemPy 1.0：开源随机地质建模与反演，《地球科学模型开发》12（2019）1–32。[65] C. J. Permann、D. R. Gaston、D. Andrˇs、R. W. Carlsen、F. Kong、A. D. Lindsay、J. M. Miller、J. W. Peterson、A. E. Slaughter、R. H. Stogner，MOOSE：实现大规模并行多物理场模拟，《SoftwareX》11（2020）100430。

[66] W. Dahmen, C. Plesken, G. Welper, 《双重贪心算法：输运主导问题的约化基方法》, ESAIM：数学建模与数值分析-Mod´elisation Math´ematique et Analyse Num´erique 48（2014年）623–663.[67] N. Cressie, 《克里金法的起源》, 《数学地质学》22（1990年）239–252.[68] F. L. Holgate, E. J. Gerner, OzTemp井温度数据, 澳大利亚地球科学局 http://www. ga. gov. au 目录（2010年）.[69] Y. Maday, T. Anthony, J. D. Penn, M. Yano, PBDW状态估计：含噪观测；配置自适应背景空间；物理解释, ESAIM：会议录与综述 50（2015年）144–168.[70] T. Taddei, 《数据同化的模型降阶方法：状态估计与结构健康监测》, 博士论文, 麻省理工学院, 2017年。