arXiv:2601.04473v1 [math.ST] 8 Jan 2026

学习伪微分算子的收敛速率

Jiaheng Chen 和 Daniel Sanz-Alonso

芝加哥大学

摘要

本文确立了学习椭圆伪微分算子的收敛速率。椭圆伪微分算子是偏微分方程和数理物理中的一类基本算子。在小波–伽辽金框架下，我们将这一算子类上的学习表述为一个具有多尺度稀疏性的结构化无限维回归问题。基于这一结构，我们提出了一种稀疏且兼顾数据与计算效率的估计量，它利用了一种针对该学习任务设计的新型矩阵压缩方案，以及一种用于平衡逼近误差和估计误差的嵌套支撑策略。除了获得该估计量的收敛速率外，我们还证明，学习得到的算子能够导出一种高效且稳定的伽辽金求解器，其数值误差与统计精度相匹配。因此，我们的研究有助于推动算子学习、数据驱动求解器和小波方法在科学计算中的融合。

关键词。算子学习；伪微分算子；小波**数学主题分类代码**。62G05；
65T60；35S05

1 引言

算子学习是科学计算、偏微分方程（PDEs）与机器学习交叉领域中的一种新兴范式 [96, 5, 104, 103, 116, 88, 97, 17, 90, 133]。在监督学习场景中，任务可以表述如下：对于未知算子 $\mathcal{A}$，我们观测到满足以下条件的含噪输入–输出对$\{(u_i,\ f_i)\}_{i=1}^N$：

$$f_i = \mathcal{A}u_i + w_i, \qquad 1 \le i \le N, \tag{1.1}$$

其中$\{w_i\}_{i=1}^N$ 表示噪声。我们的目标是在给定误差度量以及关于输入、输出和噪声的建模选择下，根据数据准确且高效地估计 $\mathcal{A}$ 。

本文为学习椭圆伪微分算子（PDOs）建立了收敛速率。椭圆伪微分算子是分析学和偏微分方程中的一大基础算子类 [74, 125, 136, 145]。伪微分算子涵盖微分算子和积分算子，并能自然地刻画非局部性和病态性；典型例子包括格林算子、狄利克雷到诺伊曼映射以及边界积分算子。我们假设输入$\{u_i\}_{i=1}^N$ 和噪声项$\{w_i\}_{i=1}^N$ 均为具有指定索博列夫光滑性的高斯随机函数（包括 Matérn 类），并在 Sobolev空间之间的算子范数下度量误差。

受偏微分算子在小波坐标中的（准）稀疏表示以及科学计算中的小波–伽辽金方法启发 [11, 44, 122, 34, 45, 36, 37, 68], ，我们提出了一种$\widehat{\ }$稀疏、数据与计算效率兼具的基于小波的估计量 $\mathcal{A}$，并证明，以高概率且相差至多多对数因子，

$$\|\widehat{\mathcal{A}} - \mathcal{A}\|_{H^t \to H^{-t'}} \lesssim N^{-\frac{1}{2+\rho}}.$$

指数 $\rho \geq 0$ 取决于误差度量 (t, t') 中的索博列夫指标 $\|\cdot\|_{H^t \to H^{-t'}}$、算子的阶 $\mathcal{A}$、空间维数、输入与噪声的光滑性，以及小波正则性/逼近阶参数。我们的理论还涵盖无噪声情形（$w_i \equiv 0$ 式（1.1）中），在这种情形下，估计量达到超代数收敛。最后，我们证明 $\mathcal{A}$ 可得到一种高效且稳定的小波–伽辽金求解器，其数值误差与统计精度相匹配，从而在科学计算中架起算子学习、数据驱动偏微分方程求解器与小波方法之间的桥梁。

1.1 主要贡献与提纲

- **在小波–伽辽金框架下学习偏微分算子。** 我们在小波–伽辽金框架下构建偏微分算子学习问题，将算子学习任务转化为对算子双无限小波矩阵表示的估计。通过这种方式，我们将连续算子学习问题化为结构化无限维矩阵回归问题。

- **具有多尺度稀疏性的结构化无限维回归。** 我们利用偏微分算子在小波坐标下的（准）稀疏表示，提出一种稀疏且兼具数据与计算效率的估计量，用于未知算子的小波–伽辽金离散。该方法包含两个关键要素：

  (i) 面向学习的矩阵压缩。 我们引入一种针对算子学习误差度量设计的新型压缩方案，确定一个目标稀疏支撑集，以捕获达到所需精度所必需的小波系数。与科学计算中的经典方法相比，我们的目标导向压缩方案能够提供更精确的学习保证。

  (ii) 嵌套支撑回归。 为减轻在目标支撑集上进行回归时由数据引起的遗漏变量偏差，我们在经过谨慎扩大的回归支撑集上拟合每一列，然后再限制回目标支撑集。我们对支撑集的扩展进行了校准，使遗漏变量偏差降低至不超过截断误差与压缩误差的量级，同时保持方差处于相同量级。这种嵌套支撑回归方法为平衡数据引起的偏差与截断和压缩带来的逼近误差提供了一种有原则的方法，对于其他结构化学习问题也可能具有独立的研究价值。

- 椭圆偏微分算子的学习收敛速率。在对输入、噪声和小波选择作出一般假设的条件下，我们的第一个主要结果——定理 5.1——为学习椭圆偏微分算子确立了一个近乎最优的高概率误差界，其阶为 $N^{-\frac{1}{2+\rho}}$（相差多对数因子）。该误差界呈现出一个参数化区域：当误差度量足够弱时，达到速率 $N^{-1/2}$（即 $\rho = 0$）；否则为 $\rho > 0$，从而得到严格的非参数率。该估计量在小波坐标中具有最优稀疏性，仅含有 $\mathcal{O}(2^{Jn})$ 个非零元素，且可在近乎 $\mathcal{O}(N2^{Jn})$ 的时间内计算，其中 2^{Jn} 是所学习的伽辽金矩阵的维数，N 是样本量；在偏差-方差最优选择截断层级 J（作为 N 的函数）时，这对应于 $\mathcal{O}\big(N^{\frac{n}{2+\rho}}()(_{t+t'-r})$ 个非零系数，总运行时间为 $\mathcal{O}\big(N^{1+\frac{n}{2+\rho}}()(_{t+t'-r})\big)$，其中$n$ 是空间维数，r 是偏微分算子的阶。定理 5.1 的证明依赖于精细的多尺度误差分析，该分析控制了截断误差和压缩误差，以及遗漏变量偏差和方差。一项关键的技术贡献是将精细的概率分析与一种由我们新提出的、面向学习的矩阵压缩方案所诱导的分块多尺度稀疏模式的精细刻画相结合。我们的理论也涵盖无噪声情形，此时估计量达到超代数收敛速率。作为一般结果的推论，我们在例 5.7 中获得了据我们所知，首个关于从一般维数下的通用含噪高斯数据中学习椭圆型偏微分方程的格林函数时，统计精度与计算成本之间显式且近乎最优的权衡

且适用于一般微分阶数。我们还将无噪声情形下已有的精度-成本权衡改进为近乎最优的权衡。

- **数据驱动偏微分方程求解器的收敛速率**。学习得到的稀疏算子可构造出高效且稳定的小波–伽辽金求解器。我们的第二个主要结果——定理 6.2——表明，该求解器的数值误差继承了算子估计量的统计精度（至多相差多对数因子），因此会随样本量 N以代数速率衰减。定理6.2 由此为科学计算中的算子学习、数据驱动偏微分方程求解器和小波–伽辽金方法搭建了一个有理论依据的桥梁。

本文结构安排如下。在引言余下部分讨论相关工作并介绍符号约定之后，我们将在第 2 节形式化定义算子学习问题。随后，我们将在第3 节的小波–伽辽金框架下阐述算子学习。第 4 节将通过面向学习的矩阵压缩和嵌套支撑回归定义我们的估计量。第 5 节和第 6 节将分别确立学习椭圆偏微分算子和数据驱动偏微分方程求解器的收敛速率。第 7 节将以结论和未来工作方向作结。

1.2 相关工作

算子学习 关于算子学习的文献浩如烟海且迅速增长，全面综述已超出本文的范围。具有代表性的例子包括基于神经网络的方法，例如傅里叶神经算子 [96, 116, 88], DeepONets [103, 104, 94], 和物理信息神经算子 [97, 80]；基于核的方法 [128, 10, 77,102, 147, 150, 79]；以及随机特征方法 [111, 112, 98, 149]。相关研究方向还包括学习格林函数 [16, 128, 121, 15, 14, 142], 面向动力系统的算子学习（例如 Koopman 算子和生成元）[19, 86, 85, 42, 84, 101], 结构化协方差与精度算子估计 [3, 4, 2, 29], 以及在逆问题中的应用 [108, 65, 113]。从理论角度来看，近期人们对逼近理论基础 [31, 103, 87, 94, 55, 88, 91, 95, 92, 93, 53], 以及算子学习问题的统计和数据复杂度分析 [109, 54, 78, 89, 100, 132, 118, 1]都给予了极大关注。下文将聚焦于与本文最为相关的研究，并请读者参阅近期综述以了解更广泛的观点 [17, 90, 133]。

论文 [54] 研究了已知基下可对角化的算子在含噪数据条件下的线性算子学习问题，从而将问题归约为特征值学习，并在贝叶斯框架下建立了后验收缩速率。研究 [78] 探讨了如何通过岭回归型估计量学习两个无限维索博列夫再生核希尔伯特空间之间的希尔伯特–施密特算子所面临的统计极限，并在希尔伯特–施密特范数下推导出匹配的上界和下界；其分析归结为对一个具有多项式型逐项衰减特征的无限维系数矩阵进行估计。与 [54, 78], 不同，我们研究的是索博列夫空间之间椭圆偏微分算子的算子范数学习。相比以更“平均”的方式度量误差，这种最坏情况保证与数据驱动偏微分方程求解器的稳定性和误差分析直接兼容，也与经典小波–伽辽金方法自然契合。从技术角度看，我们的问题本质上是非对角的：偏微分算子的双无限小波矩阵呈现多尺度（准）稀疏结构，而非逐项多项式衰减特征，因此形成了一个高度结构化的无限维回归问题。

另一相关研究方向 [16, 14] 通过随机数值线性代数恢复椭圆型偏微分方程的格林函数（解算子），并利用充分分离的区域上的低秩结构以及格林函数的非对角衰减，获得了随样本量呈指数收敛的结果。该研究设定无噪声（即 $w_i \equiv 0$ 在（1.1）中），并依赖于精心设计、与格林函数层次化低秩结构相匹配的输入；而我们的框架涵盖了统一的积分算子类和微分算子类，并允许含噪观测以及通用的高斯过程输入数据。此外，他们的分析

仅针对三维情形，其误差界依赖于解算子的希尔伯特–施密特范数；在更高维度下，该范数可能为无穷大，因为格林函数可能不是平方可积的。相比之下，我们的小波框架以统一方式处理一般维度，并采用一般的索博列夫空间到索博列夫空间的算子范数。近期，[118] 通过经验过程技术，根据逼近性质和度量熵界，针对一般算子类上的最小二乘经验风险极小化器建立了普适的收敛保证。相比之下，我们聚焦于一个具有物理意义、受偏微分方程启发的算子类，并通过显式的基于小波的多尺度回归分析推导收敛速率，该分析利用了偏微分算子在小波坐标中的可压缩性；此外，我们的估计量稀疏且计算高效。

我们还提及受偏微分算子启发或旨在表示/逼近偏微分算子的神经网络架构 [64, 20, 66, 148, 124, 30]。我们的贡献与此互为补充：据我们所知，我们首次为偏微分算子的学习提供了严格的统计收敛速率。虽然可以从函数型线性回归的角度理解我们的估计量 [117, 67], ，但其关键创新在于，系数结构受偏微分算子小波表示的先验算子分析界所支配，由此得到一种多尺度可压缩模式，我们在统计和计算方面都加以利用。

小波–伽辽金方法小波方法在科学计算和数值分析中有着悠久的历史；例如，参见 [105, 11, 7, 52, 50, 48, 49, 71, 44, 36, 33, 70]。相关书籍和综述可参见 [52, 44, 34, 35, 122, 131]。近年来，相关思想也被应用于其他领域，例如高斯随机场的多层逼近[72, 68, 6]。小波的三个特性在数值分析中尤为重要。首先，小波系统能够刻画索博列夫空间（以及 Besov 空间）：在适当构造下，它们构成一系列索博列夫空间的里斯基基。这一性质使得通过伽辽金方法在小波基上离散化的微分算子和积分算子具有简单的对角预条件子 [46, 76, 44]。其次，消失矩（消去性质）使小波能够利用算子光滑性，从而为广泛类别的算子提供稀疏或可压缩表示 [11, 50, 48, 49, 44, 129], ，且适用于标准形式和非标准形式。第三，小波支持自适应性和非线性逼近，从而能够对函数进行最优有限项逼近和稀疏表示 [57, 56, 36, 33]。这些原理共同构成了快速、自适应、多尺度数值算法的基础，使其能够利用近似稀疏性。除数值分析外，小波思想也长期影响着应用数学和数据科学的诸多领域，包括通过阈值化进行统计估计 [60, 61, 62, 58], 、图像处理中的数据压缩（例如 JPEG 2000） [38, 134, 126, 135], 、压缩感知 [59, 27], ，以及近年来与神经网络的关联 [18, 106]。

经典小波文献中有许多工作致力于数值模拟和分析学，而本文则从小波视角研究算子学习，旨在建立从含噪输入–输出数据对中学习未知算子的统计收敛速率。我们的估计量以算子的双无限小波矩阵表示为目标，并充分利用对角预条件以及偏微分算子在小波坐标中的（准）稀疏性，同时呈现出经典小波–伽辽金偏微分方程求解器中所没有的若干特征。我们强调以下几点，以说明本文与经典小波–伽辽金方法的区别。首先，有别于 [122, 45, 68], 中使用的矩阵压缩方案，我们提出一种面向学习的压缩方案，以适合统计学习任务的方式确定需要估计的矩阵元素。其次，我们提出嵌套支撑回归策略来构造估计量，并减轻数据导致的遗漏变量偏差——这种效应在确定性数值离散中并不存在。此外，我们还加入一个对称化步骤，利用未知算子的对称性来降低方差。第三，我们的误差分析需要对小波矩阵的分块稀疏结构进行更精细的刻画：除了全局层面的非零元素数（nnz）之外，

我们还量化了对应于给定尺度对的小波系数块内的稀疏性。这种改进对于算子学习至关重要，因为在算子范数下，估计量的方差会跨尺度累积，而仅靠全局 nnz 无法刻画这一点。最后，在第6 节中，我们讨论将学习得到的稀疏算子用作代理的、数据驱动的偏微分方程求解器，进一步阐明其与经典小波–伽辽金方法之间的联系与区别。总的来说，本文有助于推动算子学习、数据驱动偏微分方程求解器与小波–伽辽金方法之间的融合。

1.3 记号

全文中，M 表示一个维数为 n 、浸入欧几里得空间 R^D 的光滑、闭、连通且可定向黎曼流形，其中 $D > n.$ 。实值函数 $\phi : \mathrm{M} \to \mathrm{R}$ 的支撑集记为 $\mathrm{supp}(\phi) := \overline{\{x \in \mathrm{M} : \phi(x) \neq 0\}}$，其中闭包取于 M 中。若对某个子集 $\mathrm{M}' \subset \mathrm{M}$，存在紧集 M'' 使得 $\mathrm{M}' \subset \mathrm{M}'' \subset \mathrm{M}$，则称 M' 紧包含于 M，并记作 $\mathrm{M}' \Subset \mathrm{M}$。M 上所有光滑实值函数构成的空间记为 $C^\infty(\mathrm{M})$，而 $C_0^\infty(\mathrm{M}) \subset C^\infty(\mathrm{M})$ 表示支撑集满足 $\mathrm{supp}(\phi) \Subset \mathrm{M}$ 的光滑函数 ϕ 所构成的子空间。空间 $L^2(\mathrm{M})$ 由关于 M 上内蕴测度平方可积的所有函数构成，其内积记为 $(\cdot,\cdot)_{L^2(\mathrm{M})}$。M 上的拉普拉斯–贝尔特拉米算子记为 $\Delta_{\mathcal{M}}$。对任意 $s \in [0,\infty)$，记 $H^s(\mathrm{M})$ 为 M 上 s 阶的 Sobolev空间，它通过拉普拉斯–贝尔特拉米算子的谱分解定义；其范数记为 $\|\cdot\|_{H^s(\mathrm{M})}$。对于$s > 0$，记 $H^{-s}(\mathrm{M})$ 为 $H^s(\mathrm{M})$ 的对偶空间，并用 $\cdot,\cdot$ 表示 $H^s(\mathrm{M})$ 与 $H^{-s}(\mathrm{M})$ 之间的对偶配对。为简化记号，在不会引起歧义时，我们常省略 Sobolev空间对 M 的依赖，以及对偶配对对 s 的依赖。有时，我们会在有界开区域上而非流形上工作，此时将明确指出函数空间对该区域的依赖。

对于向量 $x \in \mathrm{R}^m$，我们用 $\|x\|_2$ 表示其欧几里得范数，用 $\|x\|_{\max}$ 表示其逐项最大范数。对于矩阵 $B \in \mathrm{R}^{m_1 \times m_2}$，我们用 $\|B\|$表示其谱范数（算子范数），用 $\|B\|_1$表示其矩阵 ℓ_1 范数（列绝对值之和的最大值），用 $\|B\|_\infty$ 表示其矩阵 ℓ_∞ 范数（行绝对值之和的最大值），并用 $\|B\|_{\max}$ 表示其逐项最大范数。更一般地，对于由集合 Λ_1、Λ_2 （可能是无限集）索引的矩阵$B \in \mathrm{R}^{\Lambda_1 \times \Lambda_2}$，我们将 B 视为从Λ_2到Λ_1的线性算子$\ell^2(\Lambda_2) \to \ell^2(\Lambda_1)$，并用 $\|B\|$表示其算子范数。对于对称半正定矩阵，我们分别用 $\sigma_{\min}(B)$ 和 $\sigma_{\max}(B)$表示其最小和最大特征值。对于有限集 Λ，我们用 $|\Lambda|$ 表示其基数。

给定两个正数列 $\{a_k\}$ 和 $\{b_k\}$，若存在一个与 k 无关的常数$c > 0$，使得对所有 k 均有 $a_k \le c\, b_k$，则记作 $a_k \lesssim b_k$。若 $a_k \lesssim b_k$ 和 $b_k \lesssim a_k$ 均成立，则记作 $a_k \asymp b_k$。若常数 c 依赖于某个参数 τ，则记作 $a_k \lesssim_\tau b_k$，$b_k \lesssim_\tau a_k$，并用 $a_k \asymp_\tau b_k$ 表示这种依赖关系。

2 问题设定

Th本节介绍偏微分算子的背景知识，并将我们的算子学习问题形式化。 .

2.1 伪微分算子

设 $\mathcal{D} \subset \mathrm{R}^n$ 是一个有界开区域。对于 $r \in \mathrm{R}$，阶为 r 的伪微分算子类 $\mathrm{OPS}^r(\mathcal{D})$（偏微分算子）由所有线性映射构成

$$\mathcal{A} : C_0^\infty(\mathcal{D}) \to C^\infty(\mathcal{D}), \quad (\mathcal{A}u)(x) := \frac{1}{(2\pi)^n} \int_{\mathbb{R}^n} e^{ix\cdot\xi} a(x,\xi) u^{\mathrm{ft}}(\xi) d\xi,$$

其中，$u^{\mathrm{ft}}(\xi) := \int_{\mathcal{D}} u(x) e^{-ix\cdot\xi} dx$ 表示 u的傅里叶变换，而 $a(x,\ \xi)$ 是属于 Hörmander 类 $\mathrm{S}^r(\mathcal{D})$ 的符号。符号类 $\mathrm{S}^r(\mathcal{D})$ 由所有函数组成

$a \in C^\infty(\mathcal{D} \times \mathrm{R}^n)$，使得对于所有 $K \Subset \mathcal{D}$ 以及每一对多重指标 α，$\beta \in \mathrm{N}_0^n$，存在一个常数 $C_{K,\alpha,\beta} > 0$，满足

$$|\partial_\xi^\alpha \partial_x^\beta a(x,\xi)| \le C_{K,\alpha,\beta}(1+|\xi|)^{r-|\alpha|}, \quad \text{for all } x \in K,\ \xi \in \mathbb{R}^n.$$

这里，N_0^n表示非负整数的 n 元组，而 $|\alpha| := \alpha_1 + \cdots + \alpha_n$. 直观而言，阶$r \in \mathrm{R}$ 刻画了 $\mathrm{OPS}^r(\mathcal{D})$ 类中算子的正则性：正阶对应微分算子，而负阶对应积分（平滑）算子。

对于流形 M，类 $\mathrm{OPS}^r(\mathrm{M})$ 按照通常的方式，通过局部坐标 [136]提升到 M 上来定义。具体而言，若线性算子 $\mathrm{B} : C^\infty(\mathrm{M}) \to C^\infty(\mathrm{M})$ 满足：对于相对于 M 的图册 $\{\mathrm{M}_i, \gamma_i\}_{i=1}^M$ 的任意有限光滑单位分解 $\{\chi_i \in C_0^\infty(\mathrm{M}_i) : i = 1, \dots, M\}$，所有变换后的算子

$$f \mapsto B_{i,i'} f := [(\mathcal{B}[\chi_i (f \circ \gamma_i^{-1})])\chi_{i'}] \circ \gamma_{i'}$$

对所有 $i, i' = 1, \dots, M$均属于 $\mathrm{OPS}^r\big(\gamma_i^{-1}(M_i)\big)$，则称 B 属于 $\mathrm{OPS}^r(\mathrm{M})$。因此，$\mathrm{OPS}^r(M)$ 是内蕴定义的，且不依赖于图册 [74, 136]的选取。更多详情请参见 [68, 附录 A.2。2] 本文中，我们始终假设M 是一个 n 维光滑闭流形，例如$M = \mathrm{T}^n$（n维环面）和$M = \mathrm{S}^n$（n维球面）；通过施加适当的边界条件，可以将我们的结果推广到带边界的区域。

对于 t，$t' > r/2$，我们将考虑 $\mathcal{A} \in \mathrm{OPS}^r$ 的算子范数，其定义为

$$\|\mathcal{A}\|_{H^t \to H^{-t'}} := \sup_{\|w\|_{H^t} \le 1} \|\mathcal{A}w\|_{H^{-t'}} = \sup_{\|w\|_{H^t} \le 1, \|v\|_{H^{t'}} \le 1} |\langle \mathcal{A}w, v\rangle|, \tag{2.1}$$

其中，对偶配对 $\langle \cdot, \cdot \rangle$ 取于 $H^{t'}$ 和 $H^{-t'}$ 之间。

2.2 算子学习

考虑统计模型

$$f_i = \mathcal{A}u_i + w_i, \qquad 1 \le i \le N, \tag{2.2}$$

其中，$\mathcal{A}$ 是未知算子，$\{(u_i,\ f_i)\}_{i=1}^N$ 是给定数据对，$\{w_i\}_{i=1}^N$ 是噪声项。我们关注以下算子学习问题：

给定数据对 $\{(u_i,\ f_i)\}_{i=1}^N$，在范数（2.1）下估计未知算子 $\mathcal{A}$ 。

我们将在关于未知算子 $\mathcal{A}$、输入函数 $\{u_i\}_{i=1}^N$ 以及噪声项 $\{w_i\}_{i=1}^N$ 的以下标准假设下开展研究。

假设 2.1（算子、输入函数和噪声）。 ***(i)*** 算子假设：对于某个给定的 $r \in \mathbb{R}$ $\mathcal{A} \in \mathrm{OPS}^r(M)$。此外，$\mathcal{A}$ 是自伴且正定的，其含义是

$$\langle \mathcal{A}v, v\rangle \gtrsim \|v\|_{H^{r/2}}^2, \quad \forall\ v \in H^{r/2}.$$

(ii) 输入函数假设：输入函数 $\{u_i\}_{i=1}^N$ 是定义在 M 上、协方差算子为 $\mathcal{C}_u \in \mathrm{OPS}^{-2r_1}(M)$ 的中心化高斯过程的独立同分布样本，其中 $r_1 > n/2 + \max\{0, r\}$ 为某个给定值。算子 $\mathcal{C}_u$ 是自伴且正定的，满足

$$\langle \mathcal{C}_u v, v\rangle \gtrsim \|v\|_{H^{-r_1}}^2, \quad \forall\ v \in H^{-r_1}.$$

(iii) 噪声假设：噪声项 $\{w_i\}_{i=1}^N$ 是来自 M 上中心化高斯过程的独立同分布样本，与 $\{u_i\}_{i=1}^N$相互独立，其协方差算子为 $\mathcal{C}_w \in \mathrm{OPS}^{-2r_2}(M)$，其中 $r_2 > n/2$ 为给定值。算子 $\mathcal{C}_w$ 是自伴且正定的，满足

$$\langle \mathcal{C}_w v, v\rangle \gtrsim \|v\|_{H^{-r_2}}^2, \quad \forall\ v \in H^{-r_2}.$$

我们的理论也涵盖无噪声数据情形，其中 $f_i = \mathcal{A}u_i$，$1 \le i \le N$；参见注5.6。请注意，我们假设参数 r，r_1，r_2（它们决定 $\mathcal{A}$ 以及输入/噪声协方差算子的阶）是已知的。在实践中，可能需要根据数据估计这些参数（例如参见 [68, 第 5 节4] 和 [83]）。研究不依赖于这些干扰参数的自适应估计量，是未来工作中一个有趣的方向。

注 2.2（假设 2.1 的解释与示例）。算子 $\mathcal{A}$。根据假设 2.1 (i)，对于每个 $s \in \mathbb{R}$，算子 $\mathcal{A}$ 是从 $H^s(M)$ 到 $H^{s-r}(M)$ 的连续双射；参见 [68,命题 11]。此外，对于任意实指数 ϱ，其分数次幂满足 $\mathcal{A}^\varrho \in \mathrm{OPS}^{\varrho r}(M)$；参见 [68,命题 12]。

薛定谔（哈密顿）椭圆算子及其分数次幂构成了一类典型示例：

$$\mathcal{H} := -\Delta_{\mathcal{M}} + V(x), \qquad \mathcal{A} := \mathcal{H}^\varrho,$$

其中，势函数 $V \in C^\infty(M)$ 一致正，即 $V(x) \ge v_{\min} > 0$。这包括微分算子（$\varrho > 0$）和积分/平滑算子（$\varrho < 0$）。

假设 2.1 (i) 也适用于一大类源自偏微分方程的自伴椭圆边界积分算子。特别是，当 $M = \partial\mathcal{D}$ 是一个$(n+1)$维区域 $\mathcal{D} \subset \mathbb{R}^{n+1}$的闭边界曲面时，这类算子包括拉普拉斯算子的第一类边界积分方程，以及与纳维–拉梅系统和斯托克斯系统相关的边界积分表述；进一步背景请参见 [45, 70, 44]。我们认为，本文的结果可推广至更广泛的伪微分算子和边界积分算子，包括非自伴和/或非强椭圆的算子。这些推广留待未来工作研究。

另一个相关但略有不同的例子是可逆扩散的无穷小生成元。对于过阻尼朗之万扩散 $dX_t = -\nabla U(X_t)\,dt + \sqrt{2/\eta}\,dW_t$，生成元 $\mathcal{G} = -\nabla U \cdot \nabla + \eta^{-1}\Delta$ 是一个二阶椭圆微分算子（因此也是一个偏微分算子），但通常在 $L^2(M)$ 上并非自伴。相反，它在加权空间$L^2_\pi(M)$ 中是对称的，其中 $\pi(x) \propto e^{-\eta U(x)}$ 是平稳测度。事实上，

$$\langle f, (-\mathcal{G})g\rangle_\pi = \eta^{-1}\int_{\mathcal{M}} \nabla f \cdot \nabla g\, \pi\, dx = \langle (-\mathcal{G})f, g\rangle_\pi.$$

因此，$-\mathcal{G}$ 在 $L^2_\pi(M)$ 上是自伴且半正定的（其零空间由常数组成）。关于如何运用本文提出的技术来学习这类算子的详细研究，将留待未来工作。

输入函数 u 和噪声 w。假设2.1 (ii)–(iii) 指出，输入函数$\{u_i\}_{i=1}^N$ 和噪声项$\{w_i\}_{i=1}^N$ 是中心化高斯过程，分别独立地采样自 $u_i \sim \mathcal{N}(0,\ \mathcal{C}_u)$ $w_i \sim \mathcal{N}(0,\ \mathcal{C}_w)$。关于 $\mathcal{C}_u$ 和 $\mathcal{C}_w$ 的假设意味着，u_i 和 w_i 均可表示为由空间白噪声驱动的着色算子方程的解，其着色算子属于 Hörmander 类：

$$\mathcal{L}_u u = \mathcal{W}_u, \qquad \mathcal{L}_w w = \mathcal{W}_w,$$

其中 $L_u = \mathcal{C}_u^{-1/2} \in \mathrm{OPS}^{r_1}(M)$、$L_w = \mathcal{C}_w^{-1/2} \in \mathrm{OPS}^{r_2}(M)$以及 $\mathcal{W}_u$、$\mathcal{W}_w$ 是在 $L^2(M)$上的相互独立的高斯白噪声；参见 [68, (2.1)，命题 1]。这一设定包含了广泛用于建模和实际应用的 Whittle–Matérn 型高斯随机函数 [143, 107, 144, 99, 127]。例如，可以取

$$\mathcal{L}_u := (\tau_u(x)I - \nabla\cdot(\kappa_u(x)\nabla))^{r_1/2}, \qquad \mathcal{L}_w := (\tau_w(x)I - \nabla\cdot(\kappa_w(x)\nabla))^{r_2/2},$$

其中 τ_u，$\tau_w \in C^\infty(M)$ 一致正，即 $\tau_u(x) \ge \tau_{u,\ \min} > 0$ 且 $\tau_w(x) \ge \tau_{w,\ \min} > 0$，而 κ_u，κ_w 光滑且一致椭圆（例如光滑的对称正定矩阵场）；参见 [120] 和 [68,附录 C]。

参数 r_1，r_2 控制 Sobolev正则性。根据 Weyl定律，$\mathcal{C}_u$ 和 $\mathcal{C}_w$ 的谱渐近，参见 [43, 72, 68]，可知

$$u_i \in H^s(\mathcal{M}) \quad \text{for all } s < r_1 - n/2 \quad \text{a.s.}, \qquad w_i \in H^s(\mathcal{M}) \quad \text{for all } s < r_2 - n/2 \quad \text{a.s.}$$

因此，条件 $r_1 > n/2$ 和 $r_2 > n/2$ 确保 u_i，$w_i \in L^2(\mathcal{M})$ 几乎必然成立。此外，由于 $\mathcal{A} : H^s(\mathcal{M})$ 与 $\to H^{s-r}(\mathcal{M})$ 对每个 $s \in \mathbb{R}$ 都是同构，因此我们有

$$\mathcal{A}u_i \in H^{s-r}(\mathcal{M}) \quad \text{for all } s < r_1 - n/2 \quad \text{a.s.},$$

所以 $\mathcal{A}u_i \in L^2(\mathcal{M})$，只要 $r_1 - n/2 - r > 0$。合并这些观测，在以下假设下

$$r_1 > n/2 + \max\{0, r\}, \qquad r_2 > n/2,$$

统计模型（2.2）中的所有函数——即 u_i，w_i，以及 $f_i = \mathcal{A}u_i + w_i$——对于每个 $L^2(\mathcal{M})$ 几乎必然属于 $1 \le i \le N$。□

3 小波-矩阵表述：算子学习

在本节中，我们将算子学习问题重新表述为学习双无限小波矩阵表示。为此，我们使用双正交小波系统，将连续算子-函数模型（2.2）离散化为离散的无限维矩阵-向量模型。

3.1 双正交小波

这里简要概述本文所用的双正交小波框架；更多细节、性质，以及相关多分辨率分析（MRA）的构造概要，可参见附录 A.1。令

$$\Psi = \{\psi_{j,k} : j \ge j_0,\, k \in \nabla_j\}, \qquad \widetilde{\Psi} = \{\widetilde{\psi}_{j,k} : j \ge j_0,\, k \in \nabla_j\}$$

表示 $L^2(\mathcal{M})$ 的一对双正交小波基，其中 $|\nabla_j| \asymp 2^{jn}$. 。指标 $\lambda = (j,k)$ 编码了尺度 (j) 和位置 (k) 的信息。我们引入小波指标集

$$\mathcal{J} := \{\lambda = (j,k) : j \ge j_0,\, k \in \nabla_j\}.$$

任意函数 $u \in L^2(\mathcal{M})$ 都可在原始基和对偶基下展开为

$$u = \mathbf{u}^\top \Psi = \sum_{\lambda \in \mathcal{J}} u_\lambda \psi_\lambda, \qquad u = \widetilde{\mathbf{u}}^\top \widetilde{\Psi} = \sum_{\lambda \in \mathcal{J}} \widetilde{u}_\lambda \widetilde{\psi}_\lambda,$$

~其中 $u_\lambda := \langle u, \psi_\lambda \rangle$ 以及 $\widetilde{u}_\lambda := \langle u, \psi_\lambda \rangle$。

双无限小波矩阵表示 $\mathbf{A} \in \mathbb{R}^{\mathcal{J} \times \mathcal{J}}$，对应于 $\mathcal{A} \in \mathsf{OPS}^r(\mathcal{M})$，定义为

$$\mathbf{A}_{\lambda,\lambda'} := \langle \mathcal{A}\psi_{\lambda'}, \psi_\lambda \rangle, \quad \forall\, \lambda, \lambda' \in \mathcal{J}.$$

$\mathcal{A}$ 在 u 上的作用可以用对偶基表示为

$$\mathcal{A}u = \sum_{\lambda \in \mathcal{J}} (\mathbf{A}\mathbf{u})_\lambda \widetilde{\psi}_\lambda.$$

利用双正交性，$\mathcal{A}u$ 与 $v = \mathbf{v}^\top \Psi = \sum_{\lambda \in \mathcal{J}} v_\lambda \psi_\lambda$ 的配对为

$$\langle \mathcal{A}u, v \rangle = \left\langle \sum_{\lambda \in \mathcal{J}} (\mathbf{A}\mathbf{u})_\lambda \widetilde{\psi}_\lambda, \sum_{\lambda \in \mathcal{J}} v_\lambda \psi_\lambda \right\rangle = \langle \mathbf{A}\mathbf{u}, \mathbf{v} \rangle = \mathbf{v}^\top \mathbf{A}\mathbf{u}.$$

因此， $\mathcal{A}$ 具有如下（形式上的）小波展开

$$\mathcal{A} = \sum_{\lambda,\lambda' \in \mathcal{J}} \mathbf{A}_{\lambda,\lambda'} \widetilde{\psi}_\lambda \otimes \widetilde{\psi}_{\lambda'} = \sum_{\lambda,\lambda' \in \mathcal{J}} \langle \mathcal{A}\psi_{\lambda'}, \psi_\lambda \rangle \widetilde{\psi}_\lambda \otimes \widetilde{\psi}_{\lambda'}. \tag{3.1}$$

算子范数 $\mathcal{A} \in \mathsf{OPS}^r$ （定义于（2.1））等价于其双无限矩阵表示$\mathbf{A}$的一个离散范数。连续范数与离散范数之间的这种等价关系，以及相关的预条件思想，将在我们的理论中发挥关键作用。具体而言，对于每个指标 $\lambda =$ （j， k），令 $|\lambda|$ ： $= j$ 并定义双无限对角缩放矩阵

$$\mathbf{D}^s := \operatorname{diag}\Big(2^{s|\lambda|} : \lambda \in \mathcal{J}\Big), \qquad s \in \mathbb{R}.$$

对于 $w = \mathbf{w}^\top \Psi \in H^t$（$M$）和以及 $v = \mathbf{v}^\top \Psi \in H^{t'}$（$M$），索博列夫范数的小波刻画（引理A.1 （iv））给出

$$\|w\|_{H^t}^2 \asymp \|\mathbf{D}^t \mathbf{w}\|_2^2, \qquad \|v\|_{H^{t'}}^2 \asymp \|\mathbf{D}^{t'} \mathbf{v}\|_2^2, \qquad \text{if } t, t' \in (-\widetilde{\gamma}, \gamma),$$

其中 γ 且 $\widetilde{\gamma}$ 是原始/对偶正则性参数。因此，范数（2.1）可表示为

$$\begin{aligned}
\|\mathcal{A}\|_{H^t \to H^{-t'}} &= \sup_{\|w\|_{H^t} \le 1, \|v\|_{H^{t'}} \le 1} |\langle \mathcal{A}w, v \rangle| \\
&\asymp \sup_{\|\mathbf{D}^t \mathbf{w}\|_2 \le 1, \|\mathbf{D}^{t'} \mathbf{v}\|_2 \le 1} |\langle \mathbf{A}\mathbf{w}, \mathbf{v} \rangle| \\
&= \sup_{\|\mathbf{D}^t \mathbf{w}\|_2 \le 1, \|\mathbf{D}^{t'} \mathbf{v}\|_2 \le 1} |\langle \mathbf{D}^{-t'} \mathbf{A} \mathbf{D}^{-t} \mathbf{D}^t \mathbf{w}, \mathbf{D}^{t'} \mathbf{v} \rangle| \\
&= \|\mathbf{D}^{-t'} \mathbf{A} \mathbf{D}^{-t}\|,
\end{aligned} \tag{3.2}$$

其中，最后一行中的$\|\cdot\|$ 表示从 ℓ^2（$\mathcal{J}$）到 ℓ^2（$\mathcal{J}$）的算子范数。注意，根据我们始终采用的假设2.1， $\mathcal{A}$是自伴的，$\mathbf{A}$ 因此是对称的。故在下文中不失一般性地假设$t \le t'$ ；参见注释5.2。

如[44]中所述，并在附录 A.1中概述，一个双正交系统 $(\Psi, \widetilde{\Psi})$ 可由四个参数刻画

$$\gamma,\ \widetilde{\gamma} \quad \text{(primal/dual regularity)}, \qquad d,\ \widetilde{d} \quad \text{(primal/dual approximation order)}.$$

参数 γ、$\widetilde{\gamma}$控制原始与对偶小波的 Sobolev正则性，而 d、d对应于消失矩的数量，并决定逼近精度。为方便起见，我们现在汇总所需的小波参数的全部条件。除对小波参数施加约束外，以下假设还规定了参数 σ 的允许范围，该参数将在我们的估计过程中使用，并与 $\mathcal{A}$ 的小波系数跨尺度分离时的非对角衰减有关，正如命题 A.2 (I) (ii) 中将变得 स्पष्ट。

假设 3.1 （小波）。 给定一对（Ψ、 Ψ）双正交小波系统，其正则性为（γ、 $\widetilde{\gamma}$），逼近阶为（d、 d），满足：

(i) $\frac{r}{2}, r_1, -r_2, t, t' \in (-\widetilde{\gamma}, \gamma)$。 *(ii)* $d >$ 最大值$\{t, t'\}$， $d > -n/2 - r/2$。 *(iii)* 最小值$\Big\{\gamma - \frac{r}{2}, \widetilde{\gamma} + \frac{r}{2}, \widetilde{d} + \frac{n}{2} + \frac{r}{2}\Big\} > \sigma >$最大值$\Big\{\frac{n}{2} +$最大值$\{t, t'\} - \frac{r}{2}, \frac{3n}{2} - t + \frac{r}{2}, \frac{t' +}{}$最大值$\{t', \frac{r_1 - r}{+t-r} n\}\}$最小值$\{t', r_1\}$。

在全文中，我们将参数 n、 r、 r_1、 r_2、 t、 t' 视为固定参数。通过选择具有足够光滑性且具有足够多消失矩的双正交小波，即取 γ、 $\widetilde{\gamma}$、 d、 d 相对于这些固定参数足够大，可以满足假设 3.1 中的条件 (i)–(iii)。

3.2 双无限矩阵框架中的算子学习

回顾统计模型（2.2）。在对偶基下，对于 $1 \le i \le N$，我们可以写成

$$f_i = \widetilde{\mathbf{f}}_i^\top \widetilde{\Psi} = \sum_{\lambda \in \mathcal{J}} \widetilde{\langle f_i, \psi_\lambda \rangle} \psi_\lambda, \quad \mathcal{A}u_i = \sum_{\lambda \in \mathcal{J}} \widetilde{(\mathbf{A}\mathbf{u}_i)_\lambda} \psi_\lambda, \quad w_i = \widetilde{\mathbf{w}}_i^\top \widetilde{\Psi} = \sum_{\lambda \in \mathcal{J}} \widetilde{\langle w_i, \psi_\lambda \rangle} \psi_\lambda. \tag{3.3}$$

因此，(2.2) 可表示为以下无限维矩阵-向量形式：

$$\widetilde{\mathbf{f}}_i = \mathbf{A}\mathbf{u}_i + \widetilde{\mathbf{w}}_i, \qquad 1 \le i \le N,$$

其中，$\mathbf{A} \in \mathrm{R}^{\mathcal{J} \times \mathcal{J}}$ 是 $\mathcal{A}$的双无限矩阵表示，而 $\widetilde{\mathbf{f}}_i, \widetilde{\mathbf{w}}_i \in \mathrm{R}^{\mathcal{J}}$ 分别是输出函数 f_i 和噪声函数 w_i的小波系数。简写为，

$$\mathbf{F} = \mathbf{U}\mathbf{A} + \mathbf{W}, \tag{3.4}$$

其中，我们利用了 $\mathcal{A}$ 是自伴的这一点，因此$\mathbf{A}$是对称的，并定义

$$\mathbf{F} := [\widetilde{\mathbf{f}}_1, \ldots, \widetilde{\mathbf{f}}_N]^\top \in \mathbb{R}^{N \times \mathcal{J}}, \quad \mathbf{U} := [\mathbf{u}_1, \ldots, \mathbf{u}_N]^\top \in \mathbb{R}^{N \times \mathcal{J}}, \quad \mathbf{W} := [\widetilde{\mathbf{w}}_1, \ldots, \widetilde{\mathbf{w}}_N]^\top \in \mathbb{R}^{N \times \mathcal{J}}.$$

因此，我们得到了以下双无限矩阵学习问题：

> 给定$\mathbf{U}$ 和$\mathbf{F}$，在范数下估计未知的双无限矩阵$\mathbf{A}$ (3.2)。

这个双无限矩阵学习问题将在我们的算子学习理论中发挥关键作用：我们将根据 $\mathcal{A}$ 的双无限矩阵估计量 $\mathbf{A}$构造其估计量，并在分析学中利用范数（2.1）和（3.2）的等价性。我们将第4 节介绍估计量的构造，并在第5节进行误差分析。估计量的构造和误差分析都利用了假设$\mathbf{A}$、数据矩阵$\mathbf{U}$以及$\widetilde{\ }$噪声矩阵$\mathbf{W}$ 所带来的双无限矩阵的以下性质，2.1 以及由小波离散化(Ψ, Ψ)所诱导的性质。这些性质是标准的，[44, 45, 122, 68], 但为求完整，我们在附录A.2中给出简要的形式化推导。

命题 3.2（关于 $\mathbf{A}$、$\mathbf{U}$、$\mathbf{W}$的性质）。在假设 *2.1*和假设 *3.1* (*i*) - (*ii*) 下，矩阵 $\mathbf{A}$、$\mathbf{U}$和 $\mathbf{W}$ 满足：

(*I*) $\boldsymbol{A}$ **的近似稀疏性：**对于 $(j, k) \in \mathcal{J}$，我们定义 $S_{j,k} := \text{凸包}\big(\text{支撑集}(\psi_{j,k})\big) \subset M$ 为 $\psi_{j的k}$支撑集的凸包。

(*i*) 对于所有 $\lambda = (j, k)$ $\lambda' = (j', k') \in \mathcal{J}$满足 距离$(S_{j,k}, S_{j',k'}) \gtrsim 2^{-\min\{j,j'\}}$，

$$|\mathbf{A}_{\lambda,\lambda'}| \lesssim 2^{-(j+j')(\widetilde{d}+n/2)} \mathrm{dist}(S_{j,k}, S_{j',k'})^{-(n+r+2\widetilde{d})}.$$

(*ii*) 对于所有 $\lambda = (j, k)$ $\lambda' = (j', k') \in \mathcal{J}$满足 距离$(S_{j,k}, S_{j',k'})$ $\gtrsim 2^{-\min\{j,j'\}}$，以及任意 $0 < \sigma <$ 最小值$\{\gamma - r/2, \widetilde{\gamma} + r/2\}$，

$$|\mathbf{A}_{\lambda,\lambda'}| \lesssim 2^{(j+j')r/2} 2^{-\sigma|j-j'|}.$$

(II) 对角预条件：存在常数 $0 < c_- < c_+ < \infty$ 使得 ：

(i) $\mathbf{A}$ 是 $\ell^2(\mathcal{J})$上的对称正定算子，并且

$$c_- \le \sigma_{\min}(\mathbf{D}^{-r/2}\mathbf{A}\mathbf{D}^{-r/2}) \le \sigma_{\max}(\mathbf{D}^{-r/2}\mathbf{A}\mathbf{D}^{-r/2}) \le c_+.$$

(ii) 输入数据的总体协方差矩阵，$\widetilde{\mathbf{C}}_u := \mathbb{E}\big[\mathbf{U}^\top\mathbf{U}/N\big]$，是定义在 $\ell^2(\mathcal{J})$ 上的对称、正定紧算子。此外，

$$c_- \le \sigma_{\min}(\mathbf{D}^{r_1}\widetilde{\mathbf{C}}_u\mathbf{D}^{r_1}) \le \sigma_{\max}(\mathbf{D}^{r_1}\widetilde{\mathbf{C}}_u\mathbf{D}^{r_1}) \le c_+.$$

(iii) 噪声的总体协方差矩阵，$\mathbf{C}_w := \mathbb{E}\big[\mathbf{W}^\top \mathbf{W}/N\big]$，是定义在 $\ell^2(\mathcal{J})$ 上的对称、正定紧算子。此外，

$$c_- \le \sigma_{\min}(\mathbf{D}^{r_2}\mathbf{C}_w\mathbf{D}^{r_2}) \le \sigma_{\max}(\mathbf{D}^{r_2}\mathbf{C}_w\mathbf{D}^{r_2}) \le c_+.$$

注 3.3. 合并命题3.2 (I) (i)–(ii) 与假设 3.1 (ii)，我们得到以下对所有 $\lambda = (j,k)$ $\lambda' = (j',k') \in \mathcal{J}$ 的一致界：

$$|\mathbf{A}_{\lambda,\lambda'}| \lesssim 2^{(j+j')r/2} \cdot 2^{-\sigma|j-j'|}\Big(1 + 2^{\min\{j,j'\}}\mathrm{dist}(S_{j,k}, S_{j',k'})\Big)^{-(n+r+2\widetilde{d})}.$$

这类估计在小波–Galerkin 文献中是经典结果，也是构建高效数值方案的基础；例如参见 [44, (7.11), (9.27)] 和 [36,(2.28),(2.29)]。 □

4 估计量的构造

本节分两个阶段构造未知算子 $\mathcal{A}$ 的估计量。第一阶段，我们利用命题 先验 估计 3.2 识别 $\mathcal{A}$. 的双无限矩阵表示中的一组显著元素。第二阶段，我们利用数据估计这些显著元素。最终的 $\mathcal{A}$ 估计量将具有如下形式

$$\widehat{\mathcal{A}} := \sum_{(\lambda,\lambda')\in \mathrm{supp}(J,t,t')} \widehat{\mathbf{A}}_{\lambda,\lambda'}\, \widetilde{\psi}_\lambda \otimes \widetilde{\psi}_{\lambda'}. \tag{4.1}$$

我们将依次介绍这两个阶段。首先，在小节 4.1中，我们介绍截断与压缩过程，该过程是定义显著元素集合 $\widehat{\mathrm{supp}(J,t,t')}$的基础，这些元素将被估计。接着，在小节 4.2 中，我们利用嵌套支撑回归策略，为 $(\lambda,\lambda') \in \mathrm{supp}(J,t,t')$ 定义逐元素估计 $\mathbf{A}_{\lambda,\lambda'}$ 。

4.1 第一阶段：截断与压缩

本小节介绍要估计的显著元素集合 $\mathrm{supp}(J,t,t')$。核心思想是，任何满足命题 *3.2* 的双无限矩阵 $\mathbf{A}$ ，都可以通过截断为有限维矩阵并随后进行矩阵压缩来很好地近似。集合$\mathrm{supp}(J,t,t')$ 对应于压缩矩阵的非零元素，且与$\mathbf{A}$无关。 我们强调，截断和压缩过程不使用给定数据，而是基于 先验 估计。

矩阵截断通过舍弃细尺度来裁剪算子，具体定义如下：

定义 4.1 (矩阵截断)。 对于 $J > j_0$，考虑层级不超过 J 的小波指标集合，定义为

$$\Lambda_J := \{(j,k) : j_0 \le j \le J, k \in \nabla_j\}.$$

给定一个 双无限矩阵 $\mathbf{A}$，我们定义其先验截断 $\mathbf{A}^{\Lambda_J} := (\mathbf{A}_{\lambda,\lambda'})_{\lambda,\ \lambda'\in\Lambda_J} \in \mathbf{R}^{\Lambda_J\times\Lambda_J}$.

在后续推导中，截断会导致对 $\mathcal{A}$ 的估计产生偏差，因为我们只会估计$\mathbf{A}$ 截至尺度层级 J. 的 J 样本量相关取值N 将通过偏差-方差权衡来确定。我们将在下文的命题5.9 中刻画矩阵截断引起的误差。

压缩利用了偏微分算子在小波基中的双无限矩阵表示$\mathbf{A}$ 的稀疏性$\mathcal{A}$ 。特别地，命题3.2 (I) 表明，当小波的支撑集 $\psi_\lambda, \psi_{\lambda'}$彼此分离或其尺度不同时，矩阵系数$\mathbf{A}_{\lambda,\lambda'} = \langle \mathcal{A}\psi_{\lambda'}, \psi_\lambda\rangle$ 会快速衰减，这促使我们给出以下定义。回忆一下，对于 $(j,k) \in \mathcal{J}$,$S_{j,k} =$ ，convhull(supp $(\psi_{j,k})$) 表示 $\psi_{j,k}$的支撑集的凸包。

定义 4.2 (矩阵压缩**)。** 给定 J, t, t', r, n, σ 以及一个双正交小波对$(\Psi, \Psi)^{\sim}$，其正则性为 γ, $\widetilde{\gamma}$，逼近阶为 d, d，定义支撑集

$$\begin{aligned}
&\operatorname{supp}(J,t,t') = \operatorname{supp}\big(J,t,t',r,n,\sigma,\widetilde{d}\big)\\
&:= \Big\{(\lambda,\lambda') \in \Lambda_J \times \Lambda_J : \operatorname{dist}(S_{j,k}, S_{j',k'}) \le \tau_{jj'}, j \le \frac{t+t'-r}{\sigma-\frac{n}{2}+t'-\frac{r}{2}} J + \frac{\sigma-\frac{n}{2}-(t-\frac{r}{2})}{\sigma-\frac{n}{2}+t'-\frac{r}{2}} j',\\
&\quad j' \le \frac{t+t'-r}{\sigma-\frac{n}{2}+t-\frac{r}{2}} J + \frac{\sigma-\frac{n}{2}-(t'-\frac{r}{2})}{\sigma-\frac{n}{2}+t-\frac{r}{2}} j\Big\}, \qquad (4.2)
\end{aligned}$$

其中，阈值 $\tau_{jj'}$ 选为

$$\tau_{jj'} := a \max\Big\{2^{-\min\{j,j'\}}, 2^{(J(t+t'-r)-jt'-j't-(j+j')\widetilde{d})/(2\widetilde{d}+r)}\Big\} \qquad (4.3)$$

其中 $a > 1$为某个足够大的常数。

与支撑集(J, t, t') 相关联的指示矩阵为

$$\mathbf{M}_{(J,t,t')} := \mathbf{1}_{\operatorname{supp}(J,t,t')} \in \{0,1\}^{\Lambda_J \times \Lambda_J}, \qquad (4.4)$$

其中，当且仅当$(\mathbf{M}_{(J,t,t')})_{\lambda,\lambda'} = 1$ 且$(\lambda, \lambda') \in$ 属于支撑集(J, t, t')时，该项成立。

先验压缩矩阵$\mathbf{A}^{\varepsilon}_{\Lambda_J} \in \mathbb{R}^{\Lambda_J \times \Lambda_J}$ 定义为

$$\mathbf{A}^{\varepsilon}_{\Lambda_J} := \mathbf{M}_{(J,t,t')} \odot \mathbf{A}_{\Lambda_J},$$

其中 $\odot$表示 *Hadamard*（逐项）积。

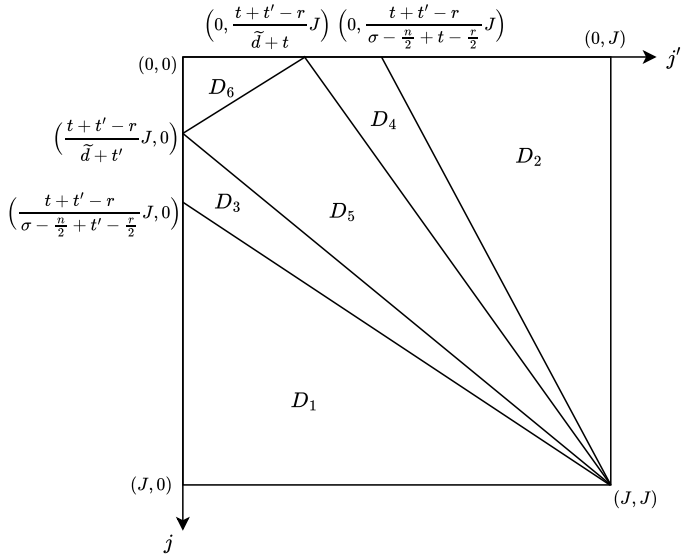

图 1：定义 4.2 中（J, t, t'）-压缩在（j, j'）平面上的示意图。指标集被划分为区域 D_1–D_6：丢弃 $D_1 \cup D_2$ 中的块；在 D_3 和 D_4 中，仅当 $\operatorname{dist}(S_{j,k}, S_{j',k'}) \le \tau_{jj'}$ with $\tau_{jj'} \asymp 2^{-j'}$ 和 $\tau_{jj'} \asymp 2^{-j}$ 时，才保留元素；在 D_5 中，仅当 $\operatorname{dist}(S_{j,k}, S_{j',k'}) \le \tau_{jj'}$ with $\tau_{jj'} \asymp 2^{(J(t+t'-r)-jt'-j't-(j+j')d)/(2d+r)}$ 时，才保留元素；D_6 不进行压缩（保留所有元素）。

在全文中，我们记 $\operatorname{supp}(J, t, t')$ 以强调其对 (J, t, t')的依赖，并省略$\widetilde{}$固定不变的参数 (r, n, σ, d)。在后续推导中，矩阵压缩会引入第二层偏差，我们将在命题 5.10中对其进行界定。

注 4.3（讨论：矩阵压缩及与 [45]的比较）。在科学计算和数值分析学中的小波方法——尤其是在边界积分方程求解器中——矩阵压缩技术对于实现（近乎）最优计算复杂度起着核心作用，也是小波–伽辽金方法 [44, 34, 37, 45, 70, 68]的关键组成部分。在本文中，我们的动机来自算子学习而非经典偏微分方程求解器，因此引入了一种专为统计估计场景设计的压缩方案。下面我们将讨论定义 4.2 中矩阵压缩的动机和关键性质，并将其与 [45]中提出的压缩方案进行比较；参见 [45, 第 7]节。

1. 矩阵压缩的关键性质。命题 5.10 表明，加权压缩误差的阶为 $2^{-J(t+t'-r)}$（相差一个因子 J），与限制为仅保留层级不超过 J 的指标所引入的截断误差 $2^{-J(t+t'-r)}$ 相匹配；参见定义 4.1 和命题 5.9。这是小波压缩方案的一项标准特征；参见 [45, 定理 9.1]和 [68, 命题 6]。此外，压缩矩阵 $\mathbf{A}^{\varepsilon}_{\Lambda_J} \in \mathbb{R}^{\Lambda_J \times \Lambda_J}$ 的非零元素总数的阶为 $\mathcal{O}(2^{Jn})$，即与小波指标的数量 $|\Lambda_J| \asymp 2^{Jn}$呈线性关系；参见命题 C.2 中的界 $\mathrm{nnz}\big(\mathbf{M}_{(J,t,t')}\big) \lesssim 2^{Jn}$ 。这一“线性复杂度”性质保证了最优稀疏近似和最优计算复杂度；参见 [45, 定理 11.1]。文献 [70] 利用这一性质获得了具有线性伸缩性的求积策略，而类似的思想也出现在文献 [68] 的协方差/精度矩阵逼近中。

2. **选择** $\tau_{jj'}$。基于 $\tau_{jj'}$ 的阈值化利用命题 3.2 (I) (i)，该命题表明，随着 ψ_λ和 $\psi_{\lambda'}$ 的支撑集彼此远离，矩阵系数 $\mathbf{A}_{\lambda,\lambda'} = \langle \mathcal{A}\psi_{\lambda'}, \psi_\lambda \rangle$ 会迅速衰减。因此，我们只保留满足 $\mathrm{dist}(S_{j,k}, S_{j',k'}) \le \tau_{jj'}$ 的元素。

与 [45, 68] （例如 [45, 第 7]节和 [68, 定义 1]）中的压缩策略相比，一个关键的概念差异在于，在我们的学习场景中，$\tau_{jj'}$ 取决于误差度量 $\|\cdot\|_{H^t \to H^{-t'}}$ 中出现的正则性 t, t'；而在 [45, 68] 中，阈值则取决于用户指定的超参数 $d' \in (d, d+r)$。对于给定的参数 t 和 t'，我们的选择通常会得到更小的 $\tau_{jj'}$ 值，从而在保持相同偏差水平 $2^{-J(t+t'-r)}$ 的同时，更积极地删除元素。更精确地说，由于 $d' > d > \max\{t, t'\}$，

$$2^{(J(t+t'-r)-jt'-j't-(j+j')\widetilde{d})/(2\widetilde{d}+r)} \le 2^{(2J(d'-r/2)-(j+j')(d'+\widetilde{d}))/(2\widetilde{d}+r)},$$

并且当 $j,\ j' < J$ 该不等式为严格不等式；参见 $\tau_{jj'}$ 中的 [68, (3.20)]。

从统计学角度来看，由于我们的估计量（稍后定义）以稀疏压缩矩阵 $\mathbf{A}^{\varepsilon}_{\Lambda_J}$ 为目标，且其支撑集为 $\mathrm{supp}(J,\ t,\ t')$，因此其方差分析在很大程度上取决于压缩方案所诱导的稀疏模式。我们的面向学习的压缩具有优势：在每个块内，它保留的系数更少（等价地说，支撑集更小），从而得到更小的方差界。我们强调，这一改进对于命题 5.12中的方差分析至关重要，因为在矩阵算子范数下，方差会跨尺度累积，而这无法仅由非零元素的全局数量来刻画。特别是，我们的压缩和 [45] 中的矩阵压缩方案都保证全局非零元素数具有最优阶 $\mathcal{O}(2^{Jn})$，即与 $|\Lambda_J|$呈线性关系。然而，要获得紧确的方差界，需要在每个块内实现近乎最优的稀疏性；这正是我们的面向学习的压缩所具有优势之处。

3. **斜率条件**。除逐项阈值化外，我们还只保留满足式（(j, j') 中两个斜率条件的 4.2）的块，这两个条件利用了命题 3.2（I）（ii）。根据~假设 3.1（iii），我们有 $\sigma - n/2 > \max\{t, t'\} - r/2$ 和 $d > \sigma - n/2 - r/2$，这确保斜率约束在（j, j'）平面上定义了一个非平凡的可容许区域；参见图 1。正如命题 5.10 的证明将会说明的，这些斜率条件经过精确设计，使得 $D_1 \cup D_2$ 中被舍弃的块对偏差的贡献至多为 $2^{-J(t+t'-r)}$。从这个意义上说，斜率条件刻意设得较为激进：它们舍弃

尽可能多的块，同时仍确保所得偏差的阶至多为 $2^{-J(t+t'-r)}$。与此同时，舍弃更多的块会减少保留的系数数量，从而有助于控制估计量的方差。

总之，阈值化水平 $\tau_{jj'}$ 的选择和斜率条件的设计遵循同一原则：在每个 (j,j') 块中尽可能少地保留元素，同时确保在 (3.2) 中加权矩阵算子范数 $\|\mathbf{D}^{-t'}(\cdot)\mathbf{D}^{-t}\|$ 下，由此产生的偏差至多为 $2^{-J(t+t'-r)}$ 阶，与定义 4.1 中引入的截断偏差相匹配。这就是 $\tau_{jj'}$ 和斜率条件为何取决于学习误差度量中预先指定的 (t,t')。相比之下， [45] 中的矩阵压缩方案旨在同时适用于一系列权重参数（达到 MRA 的逼近阶）；参见定理 9.1] 和命题 6]的 [45, [68, 。

4. **supp(**J、 t、 t'**)** 的**不对称性**。 当 $t \neq t'$时，我们的压缩本质上是不对称的。在这种情况下，$\tau_{jj'} \neq \tau_{j'j}$，斜率条件也同样是不对称的，这一点从图 1 中的划分（以 $t < t'$为例）可以看出。这种不对称性源于我们的压缩是针对学习场景中的非对称误差度量 $\|\cdot\|_{H^t \to H^{-t'}}$ 量身设计的；等价地说，它针对的是加权矩阵算子范数 $\|\mathbf{D}^{-t'}(\cdot)\mathbf{D}^{-t}\|$。相比之下， [45, 68] 中的阈值参数和矩阵压缩方案均具有对称性。

5. 关于 [**122**, **45**] 中的“第二次压缩”。我们的压缩仅使用命题 3.2 (I)。此外， [122, 45]对支撑重叠但尺度相距较远的小波所对应的元素推导出更精确的估计，然后再进行第二次阈值化。具体而言，对于 M 上的函数 f，令 sing supp$(f) := \{x \in \mathrm{M} : f$ 在 x 处不光滑$\}$，并令 $S'_{j,k} :=$ sing supp$(\psi_{j,k}) \subset \mathrm{M}$ 为$\psi_{j,k}$的奇异支撑。对于满足 $j' > j$ 且 dist$(S'_{j,k}, S_{j',k'}) \gtrsim 2^{-j'}$, [45, 的 (j,k), (j',k')， 定理 6.3]证明

$$|\mathbf{A}_{\lambda,\lambda'}| \lesssim 2^{jn/2} 2^{-j'(\widetilde{d}+n/2)} \operatorname{dist}(S'_{j,k}, S_{j',k'})^{-(r+\widetilde{d})}.$$

这促使我们进行第二次阈值化，仅保留满足以下条件的元素

$$\operatorname{dist}(S'_{j,k}, S_{j',k'}) \leq \tau'_{jj'},$$

其中 $\tau'_{jj'}$ 是一个不同于 $\tau_{jj'}$ 的参数；参见定义 1] 和第 7] 节。此处我们不采用这一改进：在我们的设定中，它似乎并未显著改善方差界；而舍弃区域 $D_1 \cup D_2$ 后，分析会更简单，且已能得到较强的方差估计（至少对于固定的 t, t'）。□

4.2 第二阶段：估计

在本小节中，我们为稀疏压缩矩阵$\mathbf{A} \in \mathbb{R}^{\Lambda_J \times \Lambda_J}$ 构造一个估计量$\mathbf{A}^{\varepsilon}_{\Lambda_J} \in \mathbb{R}^{\Lambda_J \times \Lambda_J}$.。稀疏支撑集 supp$(J,t,t')$ 将决定我们基于回归的估计量的结构：支撑集指定哪些矩阵元素参与回归，哪些元素设为零并视为偏差。

我们首先注意到，(3.4) 可以按列写出。对于任意指标 $\lambda' = (j',k') \in \mathcal{J}$，矩阵方程$\lambda'$ (3.4) 的第

$$\mathbf{F}_{\cdot,\lambda'} = \mathbf{U}\mathbf{A}_{\cdot,\lambda'} + \mathbf{W}_{\cdot,\lambda'}, \tag{4.5}$$

其中，$\mathbf{F}_{\cdot,\lambda'} \in \mathbb{R}^N$、$\mathbf{A}_{\cdot,\lambda'} \in \mathbb{R}^{\mathcal{J}}$和$\mathbf{W}_{\cdot,\lambda'} \in \mathbb{R}^N$分别表示F、A和W的第 λ'列。对于每个 λ'，这就得到一个无限维线性回归问题，其响应向量为$\mathbf{F}_{\cdot,\lambda'}$，设计矩阵为U，系数向量为$\mathbf{A}_{\cdot,\lambda'}$，噪声向量为$\mathbf{W}_{\cdot,\lambda'}$。基于(4.5)，我们的策略是通过回归逐列估计压缩矩阵$\mathbf{A}^{\varepsilon}_{\Lambda_J}$ ，同时充分利用$\mathbf{A}^{\varepsilon}_{\Lambda_J}$的稀疏支撑结构。

注意，$\mathbf{A}^{\varepsilon}_{\Lambda_J} = \mathbf{M}_{(J,t,t')} \odot \mathbf{A}^{\Lambda_J}$ 的第 λ'列的支撑集是有限集合

$$\{\lambda \in \mathcal{J} : (\lambda, \lambda') \in \mathrm{supp}(J, t, t')\}.$$

然而，直接在该指标集上进行回归可能会引入不可忽略的遗漏变量偏差，因为被忽略的分量——尽管在逼近中被省略——仍会通过数据与保留的分量相互作用；详细解释见备注4.7 。这种偏差取决于数据，且可能远大于第一阶段产生的确定性截断误差和压缩误差。为了解决这一问题，我们的回归估计量采用了一个关键思路：对于每一列 λ'，我们选择一个稍大的指标集，使其包含 $(\mathbf{A}^{\varepsilon}_{\Lambda_J})_{\cdot,\lambda'}$ 的支撑集，并将其作为回归支撑集，在这个扩大的集合上进行回归，然后再将所得估计量限制回 $\mathrm{supp}(J,t,t')$。这种嵌套支撑回归策略使我们能够有效控制遗漏变量偏差，使其至多与截断误差和压缩误差处于同一阶。

更准确地说，令 $\widetilde{J}, \widetilde{t}$、和 $\widetilde{t}'$ 为三个调节参数。附录B.1 中的引理B.2保证了以下包含关系

$$\mathrm{supp}(J, t, t') \subset \mathrm{supp}(\widetilde{J}, \widetilde{t}, \widetilde{t}') \tag{4.6}$$

在 $\widetilde{J} \geq J, \widetilde{t} \geq t$ 和 $\widetilde{t}' \geq t'$. 成立的条件下。超参数 $\widetilde{J}, \widetilde{t}, \widetilde{t}'$ 用于定义扩大的回归支撑集 $\mathrm{supp}(\widetilde{J}, \widetilde{t}, \widetilde{t}') \subset \Lambda_{\widetilde{J}} \times \Lambda_{\widetilde{J}}$，其取值将使遗漏变量偏差与截断误差和压缩误差处于同一阶。

现在，对于每个 $\lambda' \in \Lambda_J$，定义

$$\Omega_{\lambda'} := \left\{\lambda \in \mathcal{J} : (\lambda, \lambda') \in \mathrm{supp}(\widetilde{J}, \widetilde{t}, \widetilde{t}')\right\} \subset \Lambda_{\widetilde{J}}, \qquad \Omega^c_{\lambda'} := \mathcal{J} \setminus \Omega_{\lambda'}$$

并相应地分解设计矩阵和系数向量：

$$\mathbf{U} = [\mathbf{U}_{\cdot,\Omega_{\lambda'}}, \mathbf{U}_{\cdot,\Omega^c_{\lambda'}}], \quad \mathbf{A}_{\cdot,\lambda'} = \begin{bmatrix} \mathbf{A}_{\Omega_{\lambda'},\lambda'} \\ \mathbf{A}_{\Omega^c_{\lambda'},\lambda'} \end{bmatrix}. \tag{4.7}$$

这里采用如下约定：对于任意子集 $\Lambda \subset \mathcal{J}$，$\mathbf{U}_{\cdot,\Lambda} \in R^{N \times \Lambda}$ 表示由 $\mathbf{U}$ 中指标属于 Λ 的列构成的子矩阵；对于任意向量 $\mathbf{v} \in \mathbf{R}^{\mathcal{J}}$，$\mathbf{v}_\Lambda \in \mathbf{R}^{\Lambda}$ 表示其分量指标属于 Λ 的子向量。(4.6) 中 $\mathrm{supp}(J,t,t') \subset \mathrm{supp}(\widetilde{J}, \widetilde{t}, \widetilde{t}')$ 的包含关系意味着，对于每一列 λ'，向量$\mathbf{A}_{\Omega_{\lambda'},\lambda'}$ （严格）包含了我们旨在估计的 $(\mathbf{A}^{\varepsilon}_{\Lambda_J})_{\cdot,\lambda'}$ 的所有非零元素。

将式（4.7）代入式（4.5），得到

$$\mathbf{F}_{\cdot,\lambda'} = \mathbf{U}\mathbf{A}_{\cdot,\lambda'} + \mathbf{W}_{\cdot,\lambda'} = \mathbf{U}_{\cdot,\Omega_{\lambda'}}\mathbf{A}_{\Omega_{\lambda'},\lambda'} + \mathbf{U}_{\cdot,\Omega^c_{\lambda'}}\mathbf{A}_{\Omega^c_{\lambda'},\lambda'} + \mathbf{W}_{\cdot,\lambda'}. \tag{4.8}$$

假设$\mathbf{U}^{\top}_{\cdot,\Omega_{\lambda'}}\mathbf{U}_{\cdot,\Omega_{\lambda'}}$ 可逆，我们使用普通最小二乘估计量来估计$\mathbf{A}_{\Omega_{\lambda'},\lambda'}$

$$(\mathbf{U}^{\top}_{\cdot,\Omega_{\lambda'}}\mathbf{U}_{\cdot,\Omega_{\lambda'}})^{-1}\mathbf{U}^{\top}_{\cdot,\Omega_{\lambda'}}\mathbf{F}_{\cdot,\lambda'}, \tag{4.9}$$

忽略 $\mathbf{A}^{\Omega^c_{\lambda'},\lambda'}$ 的贡献。然后，我们拼接逐列回归估计，将所得矩阵限制在 $\mathrm{supp}(J,t,t')$，并利用 $\mathbf{A}$ 的对称性，构造压缩矩阵 $\widehat{\ }$ $\widehat{\ }$ $\widehat{\ }$的完整估计量 $\mathbf{A}$ 。利用 $\mathbf{A}^{\varepsilon}_{\Lambda_J \cdot}$ ，我们通过（**4.1**）定义未知算子 $\mathcal{A}$ 的最终估计量$\widehat{\ }$ $\widehat{\ }$ $\mathcal{A}$ 。完整的 A 和 $\mathcal{A}$ 的定义如下。

$\widehat{\mathcal{A}}$算法：估计量的构造

输入：

- 最大小波尺度 J和 $\widetilde{J}$；正则性参数 $t, t\prime\widetilde{t}, \widetilde{t}\prime$。
- 压缩支撑集 $\mathrm{supp}(J,t,t\prime)$，回归支撑集 $\mathrm{supp}(\widetilde{J},\widetilde{t},\widetilde{t}\prime)$。
- 数据矩阵 $\mathbf{U}_{\cdot,\Lambda_{\widetilde{J}}} \in \mathbb{R}^{N\times\Lambda_{\widetilde{J}}}$和 $\mathbf{F}_{\cdot,\Lambda_{\widetilde{J}}} \in \mathbb{R}^{N\times\Lambda_{\widetilde{J}}}$。

输出： 估计量 $\widehat{\mathcal{A}}: H^t(M) \to H^{-t'}(M)$。

步骤 1：逐列回归。
令 $\Omega_{\lambda'} := \left\{\lambda \in \mathcal{J}: (\lambda,\ \lambda') \in \mathrm{supp}(\widetilde{J},\ \widetilde{t},\ \widetilde{t}')\right\}$。通过拼接逐列回归估计，计算初步估计量 $\widehat{\mathbf{A}}^{(1)} \in \mathbb{R}^{\Lambda_{\widetilde{J}}\times\Lambda_J}$：

$$(\widehat{\mathbf{A}}^{(1)})_{\cdot,\lambda'} := \begin{bmatrix} (\mathbf{U}^\top_{\cdot,\Omega_{\lambda'}}\mathbf{U}_{\cdot,\Omega_{\lambda'}})^{-1}\mathbf{U}^\top_{\cdot,\Omega_{\lambda'}}\mathbf{F}_{\cdot,\lambda'} \\ \mathbf{0} \end{bmatrix} \in \mathbb{R}^{\Lambda_{\widetilde{J}}}, \qquad \lambda' \in \Lambda_J, \tag{4.10}$$

其中 $(\mathbf{U}^\top_{\cdot,\ \Omega_{\lambda'}}\mathbf{U}_{\cdot,\ \Omega_{\lambda'}})^{-1}\mathbf{U}^\top_{\cdot,\ \Omega_{\lambda'}}\mathbf{F}_{\cdot,\ \lambda'} \in \mathbb{R}^{\Omega_{\lambda'}}$和 $\mathbf{0} \in \mathbb{R}^{\Lambda_{\widetilde{J}}\setminus\Omega_{\lambda'}}$。

步骤 2：限制到目标稀疏模式并利用对称性。
定义估计量 $\widehat{\mathcal{A}}$逐项定义压缩矩阵 $\mathbf{A}^{\varepsilon}_{\Lambda_J} \in \mathbb{R}^{\Lambda_J\times\Lambda_J}$的估计量 $\mathbf{A} \in \mathbb{R}^{\Lambda_J\times\Lambda_J}$：

$$\widehat{\mathbf{A}}_{\lambda,\lambda'} := \begin{cases} (\widehat{\mathbf{A}}^{(1)})_{\lambda,\lambda'} & \text{if } (\lambda,\lambda') \in \mathrm{supp}(J,t,t') \text{ and } j \le j', \\ (\widehat{\mathbf{A}}^{(1)})_{\lambda',\lambda} & \text{if } (\lambda,\lambda') \in \mathrm{supp}(J,t,t') \text{ and } j > j', \\ 0 & \text{if } (\lambda,\lambda') \notin \mathrm{supp}(J,t,t'). \end{cases} \tag{4.11}$$

步骤 3：算子构造。
$\widehat{\mathcal{A}}$将估计量 $\mathcal{A}$定义为

$$\widehat{\mathcal{A}} := \sum_{(\lambda,\lambda')\in\mathrm{supp}(J,t,t')} \widehat{\mathbf{A}}_{\lambda,\lambda'}\,\widetilde{\psi}_\lambda \otimes \widetilde{\psi}_{\lambda'}. \tag{4.12}$$

注 4.4（对 $\widehat{\mathcal{A}}$ 构造的解释）。由于指标集 $\mathrm{supp}(J,\ t,\ t')$ 并非对称，压缩矩阵 $\mathbf{A}^{\varepsilon}_{\Lambda_J}$（因此我们的估计量）也不是对称的。在**步骤 1**中，我们通过按列回归构造初步估计量 $\widehat{\mathbf{A}}^{(1)}$，且未利用目标矩阵 $\mathbf{A}$的对称性。现在，我们解释**步骤 2**中对称化过程的动机：该过程利用 $\widehat{\mathbf{A}}^{(1)}$中方差较小的一侧来估计其对称对应部分，从而降低总体方差。

正如命题 5.12（见（C.19））中的方差分析所示，$(j,\ j')$ 块中 $\widehat{\mathbf{A}}^{(1)}$的一个元素的标准差阶为 $2^{jr_1-j'r_2}/\sqrt{N}$，相差对数因子。在假设 2.1下，有 $r_1 > n/2 + \overline{\max}\{0,\ r\}$ 且 $r_2 > n/2$，因此，只要 $j < j'$，

$$2^{jr_1-j'r_2} < 2^{j'r_1-jr_2}.$$

因此，对于 $j < j'$（即，对于 位于对角线 $j = j'$ 上方的(j,j')平面中的块（见图 1），$\widehat{\mathbf{A}}^{(1)}$在 (j,j')块中的元素，其方差小于转置后的 (j',j)块中对应的对称元素（该块位于对角线下方）。

这种方差不平衡可以从模型中理解

$$\mathbf{F} = \mathbf{U}\mathbf{A} + \mathbf{W}.$$

粗略地说，由于正则性假设 r_1，$r_2 > n/2$，输入矩阵 $\mathbf{U}$ 和噪声矩阵 $\mathbf{W}$ 的小波系数会随着尺度索引增大而衰减。特别是，在更细的尺度（较大的 j'）下估计 $\mathbf{A}$ 的第 λ' 列时，有效噪声水平更低。此外，较粗尺度下$\mathbf{U}$ 的元素通常具有更大的幅值，因此相应的回归具有更强的信号。综合这些效应，估计$\mathbf{A}$的对角线上方部分时的方差小于估计对角线下方元素时的方差。

受这一观察的启发，在**步骤 2** 中，我们利用目标矩阵 $\mathbf{A}$的对称性，将更准确的估计（方差较小的一侧，即对角线上方）转移到另一侧（方差较大的一侧，即对角线下方）。由于我们的目标是估计支撑在 $\underset{\Lambda_J}{\varepsilon}\mathrm{supp}(J,\ t,\ t')$上的压缩矩阵 $\mathbf{A}$，当 $t \le t'$时，下面的引理 4.5 保证，对角线下方支撑集的反射包含于对角线上方的支撑集。这一包含关系使我们能够通过对 $\widehat{\ }\mathbf{A}^{(1)}$ 进行对称化，利用方差较小的元素来定义（4.11）中的最终估计量 $\widehat{\ }\mathbf{A}$，从而相较于 $\widehat{\ }\mathbf{A}^{(1)}$降低整体方差。□

引理 4.5。 设 $\mathbf{M}_{(J,t,t')}$ 为定义 *4.2* 中引入的指示矩阵。若 $t \le t'$，则对于任意一对指标 $(\lambda,\lambda') = ((j,k),(j',k'))$ 满足 $j > j'$，有

$$(\mathbf{M}_{(J,t,t')})_{\lambda,\lambda'} = 1 \quad \Longrightarrow \quad (\mathbf{M}_{(J,t,t')})_{\lambda',\lambda} = 1.$$

注 4.6（有效数据与计算成本）。注意，在构造$\widehat{\ }$估计量 $\mathbf{A}$ 的算法中，我们只使用有限维数据矩阵

$$\mathbf{U}_{\cdot,\Lambda_{\widetilde{J}}} \in \mathbb{R}^{N\times\Lambda_{\widetilde{J}}}, \qquad \mathbf{F}_{\cdot,\Lambda_{\widetilde{J}}} \in \mathbb{R}^{N\times\Lambda_{\widetilde{J}}},$$

而非完整的系数矩阵 $\mathbf{U}$,$\mathbf{F}$。在我们的算法中，假设可以通过预言机访问 $\widetilde{\ }$输入—输出数据的小波系数，即这些量 $(\langle u_i,\psi_\lambda\rangle,\langle f_i,\psi_\lambda\rangle)$。这不会降低学习问题的一般性，因为（双）正交小波变换是可逆线性映射。关于小波–伽辽金方法的算法与实现细节，请参阅 [70, 69]。在接下来对学习算法计算成本的讨论中，我们忽略构造小波和计算这些系数的成本，并假设 $\mathbf{U}$,$\mathbf{F}$ 已知。

现在讨论该算法的计算成本。对于固定的列指标 λ'，我们计算

$$(\mathbf{U}_{\cdot,\Omega_{\lambda'}}^{\top}\mathbf{U}_{\cdot,\Omega_{\lambda'}})^{-1}\,\mathbf{U}_{\cdot,\Omega_{\lambda'}}^{\top}\mathbf{F}_{\cdot,\lambda'},$$

其中

$$\Omega_{\lambda'} := \big\{\lambda\in\mathcal{J}:\ (\lambda,\lambda')\in\mathrm{supp}(\widetilde{J},\widetilde{t},\widetilde{t}')\big\}.$$

构造格拉姆矩阵 $\mathbf{U}_{\cdot,\,\Omega_{\lambda'}}^{\top}\mathbf{U}_{\cdot,\,\Omega_{\lambda'}}$以及交叉项 $\mathbf{U}_{\cdot,\,\Omega_{\lambda'}}^{\top}\mathbf{F}_{\cdot,\,\lambda'}$的计算成本为 $\mathcal{O}\big(N|\Omega_{\lambda'}|^2\big)$，求解由此得到的线性系统（例如通过乔列斯基分解）需要 $\mathcal{O}\big(|\Omega_{\lambda'}|^3\big)$的计算成本。因此，每列的计算成本为

$$\mathcal{O}\big(N|\Omega_{\lambda'}|^2 + |\Omega_{\lambda'}|^3\big).$$

由于我们仅对指标 $\lambda'\in\Lambda_J$ 执行按列回归来构造 $\widehat{\ }\mathbf{A}^{(1)}$，对所有此类列 $\lambda'\in\Lambda_J$ 求和可得总计算成本的阶为

$$\mathcal{O}\Bigg(N\Bigg(\sum_{\lambda'\in\Lambda_J}|\Omega_{\lambda'}|^2\Bigg) + \sum_{\lambda'\in\Lambda_J}|\Omega_{\lambda'}|^3\Bigg), \tag{4.13}$$

这在很大程度上取决于回归支撑集的稀疏模式 $\mathrm{supp}(\widetilde{J},\ \widetilde{t},\ \widetilde{t}')$。在注 5.5 中，我们在特定选择 $\widetilde{J},\ \widetilde{t},\ \widetilde{t}'$ 的情况下，给出了一个更明确的界。□

注 4.7（结构化无限维线性回归）。在本注中，我们阐释按列模型（4.8），并解释估计量（4.9）误差分析背后的结构。为简化讨论，我们将（4.5）重写为一个抽象的无限维线性回归问题，

$$y = \mathbf{X}\theta + \xi,$$

其中，$y \in \mathrm{R}^N$ 是响应向量，$\mathbf{X} \in \mathrm{R}^{N\times\mathcal{J}}$ 是由无限集索引的设计矩阵 $\mathcal{J}$，$\theta \in \mathrm{R}^{\mathcal{J}}$ 是未知系数向量，而 $\xi \in \mathrm{R}^N$ 是噪声向量。目标是根据（$\mathbf{X}$ y）恢复 θ。

假设无限维向量 θ 通过先验信息呈现出某种结构，这表明其主导坐标包含在有限指标集 $\Omega \subset \mathcal{J}$中；将其补集记为 $\Omega^c := \mathcal{J}\backslash\Omega$。根据这一划分对设计矩阵和参数向量进行分解，得到

$$y = \mathbf{X}\theta + \xi = \mathbf{X}_{\cdot,\Omega}\theta_\Omega + \mathbf{X}_{\cdot,\Omega^c}\theta_{\Omega^c} + \xi.$$

若$\mathbf{X}^\top_{\cdot,\Omega}\mathbf{X}_{\cdot,\Omega}$ 可逆，则定义如下估计量：对 Ω 进行回归，并将其余坐标设为零，

$$\widehat{\theta} := \begin{pmatrix}\widehat{\theta}_\Omega \\ \mathbf{0}\end{pmatrix}, \qquad \widehat{\theta}_\Omega := (\mathbf{X}^\top_{\cdot,\Omega}\mathbf{X}_{\cdot,\Omega})^{-1}\mathbf{X}^\top_{\cdot,\Omega}y.$$

直接计算可得

$$(\mathbf{X}^\top_{\cdot,\Omega}\mathbf{X}_{\cdot,\Omega})^{-1}\mathbf{X}^\top_{\cdot,\Omega}y - \theta_\Omega = (\mathbf{X}^\top_{\cdot,\Omega}\mathbf{X}_{\cdot,\Omega})^{-1}\mathbf{X}^\top_{\cdot,\Omega}\mathbf{X}_{\cdot,\Omega^c}\theta_{\Omega^c} + (\mathbf{X}^\top_{\cdot,\Omega}\mathbf{X}_{\cdot,\Omega})^{-1}\mathbf{X}^\top_{\cdot,\Omega}\xi,$$

因此

$$\widehat{\theta} - \theta = \begin{pmatrix}(\mathbf{X}^\top_{\cdot,\Omega}\mathbf{X}_{\cdot,\Omega})^{-1}\mathbf{X}^\top_{\cdot,\Omega}\mathbf{X}_{\cdot,\Omega^c}\theta_{\Omega^c} + (\mathbf{X}^\top_{\cdot,\Omega}\mathbf{X}_{\cdot,\Omega})^{-1}\mathbf{X}^\top_{\cdot,\Omega}\xi \\ -\theta_{\Omega^c}\end{pmatrix}. \tag{4.14}$$

在指标集 Ω^c上，误差为偏差项 $-\theta_{\Omega^c}$，它是确定性的，且独立于设计$\mathbf{X}$ 和噪声 ε。由于选取 Ω 以包含 θ的显著元素，因此预期该项较小。

在指标集 Ω上，误差可分解为两项。第一项

$$(\mathbf{X}^\top_{\cdot,\Omega}\mathbf{X}_{\cdot,\Omega})^{-1}\mathbf{X}^\top_{\cdot,\Omega}\mathbf{X}_{\cdot,\Omega^c}\theta_{\Omega^c}$$

是遗漏变量偏差（OVB）项，它取决于设计的协方差结构。在经典的有限维回归中，OVB 产生于相关协变量被遗漏，其影响部分归因于纳入协变量；例如，参见 [9, 第 18] 章以及 [146, 第 3-3]章。这里，OVB 则从略有不同的角度出现：当 θ 是无限维时，任何有限维回归都必然会忽略一段（大概可以忽略不计的）尾部 θ_{Ω^c}，而当 Ω 的列相互相关时，这段尾部可能会使$\mathbf{X}$ 上的估计产生偏差。

对于具有独立同分布行的随机设计，可以将这一点精确表述。设

$$\mathbf{C} := \mathbb{E}[\mathbf{X}^\top\mathbf{X}]/N$$

表示设计矩阵每一行的总体协方差。在标准正则条件下，通常有 $\mathbf{X}^\top_{\cdot,\Omega}\mathbf{X}_{\cdot,\Omega}/N \to \mathbf{C}_{\Omega,\Omega}$ 以及 $\mathbf{X}^\top_{\cdot,\Omega}\mathbf{X}_{\cdot,\Omega^c}/N \to \mathbf{C}_{\Omega,\Omega^c}$ 当 $N \to \infty$ 时成立，因此 OVB 项收敛于

$$\mathbf{C}^{-1}_{\Omega,\Omega}\mathbf{C}_{\Omega,\Omega^c}\,\theta_{\Omega^c}.$$

一般而言，如果 θ_{Ω^c} 和/或$\mathbf{C}$ （例如，划分上的近似正交性）没有额外结构，这个极限就不为零，因此会产生真实的偏差。我们的嵌套支撑策略旨在平衡两种偏差：确定性偏差 $-\theta_{\Omega^c}$ 以及由 OVB通过设计相关性引起的 θ_{Ω^c} ；参见第5.11 节中的命题 5.2。

指标集上的第二项 Ω在（4.14）中，

$$(\mathbf{X}_{\cdot,\Omega}^{\top}\mathbf{X}_{\cdot,\Omega})^{-1}\mathbf{X}_{\cdot,\Omega}^{\top}\xi,$$

是通常的方差 项。在标准设定下，若 $\xi \sim \mathcal{N}(\mathbf{0},\ \sigma_\xi^2 I_N)$ 与$\mathbf{X}$独立，则 $\mathbf{E}\big[(\mathbf{X}_{\cdot,\ \Omega}^{\top}\mathbf{X}_{\cdot,\ \Omega})^{-1}\mathbf{X}_{\cdot,\ \Omega}^{\top}\xi\big] = 0$，并且在给定$\mathbf{X}$的条件下，

$$(\mathbf{X}_{\cdot,\Omega}^{\top}\mathbf{X}_{\cdot,\Omega})^{-1}\mathbf{X}_{\cdot,\Omega}^{\top}\xi \sim \mathcal{N}\big(0, \sigma_\xi^2(\mathbf{X}_{\cdot,\Omega}^{\top}\mathbf{X}_{\cdot,\Omega})^{-1}\big).$$

特别地，在满足$\mathbf{X}_{\cdot,\ \Omega}^{\top}\mathbf{X}_{\cdot,\ \Omega}/N \to \mathbf{C}_{\Omega,\ \Omega}$的标准条件下，该项的典型大小以 $N^{-1/2}$ 的速率衰减。

在第 5.2 小节中分析矩阵估计量（4.11）时，对于每一列 λ'，总误差都可作类似分解。不过请注意，回归支撑集 $\widetilde{\ }\widetilde{\ }\widetilde{\ }\mathrm{supp}(J,\ t,\ t')$ 严格包含压缩支撑集 $\mathrm{supp}(J,\ t,\ t')$：嵌套支撑策略中的这一扩张—限制步骤可以减轻遗漏变量偏差，但也给分析带来了额外的技术难度。此外，由于回归支撑集 $\Omega_{\lambda'}$ 和噪声水平都依赖于 λ'，我们先推导利用相应稀疏模式和噪声缩放的逐列误差界，再跨尺度汇总这些界，以控制加权算子范数下的整体矩阵误差（3.2）。□

5 算子学习的收敛速率

5.1 第一个主要结果

现在我们可以陈述并讨论第一个主要结果：

定理 5.1. 假设 *2.1* 和假设 *3.1* 成立。设 $t' \geq t > r/2$。考虑在第 $\widehat{\mathcal{A}} = \widehat{\mathcal{A}}(\{u_i, f_i\}_{i=1}^N)$ 节中定义的估计量 *4*，其参数为

$$\widetilde{J} := \left\lceil \frac{t + t' - r + \varepsilon_1}{\min\{t', r_1\} + t - r}\, J \right\rceil, \quad \widetilde{t} := t', \quad \widetilde{t'} := \max\{t', r_1\}, \quad J := \left\lceil \frac{\log_2 N}{(2+\rho)(t + t' - r)} \right\rceil, \tag{5.1}$$

其中 $\varepsilon_1 := n(t + t' - r)/(\sigma - n/2 + t - r/2)$ 且

$$\rho := 2\max\left\{ \frac{-t - r_2 + n/2}{\sigma - n/2 + t - r/2}, \frac{-t' - r_2 + n/2}{\sigma - n/2 + t' - r/2}, \frac{-t - t' + r_1 - r_2 + n}{t + t' + 2\widetilde{d}}, \frac{-t - t' + r_1 - r_2}{t + t' - r}, 0 \right\}. \tag{5.2}$$

对于任意 $\delta \in (0,\ 1)$，若 $N \gtrsim \log(1/\delta)$，则其成立的概率至少为 $1 - \delta$，

$$\|\widehat{\mathcal{A}} - \mathcal{A}\|_{H^t \to H^{-t'}} \ \lesssim\ N^{-\frac{1}{2+\rho}} \sqrt{\log\Big(\frac{N}{\delta}\Big)}\, \log N.$$

定理 5.1 确立了学习椭圆伪微分算子时的高概率误差界，其阶为 $N^{-\frac{1}{2+\rho}}$ （相差对数因子），适用于含噪数据。该结论在输入和噪声满足一般假设的条件下成立，同时需要恰当选择小波。隐含常数仅取决于表 5.1中列出的参数，我们在全文中均将这些参数视为固定值。

本小节其余部分的组织如下。注记 5.2 讨论该假设$t \leq t'$ 并解释为何它不会导致一般性损失。注记 5.3 对

指数 ρ （见（5.2））进行了详细讨论，该指数决定了收敛速率；特别地，文中强调了该速率的若干定性特征，并说明其几乎是最优的。注记 5.4 表明，估计量 $\widehat{\mathcal{A}}$ 仅具有 $\mathcal{O}(2^{Jn})$ $= \mathcal{O}\big(N^{\frac{2+\rho}{n}(t+t'-r)}\big)$ 个非零系数，从而能够为下游任务采用快速数值算法。注记 5.5 延续了注 4.6 中的计算成本讨论，并表明在参数选择（5.1）下，构造 $\mathcal{A}$ 的成本几乎为 $\mathcal{O}(N2^{Jn})$ $= \mathcal{O}\big(N^{1+\frac{2+\rho}{n}(t+t'-r)}\big)$，即关于样本量 N呈线性，并且关于学习矩阵维数 2^{Jn}也呈线性。注记 5.6 考虑无噪声情形，并表明我们的估计量达到了超代数收敛速率。最后，例 5.7 在学习椭圆型偏微分方程的格林函数这一具体情形下说明了我们的结果，并将收敛速率和计算成本分别与 [121]中的结果进行了比较。

参数	描述
N	样本量：输入—输出数据对的数量 $\{(u_i, f_i)\}_{i=1}^{N}$
n	物理区域（流形M）的维数
r	伪微分算子的阶 $\mathcal{A}$
t, t'	误差范数中的输入/输出索伯列夫指数 $\|\cdot\|_{H^t \to H^{-t'}}$
r_1, r_2	假设2.1：中输入函数/噪声的正则性 $u \in H^{(r_1-n/2)-}, w \in H^{(r_2-n/2)-}$
$\gamma, \tilde{\gamma}$	原始/对偶小波的正则性
$d, \tilde{d}$	原始/对偶小波的逼近阶
σ	控制 $\mathbf{A}$ 尺度分离衰减的指数 （命题3.2 (I) （ii）；假设3.1 （iii））

表 5.1：关键参数汇总。

注记 5.2（关于假设 $t \le t'$）。在定理 5.1 中，我们假设 $t \le t'$ 仅为简化表述，这并不失一般性。事实上，当 $t \ge t'$ 时，我们可以取估计量 $\widehat{\mathcal{A}}^*$ 的伴随算子作为估计量，其中 $\mathcal{A}$ 定义于（4.12）中。由于 $\mathcal{A}$是自伴的，

$$(\widehat{\mathcal{A}})^* - \mathcal{A} = (\widehat{\mathcal{A}} - \mathcal{A})^*.$$

此外，根据 Sobolev 范数的对偶刻画和伴随算子的定义，

$$\begin{aligned}\|(\widehat{\mathcal{A}})^* - \mathcal{A}\|_{H^t \to H^{-t'}} &= \|(\widehat{\mathcal{A}} - \mathcal{A})^*\|_{H^t \to H^{-t'}} \\ &= \sup_{\|w\|_{H^t} \le 1, \|v\|_{H^{t'}} \le 1} |\langle (\widehat{\mathcal{A}} - \mathcal{A})^* w, v\rangle| \\ &= \sup_{\|w\|_{H^t} \le 1, \|v\|_{H^{t'}} \le 1} |\langle w, (\widehat{\mathcal{A}} - \mathcal{A}) v\rangle| = \|\widehat{\mathcal{A}} - \mathcal{A}\|_{H^{t'} \to H^{-t}}.\end{aligned}$$

因此，将 5.1 定理中的 t 与 t' 互换后应用该定理，可得到 $\|(\widehat{\mathcal{A}})^* - \mathcal{A}\|_{H^t \to H^{-t'}}$的上界。由于指数 ρ 在 (5.2) 中关于 (t, t')对称，因此得到的收敛速率与 $t \le t'$情形下的收敛速率一致。 □

注 5.3（关于收敛速率的讨论）。我们对指数 ρ在（5.2）中的含义作几点说明，该指数决定了收敛速率。

1. **对** t、t' **的依赖**。指标 t、t' 量化了算子范数的强度 $\|\cdot\|_{H^t\to H^{-t'}}$ 用于衡量估计误差。粗略地说，较大的 t, t' 对应较弱的范数，因此误差度量也较宽松。当 t 和 t' 相对于 r_1, r_2, n 足够大时，即

$$\min\{t, t'\} \geq -r_2 + \frac{n}{2} \quad \text{and} \quad t + t' \geq r_1 - r_2 + n, \tag{5.3}$$

我们有 $\rho = 0$，因而达到参数率 $N^{-1/2}$（至多相差对数因子）。事实上，在条件（5.3）下，有效方差至多随 J 呈多项式增长（这将在我们的误差分析中体现出来），因此偏差-方差权衡的形式为

$$2^{-J(t+t'-r)} + \frac{\mathrm{poly}(J)}{\sqrt{N}} \quad \Longrightarrow \quad N^{-1/2}\,\mathrm{polylog}(N).$$

当（5.3）不成立时，方差随 J 呈指数增长，从而导致真正的非参数率（$\rho > 0$）：

$$2^{-J(t+t'-r)} + \frac{2^{J\rho(t+t'-r)/2}\mathrm{poly}(J)}{\sqrt{N}} \quad \Longrightarrow \quad N^{-\frac{1}{2+\rho}}\,\mathrm{polylog}(N).$$

2. 对 n, r, r_1, r_2 的依赖关系。回顾一下，n 是流形 M 的维数，r 是偏微分算子 $\mathcal{A}$ 的阶，即 $\mathcal{A} \in \mathrm{OPS}^r(\mathrm{M})$。参数 r_1 和 r_2 控制输入和噪声的正则性：几乎必然有 $u \in H^{(r_1-n/2)-}$ 和 $w \in H^{(r_2-n/2)-}$；参见注 2.2。由 (5.2) 可知，ρ 随 n, r, r_1 增大而增大，随 r_2 增大而减小。因此，维数较低的流形（较小的 n）、阶较低的算子（较小的 r）、较粗糙的输入（较小的 r_1）和/或较平滑的噪声（较大的 r_2）会带来较小的 ρ，从而获得更快的收敛速率 $N^{-\frac{1}{2+\rho}}$。

3. 对 σ, d 的依赖。参数 σ 和 d 取决于双正交小波的选择。参数 σ 量化了 A 的元素在尺度分离下的非对角衰减（尤其是对于支撑重叠的小波）；参见命题 3.2 (I) (ii)。根据假设 3.1 (iii)，该参数必须满足相应条件：其下界固定，而允许的上界随小波正则性和逼近参数（$\gamma, \tilde{\gamma}, d$）的增加而增大。特别是，通过选择逼近阶更高的更平滑小波，可以取 σ 更大的值（事实上，σ 可能与 $\gamma, \tilde{\gamma}, d$ 成正比）。当 (5.3) 不成立时，(5.2) 中前三项至少有一项严格为正。在这种情况下，增大 σ 和/或 d 会减小 ρ，从而提高收敛速率 $N^{-\frac{1}{2+\rho}}$。

4. **内在收敛速率**。在以下理想化情形中，σ 和 d 都取足够大（形式上，$\sigma, d \to \infty$），式（5.2）中的前三项会消失，且 ρ 可简化为

$$\rho = 2\max\left\{\frac{-t - t' + r_1 - r_2}{t + t' - r}, 0\right\}.$$

该指数与小波的选择无关，并决定了由（t，t'，r_1，r_2，r）所支配的内在收敛速率。直观而言，增大 σ 和 d 可以充分利用 $\mathcal{A}$ 的非对角光滑性，但速率仍由误差度量（t，t'）、输入函数正则性 r_1 与噪声正则性 r_2 之间的相互作用所支配。在此情形下，偏差-方差权衡的形式为

$$2^{-J(t+t'-r)} + \frac{2^{J\max\{-t-t'+r_1-r_2,\,0\}}}{\sqrt{N}} \quad \Longrightarrow \quad N^{-\min\left\{\frac{t+t'-r}{2(r_1-r_2-r)},\,\frac{1}{2}\right\}}. \tag{5.4}$$

量 r_1-r_2-r 可以理解为一种有效光滑度，即信号的正则性 $\mathcal{A}u \in H^{(r_1-r-n/2)}$ – 与噪声的正则性 $w \in H^{(r_2-n/2)}$ –之间的差距。 r_1-r_2-r 的值越大（ $\mathcal{A}u$ 越光滑和/或噪声 w 越粗糙），问题就越困难，学习速率也越慢。特别地，当 $r_1-r_2>t+t'$ （等价地，$r_1-r_2-r>t+t'-r$）时，我们有

$$\rho=\frac{2(-t-t'+r_1-r_2)}{t+t'-r}, \qquad \text{and hence} \qquad N^{-\frac{1}{2+\rho}}=N^{-\frac{t+t'-r}{2(r_1-r_2-r)}}.$$

最后，我们指出，因子 $2^{J\max\{-t-t'+r_1-r_2,0\}}/\sqrt{N}$ 对应于估计压缩矩阵 $(0,\ 0)$ 和 $(J,\ J)$ 块中单个元素的标准差 $\mathbf{A}^{\varepsilon}_{\Lambda_J}$。为说明这一点，考虑最简单的理想化情形：此时 $\mathbf{A}$ 恰好是对角矩阵——对于一般的偏微分算子，这一假设通常并不成立 $\mathcal{A}$ 在小波坐标中。在这种情况下， λ'第 列方程（4.5）化简为

$$\mathbf{F}_{\cdot,\lambda'}=\mathbf{U}\mathbf{A}_{\cdot,\lambda'}+\mathbf{W}_{\cdot,\lambda'}=\mathbf{A}_{\lambda',\lambda'}\mathbf{U}_{\cdot,\lambda'}+\mathbf{W}_{\cdot,\lambda'},$$

其中 $\mathbf{U}_{\cdot,\ \lambda'}\sim\mathcal{N}\big(0,\ \ (\sigma^U_{\lambda'})^2I_N\big)$ 和 $\mathbf{W}_{\cdot,\ \lambda'}\sim\mathcal{N}\big(0,\ \ (\sigma^W_{\lambda'})^2I_N\big)$ 相互独立，且根据命题 3.2 （II）（ii）–（iii）， $\sigma^U_{\lambda'}\asymp 2^{-j'r_1}$ 且 $\sigma^W_{\lambda'}\asymp 2^{-j'r_2}$。因此，恢复 $\mathbf{A}$ 的第 λ' 列可归结为估计单个标量参数 $\mathbf{A}_{\lambda',\ \lambda'}\in\mathbb{R}$。

在这个高斯位置模型中，标量最小二乘估计量（等价地，最大似然估计量）满足

$$|\widehat{\mathbf{A}}_{\lambda',\lambda'}-\mathbf{A}_{\lambda',\lambda'}|\asymp\frac{\sigma^W_{\lambda'}}{\sqrt{N}\,\sigma^U_{\lambda'}}\asymp\frac{2^{j'(r_1-r_2)}}{\sqrt{N}},$$

其中 $\sigma^U_{\lambda'}/\sigma^W_{\lambda'}$ 是信噪比。此外，这一速率是极小极大最优的（例如，可通过直接应用 Le Cam 两点论证得出）。最后，范数 $\|\cdot\|_{H^t\to H^{-t'}}$ 带来了额外的权重 $2^{-j'(t+t')}$，因此加权标准差的尺度为 *weighted*

$$\frac{2^{j'(-t-t'+r_1-r_2)}}{\sqrt{N}}.$$

当 $1/\sqrt{N}$ 时，这等于 $j'=0$；而当 $2^{J(-t-t'+r_1-r_2)}/\sqrt{N}$ 时，这等于 $j'=J$。这解释了 $2^{J\max\{-t-t'+r_1-r_2,0\}}/\sqrt{N}$ 在（5.4）中出现的原因，并表明，若不对算子或输入函数与噪声的协方差结构作出进一步假设，（5.4）中的速率便无法改进。□

注记 5.4（ $\widehat{\mathcal{A}}$的稀疏性： $\mathcal{O}(2^{Jn})$ 个非零系数）。我们的估计量 $\widehat{\mathcal{A}}$在小波坐标下是最优稀疏的，也就是说，其矩阵表示 $\widehat{\mathbf{A}}\in\mathbb{R}^{\Lambda_J\times\Lambda_J}$ 仅有 $\mathcal{O}(2^{Jn})$ 个非零元素，且该数量与 $|\Lambda_J|\asymp 2^{Jn}$呈线性关系。事实上，根据构造， $\widehat{\mathbf{A}}$的支撑集包含于 $\mathrm{supp}(J,\ t,\ t')$，而命题 C.2 保证 $\mathrm{nnz}(\widehat{\mathbf{A}})=\mathrm{nnz}\big(\mathbf{M}_{(J,t,t')}\big)\lesssim 2^{Jn}$。在（5.1）中对 J 采用偏差-方差最优选择时，我们有 $\mathcal{O}(2^{Jn})=\mathcal{O}\big(N^{\frac{n}{2+\rho+t+t'-r}}\big)$ 个非零系数位于 $\widehat{\mathbf{A}}$中。□

注 5.5（计算成本：几乎 $\mathcal{O}(N2^{Jn})$）。在本注中，我们继续讨论注 4.6 中的计算成本，并在采用（5.1）中的参数选择时，给出该成本的更精确界。由（4.13）可知，我们的估计量的计算成本取决于回归支撑集的稀疏结构 $\mathrm{supp}(J,\ t,\ t')$：

$$\mathcal{O}\bigg(N\Big(\sum_{\lambda'\in\Lambda_J}|\Omega_{\lambda'}|^2\Big)+\sum_{\lambda'\in\Lambda_J}|\Omega_{\lambda'}|^3\bigg).$$

~为简化记号，令 $J := (1+\varepsilon')J$ 并定义

$$\nu := \frac{n(\widetilde{t}+\widetilde{t}'-r)}{\sigma-n/2+\widetilde{t}'-r/2}.$$

根据命题 C.2中的逐列支撑集界，对于 $\lambda'=(j',\ k')\in\Lambda_J$，

$$|\Omega_{\lambda'}| \ \lesssim\ \widetilde{J}+2^{\nu(\widetilde{J}-j')}.$$

对于 $p\in\{2、3\}$，利用 $(x+y)^p\le 2^{p-1}(x^p+y^p)$（对于 $x,\ y>0$）以及 $|\nabla_{j'}|\asymp 2^{j'n}$，我们得到 in

$$\begin{aligned}\sum_{\lambda'\in\Lambda_J}|\Omega_{\lambda'}|^p &\lesssim \sum_{j'=0}^{J}\sum_{k'\in\nabla_{j'}}\left(\widetilde{J}+2^{\nu(\widetilde{J}-j')}\right)^p\\ &\lesssim \sum_{j'=0}^{J}2^{j'n}\left(\widetilde{J}^p+2^{p\nu(\widetilde{J}-j')}\right)\\ &\lesssim \widetilde{J}^p\sum_{j'=0}^{J}2^{j'n}\ +\ 2^{p\nu\widetilde{J}}\sum_{j'=0}^{J}2^{(n-p\nu)j'}.\end{aligned}$$

假设 $\nu<n/p$（当 σ 足够大时成立）。那么几何级数满足 $\sum_{j'=0}^{J}2^{(n-p\nu)j'}\lesssim 2^{(n-p\nu)J}$，因此

$$\sum_{\lambda'\in\Lambda_J}|\Omega_{\lambda'}|^p\lesssim J^p2^{Jn}+2^{Jn}\,2^{p\nu(\widetilde{J}-J)}\lesssim 2^{Jn\left(1+\frac{p\varepsilon'(\widetilde{t}+\widetilde{t}'-r)}{\sigma-n/2+\widetilde{t}'-r/2}\right)}.$$

特别地，

$$\sum_{\lambda'\in\Lambda_J}|\Omega_{\lambda'}|^2\lesssim 2^{Jn\left(1+\frac{2\varepsilon'(\widetilde{t}+\widetilde{t}'-r)}{\sigma-n/2+\widetilde{t}'-r/2}\right)},\qquad \sum_{\lambda'\in\Lambda_J}|\Omega_{\lambda'}|^3\lesssim 2^{Jn\left(1+\frac{3\varepsilon'(\widetilde{t}+\widetilde{t}'-r)}{\sigma-n/2+\widetilde{t}'-r/2}\right)}.$$

因此，总计算成本满足

$$N\left(\sum_{\lambda'\in\Lambda_J}|\Omega_{\lambda'}|^2\right)+\sum_{\lambda'\in\Lambda_J}|\Omega_{\lambda'}|^3\lesssim N2^{Jn(1+\varepsilon'')}+2^{Jn\left(1+\frac{3}{2}\varepsilon''\right)}\lesssim N2^{Jn(1+\varepsilon'')},\tag{5.5}$$

只要

$$N\gtrsim 2^{Jn\varepsilon''/2},\qquad \text{where}\quad \varepsilon'':=\frac{2\varepsilon'(\widetilde{t}+\widetilde{t}'-r)}{\sigma-n/2+\widetilde{t}'-r/2}>0.$$

一个充分（更强）的条件是

$$N\gtrsim 2^{\widetilde{J}n(\widetilde{t}+\widetilde{t}'-r)/(\sigma-n/2+\widetilde{t}'-r/2)}.$$

在我们的选择下，这成立：$J=\left\lceil \log_2 N^{(2+\rho)(t+t'-r)}\right\rceil$ 见（5.1）；参见（5.13）。

由于 $\varepsilon''\to 0$ 随 $\sigma\to\infty$ 变化（即，随着我们选择更平滑的双正交小波），计算成本（5.5）可以任意接近 $\mathcal{O}(N2^{Jn})$，也就是说，关于样本量 N 呈线性增长，并且除常数因子外，关于小波指标数量 $|\Lambda_J|\asymp 2^{Jn}$也呈线性增长。回顾一下，(4.11) 中的估计量满足 $\widehat{\mathbf{A}}\in\mathbf{R}^{\Lambda_J\times\Lambda_J}$。相比之下，任何显式构造 $\mathbf{U}_{\cdot,\Lambda_J}^{\top}\mathbf{U}_{\cdot,\Lambda_J}$ 的直接估计量

已经需要 $\mathcal{O}(N2^{2Jn})$的计算成本，即关于 $|\Lambda_J|$呈二次增长。此外，选择 J于 (5.1) 中的取值时，该取值可使截断、压缩和估计误差达到最优平衡，由此我们得到

$$\mathcal{O}(N2^{Jn}) = \mathcal{O}\Big(N^{1+\frac{n}{(2+\rho)(\widetilde{t}+\widetilde{t}'-r)}}\Big). \tag{5.6}$$

最后请注意，上述界限的主要贡献来自满足 $j' = J$ 的列。事实上，当 $j' = J$ 且 $k' \in \nabla_J$ 时，预计压缩矩阵 $\mathbf{A}^{\varepsilon}_{\Lambda_J}$ 的相应列仅包含 $\mathcal{O}(1)$ 个显著元素。我们推测，通过对扩展回归支撑集作出更精细的选择，可以改进界限（5.5），并达到 精确的 最优计算成本 $\mathcal{O}(N2^{Jn})$，同时保持定理 5.1 中确立的统计收敛速率；这一问题留待未来工作研究。□

注 5.6（无噪声情形：超代数收敛速率）。我们可以直接将 $\widehat{\mathcal{A}}$的分析学调整到无噪声情形。假设数据对 $\{u_i, f_i\}_{i=1}^N$ 满足

$$f_i = \mathcal{A}u_i, \qquad 1 \le i \le N, \tag{5.7}$$

因此在小波-矩阵表述中

$$\mathbf{F} = \mathbf{U}\mathbf{A}. \tag{5.8}$$

在假设 2.1 （不含 (iii)）和假设 3.1下，我们使用与（ $\mathcal{A}$ 4.12）中相同的估计量 。按照与有噪声情形相同的论证可得

$$\|\widehat{\mathcal{A}} - \mathcal{A}\|_{H^t \to H^{-t'}} \lesssim \mathsf{Error\text{-}Truncation} + \mathsf{Error\text{-}Compression} + \mathsf{Error\text{-}OVB},$$

其中方差项 误差-方差 不存在。根据命题 5.9、5.10和 5.11，若

$$N \gtrsim 2^{\widetilde{J}n(\widetilde{t}+\widetilde{t}'-r)/(\sigma-n/2+\widetilde{t}'-r/2)} + \log(1/\delta),$$

则以至少 $1-\delta$ 的概率,

$$\|\widehat{\mathcal{A}} - \mathcal{A}\|_{H^t \to H^{-t'}} \lesssim \mathsf{Error\text{-}Truncation} + \mathsf{Error\text{-}Compression} + \mathsf{Error\text{-}OVB} \lesssim J\, 2^{-J(t+t'-r)}.$$

因此，我们可以通过使样本量约束取等，选择 J ,

$$N \asymp 2^{\widetilde{J}n(\widetilde{t}+\widetilde{t}'-r)/(\sigma-n/2+\widetilde{t}'-r/2)} \quad \Longrightarrow \quad \widetilde{J} = \left\lceil \frac{\sigma - n/2 + \widetilde{t}' - r/2}{n(\widetilde{t}+\widetilde{t}'-r)} \log_2 N \right\rceil,$$

然后，利用 J 与 J 在（5.1）中的尺度关系，设定

$$J = \left\lceil \frac{\min\{t', r_1\} + t - r}{t + t' - r + \varepsilon_1} \widetilde{J} \right\rceil = \left\lceil \frac{\min\{t', r_1\} + t - r}{t + t' - r + \varepsilon_1} \cdot \frac{\sigma - n/2 + \widetilde{t}' - r/2}{n(\widetilde{t}+\widetilde{t}'-r)} \log_2 N \right\rceil. \tag{5.9}$$

采用这一选择后,

$$\|\widehat{\mathcal{A}} - \mathcal{A}\|_{H^t \to H^{-t'}} \lesssim J\, 2^{-J(t+t'-r)} \lesssim_q N^{-q} \log N, \tag{5.10}$$

其中

$$q := (t + t' - r) \cdot \frac{\min\{t', r_1\} + t - r}{t + t' - r + \varepsilon_1} \cdot \frac{\sigma - n/2 + \widetilde{t}' - r/2}{n(\widetilde{t}+\widetilde{t}'-r)}. \tag{5.11}$$

特别地， q 随 σ 线性增长（而且 σ 也可以随 γ， $\widetilde{\gamma}$， d线性增长），因此，只要选择足够光滑且逼近阶足够高的双正交小波，就可以使 q 任意大。等价地，对于任意 $q_0 > 0$，都可以选取小波使得 $q \geq q_0$，从而得到如下高概率界

$$\|\widehat{\mathcal{A}} - \mathcal{A}\|_{H^t \to H^{-t'}} \lesssim_{q_0} N^{-q_0} \log N.$$

因此，在无噪声情形下，误差会随着小波光滑度的提高而超代数衰减 N （达到对数因子的精度）。

此外，在采用 (5.9) 中的 J 时，估计矩阵 $\widehat{\mathbf{A}} \in \mathbb{R}^{\Lambda_J \times \Lambda_J}$ 的大小为 $|\Lambda_J| \asymp 2^{Jn} \asymp N^{\frac{qn}{t+t'-r}}$，且 $\widehat{\mathbf{A}}$ 包含 $\mathcal{O}(2^{Jn}) = \mathcal{O}\left(N^{\frac{qn}{t+t'-r}}\right)$ 个非零小波系数。此外，构造估计量的计算成本几乎为 $\mathcal{O}(N2^{Jn}) = \mathcal{O}\left(N^{1+\frac{qn}{t+t'-r}}\right)$，其中$q$ 如 (5.11) 所示。□

例 5.7 （学习一致椭圆型偏微分方程的格林函数）。 在上面的讨论中，所学习矩阵的大小、稀疏性层级（即非零系数的数量）以及计算成本都取决于样本量 N。这体现了非参数统计学的一个标准特征：学习问题的有效维度会随着数据量的增加而增长。由于我们的界依赖于多个参数，下面我们通过一个具体例子来说明这些界的含义。

考虑与一个阶为 $2s$ 的一致椭圆型偏微分方程相关的（经过适当归一化的）解算子（格林函数），其中 $s > 0$。在此情形下，解算子属于$\mathsf{OPS}^{-2s}(M)$，即 $r = -2s$。回顾一下，我们是在一个 n 维光滑闭流形上进行研究；将结果推广到具有适当边界条件的更一般区域也是自然的。为简化讨论，我们考虑算子范数 $\|\cdot\|_{L^2(M)\to L^2(M)}$，即令 $t = t' = 0$。

- ***No***这种情形。 由于 $r = -2s$， $t = t' = 0$，以及 $r_2 > n/2$，参数 ρ 在 (5.2) 变为

  $$\rho = \max\left\{\frac{r_1 - r_2 + n}{\widetilde{d}}, \frac{r_1 - r_2}{s}, 0\right\}.$$

  根据定理*5.1* 以及备注 *5.4* 和 *5.5*，收敛速率、矩阵 $\widehat{\mathbf{A}}$估计量 及其非零元素数（*nnz*），以及构造该估计量的计算成本，均随 N 按如下方式变化：

  $$\begin{aligned}
  &\textit{Convergence rate:} \quad \mathcal{O}\left(N^{-\frac{1}{2+\rho}} \log^{\frac{3}{2}}(N)\right).\\
  &\textit{Size and nnz of } \widehat{\mathbf{A}}\textit{:} \quad \mathcal{O}\left(N^{\frac{n}{2s(2+\rho)}}\right).\\
  &\textit{Computational cost:} \quad \mathcal{O}\left(N^{1+\frac{n}{2s(2+\rho)}}\right).
  \end{aligned}$$

  因此，要在有噪声情形下以 $\|\cdot\|_{L^2\to L^2}$ 达到目标精度容差 ϵ，即确保 $N^{-\frac{1}{2+\rho}} \log^{\frac{3}{2}} N \asymp \epsilon$，只需取

  $$\#\textit{input–output noisy data pairs:} \quad \mathcal{O}\left(\epsilon^{-(2+\rho)} \log^{\frac{3(2+\rho)}{2}}(\epsilon^{-1})\right).$$

  这进而意味着：

  $$\begin{aligned}
  &\textit{Size and nnz of } \widehat{\mathbf{A}}\textit{:} \quad \mathcal{O}\left(\epsilon^{-\frac{n}{2s}} \log^{\frac{3n}{4s}}(\epsilon^{-1})\right).\\
  &\textit{Computational cost:} \quad \mathcal{O}\left(\epsilon^{-(2+\rho)-\frac{n}{2s}} \log^{\frac{6s(2+\rho)+3n}{4s}}(\epsilon^{-1})\right).
  \end{aligned}$$

- 无噪声情形。

当 $r=-2s$ 且 $t=t'=0$时，指数 q 在 (5.10) 和 (5.11) 变为

$$q=\frac{2s}{1+\frac{n}{\sigma-n/2+s}}\cdot\frac{\sigma-n/2+r_1+s}{n(r_1+2s)}.$$

根据备注5.6，我们现在得到：

$$\begin{aligned}&\textit{Convergence rate:}\quad \mathcal{O}\big(N^{-q}\log N\big).\\&\textit{Size and nnz of }\widehat{\mathbf{A}}\textit{:}\quad \mathcal{O}\big(N^{\frac{qn}{2s}}\big).\\&\textit{Computational cost:}\quad \mathcal{O}\big(N^{1+\frac{qn}{2s}}\big).\end{aligned}$$

因此，要在无噪声情形下使 $\|\cdot\|_{L^2\to L^2}$ 下的目标精度容差达到 ϵ，即确保 $N^{-q}\log N\asymp\epsilon$，只需取

$$\#\textit{input–output noiseless data pairs:}\quad \mathcal{O}\big(\epsilon^{-\frac{1}{q}}\log^{\frac{1}{q}}(\epsilon^{-1})\big).$$

这进而意味着：

$$\begin{aligned}&\textit{Size and nnz of }\widehat{\mathbf{A}}\textit{:}\quad \mathcal{O}\big(\epsilon^{-\frac{n}{2s}}\log^{\frac{n}{2s}}(\epsilon^{-1})\big).\\&\textit{Computational cost:}\quad \mathcal{O}\big(\epsilon^{-\frac{1}{q}-\frac{n}{2s}}\log^{\frac{1}{q}+\frac{n}{2s}}(\epsilon^{-1})\big).\end{aligned}$$

我们将上述结果与 *[121]*，进行比较。该文研究如何从精心设计的无噪声向量测量中恢复离散化的（有限维）格林矩阵。他们的方法利用椭圆型偏微分方程解算子的乔列斯基分解中的近似稀疏性，其中偏微分方程的阶为 $2s$，s 是正整数；随后通过分片常数或分片仿射插值来近似连续格林函数。具体而言，要在 $\|\cdot\|_{L^2\to L^2}$ 范数下以精度 ϵ 恢复连续格林函数，*[121]* 需要使用 $\mathcal{O}\big(\log^{n+1}(\epsilon^{-1})\big)$ 个精心挑选的无噪声数据对，计算成本为 $\mathcal{O}\big(\epsilon^{-n}\log^{2n+2}(\epsilon^{-1})\big)$；参见定理 *3.5]*。主导的 $\mathcal{O}(\epsilon^{-n})$ 依赖关系源于最后的低阶插值步骤，该步骤使用恢复出的离散格林矩阵来近似连续格林函数；因此，偏微分方程的阶 $2s$ （以及椭圆平滑的强度）不会出现在主导的 ϵ 指数中。

相比之下，我们的无噪声保证表明，对于一般的高斯输入数据，要达到精度 ϵ，需要 $\mathcal{O}\big(\epsilon^{-\frac{1}{q}}\log^{\frac{1}{q}}(\epsilon^{-1})\big)$ 个数据对，计算成本为 $\mathcal{O}\big(\epsilon^{-\frac{1}{q}-\frac{n}{2s}}\log^{\frac{1}{q}+\frac{n}{2s}}(\epsilon^{-1})\big)$ 阶。通过选择逼近阶更高的更平滑小波，指数 q 可以任意增大，因此所需的数据对数量可以比关于 ϵ^{-1} 的任何固定多项式都增长得任意慢。在大 q 情形下，我们的计算成本趋近于 $\mathcal{O}\big(\epsilon^{-\frac{n}{2s}}\log^{\frac{n}{2s}}(\epsilon^{-1})\big)$；当 $s>1/2$ 时，相较于 *[121, Theorem 3.5]* 中的 $\mathcal{O}\big(\epsilon^{-n}\log^{2n+2}(\epsilon^{-1})\big)$，这改善了对 ϵ 的依赖。我们解释了为何在此情形下，计算成本 $\mathcal{O}\big(\epsilon^{-\frac{n}{2s}}\log^{\frac{n}{2s}}(\epsilon^{-1})\big)$ 在对数因子范围内是最优的。若要在 $\|\cdot\|_{L^2\to L^2}$ 中达到精度 ϵ，最多只能舍弃尺度 J 之外的信息，使得 $2^{-2Js}\asymp\epsilon$，因为忽略更细尺度所造成的不可避免偏差为 $\mathcal{O}(2^{-2Js})$。在 n 维空间中，截至尺度 J 的自由度数量为 $\mathcal{O}(2^{Jn})$ 阶，而每个自由度的学习成本至少为 $\mathcal{O}(1)$。因此，任何方法的总成本至少为

$$\mathcal{O}(2^{Jn})=\mathcal{O}\big((2^{-2Js})^{-\frac{n}{2s}}\big)=\mathcal{O}\big(\epsilon^{-\frac{n}{2s}}\big),$$

与我们的复杂度界相匹配，相差对数因子。

此外，我们的框架还适用于有噪声的观测和一般的高斯输入，在一般的索博列夫空间到索博列夫空间的算子范数下成立，并适用于分数阶椭圆型偏微分方程

（其中底层微分算子是非局部的）——这些设定不涵盖于 *[121]*。在有噪声情形下，达到 ϵ 精度的计算成本为（忽略对数因子）

$$\mathcal{O}\big(\epsilon^{-(2+\rho)-\frac{n}{2s}}\big), \quad \rho = \max\left\{\frac{r_1 - r_2 + n}{\widetilde{d}}, \frac{r_1 - r_2}{s}, 0\right\},$$

与无噪声情形下的尺度相比，这要差约 $\epsilon^{-(2+\rho)}$ 倍。一般而言， $\mathcal{O}(\epsilon^{-2})$因子不可避免：它反映了加性噪声所导致的中心极限定理尺度。特别是，要从含噪样本中以精度 ϵ估计每个自由度，通常需要 $\mathcal{O}(\epsilon^{-2})$ 个观测。因子 $\mathcal{O}\big(\epsilon^{-\frac{n}{2s}}\big)$ 计入了必须学习的自由度数量（精度所决定的尺度范围内），这一点已在上一段中讨论。据我们所知，这是首次针对从含噪数据中学习椭圆型偏微分方程的格林函数，明确给出近乎最优的统计精度与计算成本权衡。□

5.2 定理 5.1 的证明

本小节证明定理 5.1。首先，我们将总误差分解为三项：截断误差、压缩误差和估计误差。截断误差是将无限维矩阵 $\mathbf{A}$ 限制为其有限分辨率版本 $\mathbf{A}^{\Lambda_J}$所产生的偏差；压缩误差则是将 $\mathbf{A}^{\Lambda_J}$ 稀疏化为 $\mathbf{A}^{\varepsilon}_{\Lambda_J}$所产生的确定性逼近误差；参见小节 4.1。这两项均与数据无关，分别由命题 5.9和 5.10控制。估计误差是唯一依赖于数据的项，它量化了估计压缩矩阵 $\mathbf{A}^{\varepsilon}_{\Lambda_J}$时的统计不确定性。引理 5.8 进一步将其分解为遗漏变量偏差项（仍具有随机性）和方差项，它们分别由命题 5.11和 5.12界定。该引理及各命题的证明见附录 C.2、C.3和 C.4。

定理 5.1 的证明。令 $\mathrm{D}^{s}_{\Lambda_J} := (\mathrm{D}^{s}_{\lambda,\lambda'})_{\lambda,\lambda'\in\Lambda_J} \in \mathrm{R}^{\Lambda_J\times\Lambda_J}$ 表示限制在 Λ_J 上的对角权重矩阵。对于任意指标集 $\Lambda \subset \mathcal{J}$，令 $P_\Lambda : \ell^2(\mathcal{J}) \to \ell^2(\Lambda)$ 为坐标投影， $I_\Lambda := P_\Lambda^*$ 为其伴随算子（零填充嵌入）。对于式 (4.11) 中的矩阵估计量 $\widehat{\mathbf{A}} \in \mathrm{R}^{\Lambda_J\times\Lambda_J}$，以及定义 4.1 中的先验截断矩阵 $\mathrm{A}_{\Lambda_J} \in \mathrm{R}^{\Lambda_J\times\Lambda_J}$，定义它们的零填充扩展

$$\widehat{\mathbf{A}}^{\uparrow} := I_{\Lambda_J}\widehat{\mathbf{A}}P_{\Lambda_J} \in \mathbb{R}^{\mathcal{J}\times\mathcal{J}}, \qquad \mathbf{A}^{\uparrow}_{\Lambda_J} := I_{\Lambda_J}\mathbf{A}_{\Lambda_J}P_{\Lambda_J} \in \mathbb{R}^{\mathcal{J}\times\mathcal{J}}.$$

我们先从算子范数等价（3.2）开始,

$$\|\widehat{\mathcal{A}} - \mathcal{A}\|_{H^t\to H^{-t'}} \asymp \|\mathbf{D}^{-t'}(\widehat{\mathbf{A}}^{\uparrow} - \mathbf{A})\mathbf{D}^{-t}\|.$$

加上并减去 $\mathbf{A}^{\uparrow}_{\Lambda_J}$，得到

$$\begin{aligned}\|\widehat{\mathcal{A}} - \mathcal{A}\|_{H^t\to H^{-t'}} &\asymp \|\mathbf{D}^{-t'}(\widehat{\mathbf{A}}^{\uparrow} - \mathbf{A})\mathbf{D}^{-t}\| \\ &\le \|\mathbf{D}^{-t'}(\mathbf{A}^{\uparrow}_{\Lambda_J} - \mathbf{A})\mathbf{D}^{-t}\| + \|\mathbf{D}^{-t'}(\widehat{\mathbf{A}}^{\uparrow} - \mathbf{A}^{\uparrow}_{\Lambda_J})\mathbf{D}^{-t}\| \\ &= \underbrace{\|\mathbf{D}^{-t'}(\mathbf{A}^{\uparrow}_{\Lambda_J} - \mathbf{A})\mathbf{D}^{-t}\|}_{=:\text{Error-Truncation}} + \|\mathbf{D}^{-t'}_{\Lambda_J}(\widehat{\mathbf{A}} - \mathbf{A}_{\Lambda_J})\mathbf{D}^{-t}_{\Lambda_J}\|. \end{aligned} \tag{5.12}$$

将第二项分解可得

$$\|\mathbf{D}^{-t'}_{\Lambda_J}(\widehat{\mathbf{A}} - \mathbf{A}_{\Lambda_J})\mathbf{D}^{-t}_{\Lambda_J}\| \le \underbrace{\|\mathbf{D}^{-t'}_{\Lambda_J}(\mathbf{A}^{\varepsilon}_{\Lambda_J} - \mathbf{A}_{\Lambda_J})\mathbf{D}^{-t}_{\Lambda_J}\|}_{=:\text{Error-Compression}} + \underbrace{\|\mathbf{D}^{-t'}_{\Lambda_J}(\widehat{\mathbf{A}} - \mathbf{A}^{\varepsilon}_{\Lambda_J})\mathbf{D}^{-t}_{\Lambda_J}\|}_{=:\text{Error-Estimation}}.$$

根据引理 5.8，估计误差进一步分解为遗漏变量偏差和方差贡献：

$$\text{Error-Estimation} = \|\mathbf{D}_{\Lambda_J}^{-t'}(\widehat{\mathbf{A}} - \mathbf{A}_{\Lambda_J}^{\varepsilon})\mathbf{D}_{\Lambda_J}^{-t}\| \le \text{Error-OVB} + \text{Error-Var}.$$

因此，

$$\|\widehat{\mathcal{A}} - \mathcal{A}\|_{H^t \to H^{-t'}} \lesssim \text{Error-Truncation} + \text{Error-Compression} + \text{Error-OVB} + \text{Error-Var}.$$

命题 5.9 和5.10 给出了确定性界

$$\text{Error-Truncation} \lesssim 2^{-J(t+t'-r)}, \qquad \text{Error-Compression} \lesssim J2^{-J(t+t'-r)}.$$

此外，命题 5.11 和 5.12 表明，对于任意 $\delta \in (0,1)$，若

$$N \gtrsim 2^{\widetilde{J}n(\widetilde{t}+\widetilde{t}'-r)/(\sigma-n/2+\widetilde{t}'-r/2)} + \log(1/\delta), \tag{5.13}$$

则以至少为 $1-\delta$ 的概率，

$$\text{Error-OVB} \lesssim J2^{-J(t+t'-r)}, \qquad \text{Error-Var} \lesssim \sqrt{\frac{J+\log(1/\delta)}{N}} \cdot J \cdot 2^{\rho J(t+t'-r)/2},$$

其中 $\widetilde{J}$，$\widetilde{t}$，$\widetilde{t}'$ 在（5.1）中定义，而 ρ 在（5.2）中定义。

因此，在（5.13）条件下，以至少 $1-\delta$ 的概率，有

$$\|\widehat{\mathcal{A}} - \mathcal{A}\|_{H^t \to H^{-t'}} \lesssim J2^{-J(t+t'-r)} + \sqrt{\frac{J+\log(1/\delta)}{N}} \cdot J \cdot 2^{\rho J(t+t'-r)/2}.$$

为平衡偏差项和方差项，取

$$J := \left\lceil \frac{\log_2 N}{(2+\rho)(t+t'-r)} \right\rceil.$$

Then

$$2^{-J(t+t'-r)} \asymp N^{-\frac{1}{2+\rho}}, \qquad 2^{\rho J(t+t'-r)/2} \asymp N^{\frac{\rho}{2(2+\rho)}},$$

于是

$$\|\widehat{\mathcal{A}} - \mathcal{A}\|_{H^t \to H^{-t'}} \lesssim N^{-\frac{1}{2+\rho}} \sqrt{\log\left(\frac{N}{\delta}\right)} \log N.$$

接下来只需验证在此 J 取值下，样本量条件（5.13）是否成立。由于 $J \asymp \log_2 N/(2+\rho)(t+t'-r)$，我们有

$$N = 2^{(2+\rho)J(t+t'-r)} \gtrsim 2^{\widetilde{J}n(\widetilde{t}+\widetilde{t}'-r)/(\sigma-n/2+\widetilde{t}'-r/2)},$$

只要

$$(2+\rho)J(t+t'-r) > \widetilde{J}n\frac{\widetilde{t}+\widetilde{t}'-r}{\sigma-n/2+\widetilde{t}'-r/2}.$$

根据定义（5.1），$\widetilde{J}$，$\widetilde{t}$，$\widetilde{t}'$，这等价于

$$(2+\rho)J(t+t'-r) > \frac{t+t'-r+\varepsilon_1}{\min\{t',r_1\}+t-r}\, Jn\, \frac{t'+\max\{t',r_1\}-r}{\sigma-n/2+\max\{t',r_1\}-r/2},$$

其中 $\varepsilon_1 = n(t+t'-r)/(\sigma - n/2 + t - r/2)$。由于 $\rho \geq 0$，只需满足

$$2 > \frac{n + \frac{n^2}{\sigma - n/2 + t - r/2}}{\min\{t', r_1\} + t - r} \cdot \frac{t' + \max\{t', r_1\} - r}{\sigma - n/2 + \max\{t', r_1\} - r/2}.$$

注意到 $\max\{t',\ r_1\} \geq r_1 > n/2 + \max\{0,\ r\} \geq n/2 + r/2$，我们看到，上述不等式成立的一个充分条件是

$$\sigma > \max\left\{\frac{3n}{2} - t + \frac{r}{2},\ \frac{t' + \max\{t', r_1\} - r}{\min\{t', r_1\} + t - r} n\right\}.$$

这由假设3.1 （v）保证，因此样本量条件（5.13）对所有充分大的 N均成立。所声称的界随即得到。 □

在定理 5.1的证明中，我们调用引理 5.8 推导出误差分解，并分别利用命题 5.9、5.10、5.11 和 5.12来界定截断误差、压缩误差、遗漏变量偏差和方差项。这些结果的证明见附录C.2、C.3和 C.4。

引理 5.8 （误差-估计分解）。 在定理*5.1*的设定和假设下，估计误差 $\|\mathbf{D}_{\Lambda_J}^{-t'}(\widehat{\mathbf{A}} - \mathbf{A}_{\Lambda_J}^{\varepsilon})\mathbf{D}_{\Lambda_J}^{-t}\|$ 可以分解为：

$$\|\mathbf{D}_{\Lambda_J}^{-t'}(\widehat{\mathbf{A}} - \mathbf{A}_{\Lambda_J}^{\varepsilon})\mathbf{D}_{\Lambda_J}^{-t}\| \leq \mathsf{Error\text{-}OVB} + \mathsf{Error\text{-}Var},$$

其中，遗漏变量偏差和方差贡献为

$$\mathsf{Error\text{-}OVB} := \left\|\mathbf{D}_{\Lambda_J}^{-t'}\big((\mathbf{M}_{(J,t,t')})_{\mathrm{up}} \odot (\mathsf{OVB}_{\Lambda_J,\cdot})\big)\mathbf{D}_{\Lambda_J}^{-t}\right\| + \left\|\mathbf{D}_{\Lambda_J}^{-t}\big(((\mathbf{M}_{(J,t,t')})_{\mathrm{low}})^\top \odot (\mathsf{OVB}_{\Lambda_J,\cdot})\big)\mathbf{D}_{\Lambda_J}^{-t'}\right\|,$$
$$\mathsf{Error\text{-}Var} := \left\|\mathbf{D}_{\Lambda_J}^{-t'}\big((\mathbf{M}_{(J,t,t')})_{\mathrm{up}} \odot (\mathsf{Var}_{\Lambda_J,\cdot})\big)\mathbf{D}_{\Lambda_J}^{-t}\right\| + \left\|\mathbf{D}_{\Lambda_J}^{-t}\big(((\mathbf{M}_{(J,t,t')})_{\mathrm{low}})^\top \odot (\mathsf{Var}_{\Lambda_J,\cdot})\big)\mathbf{D}_{\Lambda_J}^{-t'}\right\|,$$

with

$$\mathsf{OVB} := \sum_{\lambda' \in \Lambda_J} \left(E_{\Omega_{\lambda'}}(\mathbf{U}_{\cdot,\Omega_{\lambda'}}^\top \mathbf{U}_{\cdot,\Omega_{\lambda'}})^{-1}\mathbf{U}_{\cdot,\Omega_{\lambda'}}^\top \mathbf{U}_{\cdot,\Omega_{\lambda'}^c} \mathbf{A}_{\mathbf{\Omega}_{\lambda'}^{\mathbf{c}},\lambda'}\right) e_{\lambda'}^\top \in \mathbb{R}^{\Lambda_J^\sim \times \Lambda_J},$$
$$\mathsf{Var} := \sum_{\lambda' \in \Lambda_J} \left(E_{\Omega_{\lambda'}}(\mathbf{U}_{\cdot,\Omega_{\lambda'}}^\top \mathbf{U}_{\cdot,\Omega_{\lambda'}})^{-1}\mathbf{U}_{\cdot,\Omega_{\lambda'}}^\top \mathbf{W}_{\cdot,\lambda'}\right) e_{\lambda'}^\top \in \mathbb{R}^{\Lambda_J^\sim \times \Lambda_J}.$$

这里 $\mathsf{OVB}_{\Lambda_J,\cdot}$ 以及 $\mathsf{Var}_{\Lambda_J,\cdot}$ 表示 OVB 和 Var 在指标为 Λ_J 的行上的限制，并且$\mathbf{M}_{(J,t,t')}$ 是与 $\mathrm{supp}(J,\ t,\ t')$ 相关联的指示矩阵，其定义见 (4.4)。

命题 5.9 （误差截断-截断）。 假设假设 *2.1* （*i*）和假设 *3.1*成立。对于 $t,\ t' \geq r/2$,

$$\mathsf{Error\text{-}Truncation} = \|\mathbf{D}^{-t'}(\mathbf{A}_{\Lambda_J}^{\uparrow} - \mathbf{A})\mathbf{D}^{-t}\| \lesssim 2^{-J(t+t'-r)}.$$

命题 5.10 （误差压缩-压缩）。 假设假设 *2.1* （*i*）和假设*3.1*成立。对于 $\mathbf{A}_{\Lambda_J}^{\varepsilon}$（定义见 *4.2*）以及 $t_\ell,\ t_r \in [r/2, \gamma)$ ，

$$\|\mathbf{D}_{\Lambda_J}^{-t_l}(\mathbf{A}_{\Lambda_J}^{\varepsilon} - \mathbf{A}_{\Lambda_J})\mathbf{D}_{\Lambda_J}^{-t_r}\| \lesssim J 2^{-J\left(\min\{t',t_l\}+\min\{t,t_r\}-r\right)}.$$

特别地，对于 $t, t' \geq r/2$,

$$\mathsf{Error\text{-}Compression} = \|\mathbf{D}_{\Lambda_J}^{-t'}(\mathbf{A}_{\Lambda_J}^{\varepsilon} - \mathbf{A}_{\Lambda_J})\mathbf{D}_{\Lambda_J}^{-t}\| \lesssim J\, 2^{-J(t+t'-r)}.$$

命题 5.11 (误差界 误差-OVB)。 在定理 *5.1* 的设定和假设下，以下结论成立。对于任意 $\delta \in (0,1)$，如果

$$N \gtrsim 2^{\widetilde{J}n(\widetilde{t}+\widetilde{t}'-r)/(\sigma-n/2+\widetilde{t}'-r/2)} + \log(1/\delta),$$

那么，以至少 $1-\delta$ 的概率，

$$\text{Error-OVB} \lesssim J\, 2^{-J(t+t'-r)}.$$

命题 5.12 (误差-方差界)。 在定理 *5.1* 的条件和假设下，以下结论成立。对于任意 $\delta \in (0,\ 1)$，若

$$N \gtrsim 2^{\widetilde{J}n(\widetilde{t}+\widetilde{t}'-r)/(\sigma-n/2+\widetilde{t}'-r/2)} + \log(1/\delta),$$

那么，以至少 $1-\delta$ 的概率，

$$\text{Error-Var} \lesssim \sqrt{\frac{J+\log(1/\delta)}{N}} \cdot J \cdot 2^{\rho J(t+t'-r)/2},$$

其中 ρ是式(5.2)中定义的指数。

6 数据驱动偏微分方程求解器的收敛速率

在经典的小波–伽辽金框架中，通常期望有效的矩阵压缩策略能够保持底层 Galerkin 格式的最优收敛阶。在这种情形下，不涉及数据，且算子 $\mathcal{A}$ 完全已知。相比之下，在我们的设定中，算子 $\mathcal{A}$ 未知，必须从数据中学习。这引出了一个问题：$\widehat{}$学习得到的算子 $\mathcal{A}$ 能否为此前未见过的右端项 f. 提供有效的偏微分方程求解器

本节研究将$\widehat{}$学习得到的稀疏算子 $\mathcal{A}$ 代入椭圆伪微分方程后所得数值解的收敛速率。我们证明，该速率$\widehat{}$继承了 $\mathcal{A}$ 的统计精度，与经典小波—伽辽金方法的收敛速率类似，同时也呈现出数据驱动设定所特有的新特点——其中最显著的一点是，在学习阶段，最大小波尺度被选为样本量的函数。这一分析凸显了算子学习对下游偏微分方程求解器的实用价值。

小节 6.1 给出误差界，小节6.2 则包含其证明。

6.1 第二个主要结果

在给出结果之前，我们对定理 $\widehat{}$5.1 中的估计量稍作修改，使所得矩阵估计量 $\mathbf{A}$ 满足 V_J 椭圆性条件，这对于建立数值解的收敛速率至关重要；参见 [45]。

回顾定理 5.1的证明，合并命题 5.11、5.12和5.10，我们证明了估计量 $\widehat{}\mathbf{A} \in \mathbb{R}^{\Lambda_J \times \Lambda_J}$ 满足以下界。令 $\delta \in (0,\ 1)$。若

$$N \gtrsim 2^{\widetilde{J}n(\widetilde{t}+\widetilde{t}'-r)/(\sigma-n/2+\widetilde{t}'-r/2)} + \log(1/\delta),$$

那么，以至少 $1-\delta$ 的概率，

$$\begin{aligned}\|\mathbf{D}_{\Lambda_J}^{-t'}(\widehat{\mathbf{A}} - \mathbf{A}_{\Lambda_J})\mathbf{D}_{\Lambda_J}^{-t}\| &\le \text{Error-OVB} + \text{Error-Var} + \text{Error-Compression} \\ &\lesssim J\, 2^{-J(t+t'-r)} + \sqrt{\frac{J+\log(1/\delta)}{N}} \cdot J \cdot 2^{\rho J(t+t'-r)/2},\end{aligned}$$

其中 ρ 是式（5.2）中定义的指数。

现在我们断言，对定理 5.1 中的估计量稍作修改并调整分析学后，可以得到以下更精细的界：存在某个常数 $\kappa > 0$，使得以至少 $1-\delta$的概率，

$$\|\mathbf{D}_{\Lambda_J}^{-t'}(\widehat{\mathbf{A}} - \mathbf{A}_{\Lambda_J})\mathbf{D}_{\Lambda_J}^{-t}\| \lesssim \varepsilon 2^{-J(t+t'-r)} + \sqrt{\frac{\log(1/\delta)}{N}} \cdot (J/\varepsilon)^{\kappa} \cdot 2^{\rho J(t+t'-r)/2}. \tag{6.1}$$

这里 $\varepsilon > 0$ 是一个可调超参数，而 ρ 与（5.2）中相同。选择 $J = J_*$ 使得

$$\varepsilon 2^{-J_*(t+t'-r)} \asymp \sqrt{\frac{\log(1/\delta)}{N}} \cdot (J_*/\varepsilon)^{\kappa} \cdot 2^{\rho J_*(t+t'-r)/2}, \tag{6.2}$$

因此我们得出结论：若 $N \gtrsim \log(1/\delta)$，则以至少 $1-\delta$ 的概率，

$$\begin{aligned}\|\mathbf{D}_{\Lambda_{J_*}}^{-t'}(\widehat{\mathbf{A}} - \mathbf{A}_{\Lambda_{J_*}})\mathbf{D}_{\Lambda_{J_*}}^{-t}\| &\lesssim \varepsilon 2^{-J_*(t+t'-r)} + \sqrt{\frac{\log(1/\delta)}{N}} \cdot (J_*/\varepsilon)^{\kappa} \cdot 2^{\rho J_*(t+t'-r)/2} \\ &\asymp \varepsilon 2^{-J_*(t+t'-r)} \\ &\asymp_{\mathrm{polylog}} N^{-\frac{1}{2+\rho}}.\end{aligned} \tag{6.3}$$

这里我们使用记号 $\asymp_{\mathrm{polylog}}$ 来隐藏 N 和 $1/\delta$ 中的多对数因子。

注 6.1（增强界的修正（6.1））。我们简要说明如何修改$\widehat{\ }$估计量$\mathbf{A}$ 以获得界（6.1）。我们通过修改定义支撑集^{new}（J，t，t'）中的阈值参数$\tau_{jj'}$ 以及斜率条件，引入一个新的支撑集 supp（J，t，t'），具体做法是将（4.2）中的定义修改如下：

$$\begin{aligned}\mathrm{supp}^{\mathrm{new}}(J,t,t') := \Big\{ &(\lambda,\lambda') \in \Lambda_J \times \Lambda_J : \mathrm{dist}(S_{j,k}, S_{j',k'}) \le \tau_{jj'}^{\mathrm{new}}, \\ &j \le \frac{t+t'-r}{\sigma - \frac{n}{2} + t' - \frac{r}{2}} J + \frac{\sigma - \frac{n}{2} - (t - \frac{r}{2})}{\sigma - \frac{n}{2} + t' - \frac{r}{2}} j' + \frac{\log_2(J/\varepsilon)}{\sigma - \frac{n}{2} + t' - \frac{r}{2}}, \\ &j' \le \frac{t+t'-r}{\sigma - \frac{n}{2} + t - \frac{r}{2}} J + \frac{\sigma - \frac{n}{2} - (t' - \frac{r}{2})}{\sigma - \frac{n}{2} + t - \frac{r}{2}} j + \frac{\log_2(J/\varepsilon)}{\sigma - \frac{n}{2} + t - \frac{r}{2}} \Big\},\end{aligned} \tag{6.4}$$

其中，新的阈值参数选择为

$$\tau_{jj'}^{\mathrm{new}} \asymp \max\left\{ 2^{-\min\{j,j'\}}, 2^{\frac{J(t+t'-r) - jt' - j't - (j+j')\widetilde{d} + \log_2(J/\varepsilon)}{2\widetilde{d}+r}} \right\}. \tag{6.5}$$

$\widehat{\ }$然后我们使用这个更新后的支撑集$\mathbf{A}$ 定义$\mathrm{supp}^{\mathrm{new}}$（$J$，$t$，$t'$）。我们将此修正估计量对应的（6.1）的证明推迟至附录D。□

下一个定理量化了通过将$\mathcal{A}$ 替换为满足（6.3）的学习所得稀疏算子后，偏微分方程解的收敛速率。

定理 6.2。假设假设*2.1* 和*3.1* 成立，并假设算子 $\mathcal{A}$ 未知。令 $f \in H^{-r/2}(M)$，并令 u 为 $\mathcal{A}u = f$的精确解。令 $\widehat{\ }\mathbf{A} \in \mathbb{R}^{\Lambda_{J_*} \times \Lambda_{J_*}}$ 为满足（6.3）的数据驱动稀疏估计量，并令 $\widehat{\ }\mathbf{u}$ 满足

$$\widehat{\mathbf{A}}\widehat{\mathbf{u}} = \widetilde{\mathbf{f}}_{\Lambda_{J_*}},$$

其中，$\widetilde{\ }\mathbf{f}^{\Lambda_{J_*}} := \big(\langle f, \psi_\lambda \rangle\big)_{\lambda \in \Lambda_{J_*}}$ 表示截断至不超过 J_*尺度的 f 小波系数向量。定义

$$\widehat{u} := \widehat{\mathbf{u}}^\top \Psi = \sum_{\lambda \in \Lambda_{J_*}} \widehat{u}_\lambda \psi_\lambda.$$

那么，以至少 $1-\delta$ 的概率，对于任意 $\alpha \in [0, t'-r/2]$,

$$\|\widehat{u}-u\|_{H^{r/2-\alpha}} \lesssim 2^{-J_*(t+\alpha-r/2)}\|u\|_{H^t} \asymp_{\mathrm{polylog}} N^{-\frac{t+\alpha-r/2}{(2+\rho)(t+t'-r)}}\|u\|_{H^t}, \tag{6.6}$$

其中 ρ是 (5.2) 中定义的指数。

备注 6.3（关于定理 6.2 的讨论）。在定理 6.2 中，选择 $\alpha = t'-r/2$ 可得到最快的收敛速率

$$\|\widehat{u}-u\|_{H^{r-t'}} \lesssim 2^{-J_*(t+t'-r)}\|u\|_{H^t} \asymp_{\mathrm{polylog}} N^{-\frac{1}{2+\rho}}\|u\|_{H^t}.$$

这与定理 $\mathcal{A}$ 5.1 中所建立的未知算子 $\widehat{\ }$ 的学习收敛速率相符。因此，数据驱动偏微分方程求解器继承了估计量 $\mathcal{A}$ 的统计精度。

将此结果与压缩小波伽辽金格式的经典确定性理论进行比较很有启发性 [45]。当算子 $\mathcal{A}$（或其小波离散化$\mathbf{A}$）已知精确时，[45, 定理 10.2–10.3] 表明，所提出的矩阵压缩策略保留了伽辽金方法的最优收敛速率：对于 $J > j_0$，若 u_J 是压缩系统的解，则对任意 $\alpha \in [0, d-r/2]$,

$$\|u_J-u\|_{H^{r/2-\alpha}} \lesssim 2^{-J(d+\alpha-r/2)}\|u\|_{H^d}, \tag{6.7}$$

其中 d 是原小波基的逼近阶 Ψ；另见 [45, 引理 3.1]。这表明，在确定性情形下，压缩通过对 $\mathbf{A}$ 进行稀疏化来降低计算成本，同时保持底层 Galerkin 格式的逼近精度。

我们的设定与这一经典框架在几个关键方面有所不同：

1. **未知算子**。 算子 $\mathcal{A}$ 不可用，必须从数据中学习。$\widehat{\ }$估计量 $\widehat{\mathcal{A}}$ 旨在控制误差 $\|\mathcal{A}-\widehat{\mathcal{A}}\|_{H^t\to H^{-t'}}$。因此，数值解 $\widehat{u}$ 不仅反映离散化误差和压缩误差，还反映统计误差（遗漏变量偏差和方差）。由此得到的收敛速率取决于学习阶段使用的样本量 N，而样本量决定了 $\widehat{\mathcal{A}}$逼近 $\mathcal{A}$的$\widehat{\ }$精度。

2. **固定的正则性参数**。 注意，(6.6)的结构形式与(6.7)相似。在我们的设定中，指标 t, t' 固定不变，并决定学习 $\mathcal{A}$ 所采用的度量，进而直接控制偏微分方程解所能达到的精度。相比之下，确定性界(6.7)取决于小波逼近阶 d，而在数据驱动设定中，相关的索博列夫正则性尺度范围由算子学习问题通过选择 t, t'.

3. **数据依赖的分辨率层级**。 最大小波尺度 J_* 根据样本量 N选取，以平衡截断、压缩和统计估计误差。这与经典伽辽金格式形成鲜明对比，在后者中， J 是一个纯粹的数值细化参数，与数据无关。

总之，定理 6.2 表明，即使在数据驱动情形下，算子 $\mathcal{A}$必须经过学习，所得的偏微分方程求解器仍能达到与算子学习近乎最优统计速率相当的收敛速率。此外，界（6.6）呈现出与经典逼近结果（6.7）类似的结构形式，从而将带矩阵压缩的经典小波–伽辽金方法推广到了统计情形。□

6.2 定理 6.2 的证明

定理 *6.2* 的证明。 证明分为三个步骤。在 **步骤 1** 中，我们验证估计量 $\widehat{\mathcal{A}}$ 是 V_{J_*}-椭圆的。基于此，在 **步骤 2** 中，我们通过应用 Strang 的第一 $\alpha = 0$

引理 [32]。在 **步骤 3** 中，我们处理 $\alpha \in (0, t'-r/2]$ 的情形，使用 $\alpha = 0$ 在**步骤 2**中建立的估计。

步骤 1： V_{J_*} **估计量** $\mathcal{A}$**的椭圆性。**

给定 $\widehat{\mathbf{A}} \in \mathrm{R}^{\Lambda_{J_*} \times \Lambda_{J_*}}$，我们将估计量 $\widehat{\mathcal{A}}$ 定义为

$$\widehat{\mathcal{A}} := \sum_{(\lambda,\lambda') \in \Lambda_{J_*} \times \Lambda_{J_*}} \widehat{\mathbf{A}}_{\lambda,\lambda'} \, \widetilde{\psi}_\lambda \otimes \widetilde{\psi}_{\lambda'}. \tag{6.8}$$

由（6.3）可得

$$\begin{aligned} \|\mathbf{D}_{\Lambda_{J_*}}^{-r/2}(\widehat{\mathbf{A}} - \mathbf{A}_{\Lambda_{J_*}})\mathbf{D}_{\Lambda_{J_*}}^{-r/2}\| &= \|\mathbf{D}_{\Lambda_{J_*}}^{t-r/2}\,\mathbf{D}_{\Lambda_{J_*}}^{-t}(\widehat{\mathbf{A}} - \mathbf{A}_{\Lambda_{J_*}})\mathbf{D}_{\Lambda_{J_*}}^{-t}\,\mathbf{D}_{\Lambda_{J_*}}^{t-r/2}\| \\ &\le \|\mathbf{D}_{\Lambda_{J_*}}^{t-r/2}\| \cdot \|\mathbf{D}_{\Lambda_{J_*}}^{-t}(\widehat{\mathbf{A}} - \mathbf{A}_{\Lambda_J})\mathbf{D}_{\Lambda_{J_*}}^{-t}\| \cdot \|\mathbf{D}_{\Lambda_{J_*}}^{t-r/2}\| \\ &\lesssim 2^{2J_*(t-r/2)} \cdot \varepsilon 2^{-J_*(2t-r)} = \varepsilon. \end{aligned}$$

此外，

$$\|\mathbf{D}_{\Lambda_{J_*}}^{-t}(\widehat{\mathbf{A}} - \mathbf{A}_{\Lambda_{J_*}})\mathbf{D}_{\Lambda_{J_*}}^{-r/2}\| \le \|\mathbf{D}_{\Lambda_{J_*}}^{-t}(\widehat{\mathbf{A}} - \mathbf{A}_{\Lambda_{J_*}})\mathbf{D}_{\Lambda_J}^{-t}\| \cdot \|\mathbf{D}_{\Lambda_{J_*}}^{t-r/2}\| \lesssim \varepsilon 2^{-J_*(t-r/2)}.$$

根据命题 3.2 (II) (i)，存在 $c_- > 0$ 使得 $\sigma_{\min}(\mathbf{D}^{-r/2}\mathbf{A}\mathbf{D}^{-r/2}) \ge c_-$. 由于 $\mathbf{A}^{\Lambda_{J_*}}$ 的谱包含于 $\mathbf{A}$的谱中，因此可得

$$\sigma_{\min}(\mathbf{D}_{\Lambda_{J_*}}^{-r/2}\mathbf{A}_{\Lambda_{J_*}}\mathbf{D}_{\Lambda_{J_*}}^{-r/2}) \ge \sigma_{\min}(\mathbf{D}^{-r/2}\mathbf{A}\mathbf{D}^{-r/2}) \ge c_-.$$

那么，通过选取 $\varepsilon > 0$ 足够小，我们得到（ V_{J_*}-椭圆性，见 $\mathcal{A}$ （6.8））：对于任意 $u_{J_*} \in V_{J_*} \subset H^{r/2}(M)$,

$$\begin{aligned} &\langle \widehat{\mathcal{A}} u_{J_*}, u_{J_*} \rangle = \langle \widehat{\mathbf{A}}\mathbf{u}_{J_*}, \mathbf{u}_{J_*} \rangle \\ &= \langle \mathbf{D}_{\Lambda_{J_*}}^{-r/2}\mathbf{A}_{\Lambda_{J_*}}\mathbf{D}_{\Lambda_{J_*}}^{-r/2}\mathbf{D}_{\Lambda_{J_*}}^{r/2}\mathbf{u}_{J_*}, \mathbf{D}_{\Lambda_{J_*}}^{r/2}\mathbf{u}_{J_*} \rangle - \langle \mathbf{D}_{\Lambda_{J_*}}^{-r/2}(\mathbf{A}_{\Lambda_{J_*}} - \widehat{\mathbf{A}})\mathbf{D}_{\Lambda_{J_*}}^{-r/2}\mathbf{D}_{\Lambda_{J_*}}^{r/2}\mathbf{u}_{J_*}, \mathbf{D}_{\Lambda_{J_*}}^{r/2}\mathbf{u}_{J_*} \rangle \\ &\ge \sigma_{\min}(\mathbf{D}_{\Lambda_{J_*}}^{-r/2}\mathbf{A}_{\Lambda_{J_*}}\mathbf{D}_{\Lambda_{J_*}}^{-r/2}) \cdot \|\mathbf{D}_{\Lambda_{J_*}}^{r/2}\mathbf{u}_{J_*}\|^2 - \|\mathbf{D}_{\Lambda_{J_*}}^{-r/2}(\widehat{\mathbf{A}} - \mathbf{A}_{\Lambda_{J_*}})\mathbf{D}_{\Lambda_{J_*}}^{-r/2}\| \cdot \|\mathbf{D}_{\Lambda_{J_*}}^{r/2}\mathbf{u}_{J_*}\|^2 \\ &\ge (c_- - \varepsilon)\|\mathbf{D}_{\Lambda_{J_*}}^{r/2}\mathbf{u}_{J_*}\|^2 \\ &\asymp \|u_{J_*}\|_{H^{r/2}}, \end{aligned} \tag{6.9}$$

其中最后一步由引理 A.1 (iv) 和 $r/2 \in (-\widetilde{\gamma}, \gamma)$得出。

步骤 2： $\alpha = 0$。我们先证明 $\alpha = 0$ 的情形。利用（6.9），应用 Strang 第一引理 [32,定理 4.1。1]可得

$$\|\widehat{u} - u\|_{H^{r/2}} \lesssim \inf_{v_{J_*} \in V_{J_*}} \left\{ \|v_{J_*} - u\|_{H^{r/2}} + \sup_{w_J \in V_{J_*}, w_{J_*} \neq 0} \frac{|\langle (\widehat{\mathcal{A}} - \mathcal{A})v_{J_*}, w_{J_*} \rangle|}{\|w_{J_*}\|_{H^{r/2}}} \right\}. \tag{6.10}$$

由（6.3），可得

$$|\langle (\widehat{\mathcal{A}} - \mathcal{A})v_{J_*}, w_{J_*} \rangle| \lesssim \varepsilon 2^{-J_*(t+t'-r)} \|v_{J_*}\|_{H^t} \|w_{J_*}\|_{H^{t'}}.$$

此外，引理 A.1 (ii) 表明，对于 $r/2 < t' \le \gamma$,

$$\|w_{J_*}\|_{H^{t'}} \lesssim 2^{J_*(t'-r/2)} \|w_{J_*}\|_{H^{r/2}}.$$

因此，将 $v_{J_*} = Q_{J_*}u \in V_{J_*}$ 代入（6.10）可得

$$\begin{aligned}\|\widehat{u}-u\|_{H^{r/2}} &\lesssim \|u-Q_{J_*}u\|_{H^{r/2}} + \sup_{w_{J_*}\in V_{J_*}, w_{J_*}\neq 0} \frac{|\langle(\widehat{\mathcal{A}}-\mathcal{A})Q_{J_*}u, w_{J_*}\rangle|}{\|w_{J_*}\|_{H^{r/2}}} \\ &\lesssim \|u-Q_{J_*}u\|_{H^{r/2}} + 2^{-J_*(t+t'-r)}\cdot 2^{J_*(t'-r/2)}\|Q_{J_*}u\|_{H^t} \\ &\lesssim 2^{-J_*(t-r/2)}\|u\|_{H^t},\end{aligned}$$

其中最后一步由以下事实得到：$\|Q_{J_*}u\|_{H^t} \le \|u\|_{H^t}$，以及 $\|u-Q_{J_*}u\|_{H^{r/2}} \lesssim 2^{-J_*(t-r/2)}\|u\|_{H^t}$，依据~引理 A.1 (ii)，只要满足 $-d \le r/2 < t \le d$、$r/2 < \gamma$，以及 $-\widetilde{\gamma} < t$。

步骤 3： $\alpha \in (0, t'-r/2]$。我们从以下内容开始

$$\|\widehat{u}-u\|_{H^{r/2-\alpha}} = \sup_{g\in H^{\alpha-r/2}} \frac{|\langle \widehat{u}-u, g\rangle|}{\|g\|_{H^{\alpha-r/2}}}.$$

由于 $\mathcal{A}: H^{\alpha+r/2} \to H^{\alpha-r/2}$ 是一个同构，令 $v \in H^{\alpha+r/2}$ 满足 $\mathcal{A}v = g$，则

$$\|\widehat{u}-u\|_{H^{r/2-\alpha}} = \sup_{v\in H^{\alpha+r/2}} \frac{|\langle \widehat{u}-u, \mathcal{A}v\rangle|}{\|\mathcal{A}v\|_{H^{\alpha-r/2}}} \asymp \sup_{v\in H^{\alpha+r/2}} \frac{|\langle \mathcal{A}(\widehat{u}-u), v\rangle|}{\|v\|_{H^{\alpha+r/2}}}. \tag{6.11}$$

~利用正交性 $\langle \mathcal{A}\widehat{u}, Q_{J_*}v\rangle = \langle Q_{J_*}f, Q_{J_*}v\rangle = \langle f, Q_{J_*}v\rangle = \langle \mathcal{A}u, Q_{J_*}v\rangle$，我们可以分解

$$\begin{aligned}\langle \mathcal{A}(\widehat{u}-u), v\rangle &= \langle \mathcal{A}(\widehat{u}-u), v-Q_{J_*}v\rangle + \langle \mathcal{A}(\widehat{u}-u), Q_{J_*}v\rangle \\ &= \langle \mathcal{A}(\widehat{u}-u), v-Q_{J_*}v\rangle + \langle (\mathcal{A}-\widehat{\mathcal{A}})\widehat{u}, Q_{J_*}v\rangle.\end{aligned} \tag{6.12}$$

(6.12) 右端项中的第一项有如下界

$$|\langle \mathcal{A}(\widehat{u}-u), v-Q_{J_*}v\rangle| \lesssim \|\widehat{u}-u\|_{H^{r/2}}\|v-Q_{J_*}v\|_{H^{r/2}} \lesssim 2^{-J_*(t+\alpha-r/2)}\|u\|_{H^t}\|v\|_{H^{\alpha+r/2}}, \tag{6.13}$$

其中，在第二个不等式中，我们使用了在~步骤 2 中得到的界 $\|\widehat{u}-u\|_{H^{r/2}} \lesssim 2^{-J_*(t-r/2)}\|u\|_{H^t}$，以及由引理 A.1 (ii) 得到的 $\|v-Q_{J_*}v\|_{H^{r/2}} \lesssim 2^{-J_*\alpha}\|v\|_{H^{\alpha+r/2}}$，前提是 $-d \le r/2 \le \alpha + r/2 \le d$，$r/2 < \gamma$ 和 $-\widetilde{\gamma} < \alpha + r/2$。这些条件由 $r/2 \in (-\widetilde{\gamma}, \gamma)$，$\alpha \in (0, t'-r/2]$ 和 $t' \le d$ 保证。

对于 (6.12) 中的第二项，

$$\begin{aligned}&|\langle (\widehat{\mathcal{A}}-\mathcal{A})\widehat{u}, Q_{J_*}v\rangle| \\ &\le |\langle (\widehat{\mathcal{A}}-\mathcal{A})(\widehat{u}-Q_{J_*}u), Q_{J_*}v\rangle| + |\langle (\widehat{\mathcal{A}}-\mathcal{A})Q_{J_*}u, Q_{J_*}v\rangle| \\ &\le \|\widehat{\mathcal{A}}-\mathcal{A}\|_{H^{r/2}\to H^{-t'}}\|\widehat{u}-Q_{J_*}u\|_{H^{r/2}}\|Q_{J_*}v\|_{H^{t'}} + \|\widehat{\mathcal{A}}-\mathcal{A}\|_{H^t\to H^{-t'}}\|Q_{J_*}u\|_{H^t}\|Q_{J_*}v\|_{H^{t'}} \\ &\lesssim 2^{-J_*(t'-r/2)}\cdot 2^{-J_*(t-r/2)}\cdot 2^{J_*(t'-\alpha-r/2)}\|u\|_{H^t}\|v\|_{H^{\alpha+r/2}} \\ &\quad + 2^{-J_*(t+t'-r)}\cdot 2^{J_*(t'-\alpha-r/2)}\|u\|_{H^t}\|v\|_{H^{\alpha+r/2}} \\ &\asymp 2^{-J_*(t+\alpha-r/2)}\|u\|_{H^t}\|v\|_{H^{\alpha+r/2}},\end{aligned} \tag{6.14}$$

其中我们用到了以下不等式：

$$\|\widehat{\mathcal{A}}-\mathcal{A}\|_{H^{r/2}\to H^{-t'}} \lesssim 2^{-J_*(t'-r/2)}, \qquad \|\widehat{\mathcal{A}}-\mathcal{A}\|_{H^t\to H^{-t'}} \lesssim 2^{-J_*(t+t'-r)},$$

$$\|Q_{J_*}v\|_{H^{t'}} \lesssim 2^{J_*(t'-\alpha-r/2)}\|v\|_{H^{\alpha+r/2}}, \quad \text{if } \alpha + r/2 \le t' \le \gamma,$$

and

$$\|\widehat{u}-Q_{J_*}u\|_{H^{r/2}} \le \|\widehat{u}-u\|_{H^{r/2}} + \|u-Q_{J_*}u\|_{H^{r/2}} \lesssim 2^{-J_*(t-r/2)}\|u\|_{H^t}.$$

合并（6.11）、（6.12）、（6.13）和（6.14）可得，对于任意 $\alpha \in (0, t'-r/2]$,

$$\|\widehat{u}-u\|_{H^{r/2-\alpha}} \asymp \sup_{v\in H^{\alpha+r/2}} \frac{|\langle \mathcal{A}(\widehat{u}-u), v\rangle|}{\|v\|_{H^{\alpha+r/2}}} \lesssim 2^{-J_*(t+\alpha-r/2)}\|u\|_{H^t} \asymp_{\mathrm{polylog}} N^{-\frac{t+\alpha-r/2}{(2+\rho)(t+t'-r)}}\|u\|_{H^t},$$

其中最后一步由 J_* 在（6.2）中的取值可得。证明至此完成。 □

7 结论、讨论与未来方向

本文建立了从含噪（及无噪）数据中学习椭圆伪微分算子的收敛速率。在小波–伽辽金框架下，我们将学习任务表述为一个具有多尺度稀疏性的结构化无限维回归问题。基于这一结构，我们提出了一种稀疏且兼顾数据与计算效率的估计量，它结合了面向学习的矩阵压缩方案与嵌套支撑回归策略，以平衡逼近误差和估计误差。除了获得该估计量的收敛速率外，我们还证明，学习得到的算子可导出高效且稳定的伽辽金求解器，其数值误差继承了该估计量的统计精度。因此，我们的结果有助于在科学计算中结合算子学习、数据驱动求解器与小波方法。

W我们最后总结了本研究引出的若干开放问题和未来方向 .

椭圆性、数据/噪声分布、计算与误差度量 本文假设 $\mathcal{A}$ 是强椭圆的，而命题 3.2中的小波坐标估计（I）对椭圆类以外的偏微分算子也更普遍地成立。放宽椭圆性假设，并弄清学习任务所需的较弱条件，将很有意义。为简化表述，我们还聚焦于输入和噪声均为高斯随机函数的情形。由于我们的误差分析主要依赖于命题 3.2 （II）（ii）–（iii）中的协方差性质，因此可以自然地预期，估计过程及其带来的收敛保证都能推广到非高斯情形——例如，输入分布和噪声分布满足适当的有限矩（或次高斯/次指数）假设的情况。在计算方面，我们在备注 5.5 末尾指出，若能在保持相同统计速率的同时，将运行时间改进至精确的 $\mathcal{O}(N2^{J_n})$，将很有意义。最后，关于误差度量，本文始终使用算子范数 $\|\cdot\|_{H^t\to H^{-t'}}$衡量估计精度。也可以考虑测试分布下的预测/泛化准则。例如，对于高斯测试输入 $v\sim\mathcal{N}(\mathbf{0},\ \mathcal{C}_v)$，可以研究

$$\mathbb{E}_v\big\|(\widehat{\mathcal{A}}-\mathcal{A})v\big\|_{H^{-t'}}.$$

~假设小波表示 $\mathbf{C}_v$ 的 $\mathcal{C}_v$ 满足命题 3.2 （II）（ii）的类似条件，即对于某个 $t\in\mathbb{R}$,

$$c_- \le \sigma_{\min}\big(\mathbf{D}^t\widetilde{\mathbf{C}}_v\mathbf{D}^t\big) \le \sigma_{\max}\big(\mathbf{D}^t\widetilde{\mathbf{C}}_v\mathbf{D}^t\big) \le c_+.$$

则通过标准的协方差计算可得

$$\begin{aligned}\mathbb{E}_v\big\|(\widehat{\mathcal{A}}-\mathcal{A})v\big\|_{H^{-t'}}^2 &= \mathrm{Tr}\Big(\mathbf{D}^{-t'}(\widehat{\mathbf{A}}-\mathbf{A})\mathbf{D}^{-t}\big(\mathbf{D}^t\widetilde{\mathbf{C}}_v\mathbf{D}^t\big)\big(\mathbf{D}^{-t'}(\widehat{\mathbf{A}}-\mathbf{A})\mathbf{D}^{-t}\big)^\top\Big)\\ &\le \big\|\mathbf{D}^t\widetilde{\mathbf{C}}_v\mathbf{D}^t\big\|\,\|\mathbf{D}^{-t'}(\widehat{\mathbf{A}}-\mathbf{A})\mathbf{D}^{-t}\|_{\mathrm{HS}}^2\\ &\lesssim \|\mathbf{D}^{-t'}(\widehat{\mathbf{A}}-\mathbf{A})\mathbf{D}^{-t}\|_{\mathrm{HS}}^2,\end{aligned}$$

其中，$\|\cdot\|_{\mathrm{HS}}$ 表示算子 $\ell^2(\mathcal{J})\to\ell^2(\mathcal{J})$上的希尔伯特–施密特范数。本文提出的多尺度技术也可用于在此类希尔伯特–施密特型准则下获得收敛速率；详细讨论留待未来工作。

自适应算子估计与学习 未来研究的一个重要方向是 自适应算子估计与学习。我们的学习方法背后的指导原则之一，是能够获得 $\mathcal{A}$（命题 3.2 （I））的小波坐标衰减估计，据此设计面向学习的压缩与回归过程，利用多尺度稀疏性，并致力于仅估计显著元素，从而平衡逼近（偏差）与估计（方差）误差。本文始终假设 $\mathcal{A}$ 的阶已知。因此，一个自然的问题是，能否开发出适用于未知阶、进而适用于未知稀疏模式的方法。具体而言，能否像定义 4.2中那样，不通过显式设定阈值参数$\tau_{jj'}$ 和斜率条件来选择支撑集，而是以自适应的方式从数据中学习有效支撑（即显著坐标）？

相关问题在科学计算和数值分析学中颇为经典，其中自适应方法——包括自适应小波格式 [36, 33, 40, 131] 和自适应有限元方法 [110, 8, 12, 130, 13]——在实现偏微分方程解的最优有限项逼近和稀疏表示方面发挥着核心作用。在算子学习的背景下，探索与未知算子结构相适应的类似自适应性概念，并研究适用于算子类的非线性逼近概念，是很自然的。

从统计学角度来看，稀疏性诱导正则化，尤其是 Lasso 及其变体，在现代高维统计学中占据核心地位 [137, 152, 21, 141,151]。对于本文所研究的算子学习问题，若能建立涵盖变量选择、支撑恢复和估计的稀疏预言机不等式，将会非常有意义。

学习其他结构化算子类本文通过利用先验结构，建立了在小波坐标下学习椭圆偏微分算子的收敛速率；该结构可产生多尺度可压缩性，并使无限维情形下的基于稀疏性的估计成为可能。本文发展的方法论和分析学可能扩展到具有类似结构化表示的其他算子类。例如，傅里叶积分算子（傅里叶积分算子）在波传播和双曲型偏微分方程中发挥着核心作用 [73, 138, 63, 75, 81, 148, 142], ，而曲波为广泛的傅里叶积分算子族提供了近乎最优的稀疏表示 [25, 26, 22, 24, 23], ，这表明在适当的曲波坐标下，类似的多尺度、基于稀疏性的方法或许是可行的。相关的结构化算子也出现在抛物型问题中 [140, 123, 28]。更广泛地说，也是更具挑战性的是，人们可以尝试学习结构化非线性算子；一个典型例子是线性二阶散度型椭圆偏微分方程的参数到解映射[39, 41, 118]。

算子学习的贝叶斯表述 我们已经提到， [54] 在贝叶斯框架下，为已知基中可对角化的线性算子学习建立了后验收缩速率。更广泛地说，贝叶斯表述能够量化所学习算子的不确定性，并将其传播至下游数值预测，从而为偏微分方程解和相关泛函构造后验可信集。一个关键问题是，能否利用支撑我们频率学派速率的同一多尺度可压缩性，为结构化算子类设计计算上可处理的先验（例如小波坐标中的稀疏性先验或基于收缩的先验），并建立相应的后验收缩保证。我们认为，构建一个通用的算子学习贝叶斯框架——涵盖对角化情形以外更广泛的结构化算子类，并给出收缩速率（理想情况下，还能保证可信集具有频率学派覆盖性质）——是未来工作中一个很有前景的方向。

致谢

作者的研究部分得到美国国家科学基金会 CAREER 奖 DMS-2237628 的资助。作者感谢 Omar Al-Ghattas 对稿件提出的宝贵反馈和意见。

参考文献

[1] *B. Adcock*、*G. Maier* 和 *R. Parhi*，迈向算子学习的尖锐极小极大风险界，*arXiv* 预印本 *arXiv:2512.17805*，（*2025*）。[2] *O. Al-Ghattas*、*J. Chen*、*D. Sanz-Alonso* 和 *N. Waniorek*，结构化协方差算子的最优估计，*arXiv* 预印本 *arXiv:2408.02109*，（*2024*）。[3] ,协方差算子估计：稀疏性、长度尺度与集合卡尔曼滤波器，*Bernoulli*，*31*（*2025*），第 *2377–2402* 页。[4] *O. Al-Ghattas* 和 *D. Sanz-Alonso*，通过自适应阈值化进行协方差算子估计，*Stochastic Processes and their Applications*，（*2025*），第 *104705* 页。[5] *A. Anandkumar*、*K. Azizzadenesheli*、*K. Bhattacharya*、*N. Kovachki*、*Z. Li*、*B. Liu* 和 *A. Stuart*，神经算子：用于偏微分方程的图核网络，载于 *ICLR 2020* 深度神经模型与微分方程集成研讨会，*2020*。[6] *M. Bachmayr* 和 *A. Cohen*，随机场的多层表示与随机偏微分方程解的稀疏近似，载于《多尺度、非线性与自适应近似 *II*》，施普林格，*2024*，第 *25–54* 页。[7] *E. Bacry*、*S. Mallat* 和 *G. Papanicolaou*，一种基于小波的偏微分方程时空自适应数值方法，*ESAIM*：数学建模与数值分析（*M2AN*），*26*（*1992*），第 *793–834* 页。[8] *W. Bangerth* 和 *R. Rannacher*，微分方程的自适应有限元方法，施普林格科学与商业媒体，*2003*。[9] *H. Barreto* 和 *F. Howland*，《计量经济学导论：使用 *Microsoft Excel* 进行蒙特卡洛模拟》，剑桥大学出版社，*2005*。[10] *P. Batlle*、*M. Darcy*、*B. Hosseini* 和 *H. Owhadi*，核方法在算子学习方面具有竞争力，《计算物理学杂志》，*496*（*2024*），第 *112549* 页。[11] *G. Beylkin*、*R. Coifman* 和 *V. Rokhlin*，快速小波变换与数值算法 *I*，《纯粹与应用数学通讯》，*44*（*1991*），第 *141–183* 页。[12] *P. Binev*、*W. Dahmen* 和 *R. DeVore*，具有收敛速率的自适应有限元方法，数值数学，*97*（*2004*），第 *219–268* 页。[13] *A. Bonito*、*C. Canuto*、*R. H. Nochetto* 和 *A. Veeser*，自适应有限元方法，数值学报，*33*（*2024*），第 *163–485* 页。[14] *N. Boullé*、*D. Halikias* 和 *A. Townsend*，椭圆型偏微分方程学习具有可证明的数据效率，美国国家科学院院刊，*120*（*2023*），第 *e2303904120* 页。[15] *N. Boullé*、*S. Kim*、*T. Shi* 和 *A. Townsend*，学习含时偏微分方程相关的格林函数，机器学习研究杂志，*23*（*2022*），第 *1–34* 页。[16] *N. Boullé* 和 *A. Townsend*，利用随机化线性代数学习椭圆型偏微分方程，《计算数学基础》，*23*（*2023*），第 *709–739* 页。[17] ，算子学习数学指南，载于 *Handbook of Numerical Analysis*，第 *25* 卷，爱思唯尔，*2024*，第 *83–125* 页。[18] *J. Bruna* 和 *S. Mallat*，不变散射卷积网络，*IEEE*模式分析与机器智能汇刊，*35*（*2013*），第 *1872–1886* 页。[19] *S. L. Brunton*、*M. Budišić*、*E. Kaiser* 和 *J. N. Kutz*，动力系统的现代 *Koopman* 理论，*SIAM*评论，（*2022*）。

[20] T. A. Bubba、M. Galinier、M. Lassas、M. Prato、L. Ratti 和 S. Siltanen，用于伪微分算子逆问题的深度神经网络：有限角度断层成像中的应用，《SIAM 成像科学杂志》，14（2021），第 470–505 页。[21] P. Bühlmann 和 S. Van De Geer，高维数据统计学：方法、理论与应用，施普林格科学与商业媒体，2011。[22] E. Candès 和 L. Demanet，曲波与傅里叶积分算子，《数学报告》，336（2003），第 395–398 页。[23] E. Candès、L. Demanet、D. Donoho 和 L. Ying，快速离散曲波变换，《多尺度建模与仿真》，5（2006），第 861–899 页。[24] E. J. Candès 和 L. Demanet，波传播算子的曲波表示是最优稀疏的，《纯粹与应用数学通讯》，58（2005），第 1472–1528 页。[25] E. J. Candès 和 D. L. Donoho，曲波——一种对具有边缘的对象出奇有效的非自适应表示，载于《曲线与曲面拟合：圣马洛 1999》，A. Cohen、C. Rabut 和 L. L. Schumaker 编，范德比尔特大学出版社，纳什维尔，田纳西州，2000，第 105–120 页。[26] E. J. Candès 和 D. L. Donoho，新的曲波紧框架及具有分片 C^2 奇异性的对象的最优表示，《纯粹与应用数学通讯：库朗数学科学研究所发行的期刊》，57（2004），第 219–266 页。[27] E. J. Candès 和 M. B. Wakin，压缩采样简介，《IEEE信号处理杂志》，25（2008），第 21–30 页。[28] N. Chegini 和 R. Stevenson，抛物型问题的自适应小波格式：稀疏矩阵与数值结果，《SIAM 数值分析杂志》，49（2011），第 182–212 页。[29] J. Chen 和 D. Sanz-Alonso，高斯过程的精度与乔列斯基因子估计，《SIAM/ASA 不确定性量化杂志》，13（2025），第 1085–1115 页。[30] K. Chen、J. Lai 和 C. Wang，用于逆偏微分方程算子的伪微分积分自编码器网络，《逆问题》，40（2024），第 105004 页。[31] T. Chen 和 H. Chen，具有任意激活函数的神经网络对非线性算子的通用逼近及其在动力系统中的应用，《IEEE 神经网络汇刊》，6（1995），第 911–917 页。[32] P. G. Ciarlet，椭圆型问题的有限元法，工业与应用数学学会，2002。[33] Cohen、Dahmen 和 DeVore，自适应小波方法 *II*——超越椭圆情形，《计算数学基础》，2（2002），第 203–245 页。[34] A. Cohen，数值分析中的小波方法，《数值分析手册》，7（2000），第 417–711 页。[35] ，《小波方法的数值分析》，《数学及其应用研究》第 32 卷，爱思唯尔，阿姆斯特丹，2003。[36] A. Cohen、W. Dahmen 和 R. DeVore，椭圆算子方程的自适应小波方法：收敛速率，《计算数学》，70（2001），第 27–75 页。[37] , 数值模拟中的自适应小波技术，《计算力学百科全书》，1（2004），第 157–197 页。[38] A. Cohen、I. Daubechies 和 J.-C. Feauveau，紧支撑小波的双正交基，《纯粹与应用数学通讯》，45 （1992），第 485–560 页。

[39] *A. Cohen* 和 *R. DeVore*，高维参数偏微分方程的逼近，《数值学报》，*24*（*2015*），第 *1–159* 页。[40] *A. Cohen*、*M. Hoffmann* 和 *M. Reiss*，线性逆问题的自适应小波伽辽金方法，《*SIAM* 数值分析杂志》，*42*（*2004*），第 *1479–1501* 页。[41] *A. Cohen* 和 *G. Migliorati*，具有对数正态系数的椭圆型偏微分方程的近最优逼近方法，《计算数学》，*92*（*2023*），第 *1665–1691* 页。[42] *M. J. Colbrook* 和 *A. Townsend*，动力系统 *Koopman* 算子谱性质的严格数据驱动计算，《纯粹与应用数学通讯》，*77*（*2024*），第 *221–283* 页。[43] *S. G. Cox* 和 *K. Kirchner*，广义 *Whittle–Matérn* 场在 *Sobolev* 和 *Hölder* 空间中的正则性与收敛性分析，《数值数学》，*146*（*2020*），第 *819–873* 页。[44] *W. Dahmen*，算子方程的小波与多尺度方法，《数值学报》，*6*（*1997*），第 *55–228* 页。[45] *W. Dahmen*、*H. Harbrecht* 和 *R. Schneider*，边界积分方程的压缩技术——渐近最优复杂度估计，《*SIAM* 数值分析杂志》，*43*（*2006*），第 *2251–2271* 页。[46] *W. Dahmen* 和 *A. Kunoth*，多层预条件，《数值数学》，*63*（*1992*），第 *315–344* 页。[47] *W. Dahmen*、*A. Kunoth* 和 *K. Urban*，区间上的双正交样条小波——稳定性与矩条件，《应用与计算调和分析》，*6*（*1999*），第 *132–196* 页。[48] *W. Dahmen*、*S. Prößdorf* 和 *R. Schneider*，伪微分方程的小波逼近方法 *II*：矩阵压缩与快速求解，《计算数学进展》，*1*（*1993*），第 *259–335* 页。[49] *W. Dahmen*、*S. Prössdorf* 和 *R. Schneider*，光滑闭流形上伪微分方程的多尺度方法，载于《小波分析及其应用》，第 *5* 卷，爱思唯尔，*1994*，第 *385–424* 页。[50] *W. Dahmen*、*S. Prössdorf* 和 *R. Schneider*，伪微分方程的小波逼近方法：*I.* 稳定性与收敛性，《数学杂志》，*215*（*1994*），第 *583–620* 页。[51] *W. Dahmen* 和 *R. Schneider*，流形上的小波 *I*：构造与区域分解，《*SIAM* 数学分析杂志》，*31*（*1999*），第 *184–230* 页。[52] *I. Daubechies* 编，《小波的不同视角：美国数学会短期课程，*1993* 年 *1* 月 *11–12* 日，圣安东尼奥，得克萨斯州》，《应用数学专题讨论会论文集》第 *47* 卷，美国数学会，普罗维登斯，罗得岛州，*1993*。[53] *M. V. de Hoop*、*N. B. Kovachki*、*M. Lassas* 和 *N. H. Nelsen*，电阻抗断层成像逆映射的扩展与神经算子逼近，*arXiv* 预印本 *arXiv:2511.20361*，（*2025*）。[54] *M. V. de Hoop*、*N. B. Kovachki*、*N. H. Nelsen* 和 *A. M. Stuart*，从含噪数据中学习线性算子的收敛速率，《*SIAM/ASA* 不确定性量化杂志》，*11*（*2023*），第 *480–513* 页。[55] *B. Deng*、*Y. Shin*、*L. Lu*、*Z. Zhang* 和 *G. E. Karniadakis*，由对流–扩散方程产生的算子的 *DeepONets* 学习逼近率，《神经网络》，*153*（*2022*），第 *411–426* 页。[56] *R*

. A. DeVore ，非线性逼近 ，数值学报 ，7 (1998)，pp. 51–150 .

[57] R. A. DeVore、B. Jawerth 和 V. Popov，小波分解的压缩，《美国数学杂志》，114（1992），第 737–785 页。[58] D. L. Donoho，通过软阈值处理去噪，《IEEE 信息论汇刊》，41（2002），第 613–627 页。[59] , 压缩感知，《IEEE 信息论汇刊》，52（2006），第 1289–1306 页。[60] D. L. Donoho 和 I. M. Johnstone，通过小波收缩实现理想的空间自适应;—《生物统计学》，81（1994），第 425–455 页。[61] , 通过小波收缩适应未知光滑性，《美国统计学会会刊》，90（1995），第 1200–1224 页。[62] D. L. Donoho、I. M. Johnstone、G. Kerkyacharian 和 D. Picard，通过小波阈值处理进行密度估计，《统计年刊》（1996），第 508–539 页。[63] J. J. Duistermaat，傅里叶积分算子，《数学进展》第 130 卷，Birkhäuser，1996。[64] Y. Fan、C. O. Bohorquez 和 L. Ying，*Bcr-net*：一种基于非标准小波形式的神经网络，《计算物理学杂志》，384（2019），第 1–15 页。[65] Z. Gao、L. Yan 和 T. Zhou，面向无限维贝叶斯反问题的自适应算子学习，SIAM/ASA 不确定性量化杂志，12（2024），第 1389–1423 页。[66] G. Gupta、X. Xiao 和 P. Bogdan，用于微分方程的基于多小波的算子学习，Advances in Neural Information Processing Systems，34（2021），第 24048–24062 页。[67] N. Gupta、S. Sivananthan 和 B. K. Sriperumbudur，采用一般正则化的函数型线性回归的最优收敛率，《应用与计算调和分析》，76（2025），第 101745 页。[68] H. Harbrecht、L. Herrmann、K. Kirchner 和 C. Schwab，高斯随机场的多层逼近：协方差压缩、估计与空间预测，《计算数学进展》，50（2024），第 101 页。[69] H. Harbrecht 和 M. Multerer，小波坐标下非局部算子的快速直接求解器，《计算物理学杂志》，428（2021），第 110056 页。[70] H. Harbrecht 和 R. Schneider，边界积分方程的小波 *Galerkin* 格式——实现与求积，SIAM Journal on Scientific Computing，27（2006），第 1347–1370 页。[71] A. Harten，用于双曲守恒律数值求解的多分辨率算法，《纯粹与应用数学通讯》，48（1995），第 1305–1342 页。[72] L. Herrmann、K. Kirchner 和 C. Schwab，高斯随机场的多层逼近：快速模拟，《应用科学中的数学模型与方法》，30（2020），第 181–223 页。[73] L. Hörmander，傅里叶积分算子 *I*，《数学学报》，127（1971），第 79–183 页。[74] ,线性偏微分算子分析 *III*：伪微分算子，《数学经典》，施普林格，柏林、海德堡，2007。[75] ,线性偏微分算子分析 *IV*：傅里叶积分算子，《数学经典》，施普林格，柏林、海德堡，2009。[76] S. Jaffard，用于快速求解椭圆型问题的小波方法，SIAM 数值分析杂志，29（1992），第 965–986 页。

[77] *Y. Jalalian, J. F. O. Ramirez, A. Hsu, B. Hosseini,* 和 *H. Owhadi*，《用于学习微分方程及其解算子的数据高效核方法：算法与误差分析》，*arXiv* 预印本 *arXiv:2503.01036*，（*2025*）。[78] *J. Jin, Y. Lu, J. Blanchet,* 和 *L. Ying*，《通过多层级训练实现极小极大最优核算子学习》，载于第十一届国际学习表征会议，*2022*。[79] *L. Kang* 和 *M. Xu*，《高维输入下高斯过程模型的最优核学习》，*arXiv* 预印本 *arXiv:2502.16617*，（*2025*）。[80] *G. E. Karniadakis, I. G. Kevrekidis, L. Lu, P. Perdikaris, S. Wang,* 和 *L. Yang*，《物理信息机器学习》，*Nature Reviews Physics*，*3*（*2021*），第*422–440*页。[81] *Y. Khoo* 和 *L. Ying*，《*Switchnet*：用于正向和逆散射问题的神经网络模型》，*SIAM Journal on Scientific Computing*，*41*（*2019*），第*A3182–A3201*页。[82] *V. Koltchinskii* 和 *K. Lounici*，《样本协方差算子的集中不等式与矩界》，*Bernoulli*，*23*（*2017*），第*110–133*页。[83] *M. Korte-Stapff, T. Karvonen,* 和 *É. Moulines*，《闭黎曼流形上 *Whittle-Matérn*过程的光滑性估计》，*Stochastic Processes and their Applications*，（*2025*），第*104685*页。[84] *V. Kostic, H. Halconruy, T. Devergne, K. Lounici,* 和 *M. Pontil*，《学习随机扩散过程的无穷小生成元》，*Advances in Neural Information Processing Systems*，*37*（*2024*），第*137806–137846*页。[85] *V. Kostic, K. Lounici, P. Novelli,* 和 *M. Pontil*，《*Koopman*算子学习的尖锐谱速率》，*Advances in Neural Information Processing Systems*，*36*（*2023*），第*32328–32339*页。[86] *V. Kostic, P. Novelli, A. Maurer, C. Ciliberto, L. Rosasco,* 和 *M. Pontil*，《在再生核希尔伯特空间中通过*Koopman*算子回归学习动力系统》，*Advances in Neural Information Processing Systems*，*35*（*2022*），第*4017–4031*页。[87] *N. Kovachki, S. Lanthaler,* 和 *S. Mishra*，《关于傅里叶神经算子的通用逼近与误差界》，机器学习研究杂志，*22*（*2021*），第*1–76*页。[88] *N. Kovachki, Z. Li, B. Liu, K. Azizzadenesheli, K. Bhattacharya, A. Stuart,* 和 *A. Anandkumar*，《神经算子：学习函数空间之间的映射及其在偏微分方程中的应用》，机器学习研究杂志，*24*（*2023*），第*1–97*页。[89] *N. B. Kovachki, S. Lanthaler,* 和 *H. Mhaskar*，《算子学习的数据复杂度估计》，*arXiv* 预印本 *arXiv:2405.15992*，（*2024*）。[90] *N. B. Kovachki, S. Lanthaler,* 和 *A. M. Stuart*，《算子学习：算法与分析学》，*Handbook of Numerical Analysis*，*25*（*2024*），第*419–467*页。[91] *S. Lanthaler*，《使用*PCA-Net*进行算子学习：复杂度上界与下界》，机器学习研究杂志，*24*（*2023*），第*1–67*页。[92] ,《*Lipschitz*算子的算子学习：信息论视角》，*arXiv* 预印本 *arXiv:2406.18794*，（*2024*）。[93] *S. Lanthaler, Z. Li,* 和 *A. M. Stuart*，《非局部性与非线性蕴含算子学习的通用性》，*Constructive Approximation*，*62*（*2025*），第*261–303*页。[94] *S. Lanthaler, S. Mishra,* 和 *G. E. Karniadakis*，《*DeepONets*的误差估计：无限维空间中的深度学习框架》，*Transactions of Mathematics and Its Applications*，*6*（*2022*），第*tnac001*页。

[95] *S. Lanthaler* 和 *A. M. Stuart*，《算子学习的参数复杂度》，*arXiv* 预印本 *arXiv:2306.15924*，（*2023*）。[96] *Z. Li, N. Kovachki, K. Azizzadenesheli, B. Liu, K. Bhattacharya, A. Stuart,* 和 *A. Anandkumar*，《用于参数化偏微分方程的傅里叶神经算子》，*arXiv* 预印本 *arXiv:2010.08895*，（*2020*）。[97] *Z. Li, H. Zheng, N. Kovachki, D. Jin, H. Chen, B. Liu, K. Azizzadenesheli,* 和 *A. Anandkumar*，《用于学习偏微分方程的物理信息神经算子》，*ACM/IMS* 数据科学杂志，*1*（*2024*），第*1–27*页。[98] *C. Liao, D. Needell,* 和 *H. Schaeffer*，《用于*Sobolev*空间中算子学习的柯西随机特征》，*arXiv* 预印本 *arXiv:2503.00300*，（*2025*）。[99] *F. Lindgren, H. Rue,* 和 *J. Lindström*，《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*（*2011*），第*423–498*页。[100] *H. Liu, H. Yang, M. Chen, T. Zhao,* 和 *W. Liao*，《无限维空间之间算子的深度非参数估计》，机器学习研究杂志，*25*（*2024*），第*1–67*页。[101] *L. Llamazares-Elias, S. Llamazares-Elias, J. Latz,* 和 *S. Klus*，《*Koopman*算子与生成元的数据驱动逼近：收敛速率与误差界》，*arXiv* 预印本 *arXiv:2405.00539*，（*2024*）。[102] *D. Long, N. Mrvaljević, S. Zhe,* 和 *B. Hosseini*，《用于学习微分方程及其解算子的核框架》，*Physica D*：非线性现象，*460*（*2024*），第*134095*页。[103] *L. Lu, P. Jin, G. Pang, Z. Zhang,* 和 *G. E. Karniadakis*，《基于算子通用逼近定理的*DeepONet*非线性算子学习》，《自然·机器智能》，*3*（*2021*），第*218–229*页。[104] *L. Lu, X. Meng, Z. Mao,* 和 *G. E. Karniadakis*，《*Deepxde*：用于求解微分方程的深度学习库》，*SIAM*评论，*63*（*2021*），第*208–228*页。[105] *Y. Maday, V. Perrier,* 和 *J.-C. Ravel*，《基于小波基的动态自适应方法，用于偏微分方程的逼近》，巴黎科学院报告，*I*辑，*312*（*1991*），第*405–410*页。[106] *S. Mallat*，《理解深度卷积网络》，皇家学会哲学汇刊*A*辑：数学、物理与工程科学，*374*（*2016*），第*20150203*页。[107] *B. Matérn*，《空间变异：随机模型及其在森林调查和其他抽样调查中若干问题上的应用》，*Meddelanden fran Statens Skogsforskningsinstitut*，*49*（*1960*）。[108] *R. Molinaro, Y. Yang, B. Engquist,* 和 *S. Mishra*，《用于求解偏微分方程逆问题的神经逆算子》，载于第*40*届国际机器学习会议论文集，*2023*，第*25105–25139*页。[109] *M. Mollenhauer, N. Mücke,* 和 *T. Sullivan*，《线性算子学习：作为适定非紧逆问题的无限维回归》，*arXiv* 预印本 *arXiv:2211.08875*，（*2022*）。[110] *P. Morin, R. H. Nochetto,* 和 *K. G. Siebert*，《自适应有限元方法的收敛性》，*SIAM*评论，*44*（*2002*），第*631–658*页。[111] *N. H. Nelsen* 和 *A. M. Stuart*，《巴拿赫空间之间输入输出映射的随机特征模型》，*SIAM Journal on Scientific Computing*，*43*（*2021*），第*A3212–A3243*页。

[112] ——, 使用随机特征进行算子学习：一种科学计算工具，《*SIAM*评论》，*66*（*2024*），第*535–571*页。[113] *N. H. Nelsen* 和 *Y. Yang*，算子学习与逆问题相遇：概率视角，*arXiv* 预印本 *arXiv:2508.20207*，（*2025*）。[114] *H. Nguyen* 和 *R. Stevenson*，流形上的有限元小波，《*IMA*数值分析期刊》，*23*（*2003*），第*149–173*页。[115] ，定量性质得到改进的有限元小波，《计算与应用数学杂志》，*230*（*2009*），第*706–727*页。[116]*J. Pathak*、*S. Subramanian*、*P. Harrington*、*S. Raja*、*A. Chattopadhyay*、*M. Mardani*、*T. Kurth*、*D. Hall*、*Z. Li*、*K. Azizzadenesheli* 等，*FourCastNet*：一种使用自适应傅里叶神经算子的全球数据驱动高分辨率天气模型，*arXiv* 预印本 *arXiv:2202.11214*，（*2022*）。[117] *J. O. Ramsay* 和 *C. Dalzell*，函数型数据分析的一些工具，《皇家统计学会会刊*B*辑：统计方法学》，*53*（*1991*），第*539–561*页。[118] *N. Reinhardt*、*S. Wang* 和 *J. Zech*，神经算子的统计学习理论，*arXiv* 预印本 *arXiv:2412.17582*，（*2024*）。[119] *N. Rekatsinas* 和 *R. Stevenson*，二次有限元小波*Riesz*基，《国际小波、多分辨率与信息处理杂志》，*16*（*2018*），第*1850033*页。[120] *D. Sanz-Alonso* 和 *R. Yang*，*Matérn*场的随机偏微分方程方法：图表示，《统计科学》，*37*（*2022*），第*519–540*页。[121] *F. Schäfer* 和 *H. Owhadi*，利用矩阵-向量乘积对椭圆方程求解器进行稀疏恢复，《*SIAM Journal on Scientific Computing*》，*46*（*2024*），第*A998–A1025*页。[122] *R. Schneider*，《多尺度与小波矩阵压缩：基于分析的方法，用于高效求解大型稠密方程组》，施普林格出版社，*2013*。[123] *C. Schwab* 和 *R. Stevenson*，用于抛物型演化问题的时空自适应小波方法，《计算数学》，*78*（*2009*），第*1293–1318*页。[124] *J. Shin*、*J. Lee* 和 *H. Hwang*，伪微分神经算子：用于学习偏微分方程解算子的广义傅里叶神经算子，*arXiv* 预印本 *arXiv:2201.11967*，（*2022*）。[125] *M. A. Shubin*，《伪微分算子与谱理论》，第*57*卷，施普林格，*1987*。[126] *A. Skodras*、*C. Christopoulos* 和 *T. Ebrahimi*，*JPEG 2000*静态图像压缩标准，《*IEEE*信号处理杂志》，*18*（*2001*），第*36–58*页。[127] *M. L. Stein*，《空间数据插值：克里金法的一些理论》，施普林格，*2012*。[128] *G. Stepaniants*，在再生核希尔伯特空间中学习偏微分方程，《机器学习研究杂志》，*24*（*2023*），第*1–72*页。[129] *R. Stevenson*，关于小波坐标中算子的可压缩性，《*SIAM* 数学分析杂志》，*35*（*2004*），第*1110–1132*页。[130] ，一种最优自适应有限元方法，《*SIAM* 数值分析杂志》，*42*（*2005*），第*2188–2217*页。[131] ，求解算子方程的自适应小波方法：综述，《多尺度、非线性与自适应逼近：献给 *Wolfgang Dahmen* 六十华诞》，（*2009*），第*543–597*页。[132] *U. Subedi* 和 *A. Tewari*，在学习傅里叶线性算子时控制统计误差、离散化误差和截断误差，*arXiv* 预印本 *arXiv:2408.09004*，（*2024*）。

[133] ——, 算子学习：统计视角，《统计学及其应用年度评论》，*13*（*2025*）。[134] *D. Taubman*，使用*EBCOT*实现高性能可扩展图像压缩，《*IEEE*图像处理汇刊》，*9*（*2000*），第*1158–1170*页。[135] *D. S. Taubman* 和 *M. W. Marcellin*，《*JPEG2000*：图像压缩基础、标准与实践》，克鲁维尔学术出版社，*2002*。[136] *M. Taylor*，伪微分算子，《伪微分算子》，（*2006*），第*19–44*页。[137] *R. Tibshirani*，通过*Lasso*进行回归收缩与变量选择，《皇家统计学会会刊*B*辑：统计方法学》，*58*（*1996*），第*267–288*页。[138] *F. Trèves*，《伪微分算子与傅里叶积分算子导论》第*2*卷：傅里叶积分算子，第*2*卷，施普林格科学与商业媒体，*1980*。[139] *R. Vershynin*，《高维概率：数据科学中的导论与应用》，第*47*卷，剑桥大学出版社，*2018*。[140] *T. von Petersdorff* 和 *C. Schwab*，抛物型积分微分方程的小波离散化，《*SIAM* 数值分析杂志》，*41*（*2003*），第*159–180*页。[141] *M. J. Wainwright*，《高维统计学：非渐近视角》，第*48*卷，剑桥大学出版社，*2019*。[142] *C. Wang* 和 *A. Townsend*，双曲型偏微分方程的算子学习，《机器学习研究杂志》，*26*（*2025*），第*1–44*页。[143] *P. Whittle*，关于平面上的平稳过程，《生物统计学》，（*1954*），第*434–449*页。[144] *C. K. I. Williams* 和 *C. E. Rasmussen*，《机器学习中的高斯过程》，第*2*卷，麻省理工学院出版社，剑桥（马萨诸塞州），*2006*。[145] *M.-W. Wong*，《伪微分算子导论》，第*6*卷，世界科技出版公司，*2014*。[146] *J. M. Wooldridge*，《计量经济学导论：现代方法》，圣智学习，第*6*版，*2016*。[147] *J.-Q. Yang* 和 *L. Shi*，用于非线性算子学习的基于核的随机逼近框架，*arXiv* 预印本 *arXiv:2509.11070*，（*2025*）。[148] *L. Ying*，用深度学习求解逆问题，载于《国际数学家大会论文集》，第*7*卷，*2022*，第*5154–5175*页。[149] *X. Yu* 和 *H. Schaeffer*，*Sobolev*空间中用于算子学习的正则化随机傅里叶特征与有限元重建，*arXiv* 预印本 *arXiv:2512.17884*，（*2025*）。[150] *S. Zhang*、*X. Wang* 和 *F. Lu*，算子中核学习的极小极大速率，*arXiv* 预印本 *arXiv:2502.20368*，（*2025*）。[151] *S. Zhou*，用于高维变量选择的阈值化套索回归，《统计数学研究所年刊》，（*2025*），第*1–47*页。[152] *H. Zou*，自适应套索回归及其预言机性质，《美国统计学会会刊》，*101*（*2006*），第*1418–1429*页。

A 辅助材料部分3

A.1 背景：双正交小波与多分辨率分析

设$\{V_j\}_{j>j_0}$为一列嵌套的线性有限维子空间 $V_j \subset V_{j+1} \subset \cdots \subset L^2(\mathcal{M})$。若该族$\{V_j\}_{j>j_0}$具有正则性 $\gamma > 0$ 和（逼近）阶 $d \in \mathbb{N}$，则称

$$\gamma = \sup\{s \in \mathbb{R} : V_j \subset H^s(\mathcal{M}),\ \forall\, j > j_0\},$$
$$d = \sup\left\{s \in \mathbb{R} : \inf_{v_j \in V_j} \|v - v_j\|_{L^2(\mathcal{M})} \lesssim 2^{-js}\|v\|_{H^s(\mathcal{M})},\ \forall\, v \in H^s(\mathcal{M}),\ \forall\, j > j_0\right\}.$$

我们假设子空间$\{V_j\}_{j>j_0}$是 $H^{r/2}(\mathcal{M})$-协调的，即对于某个固定的阶 $r \in \mathbb{R}$，有 $\gamma > \max\{0, r/2\}$。

此外，我们假设 $\dim(V_j) = O(2^{nj})$，并且对于每个 $j > j_0$，空间 V_j 由单尺度基 Φ_j张成，即

$$V_j = \operatorname{span}\Phi_j, \quad \text{where } \Phi_j := \{\phi_{j,k} : k \in \Delta_j\}, \quad \forall\, j > j_0.$$

这里，指标集 Δ_j 描述了 Φ_j中元素的空间局部化。与空间$\{V_j\}_{j>j_0}$类似，我们不失一般性地假设有限指标集$\{\Delta_j\}_{j>j_0}$是嵌套的，$\Delta_j \subset \Delta_{j+1} \subset \cdots$。对于每个 $j > j_0$，我们将对偶单尺度基定义为

$$\widetilde{\Phi}_j := \{\widetilde{\phi}_{j,k} : k \in \Delta_j\}, \quad \text{with } \langle \phi_{j,k}, \widetilde{\phi}_{j,k'}\rangle = \delta_{k,k'},\ \forall\, k, k' \in \Delta_j.$$

~ ~ ~ ~向量空间 $V_j := \operatorname{span}\ \Phi_j$, $j > j_0$, 也是嵌套的，$V_j \subset V_{j+1} \subset \cdots \subset L^2(\mathcal{M})$，而~族 $\{V_j\}_{j>j_0}$ 具有正则性 ~$\tilde{\gamma} > 0$ 和逼近阶 ~d。例如，设原基 ~Φ_j 由阶为 ~d 的单变量 B 样条的张量积生成，并设~对偶基 ~Φ_j的阶为 $d \geq d$，其中 $d + d$为偶数。则 $V_j = \operatorname{span}\Phi_j$和 $V_j = \operatorname{span}\ \Phi_j$~的逼近阶 d 和 d分别为.....此外，相应的正则性指标~满足 $\gamma = d - 1/2$，而 $\tilde{\gamma} > 0$与 d成正比增长。详细构造参见 [47, 51, 114, 115, 119]。

鉴于 Φ_j的双正交性，$\widetilde{\Phi}_j$，我们定义典范投影算子

$$Q_j v := \sum_{k \in \Delta_j} \langle v, \widetilde{\phi}_{j,k}\rangle \phi_{j,k}, \qquad Q_j^* v := \sum_{k \in \Delta_j} \langle v, \phi_{j,k}\rangle \widetilde{\phi}_{j,k}, \quad \forall\, v \in L^2(\mathcal{M}),$$

与多分辨率序列相关联的$\{V_j\}_{j>j_0}$、$\{\widetilde{V}_j\}_{j>j_0}$。此外，$L^2(\mathcal{M})$有界性 Q_j 蕴含 Jackson 不等式和 Bernstein 不等式；参见下文引理A.1 （ii）。

要定义多分辨率分析（MRA），我们首先引入指标集 $\nabla_j :=\sim\Delta_{j+1}\backslash\Delta_j$，$j > j_0$。给定单尺度基 Φ_j 和 Φ_j，可以构造双正交补基

$$\Psi_j = \{\psi_{j,k} : k \in \nabla_j\} \quad \text{and} \quad \widetilde{\Psi}_j = \{\widetilde{\psi}_{j,k} : k \in \nabla_j\}, \qquad j > j_0,$$

满足双正交关系

$$\langle \psi_{j,k}, \widetilde{\psi}_{j,k}\rangle = \delta_{(j,k),(j',k')} = \begin{cases} 1 & \text{if } j = j' \text{ and } k = k', \\ 0 & \text{otherwise,} \end{cases}$$

使得

$$\operatorname{diam}\big(\operatorname{supp}(\psi_{j,k})\big) \asymp 2^{-j}, \quad j > j_0.$$

我们将.....称为 [114, 115, 119] 特定构造。此外，我们采用以下约定~ ~$\Psi_{j_0} := \Phi_{j_0+1}$，$\Psi_{j_0} := \Phi_{j_0+1}$，以及 $\nabla_{j_0} := \Delta_{j_0+1}$。
对于 $j > j_0$，定义 $W_j := \operatorname{span} \Psi_j$ and $\widetilde{W}_j := \operatorname{span}$ ~
Ψ_j。双正交性意味着，对于所有 $j > j_0$，

$$V_{j+1} = W_j \oplus V_j, \quad \widetilde{V}_{j+1} = \widetilde{W}_j \oplus \widetilde{V}_j, \quad \widetilde{V}_j \perp W_j, \quad V_j \perp \widetilde{W}_j.$$

~因此，V_J 和 V_J 可分别写成补空间 W_j、
~ ~ ~W_j、$j_0 \le j < J$的直和，采用约定 $W_{j_0} := V_{j_0+1}$ and $W_{j_0} := V_{j_0+1}$。按照约定~$Q_{j_0} = Q^*_{j_0} := 0$，对于 $v_J \in V_J$ 以及对于 $\widetilde{v}_J \in V_J$，有

$$v_J = \sum_{j=j_0}^{J-1} (Q_{j+1} - Q_j) v_J, \qquad \widetilde{v}_J = \sum_{j=j_0}^{J-1} (Q^*_{j+1} - Q^*_j) \widetilde{v}_J,$$

其中

$$(Q_{j+1} - Q_j) v = \sum_{k \in \nabla_j} \langle v, \widetilde{\psi}_{j,k} \rangle \psi_{j,k}, \qquad (Q^*_{j+1} - Q^*_j) v = \sum_{k \in \nabla_j} \langle v, \psi_{j,k} \rangle \widetilde{\psi}_{j,k}.$$

由此可知，取粗尺度单尺度基与补基的并集，即可得到一对双正交对偶小波基，

$$\Psi = \bigcup_{j \ge j_0} \Psi_j, \qquad \widetilde{\Psi} = \bigcup_{j \ge j_0} \widetilde{\Psi}_j.$$

~我们分别将 Ψ 和 Ψ 称为原始和对偶多分辨率分析。此处及下文中，假设 Ψ 和 $\widetilde{\Psi}$ 中的所有基函数均在M中归一化。
下一个引理汇总了双正交小波系统（Ψ，~
Ψ）的若干关键性质，这些性质将在算子学习分析中贯穿使用；参见 [44, 47, 36, 68]。

~**引理 A.1**（双正交小波的性质）。设（Ψ，Ψ）是~M上的一个双正交小波系统，其参数为（γ，$\widetilde{\gamma}$，d，d）。则有以下结论。

（*i*）基数与局部性。对于每个 $j \ge j_0$，层级 j 上的小波数量满足$|\nabla_j| \asymp 2^{jn}$。此外，支撑集具有如下局部化性质：$\operatorname{diam}(\operatorname{supp}\psi_{j,\,k}) \asymp 2^{-j}$，并且存在一个常数 M，使得对于每个 $k \in \nabla_j$，至多有 M 个指标 $k' \in \nabla_j$满足$\operatorname{meas}\big(\operatorname{supp}(\psi_{j,\,k}) \cap \operatorname{supp}(\psi_{j,\,k'})\big) \ne 0.$

(ii) 逼近与稳定性。令 Q_j 表示到逼近空间 V_j 上的典范投影算子。我们取 $Q_{j_0} = 0$，并且对于 $j > j_0$，

$$Q_j v = \sum_{j'=j_0}^{j-1} \sum_{k \in \nabla_{j'}} \langle v, \widetilde{\psi}_{j',k} \rangle \psi_{j',k}, \quad \forall\, v \in L^2(\mathcal{M}).$$

Jackson 不等式和 *Bernstein* 不等式成立：

$$\|v - Q_j v\|_{H^s(\mathcal{M})} \lesssim 2^{-j(t-s)} \|v\|_{H^t(\mathcal{M})}, \quad \forall\, v \in H^t(\mathcal{M}),$$

~对所有 $-d \le s \le t \le d$、$s < \gamma$、$-\widetilde{\gamma} < t$，以及

$$\|Q_j v\|_{H^s(\mathcal{M})} \lesssim 2^{j(s-t)} \|Q_j v\|_{H^t(\mathcal{M})}, \quad \forall\, v \in H^t(\mathcal{M}),$$

对于所有 $t \le s \le \gamma$。

*(iii)*消失矩。对于任意足够光滑的函数 v，

$$|\langle v,\psi_{j,k}\rangle| \lesssim 2^{-j(\widetilde{d}+n/2)}|v|_{W^{\widetilde{d},\infty}(\mathrm{supp}(\psi_{j,k}))}, \quad |v|_{W^{\widetilde{d},\infty}(\Omega)} := \sup_{|\alpha|=\widetilde{d},x\in\Omega}|\partial^\alpha v(x)|.$$

(iv) 范数等价。对任意 $v\in H^t(M)$，以下等价关系成立：

$$\|v\|^2_{H^t(\mathcal{M})} \asymp \sum_{j\geq j_0}\sum_{k\in\nabla_j} 2^{2jt}|\langle v,\widetilde{\psi}_{j,k}\rangle|^2, \quad t\in(-\widetilde{\gamma},\gamma),$$
$$\|v\|^2_{H^t(\mathcal{M})} \asymp \sum_{j\geq j_0}\sum_{k\in\nabla_j} 2^{2jt}|\langle v,\psi_{j,k}\rangle|^2, \quad t\in(-\gamma,\widetilde{\gamma}).$$

A.2 A、U、W 的性质

命题 *3.2* 的证明。(I) **近似稀疏性：** (i) 中的估计在文献中广为人知 [44, 45, 122, 68]：它表明，当测试小波的支撑集彼此相距较远时，矩阵元素会迅速衰减。为方便读者，我们在此给出简短且严谨的推导。

令 $k_{\mathcal{A}}(x,y)=(2\pi)^{-n}\int_{\mathbb{R}_n} e^{i(x-y)\cdot\xi}a(x,\xi)d\xi$ 为 $\mathcal{A}\in\mathrm{OPS}^r$ 的（分布意义下的）施瓦茨核。当 $\mathrm{dist}(S_{j,k},S_{j',k'})\gtrsim 2^{-\min\{j,j'\}}$ 时，乘积 $\psi_{j,k}(x)\psi_{j',k'}(y)$ 的支撑集远离对角线 $x=y$，因此 $k_{\mathcal{A}}$ 在该集合上光滑，且可应用富比尼定理。因此，

$$\langle\mathcal{A}\psi_{j',k'},\psi_{j,k}\rangle = \iint k_{\mathcal{A}}(x,y)\psi_{j,k}(x)\psi_{j',k'}(y)dxdy = \int\left(\int k_{\mathcal{A}}(x,y)\psi_{j',k'}(y)dy\right)\psi_{j,k}(x)dx.$$

根据引理 A.1 (iii)（消失矩直到第 d阶），

$$|\langle\mathcal{A}\psi_{j',k'},\psi_{j,k}\rangle| \lesssim 2^{-j(\widetilde{d}+n/2)}\sup_{|\alpha|=\widetilde{d},x\in\mathrm{supp}(\psi_{j,k})}\left|\int\partial_x^\alpha k_{\mathcal{A}}(x,y)\psi_{j',k'}(y)dy\right|.$$

对 y 变量应用相同的论证可得

$$|\langle\mathcal{A}\psi_{j',k'},\psi_{j,k}\rangle| \lesssim 2^{-(j+j')(\widetilde{d}+n/2)}\sup_{x\in\mathrm{supp}(\psi_{j,k}),y\in\mathrm{supp}(\psi_{j',k'}),|\alpha|=|\beta|=\widetilde{d}}|\partial_x^\alpha\partial_y^\beta k_{\mathcal{A}}(x,y)|.$$

For $\mathcal{A}\in\mathrm{OPS}^r$，该核满足标准的非对角界（参见 [68, 44, 45, 122] ）

$$|\partial_x^\alpha\partial_y^\beta k_{\mathcal{A}}(x,y)|\lesssim_{\alpha,\beta}\mathrm{dist}(x,y)^{-(n+r+|\alpha|+|\beta|)}, \quad x\neq y,$$

因此

$$|\langle\mathcal{A}\psi_{j',k'},\psi_{j,k}\rangle| \lesssim 2^{-(j+j')(\widetilde{d}+n/2)}\sup_{x\in\mathrm{supp}(\psi_{j,k}),y\in\mathrm{supp}(\psi_{j',k'}),|\alpha|=|\beta|=\widetilde{d}}\mathrm{dist}(x,y)^{-(n+r+|\alpha|+|\beta|)}$$
$$= 2^{-(j+j')(\widetilde{d}+n/2)}\mathrm{dist}(S_{j,k},S_{j',k'})^{-(n+r+2\widetilde{d})}.$$

更多背景信息请参阅 [68, 命题 2] 和 [44, 122] 。

(ii) 中的估计表明，对应于支撑重叠的小波的矩阵元素会随着尺度差增大而衰减；参见 [44, 第 9 节。4], [36,(2.28)], 以及[45, 第2261]页的注记。我们依照 [44, 第 9 节。4]简要推导这一估计。

不失一般性，设 $j \geq j'$。根据假设 2.1 (i)，算子$\mathcal{A} \in \mathsf{OPS}^r(M)$ 从 $H^s(M)$ 到 $H^{s-r}(M)$ 连续且双射，对任意 $s \in \mathbb{R}$均成立；参见 [68, 命题 11]。结合柯西-施瓦茨不等式与以下界

$\|\mathcal{A}\psi_{j'k'}\|_{H^{-r/2+\sigma}} \lesssim \|\psi_{j'k'}\|_{H^{r/2+\sigma}}$，我们得到

$$|\langle \mathcal{A}\psi_{j',k'}, \psi_{j,k}\rangle| \leq \|\mathcal{A}\psi_{j',k'}\|_{H^{-r/2+\sigma}} \|\psi_{j,k}\|_{H^{r/2-\sigma}} \lesssim \|\psi_{j',k'}\|_{H^{r/2+\sigma}} \|\psi_{j,k}\|_{H^{r/2-\sigma}}. \quad \text{(A.1)}$$

若 $r/2+\sigma < \gamma$ 且 $r/2-\sigma > -\widetilde{\gamma}$，则可将引理 A.1 （iv）中的范数等价应用于（A.1）右端的每个因子。利用双正交性可得

$$\begin{aligned} |\langle \mathcal{A}\psi_{j',k'}, \psi_{j,k}\rangle| &\lesssim_{\sigma,r} \|\psi_{j',k'}\|_{H^{r/2+\sigma}} \|\psi_{j,k}\|_{H^{r/2-\sigma}} \\ &\asymp_{\sigma,r} 2^{j'(r/2+\sigma)} \cdot 2^{j(r/2-\sigma)} = 2^{(j+j')r/2-\sigma(j-j')}. \end{aligned}$$

He因此估计成立。参数 σ 必须满足 $0 < \sigma < \min\{\gamma - r/2,\ \widetilde{\gamma} + r/2\}$.

（II）**对角预条件：**

该论证遵循 [68, 命题 3 和 13], 中的思路，但为方便读者，我们给出一个自包含的证明。

~我们首先考虑 $\mathcal{C}_u$ 和 C$_u$。在假设 2.1（ii）下， $\mathcal{C}_u$ 在L^2(M)上是自伴且紧的；参见 [68, 命题 1]。因此，双无限矩阵 ~C$_u$ 定义了 $\ell^2(\mathcal{J})$上的一个对称紧算子。此外， [68, 命题 11]表明， $\mathcal{C}_u : H^{-r_1}$(M) $\to H^{r_1}$(M)是连续的，因此

$$\langle \mathcal{C}_u v, v\rangle \leq \|\mathcal{C}_w v\|_{H^{r_1}} \|v\|_{H^{-r_1}} \lesssim \|v\|^2_{H^{-r_1}}, \quad \forall\ v \in H^{-r_1}.$$

结合假设 2.1(iii)，可得范数等价关系

$$\langle \mathcal{C}_u v, v\rangle \asymp \|v\|^2_{H^{-r_1}}, \quad \forall\ v \in H^{-r_1}.$$

令 $v = {}^{\sim}\mathbf{v}^{\top}\widetilde{\Psi} = \sum_{\lambda\in\mathcal{J}} \widetilde{v}_\lambda \widetilde{\psi}_\lambda$。由于 ~$\mathbf{C}_u$ 是 $\mathcal{C}_u$ 在双正交~系统（Ψ, Ψ）中的矩阵表示，

$$\langle \mathcal{C}_u v, v\rangle = \langle \widetilde{\mathbf{C}}_u \widetilde{\mathbf{v}}, \widetilde{\mathbf{v}}\rangle.$$

根据引理 A.1 (iv)，若 $-r_1 \in (-\gamma, \widetilde{\gamma})$，则

$$\|v\|^2_{H^{-r_1}} \asymp \sum_{j\geq j_0} \sum_{k\in\nabla_j} 2^{-2jr_1} |\langle v, \psi_{j,k}\rangle|^2 = \|\mathbf{D}^{-r_1}\widetilde{\mathbf{v}}\|^2.$$

因此，对于所有满足 $\mathbf{D}^{-r_1}{}^{\sim}\mathbf{v} \in \ell^2(\mathcal{J})$ 的 ~$\mathbf{v}$，

$$\langle \widetilde{\mathbf{C}}_u \widetilde{\mathbf{v}}, \widetilde{\mathbf{v}}\rangle \asymp \|\mathbf{D}^{-r_1}\widetilde{\mathbf{v}}\|^2.$$

这意味着对角预条件矩阵具有一致谱界：

$$c_- \leq \sigma_{\min}(\mathbf{D}^{r_1}\widetilde{\mathbf{C}}_u\mathbf{D}^{r_1}) \leq \sigma_{\max}(\mathbf{D}^{r_1}\widetilde{\mathbf{C}}_u\mathbf{D}^{r_1}) \leq c_+.$$

同样的论证也可应用于第 (i) 部分中的 $\mathbf{A}$ 和第 (iii) 部分中的 $\mathbf{C}_w$ ， 前提是$r/2, -r_2 \in (-\widetilde{\gamma}, \gamma)$。

□

B 辅助材料部分 4

B.1 区域划分

数学上，图 D_1-D_6 中的六个互不相交的区域 1 定义如下:

$$
\begin{aligned}
D_1 &:= \Big\{(j,j') : J \geq j > j' \geq 0,\ j \geq \frac{t+t'-r}{\sigma-\frac{n}{2}+t'-\frac{r}{2}}J + \frac{\sigma-\frac{n}{2}-(t-\frac{r}{2})}{\sigma-\frac{n}{2}+t'-\frac{r}{2}}j'\Big\},\\
D_2 &:= \Big\{(j,j') : J \geq j' > j \geq 0,\ j' \geq \frac{t+t'-r}{\sigma-\frac{n}{2}+t-\frac{r}{2}}J + \frac{\sigma-\frac{n}{2}-(t'-\frac{r}{2})}{\sigma-\frac{n}{2}+t-\frac{r}{2}}j\Big\},\\
D_3 &:= \Big\{(j,j') : J \geq j > j' \geq 0,\ j \leq \frac{t+t'-r}{\sigma-\frac{n}{2}+t'-\frac{r}{2}}J + \frac{\sigma-\frac{n}{2}-(t-\frac{r}{2})}{\sigma-\frac{n}{2}+t'-\frac{r}{2}}j',\\
&\qquad j \geq \frac{t+t'-r}{\widetilde{d}+t'}J + \frac{\widetilde{d}+r-t}{\widetilde{d}+t'}j'\Big\},\\
D_4 &:= \Big\{(j,j') : J \geq j' > j \geq 0,\ j' \leq \frac{t+t'-r}{\sigma-\frac{n}{2}+t-\frac{r}{2}}J + \frac{\sigma-\frac{n}{2}-(t'-\frac{r}{2})}{\sigma-\frac{n}{2}+t-\frac{r}{2}}j,\\
&\qquad j' \geq \frac{t+t'-r}{\widetilde{d}+t}J + \frac{\widetilde{d}+r-t'}{\widetilde{d}+t}j\Big\},\\
D_5 &:= \Big\{(j,j') : J(t+t'-r) - jt' - j't - (j+j')\widetilde{d} < 0,\ j \leq \frac{t+t'-r}{\widetilde{d}+t'}J + \frac{\widetilde{d}+r-t}{\widetilde{d}+t'}j',\\
&\qquad j' \leq \frac{t+t'-r}{\widetilde{d}+t}J + \frac{\widetilde{d}+r-t'}{\widetilde{d}+t}j\Big\},\\
D_6 &:= \Big\{(j,j') : J(t+t'-r) - jt' - j't - (j+j')\widetilde{d} \geq 0\Big\}.
\end{aligned}
$$

B.2 支撑单调性与包含关系

引理 *B.1*（支撑单调性）。设 (J,t,t') 和 $(\widetilde{J},\widetilde{t},\widetilde{t'})$ 是两组参数，并设 $supp(J,t,t')$ 和 $supp(\widetilde{J},\widetilde{t},\widetilde{t'})$ 如定义 *4.2* 中所定义。假设 $\widetilde{J} \geq J, \widetilde{t} \geq t, \widetilde{t'} \geq t'$, $and \sigma - n/2 > max\{\widetilde{t},\widetilde{t'}\} - r/2$, $\widetilde{d} > \sigma - n/2 - r/2$. 那么，

$$\mathrm{supp}(J,t,t') \subset \mathrm{supp}(\widetilde{J},\widetilde{t},\widetilde{t'}).$$

证明。 根据定义，

$$
\begin{aligned}
\mathrm{supp}(J,t,t') = \Big\{&(\lambda,\lambda') \in \Lambda_J \times \Lambda_J : \mathrm{dist}(S_\lambda, S_{\lambda'}) \leq \tau_{jj'}, j \leq \frac{t+t'-r}{\sigma-\frac{n}{2}+t'-\frac{r}{2}}J + \frac{\sigma-\frac{n}{2}-(t-\frac{r}{2})}{\sigma-\frac{n}{2}+t'-\frac{r}{2}}j',\\
&j' \leq \frac{t+t'-r}{\sigma-\frac{n}{2}+t-\frac{r}{2}}J + \frac{\sigma-\frac{n}{2}-(t'-\frac{r}{2})}{\sigma-\frac{n}{2}+t-\frac{r}{2}}j\Big\},
\end{aligned}
$$

其中

$$\tau_{jj'} = a \max\Big\{2^{-\min\{j,j'\}}, 2^{(J(t+t'-r)-jt'-j't-(j+j')\widetilde{d})/(2\widetilde{d}+r)}\Big\}.$$

同样，

$$
\begin{aligned}
\mathrm{supp}(\widetilde{J},\widetilde{t},\widetilde{t'}) := \Big\{&(\lambda,\lambda') \in \Lambda_{\widetilde{J}} \times \Lambda_{\widetilde{J}} : \mathrm{dist}(S_\lambda, S_{\lambda'}) \leq \widetilde{\tau}_{jj'}, j \leq \frac{\widetilde{t}+\widetilde{t'}-r}{\sigma-\frac{n}{2}+\widetilde{t'}-\frac{r}{2}}\widetilde{J} + \frac{\sigma-\frac{n}{2}-(\widetilde{t}-\frac{r}{2})}{\sigma-\frac{n}{2}+\widetilde{t'}-\frac{r}{2}}j',\\
&j' \leq \frac{\widetilde{t}+\widetilde{t'}-r}{\sigma-\frac{n}{2}+\widetilde{t}-\frac{r}{2}}\widetilde{J} + \frac{\sigma-\frac{n}{2}-(\widetilde{t'}-\frac{r}{2})}{\sigma-\frac{n}{2}+\widetilde{t}-\frac{r}{2}}j\Big\},
\end{aligned}
$$

with

$$\widetilde{\tau}_{jj'} = a \max\Big\{2^{-\min\{j,j'\}}, 2^{(\widetilde{J}(\widetilde{t}+\widetilde{t}'-r)-j\widetilde{t}'-j'\widetilde{t}-(j+j')\widetilde{d})/(2\widetilde{d}+r)}\Big\}.$$

固定 $(\lambda,\lambda') \in \mathsf{supp}(J,t,t')$；我们证明 $(\lambda,\lambda') \in \mathsf{supp}(\widetilde{J},\widetilde{t},\widetilde{t}')$。由于 $J \le \widetilde{J}$，我们立即得到

$$(\lambda,\lambda') \in \Lambda_J \times \Lambda_J \quad \Longrightarrow \quad (\lambda,\lambda') \in \Lambda_{\widetilde{J}} \times \Lambda_{\widetilde{J}}.$$

接下来，注意到

$$J(t+t'-r) - jt' - j't - (j+j')\widetilde{d} \le \widetilde{J}(\widetilde{t}+\widetilde{t}'-r) - j\widetilde{t}' - j'\widetilde{t} - (j+j')\widetilde{d},$$

因此 $\tau_{jj'} \le \widetilde{\tau}_{jj'}$。由此可得，

$$\mathrm{dist}(S_\lambda, S_{\lambda'}) \le \tau_{jj'} \quad \Longrightarrow \quad \mathrm{dist}(S_\lambda, S_{\lambda'}) \le \widetilde{\tau}_{jj'}.$$

此外，

$$j \le \frac{t+t'-r}{\sigma-\frac{n}{2}+t'-\frac{r}{2}}J + \frac{\sigma-\frac{n}{2}-(t-\frac{r}{2})}{\sigma-\frac{n}{2}+t'-\frac{r}{2}}j' \quad \Longrightarrow \quad j \le \frac{\widetilde{t}+\widetilde{t}'-r}{\sigma-\frac{n}{2}+\widetilde{t}'-\frac{r}{2}}\widetilde{J} + \frac{\sigma-\frac{n}{2}-(\widetilde{t}-\frac{r}{2})}{\sigma-\frac{n}{2}+\widetilde{t}'-\frac{r}{2}}j',$$

这由下式可得

$$\begin{aligned}
&\left(\frac{\widetilde{t}+\widetilde{t}'-r}{\sigma-\frac{n}{2}+\widetilde{t}'-\frac{r}{2}}\widetilde{J} + \frac{\sigma-\frac{n}{2}-(\widetilde{t}-\frac{r}{2})}{\sigma-\frac{n}{2}+\widetilde{t}'-\frac{r}{2}}j'\right) - \left(\frac{t+t'-r}{\sigma-\frac{n}{2}+t'-\frac{r}{2}}J + \frac{\sigma-\frac{n}{2}-(t-\frac{r}{2})}{\sigma-\frac{n}{2}+t'-\frac{r}{2}}j'\right)\\
&\ge \left(\frac{\widetilde{t}+\widetilde{t}'-r}{\sigma-\frac{n}{2}+\widetilde{t}'-\frac{r}{2}}J + \frac{\sigma-\frac{n}{2}-(\widetilde{t}-\frac{r}{2})}{\sigma-\frac{n}{2}+\widetilde{t}'-\frac{r}{2}}j'\right) - \left(\frac{t+t'-r}{\sigma-\frac{n}{2}+t'-\frac{r}{2}}J + \frac{\sigma-\frac{n}{2}-(t-\frac{r}{2})}{\sigma-\frac{n}{2}+t'-\frac{r}{2}}j'\right)\\
&= \left(\frac{\widetilde{t}+\widetilde{t}'-r}{\sigma-\frac{n}{2}+\widetilde{t}'-\frac{r}{2}} - \frac{t+t'-r}{\sigma-\frac{n}{2}+t'-\frac{r}{2}}\right)(J-j') \ge 0,
\end{aligned}$$

因为 $t \le \widetilde{t}$, $t' \le \widetilde{t}'$, $J \le \widetilde{J}$，且 $\sigma - n/2 > \max\{\widetilde{t}, \widetilde{t}'\} - r/2$。完全类似的论证可得

$$j' \le \frac{t+t'-r}{\sigma-\frac{n}{2}+t-\frac{r}{2}}J + \frac{\sigma-\frac{n}{2}-(t'-\frac{r}{2})}{\sigma-\frac{n}{2}+t-\frac{r}{2}}j \quad \Longrightarrow \quad j' \le \frac{\widetilde{t}+\widetilde{t}'-r}{\sigma-\frac{n}{2}+\widetilde{t}-\frac{r}{2}}\widetilde{J} + \frac{\sigma-\frac{n}{2}-(\widetilde{t}'-\frac{r}{2})}{\sigma-\frac{n}{2}+\widetilde{t}-\frac{r}{2}}j.$$

因此，支撑集$(J,\ t,\ t') \subset$支撑集$(\widetilde{J},\ \widetilde{t},\ \widetilde{t}')$。 □

引理的证明 *4.5.* 设 $(\lambda,\lambda') = ((j,k),(j',k'))$ 满足 $j > j'$ 且假设 $(\mathbf{M}_{(J,t,t')})_{\lambda,\lambda'} = 1$。那么，根据掩码 $\mathbf{M}_{(J,t,t')}$ 的定义，以下条件成立：

- 距离条件： $\mathsf{dist}(S_{j,k},S_{j'k'}) \le \tau_{jj'}$.
- 对 $j_\prime$ 的尺度约束： $j' \le \frac{t+t'}{\sigma-\frac{n}{2}+t-\frac{r}{2}}J + \frac{\sigma-\frac{n}{2}-(t-\frac{r}{2})}{\sigma-\frac{n}{2}+t-\frac{r}{2}}j$.
- 对 j 的尺度约束： $j \le \frac{t+t'}{\sigma-\frac{n}{2}+t'-\frac{r}{2}}J + \frac{\sigma-\frac{n}{2}-(t'-\frac{r}{2})}{\sigma-\frac{n}{2}+t'-\frac{r}{2}}j'$.

要证明 $(\mathbf{M}_{(J,t,t')})_{\lambda',\lambda} = 1$，我们必须验证 $(\lambda',\lambda) = ((j',k'),(j,k))$ 也满足定义 $\mathsf{supp}(J,t,t')$ 的条件。

- 距离条件。由于$\mathrm{dist}(S_{j,k},S_{j'k'}) = \mathrm{dist}(S_{j'k'}\ S_{j,k})$，我们有

$$\mathrm{dist}(S_{j',k'},S_{j,k}) \le \tau_{jj'} \overset{(\star)}{\le} \tau_{j'j},$$

其中（⋆）由下式得出

$$\tau_{jj'} \le \tau_{j'j} \quad \Longleftrightarrow \quad (-jt' - j't) \le (-j't' - jt) \quad \Longleftrightarrow \quad (j' - j)(t' - t) \le 0.$$

由于 $j > j'$ 和 $t \le t'$，最后一个不等式成立。因此，(λ', λ) 满足距离条件。

- $'$ 的尺度约束。由于 $j_\prime < j$，因此

$$j' < j \le \frac{t + t' - r}{\sigma - \frac{n}{2} + t' - \frac{r}{2}} J + \frac{\sigma - \frac{n}{2} - (t - \frac{r}{2})}{\sigma - \frac{n}{2} + t' - \frac{r}{2}} j,$$

因此，j' 所需的上界成立。

- 的尺度约束。由假设不等式

$$j \le \frac{t + t' - r}{\sigma - \frac{n}{2} + t' - \frac{r}{2}} J + \frac{\sigma - \frac{n}{2} - (t - \frac{r}{2})}{\sigma - \frac{n}{2} + t' - \frac{r}{2}} j'$$

并利用 $t \le t'$ 这一事实，我们得到

$$j \le \frac{t + t' - r}{\sigma - \frac{n}{2} + t - \frac{r}{2}} J + \frac{\sigma - \frac{n}{2} - (t' - \frac{r}{2})}{\sigma - \frac{n}{2} + t - \frac{r}{2}} j'.$$

由于 $\mathrm{supp}(J,\ t,\ t')$ 的所有定义条件对于 $(\lambda',\ \lambda)$ 均成立，我们得出 $(\mathbf{M}_{(J,t,t')})_{\lambda',\ \lambda} = 1$，从而完成证明。 □

C 辅助材料部分 5

C.1 稀疏性估计

在本小节中，我们首先给出引理 C.1 中的若干基本衰减估计，这些结果由命题 3.2 (I) 推导而来。这些界将在对估计量进行分析时反复用到。尽管文献中的许多地方都出现过类似估计——例如，[44, 45,68]——为方便起见，我们在此给出简明且自成一体的论述。下面的命题C.2刻画了定义 4.2 中引入的稀疏模式 $\mathrm{supp}(J,\ t,\ t')$，并统计压缩矩阵中相应非零元素的数量；这在方差分析中起着核心作用。特别地，这表明该压缩实现了最优稀疏性，因为 $\mathrm{nnz}\big(\mathbf{M}_{(J,t,t')}\big) \lesssim 2^{Jn}$，与 V_J 中内在的自由度数量相匹配。

回顾一下，根据定义 4.2，阈值参数为

$$\tau_{jj'} \asymp \max\Big\{2^{-\min\{j,j'\}}, 2^{(J(t+t'-r)-jt'-j't-(j+j')\widetilde{d})/(2\widetilde{d}+r)}\Big\}.$$

引理 C.1 (基本估计)。 *(i)* 对于 $j \ge j' \ge j_0$ 和 $\alpha > \frac{n}{n+r+2\widetilde{d}}$，

$$\sum_{k \in \nabla_j} |\langle \mathcal{A}\psi_{j,k}, \psi_{j',k'}\rangle|^\alpha \lesssim 2^{\alpha(j+j')r/2} \cdot 2^{-(j-j')\alpha\sigma} 2^{(j-j')n},$$

$$\sum_{k' \in \nabla_{j'}} |\langle \mathcal{A}\psi_{j,k}, \psi_{j',k'}\rangle|^\alpha \lesssim 2^{\alpha(j+j')r/2} \cdot 2^{-(j-j')\alpha\sigma}.$$

(ii) 对于 $j \ge j' \ge j_0$ 和 $\tau_{jj'} \gtrsim 2^{-j'}$，

$$\big|\{k \in \nabla_j : \mathrm{dist}(S_{j,k}, S_{j',k'}) \le \tau_{jj'}\}\big| \lesssim 2^{jn} \tau_{jj'}^n,$$

$$\big|\{k' \in \nabla_{j'} : \mathrm{dist}(S_{j,k}, S_{j',k'}) \le \tau_{jj'}\}\big| \lesssim 2^{j'n} \tau_{jj'}^n.$$

(iii) 对于 $j \geq j' \geq j_0$、$\tau_{jj'} \gtrsim 2^{-j'}$ 和 $\alpha > \frac{n}{n+r+2\widetilde{d}}$,

$$\sum_{k\in\nabla_j:\mathrm{dist}(S_{j,k},S_{j',k'})\geq\tau_{jj'}} |\langle \mathcal{A}\psi_{j,k}, \psi_{j',k'}\rangle|^\alpha \lesssim 2^{-\alpha(j+j')(\widetilde{d}+n/2)} 2^{jn} \tau_{jj'}^{-\alpha(n+r+2\widetilde{d})+n},$$

$$\sum_{k'\in\nabla_{j'}:\mathrm{dist}(S_{j,k},S_{j',k'})\geq\tau_{jj'}} |\langle \mathcal{A}\psi_{j,k}, \psi_{j',k'}\rangle|^\alpha \lesssim 2^{-\alpha(j+j')(\widetilde{d}+n/2)} 2^{j'n} \tau_{jj'}^{-\alpha(n+r+2\widetilde{d})+n}.$$

证明。 (i) 根据命题 3.2 (I)，对于 $j \geq j' \geq j_0$,

$$|\langle \mathcal{A}\psi_{j,k}, \psi_{j',k'}\rangle| \lesssim 2^{(j+j')r/2} \cdot 2^{-(j-j')\sigma} \Big(1 + 2^{j'} \mathrm{dist}(S_{j,k}, S_{j',k'})\Big)^{-(n+r+2\widetilde{d})}.$$

其中 $0 < \sigma < \min\{\gamma - r/2,\ \widetilde{\gamma} + r/2\}$。因此,

$$\begin{aligned}
&\sum_{k\in\nabla_j} |\langle \mathcal{A}\psi_{j,k}, \psi_{j',k'}\rangle|^\alpha \\
&\lesssim 2^{\alpha(j+j')r/2} \cdot 2^{-(j-j')\alpha\sigma} \sum_{k\in\nabla_j} \Big(1 + 2^{j'} \mathrm{dist}(S_{j,k}, S_{j',k'})\Big)^{-\alpha(n+r+2\widetilde{d})} \\
&= 2^{\alpha(j+j')r/2} \cdot 2^{-(j-j')\alpha\sigma} \sum_{k\in\nabla_j:\mathrm{dist}(S_{j,k},S_{j',k'})\leq C2^{-j'}} \Big(1 + 2^{j'} \mathrm{dist}(S_{j,k}, S_{j',k'})\Big)^{-\alpha(n+r+2\widetilde{d})} \\
&\quad + 2^{\alpha(j+j')r/2} \cdot 2^{-(j-j')\alpha\sigma} \sum_{k\in\nabla_j:\mathrm{dist}(S_{j,k},S_{j',k'})\geq C2^{-j'}} \Big(1 + 2^{j'} \mathrm{dist}(S_{j,k}, S_{j',k'})\Big)^{-\alpha(n+r+2\widetilde{d})}.
\end{aligned}$$

由于 $j \geq j'$，第一个求和中至多有 $O(2^{(j-j')n})$ 项:

$$\sum_{k\in\nabla_j:\mathrm{dist}(S_{j,k},S_{j',k'})\leq C2^{-j'}} \Big(1 + 2^{j'} \mathrm{dist}(S_{j,k}, S_{j',k'})\Big)^{-\alpha(n+r+2\widetilde{d})} \lesssim 2^{(j-j')n}.$$

~此外，由于 $\alpha(n + r + 2d) > n$，我们可以用积分来估计第二个求和:

$$\begin{aligned}
&\sum_{k\in\nabla_j:\mathrm{dist}(S_{j,k},S_{j',k'})\geq C2^{-j'}} \Big(1 + 2^{j'} \mathrm{dist}(S_{j,k}, S_{j',k'})\Big)^{-\alpha(n+r+2\widetilde{d})} \\
&\lesssim 2^{-\alpha(n+r+2\widetilde{d})j'} \cdot 2^{jn} \int_{\|x\|\geq C2^{-j'}} \|x\|^{-\alpha(n+r+2\widetilde{d})} dx \\
&\lesssim 2^{-\alpha(n+r+2\widetilde{d})j'} \cdot 2^{jn} \cdot (C2^{-j'})^{-\alpha(n+r+2\widetilde{d})+n} \\
&\asymp 2^{(j-j')n}.
\end{aligned}$$

合并上述估计可得

$$\sum_{k\in\nabla_j} |\langle \mathcal{A}\psi_{j,k}, \psi_{j',k'}\rangle|^\alpha \lesssim 2^{\alpha(j+j')r/2} \cdot 2^{-(j-j')\alpha\sigma} 2^{(j-j')n}.$$

(ii) 由引理 A.1 (i)，存在固定常数 C 和 M 使得

$$\mathrm{diam}(\mathrm{supp}(\psi_{j,k})) \leq C2^{-j},$$

并且，对于每个 $k \in \nabla_j$，满足条件的指标 $k' \in \nabla_{j'}$ 至多有 M 个

$$\mathrm{meas}(\mathrm{supp}(\psi_{j,k}) \cap \mathrm{supp}(\psi_{j,k'})) \neq 0.$$

此外，由于 $\tau_{jj'} \gtrsim 2^{-\min\{j,j'\}}$，可知

$$\big|\{k \in \nabla_j : \mathrm{dist}(S_{j,k}, S_{j',k'}) \le \tau_{jj'}\}\big| \lesssim \Big(\frac{\tau_{jj'}}{2^{-j}}\Big)^n = 2^{jn}\tau_{jj'}^n,$$

同理，

$$\big|\{k' \in \nabla_{j'} : \mathrm{dist}(S_{j,k}, S_{j',k'}) \le \tau_{jj'}\}\big| \lesssim \Big(\frac{\tau_{jj'}}{2^{-j'}}\Big)^n = 2^{j'n}\tau_{jj'}^n.$$

(iii) 根据命题 3.2 (I)(i)，

$$\sum_{k\in\nabla_j:\mathrm{dist}(S_{j,k},S_{j',k'})\ge\tau_{jj'}} |\langle \mathcal{A}\psi_{j,k}, \psi_{j',k'}\rangle|^\alpha$$
$$\lesssim \sum_{k\in\nabla_j:\mathrm{dist}(S_{j,k},S_{j',k'})\ge\tau_{jj'}} 2^{-\alpha(j+j')(\widetilde{d}+n/2)}\mathrm{dist}(S_{j,k}, S_{j',k'})^{-\alpha(n+r+2\widetilde{d})}.$$

对于 $j \ge j' \ge j_0$, $\tau_{jj'} \gtrsim 2^{-j'} = 2^{-\min\{j,j'\}}$，我们可以用积分来估计该和：

$$\lesssim 2^{-\alpha(j+j')(\widetilde{d}+n/2)} 2^{jn} \tau_{jj'}^{-\alpha(n+r+2\widetilde{d})+n},$$

~其中最后一步用到了 $\alpha(n+r+2d) > n$。同理，

$$\sum_{k'\in\nabla_{j'}:\mathrm{dist}(S_{j,k},S_{j',k'})\ge\tau_{jj'}} |\langle \mathcal{A}\psi_{j,k}, \psi_{j',k'}\rangle|^\alpha \lesssim 2^{-\alpha(j+j')(\widetilde{d}+n/2)} 2^{j'n} \tau_{jj'}^{-\alpha(n+r+2\widetilde{d})+n}.$$

□

命题 C.2（supp(J,t,t') 的稀疏模式）。对于任意矩阵 B，记 $\mathrm{nnz}(B) := \#\{(i,j) : B_{ij} \neq 0\}$，$\mathrm{nnz}_{\mathrm{row}}(B) := \max_i \#\{j : B_{ij} \neq 0\}$，以及 $\mathrm{nnz}_{\mathrm{col}}(B) := \max_j \#\{i : B_{ij} \neq 0\}$。设 $\mathbf{M}_{(J,t,t')} \in \{0,1\}^{\Lambda_J \times \Lambda_J}$ 如定义 4.2 所定义。则以下估计成立。

(i) 对于$(j,j') \in D_1 \cup D_2$,

$$\mathrm{nnz}\big((\mathbf{M}_{(J,t,t')})_{j,j'}\big) = 0, \qquad \mathrm{nnz}_{\mathrm{row}}\big((\mathbf{M}_{(J,t,t')})_{j,j'}\big) = 0, \qquad \mathrm{nnz}_{\mathrm{col}}\big((\mathbf{M}_{(J,t,t')})_{j,j'}\big) = 0.$$

(ii) 对于$(j,j') \in D_3$, $\tau_{jj'} \asymp 2^{-j'}$,

$$\mathrm{nnz}\big((\mathbf{M}_{(J,t,t')})_{j,j'}\big) \lesssim 2^{jn}, \qquad \mathrm{nnz}_{\mathrm{row}}\big((\mathbf{M}_{(J,t,t')})_{j,j'}\big) \lesssim 1, \qquad \mathrm{nnz}_{\mathrm{col}}\big((\mathbf{M}_{(J,t,t')})_{j,j'}\big) \lesssim 2^{(j-j')n}.$$

(iii) 对于$(j,j') \in D_4$, $\tau_{jj'} \asymp 2^{-j}$,

$$\mathrm{nnz}\big((\mathbf{M}_{(J,t,t')})_{j,j'}\big) \lesssim 2^{j'n}, \qquad \mathrm{nnz}_{\mathrm{row}}\big((\mathbf{M}_{(J,t,t')})_{j,j'}\big) \lesssim 2^{(j'-j)n}, \qquad \mathrm{nnz}_{\mathrm{col}}\big((\mathbf{M}_{(J,t,t')})_{j,j'}\big) \lesssim 1.$$

~ ~ (iv) 对于$(j,j') \in D_5$, $\tau_{jj'} \asymp 2^{(J(t+t'-r)_- - jt' - j't - (j+j')d)/(2d+r)}$,

非零元素数$\big((\mathbf{M}_{(J,t,t')})_{j,j'}\big) \lesssim 2^{(j+j')n}\tau_{jj'}^n$, 非零元素数$_{\mathrm{row}}\big((\mathbf{M}_{(J,t,t')})_{j,j'}\big) \lesssim 2^{j'n}\tau_{jj'}^n$, 非零元素数$_{\mathrm{col}}\big((\mathbf{M}_{(J,t,t')})_{j,j'}\big) \lesssim 2^{jn}\tau_{jj'}^n$. (v) 对于 $(j,j') \in D_6$, $(A_\Lambda^\varepsilon)_{j,j'} = (A_\Lambda)_{j,j'}$,

非零元素数$\big((\mathbf{M}_{(J,t,t')})_{j,j'}\big) \lesssim 2^{(j+j')n}$, $\mathrm{nnz}_{\mathrm{row}}\big((\mathbf{M}_{(J,t,t')})_{j,j'}\big) = 2^{j'n}$, $\mathrm{nnz}_{\mathrm{col}}\big((\mathbf{M}_{(J,t,t')})_{j,j'}\big) \lesssim 2^{jn}$.

此外，

$$\mathrm{nnz}\big(\mathbf{M}_{(J,t,t')}\big) \lesssim 2^{Jn}, \quad \mathrm{nnz}_{\mathrm{row}}\big(\mathbf{M}_{(J,t,t')}\big) \lesssim 2^{Jn\frac{t+t'-r}{\sigma-n/2+t-r/2}}, \quad \mathrm{nnz}_{\mathrm{col}}\big(\mathbf{M}_{(J,t,t')}\big) \lesssim 2^{Jn\frac{t+t'-r}{\sigma-n/2+t'-r/2}}.$$

$\widetilde{\ }$所有隐含常数都可能依赖于 n, t, t', r, σ 以及小波系统 $d, \tilde{d}, \gamma, \tilde{\gamma}$，但与 j, j', J 无关。

证明。(i) 直接由以下对象的构造得出：$\mathbf{M}_{(J,t,t')}$，见定义 4.2：指示矩阵$(\mathbf{M}_{(J,t,t')})_{j,\,j'} = 0$ 当 $(j,\ j') \in D_1 \cup D_2$。对于 (ii)–(v)，使用引理 C.1：当 $j \geq j'$ 且 $\tau_{jj'} \gtrsim 2^{-j'}$时，有

$$\begin{aligned}\big|\{k \in \nabla_j : \mathrm{dist}(S_{j,k}, S_{j',k'}) \leq \tau_{jj'}\}\big| &\lesssim 2^{jn}\tau_{jj'}^n,\\ \big|\{k' \in \nabla_{j'} : \mathrm{dist}(S_{j,k}, S_{j',k'}) \leq \tau_{jj'}\}\big| &\lesssim 2^{j'n}\tau_{jj'}^n,\end{aligned}$$

再加上

$$\tau_{jj'} \asymp \begin{cases} 2^{-j'} & \text{if } (j,j') \in D_3,\\ 2^{-j} & \text{if } (j,j') \in D_4,\\ 2^{(J(t+t'-r)-jt'-j't-(j+j')\widetilde{d})/(2\widetilde{d}+r)} & \text{if } (j,j') \in D_5,\\ \Theta(1) & \text{if } (j,j') \in D_6,\end{cases}$$

由此可得每个区域的相应界。

为证明全局界，我们对所有区域中的 $(j,\ j')$ 求和，累加分块估计。具体而言，对于满足 $j' \leq \frac{t+t'-r}{\widetilde{d}+t} J$ 的第 $\lambda' = (j',\ k')$ 列，其非零元素的数量上界为

$$\begin{aligned}&\sum_{0\leq j\leq \frac{t+t'-r}{\tilde{d}+t'}J-\frac{\tilde{d}+t}{\tilde{d}+t'}j'} 2^{jn} + \sum_{\frac{t+t'-r}{\tilde{d}+t'}J-\frac{\tilde{d}+t}{\tilde{d}+t'}j'\leq j\leq \frac{t+t'-r}{\tilde{d}+t'}J+\frac{\tilde{d}+r-t}{\tilde{d}+t'}j'} 2^{jn}\cdot 2^{n(J(t+t'-r)-jt'-j't-(j+j')\widetilde{d})/(2\widetilde{d}+r)}\\ &\quad + \sum_{\frac{t+t'-r}{\tilde{d}+t'}J+\frac{\tilde{d}+r-t}{\tilde{d}+t'}j'\leq j\leq \frac{t+t'-r}{\sigma-n/2+t'-r/2}J+\frac{\sigma-n/2-(t-r/2)}{\sigma-n/2+t'-r/2}j'} 2^{(j-j')n}\\ &\lesssim 2^{n\left(\frac{t+t'-r}{\tilde{d}+t'}J-\frac{\tilde{d}+t}{\tilde{d}+t'}j'\right)} + 2^{n\frac{t+t'-r}{\tilde{d}+t'}(J-j')} + 2^{n\frac{t+t'-r}{\sigma-n/2+t'-r/2}(J-j')}\\ &\lesssim 2^{Jn(t+t'-r)/(\sigma-\frac{n}{2}+t'-\frac{r}{2})}.\end{aligned}$$

对于具有 $\frac{t+t'-r}{\widetilde{d}+t} J \leq j' \leq \frac{t+t'-r}{\sigma-\frac{n}{2}+t-\frac{r}{2}} J$ 的第 $(j',\ k')$ 列，非零元素的数量至多为

$$\begin{aligned}&\sum_{0\leq j\leq \frac{\tilde{d}+t}{\tilde{d}+r-t'}j'-\frac{t+t'-r}{\tilde{d}+r-t'}J} 1 + \sum_{\frac{\tilde{d}+t}{\tilde{d}+r-t'}j'-\frac{t+t'-r}{\tilde{d}+r-t'}J\leq j\leq \frac{t+t'-r}{\tilde{d}+t'}J+\frac{\tilde{d}+r-t}{\tilde{d}+t'}j'} 2^{jn}\cdot 2^{n(J(t+t'-r)-jt'-j't-(j+j')\widetilde{d})/(2\widetilde{d}+r)}\\ &\quad + \sum_{\frac{t+t'-r}{\tilde{d}+t'}J+\frac{\tilde{d}+r-t}{\tilde{d}+t'}j'\leq j\leq \frac{t+t'-r}{\sigma-n/2+t'-r/2}J+\frac{\sigma-n/2-(t-r/2)}{\sigma-n/2+t'-r/2}j'} 2^{(j-j')n}\\ &\lesssim \frac{\widetilde{d}+t}{\widetilde{d}+r-t'}j' - \frac{t+t'-r}{\widetilde{d}+r-t'}J + 2^{n\frac{t+t'-r}{\tilde{d}+t'}(J-j')} + 2^{n\frac{t+t'-r}{\sigma-n/2+t'-r/2}(J-j')}\\ &\lesssim 2^{Jn(t+t'-r)/(\sigma-\frac{n}{2}+t'-\frac{r}{2})}.\end{aligned}$$

对于具有 $\frac{t+t'-r}{\sigma-\frac{n}{2}+t-\frac{r}{2}}J \le j' \le J$ 的第 $(j',\ k')$ 列，非零元素的数量至多为

$$
\begin{aligned}
&\sum_{\frac{\sigma-n/2+t-r/2}{\sigma-n/2-(t'-r/2)}j'-\frac{t+t'-r}{\sigma-n/2-(t'-r/2)}J\le j\le\frac{\tilde{d}+t}{\tilde{d}+r-t'}j'-\frac{t+t'-r}{\tilde{d}+r-t'}J} 1\\
&\quad+\sum_{\frac{\tilde{d}+t}{\tilde{d}+r-t'}j'-\frac{t+t'-r}{\tilde{d}+r-t'}J\le j\le\frac{t+t'-r}{\tilde{d}+t'}J+\frac{\tilde{d}+r-t}{\tilde{d}+t'}j'} 2^{jn}\cdot 2^{n(J(t+t'-r)-jt'-j't-(j+j')\widetilde{d})/(2\widetilde{d}+r)}\\
&\quad+\sum_{\frac{t+t'-r}{\tilde{d}+t'}J+\frac{\tilde{d}+r-t}{\tilde{d}+t'}j'\le j\le\frac{t+t'-r}{\sigma-n/2+t'-r/2}J+\frac{\sigma-n/2-(t-r/2)}{\sigma-n/2+t'-r/2}j'} 2^{(j-j')n}\\
&\lesssim \frac{\widetilde{d}+t}{\widetilde{d}+r-t'}j'-\frac{t+t'-r}{\widetilde{d}+r-t'}J-\left(\frac{\sigma-\frac{n}{2}+t-\frac{r}{2}}{\sigma-\frac{n}{2}-(t'-\frac{r}{2})}j'-\frac{t+t'-r}{\sigma-\frac{n}{2}-(t'-\frac{r}{2})}J\right)\\
&\quad+2^{n\frac{t+t'-r}{\tilde{d}+t'}(J-j')}+2^{n\frac{t+t'-r}{\sigma-n/2+t'-r/2}(J-j')}\\
&\lesssim 2^{Jn(t+t'-r)/(\sigma-\frac{n}{2}+t'-\frac{r}{2})}.
\end{aligned}
$$

合并这三种情况可得

$$\mathrm{nnz}_{\mathrm{col}}\big(\mathbf{M}_{(J,t,t')}\big)\lesssim 2^{Jn(t+t'-r)/(\sigma-\frac{n}{2}+t'-\frac{r}{2})}.$$

同理,

$$\mathrm{nnz}_{\mathrm{row}}\big(\mathbf{M}_{(J,t,t')}\big)\lesssim 2^{Jn(t+t'-r)/(\sigma-\frac{n}{2}+t-\frac{r}{2})}.$$

最后,

$$
\begin{aligned}
&\mathrm{nnz}\big(\mathbf{M}_{(J,t,t')}\big)\lesssim\sum_{j'=0}^{J}2^{j'n}\cdot\mathrm{nnz}((\mathbf{M}_{(J,t,t')})_{\cdot,\lambda'})\\
&\lesssim\sum_{0\le j'\le\frac{t+t'-r}{\tilde{d}+t}J}2^{j'n}\left(2^{n\left(\frac{t+t'-r}{\tilde{d}+t'}J-\frac{\tilde{d}+t}{\tilde{d}+t'}j'\right)}+2^{n\frac{t+t'-r}{\sigma-n/2+t'-r/2}(J-j')}\right)\\
&\quad+\sum_{\frac{t+t'-r}{\tilde{d}+t}J\le j'\le\frac{t+t'-r}{\sigma-n/2+t-r/2}J}2^{j'n}\left(\frac{\widetilde{d}+t}{\widetilde{d}+r-t'}j'-\frac{t+t'-r}{\widetilde{d}+r-t'}J+2^{n\frac{t+t'-r}{\sigma-n/2+t'-r/2}(J-j')}\right)\\
&\quad+\sum_{\frac{t+t'-r}{\sigma-n/2+t-r/2}J\le j'\le J}2^{j'n}\left(\frac{\widetilde{d}+t}{\widetilde{d}+r-t'}j'-\frac{t+t'-r}{\widetilde{d}+r-t'}J-\left(\frac{\sigma-\frac{n}{2}+t-\frac{r}{2}}{\sigma-\frac{n}{2}-(t'-\frac{r}{2})}j'-\frac{t+t'-r}{\sigma-\frac{n}{2}-(t'-\frac{r}{2})}J\right)\right)\\
&\quad+\sum_{\frac{t+t'-r}{\sigma-n/2+t-r/2}J\le j'\le J}2^{j'n}\cdot 2^{n\frac{t+t'-r}{\sigma-n/2+t'-r/2}(J-j')}\\
&\lesssim 2^{Jn},
\end{aligned}
$$

其中我们使用了不等式 $\sum_{0\le j'\le J}2^{j'n}(J-j')\lesssim 2^{Jn}$。 □

C.2 截断

命题 *5.9* 的证明。 我们将 $(\cdot)_{j,j'}$ 记作 (j,j') 小波块。根据引理 C.1 (i)，结合$\alpha=1$，对任意 $j,j'\geq j_0$,

$$\begin{aligned}\|\mathbf{A}_{j,j'}\| &\leq \|\mathbf{A}_{j,j'}\|_1^{1/2}\,\|\mathbf{A}_{j,j'}\|_\infty^{1/2}\\ &\leq \sup_{k'\in\nabla_{j'}}\left(\sum_{k\in\nabla_j}|\langle\mathcal{A}\psi_{j,k},\psi_{j',k'}\rangle|\right)^{1/2}\cdot\sup_{k\in\nabla_j}\left(\sum_{k'\in\nabla_{j'}}|\langle\mathcal{A}\psi_{j,k},\psi_{j',k'}\rangle|\right)^{1/2}\\ &\lesssim \left(2^{(j+j')r/2}\cdot 2^{-|j-j'|\sigma}2^{|j-j'|n}\cdot 2^{(j+j')r/2}\cdot 2^{-|j-j'|\sigma}\right)^{1/2}\\ &\asymp 2^{(j+j')r/2-|j-j'|(\sigma-n/2)}.\end{aligned}$$

现在我们将引理 C.3 应用于 $\mathbf{D}^{-t'}(\mathbf{A}_{\Lambda_J}^{\uparrow}-\mathbf{A})\mathbf{D}^{-t}$，采用自然的 (j,j')-分块划分。令 $\mathcal{N}(\cdot)$ 表示分块范数矩阵。于是

$$\begin{aligned}\|\mathbf{D}^{-t'}(\mathbf{A}_{\Lambda_J}^{\uparrow}-\mathbf{A})\mathbf{D}^{-t}\| &\leq \|\mathcal{N}(\mathbf{D}^{-t'}(\mathbf{A}_{\Lambda_J}^{\uparrow}-\mathbf{A})\mathbf{D}^{-t})\|\\ &\leq \|\mathcal{N}(\mathbf{D}^{-t'}(\mathbf{A}_{\Lambda_J}^{\uparrow}-\mathbf{A})\mathbf{D}^{-t})\|_1^{1/2}\,\|\mathcal{N}(\mathbf{D}^{-t'}(\mathbf{A}_{\Lambda_J}^{\uparrow}-\mathbf{A})\mathbf{D}^{-t})\|_\infty^{1/2}.\end{aligned}$$

对于 $j,j'\geq j_0$,

$$\left(\mathcal{N}(\mathbf{D}^{-t'}(\mathbf{A}_{\Lambda_J}^{\uparrow}-\mathbf{A})\mathbf{D}^{-t})\right)_{j,j'}=\|(\mathbf{D}^{-t'}(\mathbf{A}_{\Lambda_J}^{\uparrow}-\mathbf{A})\mathbf{D}^{-t})_{j,j'}\|=\begin{cases}0 & \text{if } j,j'\leq J,\\ 2^{-jt'-j't}\|\mathbf{A}_{j,j'}\| & \text{otherwise.}\end{cases}$$

我们首先估计 $\|\mathcal{N}(\mathbf{D}^{-t'}(\mathbf{A}_{\Lambda_J}^{\uparrow}-\mathbf{A})\mathbf{D}^{-t})\|_1$。分为以下两种情况： $j'\leq J$和 $j'>J$:

$$\|\mathcal{N}(\mathbf{D}^{-t'}(\mathbf{A}_{\Lambda_J}^{\uparrow}-\mathbf{A})\mathbf{D}^{-t})\|_1\leq\left(\sup_{j'\leq J}\sum_{j>J}2^{-jt'-j't}\|\mathbf{A}_{j,j'}\|\right)+\left(\sup_{j'>J}\sum_{j\geq j_0}2^{-jt'-j't}\|\mathbf{A}_{j,j'}\|\right).$$

情况 1： $j'\leq J$。 利用 $\|\mathbf{A}_{j,j'}\|$的界,

$$\begin{aligned}\sup_{j'\leq J}\sum_{j>J}2^{-jt'-j't}\|\mathbf{A}_{j,j'}\| &\lesssim \sup_{j'\leq J}\sum_{j>J}2^{-jt'-j't}\cdot 2^{(j+j')r/2-|j-j'|(\sigma-n/2)}\\ &= \sup_{j'\leq J}2^{j'(\sigma-n/2-t+r/2)}\sum_{j>J}2^{-j(\sigma-n/2+t'-r/2)}\\ &\lesssim 2^{J(\sigma-n/2-t+r/2)}\cdot 2^{-J(\sigma-n/2+t'-r/2)}\\ &= 2^{-J(t+t'-r)}.\end{aligned}$$

情形 2： $j'>J$。 利用对 $\|\mathbf{A}_{j,j'}\|$ 的估计，并将总和分解为两个区域$j_0\leq j\leq j'$ 和 $j>j'$，我们得到

$$\begin{aligned}&\sup_{j'>J}\sum_{j\geq j_0}2^{-jt'-j't}\|\mathbf{A}_{j,j'}\|\lesssim\sup_{j'>J}\sum_{j\geq j_0}2^{-jt'-j't}\cdot 2^{(j+j')r/2-|j-j'|(\sigma-n/2)}\\ &=\sup_{j'>J}\left(\sum_{j_0\leq j\leq j'}+\sum_{j>j'}\right)2^{-jt'-j't+(j+j')r/2-|j-j'|(\sigma-n/2)}\\ &=\sup_{j'>J}\left(\sum_{j_0\leq j\leq j'}2^{j'(-t+r/2-\sigma+n/2)}\cdot 2^{j(\sigma-n/2-(t'-r/2))}+\sum_{j>j'}2^{j'(-t+r/2+\sigma-n/2)}\cdot 2^{-j(\sigma-n/2+t'-r/2)}\right)\\ &\lesssim\sup_{j'>J}\left(2^{j'(-t+r/2-\sigma+n/2)}\cdot 2^{j'(\sigma-n/2-(t'-r/2))}+2^{j'(-t+r/2+\sigma-n/2)}\cdot 2^{-j'(\sigma-n/2+t'-r/2)}\right)\\ &\asymp 2^{-J(t+t'-r)},\end{aligned}$$

其中我们用到了 $\sigma - n/2 > \max\{t, t'\} - r/2 \geq 0$。

合并两种情况，我们得出

$$\|\mathcal{N}(\mathbf{D}^{-t'}(\mathbf{A}^{\uparrow}_{\Lambda_J} - \mathbf{A})\mathbf{D}^{-t})\|_1 \lesssim 2^{-J(t+t'-r)}.$$

对于 $\|\mathcal{N}(\cdot)\|_\infty$ 的界同理可得，因此

$$\|\mathcal{N}(\mathbf{D}^{-t'}(\mathbf{A}^{\uparrow}_{\Lambda_J} - \mathbf{A})\mathbf{D}^{-t})\|_\infty \lesssim 2^{-J(t+t'-r)}.$$

因此，

$$\begin{aligned}\text{Error-Truncation} &= \|\mathbf{D}^{-t'}(\mathbf{A}^{\uparrow}_{\Lambda_J} - \mathbf{A})\mathbf{D}^{-t}\| \\ &\leq \|\mathcal{N}(\mathbf{D}^{-t'}(\mathbf{A}^{\uparrow}_{\Lambda_J} - \mathbf{A})\mathbf{D}^{-t})\|_1^{1/2} \, \|\mathcal{N}(\mathbf{D}^{-t'}(\mathbf{A}^{\uparrow}_{\Lambda_J} - \mathbf{A})\mathbf{D}^{-t})\|_\infty^{1/2} \\ &\lesssim 2^{-J(t+t'-r)}.\end{aligned}$$

□

C.3 压缩

命题 *5.10* 的证明。 我们先考虑以下情况： $t_\ell = t'$， $t_r = t$。回顾

$$\text{Error-Compression} = \|\mathbf{D}^{-t'}_{\Lambda_J}(\mathbf{A}^{\varepsilon}_{\Lambda_J} - \mathbf{A}_{\Lambda_J})\mathbf{D}^{-t}_{\Lambda_J}\|.$$

我们应用范数压缩不等式（引理 C.3），采用自然的 $(j,\ j')$-分块划分，并用 $(\cdot)_{j,\ j'}$ 表示 $(j,\ j')$-分块：

$$\|\mathbf{D}^{-t'}_{\Lambda_J}(\mathbf{A}^{\varepsilon}_{\Lambda_J} - \mathbf{A}_{\Lambda_J})\mathbf{D}^{-t}_{\Lambda_J}\| \leq \left(\sum_{j,j'=j_0}^{J} 2^{-2jt'-2j't}\|(\mathbf{A}^{\varepsilon}_{\Lambda_J})_{j,j'} - (\mathbf{A}_{\Lambda_J})_{j,j'}\|^2\right)^{1/2}.$$

接下来，我们分析 $\|(\mathbf{A}^{\varepsilon}_{\Lambda_J})_{j,j'} - (\mathbf{A}_{\Lambda_J})_{j,j'}\|$ 在 $j_0 \leq j, j' \leq J$ 下。

情况 1。若

$$j > j' \text{ and } j \geq \frac{t+t'-r}{\sigma - \frac{n}{2} + t' - \frac{r}{2}} J + \frac{\sigma - \frac{n}{2} - (t - \frac{r}{2})}{\sigma - \frac{n}{2} + t' - \frac{r}{2}} j',$$

则 $(\mathbf{A}^{\varepsilon}_{\Lambda_J})_{j,j'} = 0$，因此 $\|(\mathbf{A}^{\varepsilon}_{\Lambda_J})_{j,j'} - (\mathbf{A}_{\Lambda_J})_{j,j'}\| = \|(\mathbf{A}_{\Lambda_J})_{j,j'}\|$。利用矩阵范数不等式 $\|B\| \leq \|B\|_1^{1/2} \|B\|_\infty^{1/2}$ 以及引理 C.1 (i)（取 $\alpha = 1$），我们得到

$$\begin{aligned}\|(\mathbf{A}_{\Lambda_J})_{j,j'}\| &\leq \|(\mathbf{A}_{\Lambda_J})_{j,j'}\|_1^{1/2} \cdot \|(\mathbf{A}_{\Lambda_J})_{j,j'}\|_\infty^{1/2} \\ &= \sup_{k' \in \nabla_{j'}} \left(\sum_{k \in \nabla_j} |\langle \mathcal{A}\psi_{j,k}, \psi_{j',k'}\rangle|\right)^{1/2} \cdot \sup_{k \in \nabla_j} \left(\sum_{k' \in \nabla_{j'}} |\langle \mathcal{A}\psi_{j,k}, \psi_{j',k'}\rangle|\right)^{1/2} \\ &\lesssim \left(2^{(j+j')r/2} \cdot 2^{-(j-j')\sigma} 2^{(j-j')n} \cdot 2^{(j+j')r/2} \cdot 2^{-(j-j')\sigma}\right)^{1/2} \\ &= 2^{(j+j')r/2 - (j-j')(\sigma - n/2)} \\ &\leq 2^{-J(t+t'-r) + jt' + j't},\end{aligned}$$

其中，最后一个不等式由给定的斜率条件推出：

$$j \geq \frac{t+t'-r}{\sigma - \frac{n}{2} + t' - \frac{r}{2}} J + \frac{\sigma - \frac{n}{2} - (t - \frac{r}{2})}{\sigma - \frac{n}{2} + t' - \frac{r}{2}} j'.$$

情形 2。若

$$j' > j \text{ and } j' \geq \frac{t+t'-r}{\sigma-\frac{n}{2}+t-\frac{r}{2}}J + \frac{\sigma-\frac{n}{2}-(t'-\frac{r}{2})}{\sigma-\frac{n}{2}+t-\frac{r}{2}}j,$$

于是，按照相同的推理，

$$\|(\mathbf{A}^{\varepsilon}_{\Lambda_J})_{j,j'} - (\mathbf{A}_{\Lambda_J})_{j,j'}\| = \|(\mathbf{A}_{\Lambda_J})_{j,j'}\| \lesssim 2^{(j+j')r/2-(j'-j)(\sigma-n/2)} \lesssim 2^{-J(t+t'-r)+jt'+j't}.$$

情形 3。 在剩余的块中，我们将所有满足 $\operatorname{dist}(S_{j,k},S_{j',k'}) > \tau_{jj'}$ 的元素置为零。因此

$$\|(\mathbf{A}^{\varepsilon}_{\Lambda_J})_{j,j'} - (\mathbf{A}_{\Lambda_J})_{j,j'}\| \leq \|(\mathbf{A}^{\varepsilon}_{\Lambda_J})_{j,j'} - (\mathbf{A}_{\Lambda_J})_{j,j'}\|_1^{1/2} \cdot \|(\mathbf{A}^{\varepsilon}_{\Lambda_J})_{j,j'} - (\mathbf{A}_{\Lambda_J})_{j,j'}\|_\infty^{1/2},$$

with

$$\begin{aligned}\|(\mathbf{A}^{\varepsilon}_{\Lambda_J})_{j,j'} - (\mathbf{A}_{\Lambda_J})_{j,j'}\|_1 &= \sup_{k'\in\nabla_{j'}} \sum_{k\in\nabla_j:\operatorname{dist}(S_{j,k},S_{j',k'})\geq\tau_{jj'}} |\langle \mathcal{A}\psi_{j,k}, \psi_{j',k'}\rangle| \\ &\lesssim 2^{-(j+j')(\widetilde{d}+n/2)} 2^{jn} \tau_{jj'}^{-(r+2\widetilde{d})},\end{aligned}$$

and

$$\begin{aligned}\|(\mathbf{A}^{\varepsilon}_{\Lambda_J})_{j,j'} - (\mathbf{A}_{\Lambda_J})_{j,j'}\|_\infty &= \sup_{k\in\nabla_j} \sum_{k'\in\nabla_{j'}:\operatorname{dist}(S_{j,k},S_{j',k'})\geq\tau_{jj'}} |\langle \mathcal{A}\psi_{j,k}, \psi_{j',k'}\rangle| \\ &\lesssim 2^{-(j+j')(\widetilde{d}+n/2)} 2^{j'n} \tau_{jj'}^{-(r+2\widetilde{d})},\end{aligned}$$

由引理 C.1(iii)。因此，

$$\|(\mathbf{A}^{\varepsilon}_{\Lambda_J})_{j,j'} - (\mathbf{A}_{\Lambda_J})_{j,j'}\| \lesssim 2^{-(j+j')\widetilde{d}} \cdot \tau_{jj'}^{-(2\widetilde{d}+r)} \lesssim 2^{-J(t+t'-r)+jt'+j't},$$

~ ~其中，最后一个不等式用到了 $\tau_{jj'} \gtrsim 2^{(J(t+t'-r}{}_{)-jt'-j't-(j+j')d)/(2d+r)}$。

合并**情形 1–3**可得，

$$\begin{aligned}\|\mathbf{D}_{\Lambda_J}^{-t'}(\mathbf{A}^{\varepsilon}_{\Lambda_J} - \mathbf{A}_{\Lambda_J})\mathbf{D}_{\Lambda_J}^{-t}\| &\leq \Big(\sum_{j,j'=j_0}^{J} 2^{-2jt'-2j't}\|(\mathbf{A}^{\varepsilon}_{\Lambda_J})_{j,j'} - (\mathbf{A}_{\Lambda_J})_{j,j'}\|^2\Big)^{1/2} \\ &\lesssim \Big(\sum_{j,j'=j_0}^{J} 2^{-2jt'-2j't} 2^{-2J(t+t'-r)+2jt'+2j't}\Big)^{1/2} \\ &\lesssim J\, 2^{-J(t+t'-r)},\end{aligned}$$

这正是所要证明的。

对于满足 $r/2 \leq t_\ell, t_r < \gamma$ 的一般权重 t_ℓ 和 t_r，我们有

$$\begin{aligned}\|\mathbf{D}_{\Lambda_J}^{-t_\ell}(\mathbf{A}^{\varepsilon}_{\Lambda_J} - \mathbf{A}_{\Lambda_J})\mathbf{D}_{\Lambda_J}^{-t_r}\| &= \|\mathbf{D}_{\Lambda_J}^{t'-t_\ell}\,\mathbf{D}_{\Lambda_J}^{-t'}(\mathbf{A}^{\varepsilon}_{\Lambda_J} - \mathbf{A}_{\Lambda_J})\mathbf{D}_{\Lambda_J}^{-t}\,\mathbf{D}_{\Lambda_J}^{t-t_r}\| \\ &\leq \|\mathbf{D}_{\Lambda_J}^{t'-t_\ell}\| \cdot \|\mathbf{D}_{\Lambda_J}^{-t'}(\mathbf{A}^{\varepsilon}_{\Lambda_J} - \mathbf{A}_{\Lambda_J})\mathbf{D}_{\Lambda_J}^{-t}\| \cdot \|\mathbf{D}_{\Lambda_J}^{t-t_r}\| \\ &\lesssim J\, 2^{-J(t+t'-r)} \cdot 2^{J\max\{0,t'-t_\ell\}} \cdot 2^{J\max\{0,t-t_r\}} \\ &\asymp J 2^{-J\left(\min\{t',t_\ell\}+\min\{t,t_r\}-r\right)}.\end{aligned}$$

□

C.4 估计

C.4.1 误差分解

引理的证明 *5.8.* 回顾一下，估计量 $\widehat{\mathbf{A}} \in \mathbb{R}^{\Lambda_J \times \Lambda_J}$ 式（4.11）按元素定义为

$$\widehat{\mathbf{A}}_{\lambda,\lambda'} := \begin{cases} (\widehat{\mathbf{A}}^{(1)})_{\lambda,\lambda'} & \text{if } (\lambda,\lambda') \in \operatorname{supp}(J,t,t') \text{ and } j \le j', \\ (\widehat{\mathbf{A}}^{(1)})_{\lambda',\lambda} & \text{if } (\lambda,\lambda') \in \operatorname{supp}(J,t,t') \text{ and } j > j', \\ 0 & \text{if } (\lambda,\lambda') \notin \operatorname{supp}(J,t,t'). \end{cases}$$

由于 $\widehat{\mathbf{A}}^{(1)} \in \mathrm{R}^{\widetilde{\Lambda_J} \times \Lambda_J}$ 是矩形矩阵，我们引入一个中间方阵

$$\widehat{\mathbf{A}}^{(2)} := (\widehat{\mathbf{A}}^{(1)})_{\Lambda_J,\cdot} \in \mathbb{R}^{\Lambda_J \times \Lambda_J},$$

该矩阵通过将 $\widehat{\mathbf{A}}^{(1)}$ 限制到由 Λ_J 指标标记的行上得到。

为便于分析，我们根据尺度索引的顺序，将任意矩阵 $\mathbf{B} \in \mathbb{R}^{\Lambda_J \times \Lambda_J}$ 分解为其上三角部分和下三角 部分:

$$(\mathbf{B}_{\mathrm{up}})_{\lambda,\lambda'} := \begin{cases} \mathbf{B}_{\lambda,\lambda'} & j \le j', \\ 0 & j > j', \end{cases} \qquad (\mathbf{B}_{\mathrm{low}})_{\lambda,\lambda'} := \begin{cases} 0 & j \le j', \\ \mathbf{B}_{\lambda,\lambda'} & j > j'. \end{cases}$$

因此，$\mathbf{B} = \mathbf{B}_{\mathrm{up}} + \mathbf{B}_{\mathrm{low}}$。我们将此分解应用于 $\widehat{\mathbf{A}}$ $\widehat{\mathbf{A}}^{(2)}$，$\mathbf{A}_{\Lambda_J}$，$\mathbf{A}^{\varepsilon}_{\Lambda_J}$ 以及掩码矩阵$\mathbf{M}_{(J,t,t')}$。

$\widehat{\mathbf{A}}$根据构造（4.11），$\mathbf{A}$ 的上三角部分和下三角部分满足

$$\widehat{\mathbf{A}}_{\mathrm{up}} = (\mathbf{M}_{(J,t,t')})_{\mathrm{up}} \odot \widehat{\mathbf{A}}^{(2)}, \qquad \widehat{\mathbf{A}}_{\mathrm{low}} = (\mathbf{M}_{(J,t,t')})_{\mathrm{low}} \odot (\widehat{\mathbf{A}}^{(2)})^{\top}.$$

由于压缩矩阵由 $\mathbf{A}^{\varepsilon}_{\Lambda_J} = \mathbf{M}_{(J,t,t')} \odot \mathbf{A}_{\Lambda_J}$ 给出，因此同样有

$$(\mathbf{A}^{\varepsilon}_{\Lambda_J})_{\mathrm{up}} = (\mathbf{M}_{(J,t,t')})_{\mathrm{up}} \odot \mathbf{A}_{\Lambda_J}, \qquad (\mathbf{A}^{\varepsilon}_{\Lambda_J})_{\mathrm{low}} = (\mathbf{M}_{(J,t,t')})_{\mathrm{low}} \odot \mathbf{A}_{\Lambda_J}.$$

因此我们得到

$$\begin{aligned} &\|\mathbf{D}_{\Lambda_J}^{-t'}(\widehat{\mathbf{A}} - \mathbf{A}^{\varepsilon}_{\Lambda_J})\mathbf{D}_{\Lambda_J}^{-t}\| \\ &= \left\|\mathbf{D}_{\Lambda_J}^{-t'}\left(\widehat{\mathbf{A}}_{\mathrm{up}} + \widehat{\mathbf{A}}_{\mathrm{low}} - (\mathbf{A}^{\varepsilon}_{\Lambda_J})_{\mathrm{up}} - (\mathbf{A}^{\varepsilon}_{\Lambda_J})_{\mathrm{low}}\right)\mathbf{D}_{\Lambda_J}^{-t}\right\| \\ &\le \left\|\mathbf{D}_{\Lambda_J}^{-t'}\left(\widehat{\mathbf{A}}_{\mathrm{up}} - (\mathbf{A}^{\varepsilon}_{\Lambda_J})_{\mathrm{up}}\right)\mathbf{D}_{\Lambda_J}^{-t}\right\| + \left\|\mathbf{D}_{\Lambda_J}^{-t'}\left(\widehat{\mathbf{A}}_{\mathrm{low}} - (\mathbf{A}^{\varepsilon}_{\Lambda_J})_{\mathrm{low}}\right)\mathbf{D}_{\Lambda_J}^{-t}\right\| \\ &= \left\|\mathbf{D}_{\Lambda_J}^{-t'}(\mathbf{M}_{(J,t,t')})_{\mathrm{up}} \odot \left(\widehat{\mathbf{A}}^{(2)} - \mathbf{A}_{\Lambda_J}\right)\mathbf{D}_{\Lambda_J}^{-t}\right\| + \left\|\mathbf{D}_{\Lambda_J}^{-t'}(\mathbf{M}_{(J,t,t')})_{\mathrm{low}} \odot \left((\widehat{\mathbf{A}}^{(2)})^{\top} - \mathbf{A}_{\Lambda_J}\right)\mathbf{D}_{\Lambda_J}^{-t}\right\| \\ &\overset{(\star)}{=} \left\|\mathbf{D}_{\Lambda_J}^{-t'}(\mathbf{M}_{(J,t,t')})_{\mathrm{up}} \odot \left(\widehat{\mathbf{A}}^{(2)} - \mathbf{A}_{\Lambda_J}\right)\mathbf{D}_{\Lambda_J}^{-t}\right\| + \left\|\mathbf{D}_{\Lambda_J}^{-t}((\mathbf{M}_{(J,t,t')})_{\mathrm{low}})^{\top} \odot \left(\widehat{\mathbf{A}}^{(2)} - \mathbf{A}_{\Lambda_J}\right)\mathbf{D}_{\Lambda_J}^{-t'}\right\|, \end{aligned} \tag{C.1}$$

其中，$(\star)$ 是通过转置第二项并利用 $(\mathbf{A}_{\Lambda_J})^{\top} = \mathbf{A}_{\Lambda_J}$ 得到的。注意，这两项都化为涉及 $\widehat{\mathbf{A}}^{(2)} - \mathbf{A}_{\Lambda_J}$ 的类似结构误差。

回顾一下，回归支撑集 $\Omega_{\lambda'} = \left\{\lambda \in \mathcal{J} : (\lambda,\lambda') \in \text{支撑集}(\widetilde{J},\widetilde{t},\widetilde{t'})\right\}$, $\Omega^c_{\lambda'} = \mathcal{J} \setminus \Omega_{\lambda'}$,and (4.8):

$$\mathbf{F}_{\cdot,\lambda'} = \mathbf{U}\mathbf{A}_{\cdot,\lambda'} + \mathbf{W}_{\cdot,\lambda'} = \mathbf{U}_{\cdot,\Omega_{\lambda'}}\mathbf{A}_{\Omega_{\lambda'},\lambda'} + \mathbf{U}_{\cdot,\Omega^c_{\lambda'}}\mathbf{A}_{\Omega^c_{\lambda'},\lambda'} + \mathbf{W}_{\cdot,\lambda'},$$

并且，我们在 **步骤 1**(4.10) 中的初步估计量按列定义为

$$(\widehat{\mathbf{A}}^{(1)})_{\cdot,\lambda'} := \begin{bmatrix} (\mathbf{U}_{\cdot,\Omega_{\lambda'}}^\top \mathbf{U}_{\cdot,\Omega_{\lambda'}})^{-1}\mathbf{U}_{\cdot,\Omega_{\lambda'}}^\top \mathbf{F}_{\cdot,\lambda'} \\ \mathbf{0} \end{bmatrix} \in \mathbb{R}^{\Lambda_J^\sim}, \qquad \lambda' \in \Lambda_J.$$

注意到

$$\begin{aligned}(\mathbf{U}_{\cdot,\Omega_{\lambda'}}^\top \mathbf{U}_{\cdot,\Omega_{\lambda'}})^{-1}\mathbf{U}_{\cdot,\Omega_{\lambda'}}^\top \mathbf{F}_{\cdot,\lambda'} &= (\mathbf{U}_{\cdot,\Omega_{\lambda'}}^\top \mathbf{U}_{\cdot,\Omega_{\lambda'}})^{-1}\mathbf{U}_{\cdot,\Omega_{\lambda'}}^\top \Big(\mathbf{U}_{\cdot,\Omega_{\lambda'}}\mathbf{A}_{\Omega_{\lambda'},\lambda'} + \mathbf{U}_{\cdot,\Omega_{\lambda'}^c}\mathbf{A}_{\Omega_{\lambda'}^c,\lambda'} + \mathbf{W}_{\cdot,\lambda'}\Big) \\ &= \mathbf{A}_{\Omega_{\lambda'},\lambda'} + (\mathbf{U}_{\cdot,\Omega_{\lambda'}}^\top \mathbf{U}_{\cdot,\Omega_{\lambda'}})^{-1}\mathbf{U}_{\cdot,\Omega_{\lambda'}}^\top \mathbf{U}_{\cdot,\Omega_{\lambda'}^c}\mathbf{A}_{\Omega_{\lambda'}^c,\lambda'} + (\mathbf{U}_{\cdot,\Omega_{\lambda'}}^\top \mathbf{U}_{\cdot,\Omega_{\lambda'}})^{-1}\mathbf{U}_{\cdot,\Omega_{\lambda'}}^\top \mathbf{W}_{\cdot,\lambda'}.\end{aligned}$$

因此，对于每一列 $\lambda' \in \Lambda_J$,

$$(\widehat{\mathbf{A}}^{(1)})_{\cdot,\lambda'} = \begin{bmatrix} \mathbf{A}_{\Omega_{\lambda'},\lambda'} \\ \mathbf{0} \end{bmatrix} + \begin{bmatrix} (\mathbf{U}_{\cdot,\Omega_{\lambda'}}^\top \mathbf{U}_{\cdot,\Omega_{\lambda'}})^{-1}\mathbf{U}_{\cdot,\Omega_{\lambda'}}^\top \mathbf{U}_{\cdot,\Omega_{\lambda'}^c}\mathbf{A}_{\Omega_{\lambda'}^c,\lambda'} \\ \mathbf{0} \end{bmatrix} + \begin{bmatrix} (\mathbf{U}_{\cdot,\Omega_{\lambda'}}^\top \mathbf{U}_{\cdot,\Omega_{\lambda'}})^{-1}\mathbf{U}_{\cdot,\Omega_{\lambda'}}^\top \mathbf{W}_{\cdot,\lambda'} \\ \mathbf{0} \end{bmatrix}.$$

设 $E_{\Omega_{\lambda'}} : \mathrm{R}^{\Omega_{\lambda'}} \to \mathrm{R}^{\Lambda_J^\sim}$为行嵌入算子（在 $\Omega_{\lambda'}$ 之外补零）。我们定义

$$\mathsf{A} := \sum_{\lambda' \in \Lambda_J} \big(E_{\Omega_{\lambda'}}\mathbf{A}_{\Omega_{\lambda'},\lambda'}\big)e_{\lambda'}^\top \in \mathbb{R}^{\Lambda_J^\sim \times \Lambda_J},$$

$$\mathsf{OVB} := \sum_{\lambda' \in \Lambda_J} \Big(E_{\Omega_{\lambda'}}(\mathbf{U}_{\cdot,\Omega_{\lambda'}}^\top \mathbf{U}_{\cdot,\Omega_{\lambda'}})^{-1}\mathbf{U}_{\cdot,\Omega_{\lambda'}}^\top \mathbf{U}_{\cdot,\Omega_{\lambda'}^c}\mathbf{A}_{\mathbf{\Omega_{\lambda'}^c},\lambda'}\Big)e_{\lambda'}^\top \in \mathbb{R}^{\Lambda_J^\sim \times \Lambda_J},$$

and

$$\mathsf{Var} := \sum_{\lambda' \in \Lambda_J} \Big(E_{\Omega_{\lambda'}}(\mathbf{U}_{\cdot,\Omega_{\lambda'}}^\top \mathbf{U}_{\cdot,\Omega_{\lambda'}})^{-1}\mathbf{U}_{\cdot,\Omega_{\lambda'}}^\top \mathbf{W}_{\cdot,\lambda'}\Big)e_{\lambda'}^\top \in \mathbb{R}^{\Lambda_J^\sim \times \Lambda_J},$$

其中 $e_{\lambda'} \in \mathrm{R}^{\Lambda_J}$ 是标准基向量，其唯一非零元素为坐标 λ' 处的 1。因此,

$$\widehat{\mathbf{A}}^{(1)} = \mathsf{A} + \mathsf{OVB} + \mathsf{Var} \in \mathbb{R}^{\Lambda_J^\sim \times \Lambda_J}.$$

取指标为 Λ_J 的各行，我们有

$$\widehat{\mathbf{A}}^{(2)} = (\widehat{\mathbf{A}}^{(1)})_{\Lambda_J,\cdot} = \mathsf{A}_{\Lambda_J,\cdot} + \mathsf{OVB}_{\Lambda_J,\cdot} + \mathsf{Var}_{\Lambda_J,\cdot} \in \mathbb{R}^{\Lambda_J \times \Lambda_J}.$$

由于对于每个 $\lambda' \in \Lambda_J$，M_J 的t 第t' 列 λ' 的支撑集包含于 $\Omega_{\lambda'}$，且 $\mathsf{A}_{\cdot}$ 与λ' 在 $\Omega_{\lambda'}$ 上与 $(\mathbf{A}^{\Lambda_J})_{\cdot}$ 相同λ'，因此我们有

$$\mathbf{M}_{(J,t,t')} \odot (\mathsf{A}_{\Lambda_J,\cdot} - \mathbf{A}_{\Lambda_J}) = 0.$$

因此,

$$\begin{aligned}(\mathbf{M}_{(J,t,t')})_{\mathrm{up}} \odot \Big(\widehat{\mathbf{A}}^{(2)} - \mathbf{A}_{\Lambda_J}\Big) &= (\mathbf{M}_{(J,t,t')})_{\mathrm{up}} \odot (\mathsf{A}_{\Lambda_J,\cdot} + \mathsf{OVB}_{\Lambda_J,\cdot} + \mathsf{Var}_{\Lambda_J,\cdot} - \mathbf{A}_{\Lambda_J}) \\ &= (\mathbf{M}_{(J,t,t')})_{\mathrm{up}} \odot (\mathsf{OVB}_{\Lambda_J,\cdot} + \mathsf{Var}_{\Lambda_J,\cdot}). \qquad \text{(C.2)}\end{aligned}$$

此外，根据引理4.5，$((\mathbf{M}_{(J,t,t')})_{\mathrm{low}})^\top$ 的支撑集包含于$(\mathbf{M}_{(J,t,t')})_{\mathrm{up}}$的支撑集。由于我们已经有$(\mathbf{M}_{(J,t,t')})_{\mathrm{up}} \odot (\mathbf{A}_{\Lambda_J,\cdot} - \mathbf{A}_{\Lambda_J}) = 0$，因此

$$((\mathbf{M}_{(J,t,t')})_{\mathrm{low}})^\top \odot (\mathsf{A}_{\Lambda_J,\cdot} - \mathbf{A}_{\Lambda_J}) = 0.$$

因此,

$$\begin{aligned}((\mathbf{M}_{(J,t,t')})_{\mathrm{low}})^\top \odot \Big(\widehat{\mathbf{A}}^{(2)} - \mathbf{A}_{\Lambda_J}\Big) &= ((\mathbf{M}_{(J,t,t')})_{\mathrm{low}})^\top \odot (\mathsf{A}_{\Lambda_J,\cdot} + \mathsf{OVB}_{\Lambda_J,\cdot} + \mathsf{Var}_{\Lambda_J,\cdot} - \mathbf{A}_{\Lambda_J}) \\ &= ((\mathbf{M}_{(J,t,t')})_{\mathrm{low}})^\top \odot (\mathsf{OVB}_{\Lambda_J,\cdot} + \mathsf{Var}_{\Lambda_J,\cdot}). \qquad \text{(C.3)}\end{aligned}$$

合并（C.1）、（C.2）和（C.3），我们得到估计误差的如下上界：

$$
\begin{aligned}
&\|\mathbf{D}_{\Lambda_J}^{-t'}(\widehat{\mathbf{A}}-\mathbf{A}_{\Lambda_J}^{\varepsilon})\mathbf{D}_{\Lambda_J}^{-t}\| \\
&\le \left\|\mathbf{D}_{\Lambda_J}^{-t'}(\mathbf{M}_{(J,t,t')})_{\text{up}}\odot\left(\widehat{\mathbf{A}}^{(2)}-\mathbf{A}_{\Lambda_J}\right)\mathbf{D}_{\Lambda_J}^{-t}\right\|+\left\|\mathbf{D}_{\Lambda_J}^{-t}((\mathbf{M}_{(J,t,t')})_{\text{low}})^{\top}\odot\left(\widehat{\mathbf{A}}^{(2)}-\mathbf{A}_{\Lambda_J}\right)\mathbf{D}_{\Lambda_J}^{-t'}\right\| \\
&= \left\|\mathbf{D}_{\Lambda_J}^{-t'}(\mathbf{M}_{(J,t,t')})_{\text{up}}\odot(\mathsf{OVB}_{\Lambda_J,\cdot}+\mathsf{Var}_{\Lambda_J,\cdot})\mathbf{D}_{\Lambda_J}^{-t}\right\|+\left\|\mathbf{D}_{\Lambda_J}^{-t}((\mathbf{M}_{(J,t,t')})_{\text{low}})^{\top}\odot(\mathsf{OVB}_{\Lambda_J,\cdot}+\mathsf{Var}_{\Lambda_J,\cdot})\mathbf{D}_{\Lambda_J}^{-t'}\right\| \\
&\le \underbrace{\left\|\mathbf{D}_{\Lambda_J}^{-t'}\big((\mathbf{M}_{(J,t,t')})_{\text{up}}\odot(\mathsf{OVB}_{\Lambda_J,\cdot})\big)\mathbf{D}_{\Lambda_J}^{-t}\right\|+\left\|\mathbf{D}_{\Lambda_J}^{-t}\big(((\mathbf{M}_{(J,t,t')})_{\text{low}})^{\top}\odot(\mathsf{OVB}_{\Lambda_J,\cdot})\big)\mathbf{D}_{\Lambda_J}^{-t'}\right\|}_{=:\text{Error-OVB}} \\
&\quad+\underbrace{\left\|\mathbf{D}_{\Lambda_J}^{-t'}\big((\mathbf{M}_{(J,t,t')})_{\text{up}}\odot(\mathsf{Var}_{\Lambda_J,\cdot})\big)\mathbf{D}_{\Lambda_J}^{-t}\right\|+\left\|\mathbf{D}_{\Lambda_J}^{-t}\big(((\mathbf{M}_{(J,t,t')})_{\text{low}})^{\top}\odot(\mathsf{Var}_{\Lambda_J,\cdot})\big)\mathbf{D}_{\Lambda_J}^{-t'}\right\|}_{=:\text{Error-Var}}. \qquad \text{(C.4)}
\end{aligned}
$$

证明至此完成。 □

C.4.2 遗漏变量偏差

命题的证明 *5.11*。回顾（5.1）中的定义，参数 $\widetilde{J}$，$\widetilde{t}$，$\widetilde{t'}$ 定义如下

$$
\widetilde{J}:=\frac{t+t'-r+\varepsilon_1}{\min\{t',r_1\}+t-r}J,\quad \widetilde{t}:=\max\{t,t'\}=t',\quad \widetilde{t'}:=\max\{r_1,t'\}. \qquad \text{(C.5)}
$$

我们还回顾一下（误差-OVB ）中的定义C.4：

$$
\begin{aligned}
\text{Error-OVB} &= \left\|\mathbf{D}_{\Lambda_J}^{-t'}\big((\mathbf{M}_{(J,t,t')})_{\text{up}}\odot(\mathsf{OVB}_{\Lambda_J,\cdot})\big)\mathbf{D}_{\Lambda_J}^{-t}\right\|+\left\|\mathbf{D}_{\Lambda_J}^{-t}\big(((\mathbf{M}_{(J,t,t')})_{\text{low}})^{\top}\odot(\mathsf{OVB}_{\Lambda_J,\cdot})\big)\mathbf{D}_{\Lambda_J}^{-t'}\right\| \\
&= \left\|(\mathbf{M}_{(J,t,t')})_{\text{up}}\odot\left(\mathbf{D}_{\Lambda_J}^{-t'}\mathsf{OVB}_{\Lambda_J,\cdot}\mathbf{D}_{\Lambda_J}^{-t}\right)\right\|+\left\|((\mathbf{M}_{(J,t,t')})_{\text{low}})^{\top}\odot\left(\mathbf{D}_{\Lambda_J}^{-t}\mathsf{OVB}_{\Lambda_J,\cdot}\mathbf{D}_{\Lambda_J}^{-t'}\right)\right\|,
\end{aligned}
$$

其中

$$
\mathsf{OVB}=\sum_{\lambda'\in\Lambda_J}\left(E_{\Omega_{\lambda'}}(\mathbf{U}_{\cdot,\Omega_{\lambda'}}^{\top}\mathbf{U}_{\cdot,\Omega_{\lambda'}})^{-1}\mathbf{U}_{\cdot,\Omega_{\lambda'}}^{\top}\mathbf{U}_{\cdot,\Omega_{\lambda'}^c}\mathbf{A}_{\mathbf{\Omega}_{\lambda',\lambda'}^{\mathbf{c}}}\right)e_{\lambda'}^{\top}\in\mathbb{R}^{\Lambda_{\widetilde{J}}\times\Lambda_J}.
$$

定义 $B^{t',\ t}:=\mathbf{D}_{\Lambda_{\widetilde{J}}}^{-t'}\mathsf{OVB}\,\mathbf{D}_{\Lambda_J}^{-t}$ 和 $B^{t,\ t'}:=\mathbf{D}_{\Lambda_{\widetilde{J}}}^{-t}\mathsf{OVB}\mathbf{D}_{\Lambda_J}^{-t'}$，则

$$
\text{Error-OVB}=\left\|(\mathbf{M}_{(J,t,t')})_{\text{up}}\odot(B^{t',t})_{\Lambda_J,\cdot}\right\|+\left\|((\mathbf{M}_{(J,t,t')})_{\text{low}})^{\top}\odot(B^{t,t'})_{\Lambda_J,\cdot}\right\|. \qquad \text{(C.6)}
$$

为简化记号，我们写作

$$
\widehat{\Sigma}:=\frac{1}{N}\mathbf{U}^{\top}\mathbf{U},\quad \widehat{\Sigma}_{\Omega_{\lambda'},\Omega_{\lambda'}}:=\frac{1}{N}\mathbf{U}_{\cdot,\Omega_{\lambda'}}^{\top}\mathbf{U}_{\cdot,\Omega_{\lambda'}},\quad \widehat{\Sigma}_{\Omega_{\lambda'},\Omega_{\lambda'}^c}:=\frac{1}{N}\mathbf{U}_{\cdot,\Omega_{\lambda'}}^{\top}\mathbf{U}_{\cdot,\Omega_{\lambda'}^c}.
$$

我们还定义相应的总体量*

$$
\Sigma:=\mathbb{E}\left[\frac{1}{N}\mathbf{U}^{\top}\mathbf{U}\right],\quad \Sigma_{\Omega_{\lambda'},\Omega_{\lambda'}}:=\mathbb{E}\left[\frac{1}{N}\mathbf{U}_{\cdot,\Omega_{\lambda'}}^{\top}\mathbf{U}_{\cdot,\Omega_{\lambda'}}\right],\quad \Sigma_{\Omega_{\lambda'},\Omega_{\lambda'}^c}:=\mathbb{E}\left[\frac{1}{N}\mathbf{U}_{\cdot,\Omega_{\lambda'}}^{\top}\mathbf{U}_{\cdot,\Omega_{\lambda'}^c}\right].
$$

此外，根据命题 3.2 (II) (ii)，令 $\Sigma=\mathbf{D}^{r_1}\Sigma\mathbf{D}^{r_1}$ 为条件良好的设计协方差，其满足 $c_-\le\sigma_{\min}(\Sigma)\le\sigma_{\max}(\Sigma)\le c_+$ 其中 $c_-,c_+>0$为绝对常数；其经验版本 $\Sigma=\mathbf{D}^{r_1}\Sigma\mathbf{D}^{r_1}$。

* 这里我们写作 Σ来代替$\mathbf{C}_u$ 定义于命题3.2 (II) (ii)，以简化记号 n.

λ' 列的 $B^{t',\ t}$ 可写为

$$(B^{t',t})_{\cdot,\lambda'} = E_{\Omega_{\lambda'}} \mathbf{D}_{\Omega_{\lambda'}}^{-(t'-r_1)} \big(\widehat{\widetilde{\Sigma}}_{\Omega_{\lambda'},\Omega_{\lambda'}}\big)^{-1} \widehat{\widetilde{\Sigma}}_{\Omega_{\lambda'},\Omega_{\lambda'}^c} \big(\mathbf{D}_{\Omega_{\lambda'}^c}^{-r_1} \mathbf{A}_{\Omega_{\lambda'}^c,\lambda'}\big)(\mathbf{D}_{\Lambda_J}^{-t})_{\lambda',\lambda'} \in \mathbb{R}^{\Lambda_{\widetilde{J}}}, \tag{C.7}$$

其中 $(\mathbf{D}_{\Lambda_J}^{-t})_{\lambda',\ \lambda'} = 2^{-j't} \in \mathrm{R}_{\circ}$

固定 $\delta \in (0,\ 1)_{\circ}$ 根据引理 C.5 的第一部分，对于每个，如果 log()，那么以至少 的概率，矩阵，$\Omega_{\lambda'}$ 可逆并且满足

$$\|\big(\widehat{\widetilde{\Sigma}}_{\Omega_{\lambda'},\Omega_{\lambda'}}\big)^{-1}\| = 1/\sigma_{\min}(\widehat{\widetilde{\Sigma}}_{\Omega_{\lambda'},\Omega_{\lambda'}}) \lesssim 1. \tag{C.8}$$

记

$$\widehat{\widetilde{\Sigma}}_{\Omega_{\lambda'},\Omega_{\lambda'}^c} = \mathbf{D}_{\Omega_{\lambda'}}^{r_1} \widehat{\Sigma}_{\Omega_{\lambda'},\Omega_{\lambda'}^c} \mathbf{D}_{\Omega_{\lambda'}^c}^{r_1} = \frac{1}{N}\sum_{i=1}^{N} \bar{\mathbf{U}}_{i,\Omega_{\lambda'}}^{\top} \bar{\mathbf{U}}_{i,\Omega_{\lambda'}^c},$$

$^{-}$其中 $\mathbf{U}_{i,\ \Omega_{\lambda'}} := \mathbf{U}_{i,\ \Omega_{\lambda'}} \mathbf{D}_{\Omega_{\lambda'}}^{r_1},\ \mathbf{U}_{i,\ \Omega_{\lambda'}^c} := \mathbf{U}_{i,\ \Omega_{\lambda'}^c}$

$\mathbf{D}^{r_1}_{\Omega^c_{\lambda'}\circ}$ 因此，

$$\Big\|\widehat{\widetilde{\Sigma}}_{\Omega_{\lambda'},\Omega_{\lambda'}^c} \big(\mathbf{D}_{\Omega_{\lambda'}^c}^{-r_1} \mathbf{A}_{\Omega_{\lambda'}^c,\lambda'}\big)(\mathbf{D}_{\Lambda_J}^{-t'})_{\lambda',\lambda'}\Big\| = \Big\|\frac{1}{N}\sum_{i=1}^{N} \bar{\mathbf{U}}_{i,\Omega_{\lambda'}}^{\top} \Big\langle \bar{\mathbf{U}}_{i,\Omega_{\lambda'}^c}, \big(\mathbf{D}_{\Omega_{\lambda'}^c}^{-r_1} \mathbf{A}_{\Omega_{\lambda'}^c,\lambda'}\big)(\mathbf{D}_{\Lambda_J}^{-t'})_{\lambda',\lambda'} \Big\rangle\Big\|. \tag{C.9}$$

为控制（C.9），我们应用引理 C.4，其中

$$X_i = \bar{\mathbf{U}}_{i,\Omega_{\lambda'}}^{\top} \in \mathbb{R}^{|\Omega_{\lambda'}|}, \qquad Z_i = \Big\langle \bar{\mathbf{U}}_{i,\Omega_{\lambda'}^c}, \big(\mathbf{D}_{\Omega_{\lambda'}^c}^{-r_1} \mathbf{A}_{\Omega_{\lambda'}^c,\lambda'}\big)(\mathbf{D}_{\Lambda_J}^{-t'})_{\lambda',\lambda'} \Big\rangle,$$

且没有te that

$$\mu = \mathbb{E}[X_i Z_i] = \bar{\Sigma}_{\Omega_{\lambda'},\Omega_{\lambda'}^c} \big(\mathbf{D}_{\Omega_{\lambda'}^c}^{-r_1} \mathbf{A}_{\Omega_{\lambda'}^c,\lambda'}\big)(\mathbf{D}_{\Lambda_J}^{-t'})_{\lambda',\lambda'},$$

$$\sigma_Z^2 = \mathbb{E}[Z_i^2] = \Big\langle \bar{\Sigma}_{\Omega_{\lambda'}^c,\Omega_{\lambda'}^c} \big(\mathbf{D}_{\Omega_{\lambda'}^c}^{-r_1} \mathbf{A}_{\Omega_{\lambda'}^c,\lambda'}\big)(\mathbf{D}_{\Lambda_J}^{-t'})_{\lambda',\lambda'}, \big(\mathbf{D}_{\Omega_{\lambda'}^c}^{-r_1} \mathbf{A}_{\Omega_{\lambda'}^c,\lambda'}\big)(\mathbf{D}_{\Lambda_J}^{-t'})_{\lambda',\lambda'} \Big\rangle,$$

$$\Sigma_{XX} = \mathbb{E}[X_i X_i^{\top}] = \bar{\Sigma}_{\Omega_{\lambda'},\Omega_{\lambda'}}.$$

由此可知，以至少 $1 - \delta/2$ 的概率，

$$\begin{aligned}
&\Big\|\widehat{\widetilde{\Sigma}}_{\Omega_{\lambda'},\Omega_{\lambda'}^c} \big(\mathbf{D}_{\Omega_{\lambda'}^c}^{-r_1} \mathbf{A}_{\Omega_{\lambda'}^c,\lambda'}\big)(\mathbf{D}_{\Lambda_J}^{-t'})_{\lambda',\lambda'}\Big\|_2 \\
&= \Big\|\frac{1}{N}\sum_{i=1}^{N} \bar{\mathbf{U}}_{i,\Omega_{\lambda'}} \Big\langle \bar{\mathbf{U}}_{i,\Omega_{\lambda'}^c}, \big(\mathbf{D}_{\Omega_{\lambda'}^c}^{-r_1} \mathbf{A}_{\Omega_{\lambda'}^c,\lambda'}\big)(\mathbf{D}_{\Lambda_J}^{-t'})_{\lambda',\lambda'} \Big\rangle\Big\|_2 \\
&\lesssim \big(\sigma_Z \|\Sigma_{XX}\|^{1/2} + \|\mu\|_2\big)\left(\sqrt{\frac{|\Omega_{\lambda'}| + \log(4/\delta)}{N}} + \frac{|\Omega_{\lambda'}| + \log(4/\delta)}{N}\right) \\
&\overset{(\star)}{\lesssim} \|\big(\mathbf{D}_{\Omega_{\lambda'}^c}^{-r_1} \mathbf{A}_{\Omega_{\lambda'}^c,\lambda'}\big)(\mathbf{D}_{\Lambda_J}^{-t'})_{\lambda',\lambda'}\|_2 \left(\sqrt{\frac{|\Omega_{\lambda'}| + \log(4/\delta)}{N}} + \frac{|\Omega_{\lambda'}| + \log(4/\delta)}{N}\right) \\
&\lesssim \|\big(\mathbf{D}_{\Omega_{\lambda'}^c}^{-r_1} \mathbf{A}_{\Omega_{\lambda'}^c,\lambda'}\big)(\mathbf{D}_{\Lambda_J}^{-t'})_{\lambda',\lambda'}\|_2,
\end{aligned} \tag{C.10}$$

条件是 $N \gtrsim |\Omega_{\lambda'}| + \log(1/\delta)_{\circ}$ 不等式（$\star$）由以下界得出

$$\begin{aligned}
\sigma_Z = (\mathbb{E}[Z_i^2])^{1/2} &= \Big\langle \bar{\Sigma}_{\Omega_{\lambda'}^c,\Omega_{\lambda'}^c} \big(\mathbf{D}_{\Omega_{\lambda'}^c}^{-r_1} \mathbf{A}_{\Omega_{\lambda'}^c,\lambda'}\big)(\mathbf{D}_{\Lambda_J}^{-t'})_{\lambda',\lambda'}, \big(\mathbf{D}_{\Omega_{\lambda'}^c}^{-r_1} \mathbf{A}_{\Omega_{\lambda'}^c,\lambda'}\big)(\mathbf{D}_{\Lambda_J}^{-t'})_{\lambda',\lambda'} \Big\rangle^{1/2} \\
&\lesssim \|\big(\mathbf{D}_{\Omega_{\lambda'}^c}^{-r_1} \mathbf{A}_{\Omega_{\lambda'}^c,\lambda'}\big)(\mathbf{D}_{\Lambda_J}^{-t'})_{\lambda',\lambda'}\|_2, \qquad \text{since } \|\bar{\Sigma}_{\Omega_{\lambda'}^c,\Omega_{\lambda'}^c}\| \lesssim 1,
\end{aligned}$$

and

$$\|\mu\|_2 = \left\|\bar{\Sigma}_{\Omega_{\lambda'},\Omega^c_{\lambda'}}\big(\mathbf{D}^{-r_1}_{\Omega^c_{\lambda'}}\mathbf{A}_{\Omega^c_{\lambda'},\lambda'}\big)(\mathbf{D}^{-t'}_{\Lambda_J})_{\lambda',\lambda'}\right\|_2 \lesssim \|\big(\mathbf{D}^{-r_1}_{\Omega^c_{\lambda'}}\mathbf{A}_{\Omega^c_{\lambda'},\lambda'}\big)(\mathbf{D}^{-t'}_{\Lambda_J})_{\lambda',\lambda'}\|_2, \quad \text{since } \|\bar{\Sigma}_{\Omega_{\lambda'},\Omega^c_{\lambda'}}\| \lesssim 1.$$

因此，合并（C.7）、（C.8）和（C.10），我们得出结论：如果

$$N \gtrsim |\Omega_{\lambda'}| + \log(1/\delta),$$

那么，以至少 $1-\delta$ 的概率，

$$\begin{aligned}\|(B^{t',t})_{\cdot,\lambda'}\|_2 &= \left\|E_{\Omega_{\lambda'}}\mathbf{D}^{-(t'-r_1)}_{\Omega_{\lambda'}}\big(\widehat{\bar{\Sigma}}_{\Omega_{\lambda'},\Omega_{\lambda'}}\big)^{-1}\widehat{\bar{\Sigma}}_{\Omega_{\lambda'},\Omega^c_{\lambda'}}\big(\mathbf{D}^{-r_1}_{\Omega^c_{\lambda'}}\mathbf{A}_{\Omega^c_{\lambda'},\lambda'}\big)(\mathbf{D}^{-t}_{\Lambda_J})_{\lambda',\lambda'}\right\|_2 \\ &\le \|E_{\Omega_{\lambda'}}\| \cdot \|\mathbf{D}^{-(t'-r_1)}_{\Omega_{\lambda'}}\| \cdot \|\big(\widehat{\bar{\Sigma}}_{\Omega_{\lambda'},\Omega_{\lambda'}}\big)^{-1}\| \cdot \left\|\widehat{\bar{\Sigma}}_{\Omega_{\lambda'},\Omega^c_{\lambda'}}\big(\mathbf{D}^{-r_1}_{\Omega^c_{\lambda'}}\mathbf{A}_{\Omega^c_{\lambda'},\lambda'}\big)(\mathbf{D}^{-t}_{\Lambda_J})_{\lambda',\lambda'}\right\|_2 \\ &\lesssim 2^{\widetilde{J}\max\{r_1-t',0\}} \cdot \|\big(\mathbf{D}^{-r_1}_{\Omega^c_{\lambda'}}\mathbf{A}_{\Omega^c_{\lambda'},\lambda'}\big)(\mathbf{D}^{-t}_{\Lambda_J})_{\lambda',\lambda'}\|_2, \end{aligned} \tag{C.11}$$

其中，最后一个不等式用到了 $\|E_{\Omega_{\lambda'}}\| = 1$，$\|\mathbf{D}^{-(t'-r_1)_{\Omega_{\lambda'}}}\| \le 2^{\widetilde{J}\max\{r_1-t',\ 0\}}$。

设 $\mathbf{A}^{\varepsilon}_{\Lambda_{\widetilde{J}}}$ 为定义 4.2 中的压缩矩阵，其参数为 $(\widetilde{J},\ \widetilde{t},\ \widetilde{t'})$，即

$$\mathbf{A}^{\varepsilon}_{\Lambda_{\widetilde{J}}} := M_{(\widetilde{J},\widetilde{t},\widetilde{t'})} \odot \mathbf{A}_{\Lambda_{\widetilde{J}}}.$$

应用命题 5.10，取参数 $(\widetilde{J},\ \widetilde{t},\ \widetilde{t'})$，可得，对于 $r_1,\ t \in [r/2,\ \gamma)$，

$$\|\mathbf{D}^{-r_1}_{\Lambda_{\widetilde{J}}}(\mathbf{A}^{\varepsilon}_{\Lambda_{\widetilde{J}}} - \mathbf{A}_{\Lambda_{\widetilde{J}}})\mathbf{D}^{-t}_{\Lambda_{\widetilde{J}}}\| \lesssim \widetilde{J}2^{-\widetilde{J}\left(\min\{\widetilde{t'},r_1\}+\min\{\widetilde{t},t\}-r\right)} = \widetilde{J}2^{-\widetilde{J}(r_1+t-r)},$$

其中，最后一个等式成立是因为 $t = \min\{t,\ t'\} \ge t$ 且 $t' = \min\{r_1,\ t'\} \ge r_1$。

另一方面，由于 $t > r/2$ 且 $r_1 > n/2 + \max\{0,\ r\} \ge r/2$ 根据假设 2.1 (ii)，由命题 5.9，

$$\|\mathbf{D}^{-r_1}(\mathbf{A}^{\uparrow}_{\Lambda_{\widetilde{J}}} - \mathbf{A})\mathbf{D}^{-t}\| \lesssim 2^{-\widetilde{J}(r_1+t-r)}.$$

回归支撑集定义如下：

$$\Omega_{\lambda'} := \Big\{\lambda \in \mathcal{J} : (\lambda,\lambda') \in \operatorname{supp}(\widetilde{J},\widetilde{t},\widetilde{t'})\Big\}, \qquad \Omega^c_{\lambda'} := \mathcal{J} \setminus \Omega_{\lambda'}.$$

对于每个 $\lambda' \in \Lambda_J$，向量 $\mathbf{A}^{\Omega^c_{\lambda'}\lambda'}$ 恰好收集了回归中省略的 $\mathbf{A}$ 的第 λ' 列中的那些元素。根据截断 $\mathbf{A}^{\uparrow}_{\Lambda_{\widetilde{J}}}$ 和压缩掩码 $M_{(\widetilde{J},\widetilde{t},\widetilde{t'})}$ 的构造，这些省略的元素与 $\mathbf{A}^{\uparrow}_{\Lambda_{\widetilde{J}}} - \mathbf{A}$ 和 $\mathbf{A}^{\varepsilon}_{\Lambda_{\widetilde{J}}} - \mathbf{A}^{\Lambda_{\widetilde{J}}}$ 的第 λ' 列中的元素相同。因此，在两侧乘以对角矩阵加权后，

$$\|\big(\mathbf{D}^{-r_1}_{\Omega^c_{\lambda'}}\mathbf{A}_{\Omega^c_{\lambda'},\lambda'}\big)(\mathbf{D}^{-t}_{\Lambda_J})_{\lambda',\lambda'}\|_2 \le \|\mathbf{D}^{-r_1}(\mathbf{A}^{\uparrow}_{\Lambda_{\widetilde{J}}} - \mathbf{A})\mathbf{D}^{-t}\| + \|\mathbf{D}^{-r_1}_{\Lambda_{\widetilde{J}}}(\mathbf{A}^{\varepsilon}_{\Lambda_{\widetilde{J}}} - \mathbf{A}_{\Lambda_{\widetilde{J}}})\mathbf{D}^{-t}_{\Lambda_{\widetilde{J}}}\| \lesssim \widetilde{J}2^{-\widetilde{J}(r_1+t-r)}. \tag{C.12}$$

合并 (C.11) 和 (C.12) 可得

$$\begin{aligned}\|(B^{t',t})_{\cdot,\lambda'}\|_2 &\lesssim 2^{\widetilde{J}\max\{r_1-t',0\}} \cdot \|\big(\mathbf{D}^{-r_1}_{\Omega^c_{\lambda'}}\mathbf{A}_{\Omega^c_{\lambda'},\lambda'}\big)(\mathbf{D}^{-t}_{\Lambda_J})_{\lambda',\lambda'}\|_2 \\ &\lesssim 2^{\widetilde{J}\max\{r_1-t',0\}} \cdot \widetilde{J}2^{-\widetilde{J}(r_1+t-r)} \\ &= \widetilde{J}2^{-\widetilde{J}(\min\{t',r_1\}+t-r)}.\end{aligned}$$

对所有由 $\lambda' \in \Lambda_J$ 编号的列应用并集界，可得，若

$$N \gtrsim \max_{\lambda' \in \Lambda_J} |\Omega_{\lambda'}| + J + \log(1/\delta),$$

那么，以至少 $1-\delta$ 的概率，以下界对所有 $\lambda' \in \Lambda_J$ 同时成立：

$$\|(B^{t',t})_{\cdot,\lambda'}\|_2 \lesssim \widetilde{J} 2^{-\widetilde{J}(\min\{t',r_1\}+t-r)}.$$

对于任意矩阵 B，记 $\mathsf{nnz}(B) := \#\{(i,j) : B_{ij} \neq 0\}$、$\mathsf{nnz}_{\mathrm{row}}(B) := \max_i \#\{j : B_{ij} \neq 0\}$，以及 $\mathsf{nnz}_{\mathrm{col}}(B) := \max_j \#\{i : B_{ij} \neq 0\}$。根据命题 C.2,

$$\mathrm{nnz}_{\mathrm{row}}(\mathbf{M}_{(J,t,t')}) \lesssim 2^{Jn(t+t'-r)/(\sigma-\frac{n}{2}+t-\frac{r}{2})}, \quad \mathrm{nnz}_{\mathrm{col}}(\mathbf{M}_{(J,t,t')}) \lesssim 2^{Jn(t+t'-r)/(\sigma-\frac{n}{2}+t'-\frac{r}{2})}.$$

因此,

$$\begin{aligned}
\Big\|(\mathbf{M}_{(J,t,t')})_{\mathrm{up}} \odot (B^{t',t})_{\Lambda_J,\cdot}\Big\|_1 &= \max_{\lambda' \in \Lambda_J} \Big\| \Big((\mathbf{M}_{(J,t,t')})_{\mathrm{up}} \odot (B^{t',t})_{\Lambda_J,\cdot} \Big)_{\cdot,\lambda'} \Big\|_1 \\
&\le \max_{\lambda' \in \Lambda_J} \sqrt{\mathrm{nnz}_{\mathrm{col}}((\mathbf{M}_{(J,t,t')})_{\mathrm{up}})} \Big\| \Big((\mathbf{M}_{(J,t,t')})_{\mathrm{up}} \odot (B^{t',t})_{\Lambda_J,\cdot} \Big)_{\cdot,\lambda'} \Big\|_2 \\
&\le \max_{\lambda' \in \Lambda_J} \sqrt{\mathrm{nnz}_{\mathrm{col}}(\mathbf{M}_{(J,t,t')})} \cdot \|(B^{t',t})_{\cdot,\lambda'}\|_2 \\
&\lesssim 2^{Jn(t+t'-r)/2(\sigma-\frac{n}{2}+t'-\frac{r}{2})} \cdot \widetilde{J} 2^{-\widetilde{J}(\min\{t',r_1\}+t-r)}.
\end{aligned}$$

另一方面,

$$\begin{aligned}
\Big\|(\mathbf{M}_{(J,t,t')})_{\mathrm{up}} \odot (B^{t',t})_{\Lambda_J,\cdot}\Big\|_\infty &= \max_{\lambda \in \Lambda_J} \Big\| \Big((\mathbf{M}_{(J,t,t')})_{\mathrm{up}} \odot (B^{t',t})_{\Lambda_J,\cdot} \Big)_{\lambda,\cdot} \Big\|_1 \\
&\le \mathrm{nnz}_{\mathrm{row}}((\mathbf{M}_{(J,t,t')})_{\mathrm{up}}) \cdot \Big\|(\mathbf{M}_{(J,t,t')})_{\mathrm{up}} \odot (B^{t',t})_{\Lambda_J,\cdot}\Big\|_{\max} \\
&\le \mathrm{nnz}_{\mathrm{row}}(\mathbf{M}_{(J,t,t')}) \cdot \max_{\lambda' \in \Lambda_J} \|(B^{t',t})_{\cdot,\lambda'}\|_2 \\
&\lesssim 2^{Jn(t+t'-r)/(\sigma-\frac{n}{2}+t-\frac{r}{2})} \cdot \widetilde{J} 2^{-\widetilde{J}(\min\{t',r_1\}+t-r)}.
\end{aligned}$$

因此,

$$\begin{aligned}
\Big\|(\mathbf{M}_{(J,t,t')})_{\mathrm{up}} \odot (B^{t',t})_{\Lambda_J,\cdot}\Big\| &\le \Big(\Big\|(\mathbf{M}_{(J,t,t')})_{\mathrm{up}} \odot (B^{t',t})_{\Lambda_J,\cdot}\Big\|_1 \Big\|(\mathbf{M}_{(J,t,t')})_{\mathrm{up}} \odot (B^{t',t})_{\Lambda_J,\cdot}\Big\|_\infty \Big)^{1/2} \\
&\lesssim \widetilde{J} 2^{-\widetilde{J}(\min\{t',r_1\}+t-r)} \cdot 2^{\varepsilon_1 J}, \qquad \text{(C.13)}
\end{aligned}$$

其中

$$\varepsilon_1 := \frac{n(t+t'-r)}{\sigma - n/2 + t - r/2} \ge n(t+t'-r)\Big(\frac{1}{4(\sigma - n/2 + t' - r/2)} + \frac{1}{2(\sigma - n/2 + t - r/2)} \Big).$$

类似地，如果 $N \gtrsim \max_{\lambda' \in \Lambda_J} |\Omega_{\lambda'}| + J + \log(1/\delta)$，那么以至少 $1-\delta$ 的概率，以下界对所有 $\lambda' \in \Lambda_J$ 同时成立：

$$\|(B^{t,t'})_{\cdot,\lambda'}\| \lesssim \widetilde{J} 2^{-\widetilde{J}(\min\{t,r_1\}+t'-r)}.$$

使用相同的论证可得

$$\Big\|((\mathbf{M}_{(J,t,t')})_{\mathrm{low}})^\top \odot (B^{t,t'})_{\Lambda_J,\cdot}\Big\| \lesssim \widetilde{J} 2^{-\widetilde{J}(\min\{t,r_1\}+t'-r)} \cdot 2^{\varepsilon_1 J}, \qquad \text{(C.14)}$$

其中我们用到了

$$\varepsilon_1 = \frac{n(t+t'-r)}{\sigma - n/2 + t - r/2} \geq n(t+t'-r)\bigg(\frac{1}{4(\sigma - n/2 + t - r/2)} + \frac{1}{2(\sigma - n/2 + t' - r/2)}\bigg).$$

合并（C.6）、（C.13）和（C.14）可得，若 $N \gtrsim \max_{\lambda' \in \Lambda_J} |\Omega_{\lambda'}| + J + \log(1/\delta)$，则以至少 $1-\delta$ 的概率，

$$\begin{aligned}
\mathsf{Error\text{-}OVB} &= \Big\| (\mathbf{M}_{(J,t,t')})_{\mathrm{up}} \odot (B^{t',t})_{\Lambda_J,\cdot} \Big\| + \Big\| ((\mathbf{M}_{(J,t,t')})_{\mathrm{low}})^\top \odot (B^{t,t'})_{\Lambda_J,\cdot} \Big\| \\
&\lesssim \widetilde{J} 2^{-\widetilde{J}(\min\{t',r_1\}+t-r)} \cdot 2^{\varepsilon_1 J} + \widetilde{J} 2^{-\widetilde{J}(\min\{t,r_1\}+t'-r)} \cdot 2^{\varepsilon_1 J} \\
&\lesssim J\, 2^{-J(t+t'-r)},
\end{aligned}$$

~其中，最后一行由 J 的定义得出

$$\widetilde{J} = \left\lceil \frac{t+t'-r+\varepsilon_1}{\min\{t',r_1\}+t-r}\, J \right\rceil \geq \max\left\{ \frac{t+t'-r+\varepsilon_1}{\min\{t',r_1\}+t-r} J,\ \frac{t+t'-r+\varepsilon_1}{\min\{t,r_1\}+t'-r} J \right\}.$$

最后一个等式成立，因为 $t \leq t'$ 蕴含

$$\min\{t',r_1\} + t - r \leq \min\{t,r_1\} + t' - r.$$

总之，我们使用集合 supp($\widetilde{J}, \widetilde{t}, \widetilde{t'}$) 进行回归。然后，我们得出，若

$$N \gtrsim \max_{\lambda' \in \Lambda_J} |\Omega_{\lambda'}| + J + \log(1/\delta) \asymp 2^{\widetilde{J} n (\widetilde{t}+\widetilde{t'}-r)/(\sigma - \frac{n}{2} + \widetilde{t'} - \frac{r}{2})} + \log(1/\delta),$$

则以至少 $1-\delta$ 的概率，

$$\mathsf{Error\text{-}OVB} \lesssim J\, 2^{-J(t+t'-r)}. \tag{C.15}$$

□

C.4.3 方差

命题的证明 *5.12*。 回想一下

$$\begin{aligned}
\mathsf{Error\text{-}Var} &= \Big\| \mathbf{D}_{\Lambda_J}^{-t'} \big((\mathbf{M}_{(J,t,t')})_{\mathrm{up}} \odot (\mathsf{Var}_{\Lambda_J,\cdot}) \big) \mathbf{D}_{\Lambda_J}^{-t} \Big\| + \Big\| \mathbf{D}_{\Lambda_J}^{-t} \big(((\mathbf{M}_{(J,t,t')})_{\mathrm{low}})^\top \odot (\mathsf{Var}_{\Lambda_J,\cdot}) \big) \mathbf{D}_{\Lambda_J}^{-t'} \Big\| \\
&= \Big\| (\mathbf{M}_{(J,t,t')})_{\mathrm{up}} \odot \Big(\mathbf{D}_{\Lambda_J}^{-t'} \mathsf{Var}_{\Lambda_J,\cdot} \mathbf{D}_{\Lambda_J}^{-t} \Big) \Big\| + \Big\| ((\mathbf{M}_{(J,t,t')})_{\mathrm{low}})^\top \odot \Big(\mathbf{D}_{\Lambda_J}^{-t} \mathsf{Var}_{\Lambda_J,\cdot} \mathbf{D}_{\Lambda_J}^{-t'} \Big) \Big\|,
\end{aligned}$$

其中

$$\mathsf{Var} := \sum_{\lambda' \in \Lambda_J} \Big(E_{\Omega_{\lambda'}} (\mathbf{U}_{\cdot,\Omega_{\lambda'}}^\top \mathbf{U}_{\cdot,\Omega_{\lambda'}})^{-1} \mathbf{U}_{\cdot,\Omega_{\lambda'}}^\top \mathbf{W}_{\cdot,\lambda'} \Big) e_{\lambda'}^\top \in \mathbb{R}^{\Lambda_{\widetilde{J}} \times \Lambda_J}.$$

定义 $V^{t',\ t} := \mathbf{D}_{\Lambda_{\widetilde{J}}}^{-t'} \mathsf{Var}\, \mathbf{D}_{\Lambda_J}^{-t}$ 以及 $V^{t,\ t'} := \mathbf{D}_{\Lambda_{\widetilde{J}}}^{-t} \mathsf{Var} \mathbf{D}_{\Lambda_J}^{-t'}$，则

$$\mathsf{Error\text{-}Var} = \Big\| (\mathbf{M}_{(J,t,t')})_{\mathrm{up}} \odot (V^{t',t})_{\Lambda_J,\cdot} \Big\| + \Big\| ((\mathbf{M}_{(J,t,t')})_{\mathrm{low}})^\top \odot (V^{t,t'})_{\Lambda_J,\cdot} \Big\|. \tag{C.16}$$

To c控制（C.16）中的第一项，我们应用范数压缩不等式（引理 C.3 ）

采用自然的 $(j,\ j')$ 分块划分，并将 $(\cdot)_{j,\ j'}$ 记为 $(j,\ j')$ 分块。我们得到

$$\begin{aligned}
&\left\|(\mathbf{M}_{(J,t,t')})_{\mathrm{up}} \odot (V^{t',t})_{\Lambda_J,\cdot}\right\| \le \left(\sum_{j_0\le j\le j'\le J} \left\|\left(\mathbf{M}_{(J,t,t')})_{\mathrm{up}} \odot (V^{t',t})_{\Lambda_J,\cdot}\right)_{j,j'}\right\|^2\right)^{1/2} \\
&\le \left(\sum_{j_0\le j\le j'\le J} \left\|\left(\mathbf{M}_{(J,t,t')})_{\mathrm{up}} \odot (V^{t',t})_{\Lambda_J,\cdot}\right)_{j,j'}\right\|_1 \cdot \left\|\left(\mathbf{M}_{(J,t,t')})_{\mathrm{up}} \odot (V^{t',t})_{\Lambda_J,\cdot}\right)_{j,j'}\right\|_\infty\right)^{1/2} \\
&\le \left(\sum_{j_0\le j\le j'\le J} \mathrm{nnz}_{\mathrm{col}}\big((\mathbf{M}_{(J,t,t')})_{j,j'}\big) \cdot \mathrm{nnz}_{\mathrm{row}}\big((\mathbf{M}_{(J,t,t')})_{j,j'}\big) \cdot \big\|(V^{t',t})_{j,j'}\big\|_{\max}^2\right)^{1/2} \\
&= \left(\sum_{j_0\le j\le j'\le J} 2^{-2jt'-2j't} \cdot (\mathrm{count}(j,j'))^2 \cdot \|(\mathsf{Var})_{j,j'}\|_{\max}^2\right)^{1/2}, \qquad \text{(C.17)}
\end{aligned}$$

其中最后一行用到了 $(V^{t',t})_{j,j'} = 2^{-jt'-j't}$(方差)$_{j,j'}$ 和

$$\mathrm{count}(j,j') := \Big(\mathrm{nnz}_{\mathrm{col}}\big((\mathbf{M}_{(J,t,t')})_{j,j'}\big) \cdot \mathrm{nnz}_{\mathrm{row}}\big((\mathbf{M}_{(J,t,t')})_{j,j'}\big)\Big)^{1/2}.$$

根据命题 C.2，我们有

$$\mathrm{count}(j,j') \lesssim \begin{cases} 0 & \text{if } (j,j') \in D_1 \cup D_2, \\ 2^{(j-j')n/2} & \text{if } (j,j') \in D_3, \\ 2^{(j'-j)n/2} & \text{if } (j,j') \in D_4, \\ 2^{(j+j')n/2} 2^{n(J(t+t'-r)-jt'-j't-(j+j')\widetilde{d})/(2\widetilde{d}+r)} & \text{if } (j,j') \in D_5, \\ 2^{(j+j')n/2} & \text{if } (j,j') \in D_6. \end{cases} \qquad \text{(C.18)}$$

接下来，我们推导对所有 $j_0 \le j, j' \le J$，$\|$(方差)$_{j,j'}\|_{\max}$ 的高概率上界。

对于 $j_0 \le j, j' \le J$和 $\lambda' = (j', k')$，其中 $k' \in \nabla_{j'}$，令 (方差)$_{j,\lambda'}$ =(方差)$_{j,(j',k')} \in \mathbb{R}^{\nabla_j}$表示 (方差)$_{j,j'}$ 的第 λ' 列。于是，

$$\|(\mathsf{Var})_{j,j'}\|_{\max} = \max_{k'\in\nabla_{j'}} \|(\mathsf{Var})_{j,\lambda'}\|_{\max}.$$

回忆一下，方差项 Var 的第 λ' 列为

$$(\mathsf{Var})_{\cdot,\lambda'} = E_{\Omega_{\lambda'}}(\mathbf{U}_{\cdot,\Omega_{\lambda'}}^\top \mathbf{U}_{\cdot,\Omega_{\lambda'}})^{-1} \mathbf{U}_{\cdot,\Omega_{\lambda'}}^\top \mathbf{W}_{\cdot,\lambda'} \in \mathbb{R}^{\Lambda_J^\sim},$$

其中 $\mathbf{W}_{\cdot,\lambda'} \sim \mathcal{N}(0, \sigma_{\lambda'}^2 I_N)$ 与 $\mathbf{U}$ 相互独立，且 $\sigma_{\lambda'} \asymp 2^{-j'r_2}$根据命题 3.2 (II) (iii)。在给定 $\mathbf{U}$ 的条件下，

$$(\mathbf{U}_{\cdot,\Omega_{\lambda'}}^\top \mathbf{U}_{\cdot,\Omega_{\lambda'}})^{-1} \mathbf{U}_{\cdot,\Omega_{\lambda'}}^\top \mathbf{W}_{\cdot,\lambda'} \mid \mathbf{U} \sim \mathcal{N}\Big(0, \sigma_{\lambda'}^2 (\mathbf{U}_{\cdot,\Omega_{\lambda'}}^\top \mathbf{U}_{\cdot,\Omega_{\lambda'}})^{-1}\Big).$$

写出 $\widehat{\Sigma}_{\Omega_{\lambda''},\Omega_{\lambda'}} := \frac{1}{N}\mathbf{U}_{\cdot,\Omega_{\lambda''}}^\top \mathbf{U}_{\cdot,\Omega_{\lambda'}}$和 $\widehat{\widetilde{\Sigma}}_{\Omega_{\lambda''},\Omega_{\lambda'}} := \mathbf{D}_{\Omega_{\lambda''}}^{r_1} \widehat{\Sigma}_{\Omega_{\lambda''},\Omega_{\lambda'}} \mathbf{D}_{\Omega_{\lambda'}}^{r_1}$。因此，

$$(\mathbf{U}_{\cdot,\Omega_{\lambda'}}^\top \mathbf{U}_{\cdot,\Omega_{\lambda'}})^{-1} \mathbf{U}_{\cdot,\Omega_{\lambda'}}^\top \mathbf{W}_{\cdot,\lambda'} \mid \mathbf{U} \sim \mathcal{N}\Big(0, \frac{\sigma_{\lambda'}^2}{N} \widehat{\Sigma}_{\Omega_{\lambda'},\Omega_{\lambda'}}^{-1}\Big) = \mathcal{N}\Big(0, \frac{\sigma_{\lambda'}^2}{N} \mathbf{D}_{\Omega_{\lambda'}}^{r_1} \widehat{\widetilde{\Sigma}}_{\Omega_{\lambda'},\Omega_{\lambda'}}^{-1} \mathbf{D}_{\Omega_{\lambda'}}^{r_1}\Big).$$

注意，向量 (Var)$_{j,\ \lambda'} \in \mathbb{R}^{\nabla_j}$ 是通过将行指标限制在第 j个块而从 (Var)$_{\cdot,\ \lambda'}$ 中得到的子向量。因此，在给定$\mathbf{U}$的条件下，(Var)$_{j,\ \lambda'}$中的非零元素与$v_{j,\ \lambda'}$具有相同的分布，其中

$$v_{j,\lambda'} \sim \mathcal{N}\Big(0, \frac{2^{2jr_1}\sigma_{\lambda'}^2}{N} (\widehat{\widetilde{\Sigma}}_{\Omega_{\lambda'},\Omega_{\lambda'}}^{-1})_{j,j}\Big).$$

设 $\bar{v}_{j,\lambda'} := v_{j,\lambda'} \big(\frac{2^{jr_1}\sigma_{\lambda'}}{\sqrt{N}} \big)^{-1}$，则 $\bar{v}_{j,\lambda'} \sim \mathcal{N}\big(0, \big(\widehat{\Sigma}^{-1}_{\Omega_{\lambda'},\Omega_{\lambda'}}\big)_{j,j}\big)$。

固定 $\delta \in (0,1)$，并设

$$\delta_{j,\lambda'} := 2^{-j'n} J^{-2} \delta, \quad \lambda' = (j', k').$$

由于 $\Sigma_{\Omega_{\lambda'},\Omega_{\lambda'}}$ 是条件良好的，其特征值为一阶量，引理C.5 表明，只要 $N \gtrsim |\Omega_{\lambda'}| + \log(1/\delta_{j,\lambda'})$，我们以至少 $1 - \delta_{j,\lambda'}$ 的概率有，

$$\|\bar{v}_{j,\lambda'}\|_{\max} \lesssim \sqrt{\log\big(\dim(\bar{v}_{j,\lambda'})/\delta_{j,\lambda'}\big)} \lesssim \sqrt{\log\big(2^{jn}/\delta_{j,\lambda'}\big)} \lesssim \sqrt{j} + \sqrt{\log(1/\delta_{j,\lambda'})}.$$

对 $k' \in \nabla_{j'}$ 使用并集界，若 $N \gtrsim \max_{k' \in \nabla_{j'}} \big(|\Omega_{\lambda'}| + \log(1/\delta_{j,\lambda'})\big)$，则我们以至少 $1 - \sum_{k' \in \nabla_{j'}} \delta_{j,\lambda'}$ 的概率有，

$$\max_{k' \in \nabla_{j'}} \|\bar{v}_{j,\lambda'}\|_{\max} \lesssim \sqrt{j} + \sqrt{\log(1/\min_{k' \in \nabla_{j'}} \delta_{j,\lambda'})}.$$

最后，对所有 (j, j')–块应用并集界，若

$$N \gtrsim \max_{j,j'} \Big(\max_{k' \in \nabla_{j'}} \big(|\Omega_{\lambda'}| + \log(1/\delta_{j,\lambda'})\big) \Big) \asymp \max_{\lambda'} |\Omega_{\lambda'}| + J + \log(1/\delta),$$

则以概率至少为 $1 - \sum_{j,j'} \sum_{k' \in \nabla_{j'}} \delta_{j,\lambda'} \geq 1 - \delta$ 时，我们对所有 j, j' 同时有

$$\max_{k' \in \nabla_{j'}} \|\bar{v}_{j,\lambda'}\|_{\max} \lesssim \sqrt{j} + \sqrt{\log(1/\min_{k' \in \nabla_{j'}} \delta_{j,\lambda'})} \lesssim \sqrt{J + \log(1/\delta)}.$$

回顾一下，$\max_{\lambda'} |\Omega_{\lambda'}|$ 表示对 $\lambda' \in \Lambda_J$取最大值的按列回归支撑集大小。由于回归支撑集选为 $\mathrm{supp}(\widetilde{J}, \widetilde{t}, \widetilde{t}')$，其定义见（4.2），因此有

$$\max_{\lambda'} |\Omega_{\lambda'}| = \mathrm{nnz}_{\mathrm{col}}\Big(\mathrm{supp}(\widetilde{J}, \widetilde{t}, \widetilde{t}')\Big) \lesssim 2^{\widetilde{J}n(\widetilde{t}+\widetilde{t}'-r)/(\sigma - \frac{n}{2} + \widetilde{t}' - \frac{r}{2})},$$

如命题C.2所述。

因此，如果

$$N \gtrsim 2^{\widetilde{J}n(\widetilde{t}+\widetilde{t}'-r)/(\sigma - n/2 + \widetilde{t}' - r/2)} + \log(1/\delta),$$

那么，以至少 $1 - \delta$ 的概率，对所有 j, j'，同时有

$$\begin{aligned} \|(\mathsf{Var})_{j,j'}\|_{\max} &= \max_{k' \in \nabla_{j'}} \|(\mathsf{Var})_{j,\lambda'}\|_{\max} \\ &\leq \max_{k' \in \nabla_{j'}} \frac{2^{jr_1}\sigma_{\lambda'}}{\sqrt{N}} \|\bar{v}_{j,\lambda'}\|_{\max} \\ &\lesssim \frac{2^{jr_1 - j'r_2}}{\sqrt{N}} \max_{k' \in \nabla_{j'}} \|\bar{v}_{j,\lambda'}\|_{\max} \\ &\lesssim \frac{2^{jr_1 - j'r_2}}{\sqrt{N}} \cdot \sqrt{J + \log(1/\delta)}. \end{aligned} \tag{C.19}$$

令 $D_{\mathrm{up}} := \{(j, j') : 0 \leq j \leq j' \leq J\}$ 以及 $D_{\mathrm{low}} := \{(j, j') : 0 \leq j' < j \leq J\}$。合并估计（C.17）、（C.18）和（C.19）可得，若 $N \gtrsim 2^{\widetilde{J}n(\widetilde{t}+\widetilde{t}'-r)/(\sigma - n/2 + \widetilde{t}' - r/2)} + \log(1/\delta)$，

那么，以至少 $1-\delta$ 的概率，

$$\begin{aligned}
&\left\|(\mathbf{M}_{(J,t,t')})_{\mathrm{up}} \odot (V^{t',t})_{\Lambda_J,\cdot}\right\| \le \left(\sum_{j_0\le j\le j'\le J} 2^{-2jt'-2j't}\cdot(\mathrm{count}(j,j'))^2\cdot\|(\mathsf{Var})_{j,j'}\|_{\max}^2\right)^{1/2}\\
&\le\sqrt{\frac{J+\log(1/\delta)}{N}}\left(\sum_{(j,j')\in D_{\mathrm{up}}} 2^{2j(r_1-t')-2j'(r_2+t)}\cdot(\mathrm{count}(j,j'))^2\right)^{1/2}\\
&\lesssim\sqrt{\frac{J+\log(1/\delta)}{N}}\left(\sum_{(j,j')\in D_4} 2^{2j(r_1-t')-2j'(r_2+t)}\cdot 2^{(j'-j)n}+\sum_{(j,j')\in D_{\mathrm{up}}\cap D_6} 2^{2j(r_1-t')-2j'(r_2+t)}\cdot 2^{(j+j')n}\right.\\
&\quad\left.+\sum_{(j,j')\in D_{\mathrm{up}}\cap D_5} 2^{2j(r_1-t')-2j'(r_2+t)}\cdot 2^{(j+j')n}2^{2n(J(t+t'-r)-jt'-j't-(j+j')\widetilde{d})/(2\widetilde{d}+r)}\right)^{1/2}\\
&\lesssim\sqrt{\frac{J+\log(1/\delta)}{N}}\cdot J\cdot\left(1+2^{J\frac{t+t'-r}{\sigma-n/2+t-r/2}(-t-r_2+n/2)}+2^{J\frac{t+t'-r}{2\widetilde{d}+t+t'}(-t-t'+r_1-r_2+n)}+2^{J(-t-t'+r_1-r_2)}\right),
\end{aligned}\tag{C.20}$$

其中，在最后一步中我们反复使用了标准指数和估计

$$\sum_{j=k_1}^{k_2} 2^{j\alpha}\asymp\begin{cases}2^{k_2\alpha} & \text{if } \alpha>0,\\ k_2-k_1 & \text{if } \alpha=0,\\ 2^{k_1\alpha} & \text{if } \alpha<0,\end{cases}$$

此外，由于区域 D_1-D_6 由直线划分，求和中的主要贡献来自位于边界线和角点处的点对 (j,j')。这一观察将提取主阶项的问题化简为检查有限个边界／角点贡献。事实上，式（C.20）最后一行括号内的四项，分别对应于 (j,j')平面中的以下四个块：

$$(0,0),\quad \left(0,\ \frac{t+t'-r}{\sigma-\frac{n}{2}+t-\frac{r}{2}}J\right),\quad \left(\frac{t+t'-r}{2\widetilde{d}+t+t'}J,\ \frac{t+t'-r}{2\widetilde{d}+t+t'}J\right),\quad (J,J).$$

同样，对于（C.16）中的第二项，我们有

$$\begin{aligned}
&\left\|((\mathbf{M}_{(J,t,t')})_{\mathrm{low}})^\top \odot (V^{t,t'})_{\Lambda_J,\cdot}\right\| = \left\|(\mathbf{M}_{(J,t,t')})_{\mathrm{low}} \odot (V^{t,t'})^\top_{\Lambda_J,\cdot}\right\|\\
&\le\left(\sum_{j_0\le j'<j\le J} 2^{-2jt'-2j't}\cdot(\mathrm{count}(j,j'))^2\cdot\|(\mathsf{Var})_{j',j}\|_{\max}^2\right)^{1/2}\\
&\lesssim\sqrt{\frac{J+\log(1/\delta)}{N}}\left(\sum_{(j,j')\in D_{\mathrm{low}}} 2^{2j'(r_1-t)-2j(r_2+t')}\cdot(\mathrm{count}(j,j'))^2\right)^{1/2}\\
&\lesssim\sqrt{\frac{J+\log(1/\delta)}{N}}\left(\sum_{(j,j')\in D_3} 2^{2j'(r_1-t)-2j(r_2+t')}\cdot 2^{(j-j')n}+\sum_{(j,j')\in D_{\mathrm{low}}\cap D_6} 2^{2j'(r_1-t)-2j(r_2+t')}\cdot 2^{(j+j')n}\right.\\
&\quad\left.+\sum_{(j,j')\in D_{\mathrm{low}}\cap D_5} 2^{2j'(r_1-t)-2j(r_2+t')}\cdot 2^{(j+j')n}2^{2n(J(t+t'-r)-jt'-j't-(j+j')\widetilde{d})/(2\widetilde{d}+r)}\right)^{1/2}\\
&\lesssim\sqrt{\frac{J+\log(1/\delta)}{N}}\cdot J\cdot\left(1+2^{J\frac{t+t'-r}{\sigma-n/2+t'-r/2}(-t'-r_2+n/2)}+2^{J\frac{t+t'-r}{2\widetilde{d}+t+t'}(-t-t'+r_1-r_2+n)}+2^{J(-t-t'+r_1-r_2)}\right).
\end{aligned}\tag{C.21}$$

因此，合并（C.16）、（C.20）和（C.21），我们得出，如果

$$N \gtrsim 2^{\widetilde{J}n(\widetilde{t}+\widetilde{t}'-r)/(\sigma-n/2+\widetilde{t}'-r/2)} + \log(1/\delta),$$

那么，以至少 $1-\delta$ 的概率，

$$\begin{aligned}
\text{Error-Var} &= \left\| (\mathbf{M}_{(J,t,t')})_{\text{up}} \odot (V^{t',t})_{\Lambda_J,\cdot} \right\| + \left\| ((\mathbf{M}_{(J,t,t')})_{\text{low}})^\top \odot (V^{t,t'})_{\Lambda_J,\cdot} \right\| \\
&\lesssim \sqrt{\frac{J+\log(1/\delta)}{N}} \cdot J \cdot \Bigg(1 + 2^{J\frac{t+t'-r}{\sigma-n/2+t'-r/2}(-t'-r_2+n/2)} + 2^{J\frac{t+t'-r}{\sigma-n/2+t-r/2}(-t-r_2+n/2)} \\
&\quad + 2^{J\frac{t+t'-r}{\widetilde{2d}+t+t'}(-t-t'+r_1-r_2+n)} + 2^{J(-t-t'+r_1-r_2)} \Bigg) \\
&\asymp \sqrt{\frac{J+\log(1/\delta)}{N}} \cdot J \cdot 2^{\rho J(t+t'-r)/2},
\end{aligned} \tag{C.22}$$

其中

$$\rho = 2\max\left\{ \frac{-t-r_2+n/2}{\sigma-n/2+t-r/2}, \frac{-t'-r_2+n/2}{\sigma-n/2+t'-r/2}, \frac{-t-t'+r_1-r_2+n}{t+t'+2\widetilde{d}}, \frac{-t-t'+r_1-r_2}{t+t'-r}, 0 \right\}.$$

最终方差的主要贡献来自 $(j,\ j')$ 平面上的以下五个块：

$$(0,0),\quad \left(0,\ \frac{t+t'-r}{\sigma-\frac{n}{2}+t-\frac{r}{2}}J\right),\quad \left(\frac{t+t'-r}{\sigma-\frac{n}{2}+t'-\frac{r}{2}}J,\ 0\right),\quad \left(\frac{t+t'-r}{2\widetilde{d}+t+t'}J,\ \frac{t+t'-r}{2\widetilde{d}+t+t'}J\right),\quad (J,J).$$

□

C.5 技术引理

引理 C.3 (范数压缩不等式)。 设 $A \in \mathbb{R}^{p\times p}$，并设 $(I_1,\dots,I_G)$ 为 $\{1,\dots,p\}$ 的一个划分，其中 $|I_j| = p_j$ 且 $p_1+\cdots+p_G = p$。对于 $1 \le i,j \le G$，记 $A_{ij} := A[I_i, I_j] \in \mathbb{R}^{p_i\times p_j}$ 为 (i,j) 块。定义 A 的范数压缩为矩阵

$$\mathcal{N}(A) \in \mathbb{R}^{G\times G}, \qquad \big(\mathcal{N}(A)\big)_{ij} := \|A_{ij}\|,$$

其中 $\|\cdot\|$ 表示算子范数。对于任意 $A \in \mathbb{R}^{p\times p}$ 以及上述任意划分 $(I_1,\dots,I_G)$，

$$\|A\| \le \|\mathcal{N}(A)\| \le \left(\sum_{i,j=1}^{G} \|A_{ij}\|^2 \right)^{1/2}.$$

引理的证明 *C.3*。 设 $x \in \mathbb{R}^p$ 满足 $\|x\|_2 = 1$，并将 $x = (x_1;\ \dots;\ x_G)$ 分解为 $x_j \in \mathbb{R}^{p_j}$。那么 i 第个块 Ax 为 $\sum_{j=1}^G A_{ij}x_j$，因此根据三角不等式和谱范数的定义，

$$\|(Ax)_i\|_2 \ \le\ \sum_{j=1}^{G} \|A_{ij}\|\,\|x_j\|_2 \ =\ \sum_{j=1}^{G} \big(\mathcal{N}(A)\big)_{ij}\, y_j,$$

其中 $y = (y_1,\ \dots,\ y_G)^\top$，满足 $y_j := \|x_j\|_2$。因此，

$$\|Ax\|_2^2 = \sum_{i=1}^{G} \|(Ax)_i\|_2^2 \ \le\ \sum_{i=1}^{G} \Big(\sum_{j=1}^{G} \big(\mathcal{N}(A)\big)_{ij} y_j \Big)^2 = \|\mathcal{N}(A)\, y\|_2^2 \ \le\ \|\mathcal{N}(A)\|^2\, \|y\|_2^2 \ \le\ \|\mathcal{N}(A)\|^2,$$

因为 $\|y\|_2^2 = \sum_j \|x_j\|_2^2 = \|x\|_2^2 = 1$。对所有单位向量 x 取上确界，即得第一个不等式。该结论中的第二个不等式由下式推出

$$\|\mathcal{N}(A)\| \le \|\mathcal{N}(A)\|_F = \Big(\sum_{i,j=1}^{G} \|A_{ij}\|^2 \Big)^{1/2}.$$

□

引理 C.4。设 (X, Z) 服从联合均值为零的高斯分布，且 $X \in \mathbb{R}^d$ 以及 $Z \in \mathbb{R}$。记 e

$$\mu := \mathbb{E}[XZ] \in \mathbb{R}^d, \qquad \sigma_Z^2 := \mathbb{E}[Z^2], \qquad \Sigma_{XX} := \mathbb{E}[XX^\top].$$

设 $(X_i,\ Z_i)_{i=1}^N$ 是 (X, Z) 的独立同分布副本。则对于任意 $\delta \in (0,1)$，以至少 $1-\delta$ 的概率 ，

$$\left\| \frac{1}{N} \sum_{i=1}^{N} X_i Z_i - \mu \right\|_2 \le C\big(\sigma_Z \|\Sigma_{XX}\|^{1/2} + \|\mu\|_2\big) \left(\sqrt{\frac{d + \log(2/\delta)}{N}} + \frac{d + \log(2/\delta)}{N} \right),$$

其中 $C > 0$ 是一个普适常数。

引理 *C.4* 的证明。固定一个单位向量 $u \in \mathbb{S}^{d-1}$。由于 $\langle u, X\rangle$ 和 Z 都服从中心化高斯分布，且 $\|\langle u, X\rangle\|_{\psi_2} \lesssim \|\Sigma_{XX}\|^{1/2}, \|Z\|_{\psi_2} \lesssim \sigma_Z$，因此它们的乘积是次指数变量（例如参见[139,引理 2.7。7]），并且

$$\|\langle u, XZ - \mu\rangle\|_{\psi_1} \lesssim \|\langle u, X\rangle Z\|_{\psi_1} + |\langle u, \mu\rangle| \lesssim \sigma_Z \|\Sigma_{XX}\|^{1/2} + \|\mu\|_2.$$

令 $W_i(u) := \langle u, X_i Z_i - \mu\rangle$。根据独立次指数变量的标量伯恩斯坦不等式（[139, 定理 2.8。1]），对于任意 $t > 0$,

$$\mathbb{P}\left[\left| \frac{1}{N} \sum_{i=1}^{N} W_i(u) \right| \ge C(\sigma_Z \|\Sigma_{XX}\|^{1/2} + \|\mu\|_2) \left(\sqrt{\frac{t}{N}} + \frac{t}{N} \right) \right] \le 2e^{-t}. \tag{C.23}$$

令 $\mathcal{N}$为一个$\frac{1}{4}$2-网，覆盖 $\mathbb{S}^{d-1}$ 且满足 $|\mathcal{N}| \le 5^d$。对 $\mathcal{N}$以及（C.23）应用 $t = d + \log(2/\delta)$并集界，可得其成立的概率至少为 $1-\delta$,

$$\max_{u \in \mathcal{N}} \left| \frac{1}{N} \sum_{i=1}^{N} W_i(u) \right| \le C\big(\sigma_Z \|\Sigma_{XX}\|^{1/2} + \|\mu\|_2\big) \left(\sqrt{\frac{d + \log(2/\delta)}{N}} + \frac{d + \log(2/\delta)}{N} \right).$$

最后，标准的覆盖论证表明，对于任意 $v \in \mathbb{R}^d$,

$$\|v\|_2 \le 2 \max_{u \in \mathcal{N}} \langle u, v\rangle.$$

将其应用于 $v = \frac{1}{N} \sum_{i=1}^{N} X_i Z_i - \mu$，即可得到所需的欧几里得范数界，并完成证明。

□

引理 C.5. 设 $X \sim \mathcal{N}(0, \Sigma)$ 属于 $\mathbb{R}^d$，且 $cI_d \preceq \Sigma \preceq CI_d$，其中某些常数为 $0 < c \le C < \infty$；设 $X_1, \ldots, X_N$是 X的独立同分布副本。定义 $\widehat{\Sigma} := \frac{1}{N} \sum_{i=1}^{N} X_i X_i^\top$. 固定 $\delta \in (0, 1)$。若 $N \gtrsim (\frac{C}{c})^2 (d + \log(1/\delta))$，则其成立的概率至少为 $1 - \delta/2$,

$$\sigma_{\min}(\widehat{\Sigma}) \ge c/2,$$

并且特别地，$\widehat{\Sigma}$ 是可逆的。令 $\Omega \subset \{1, \ldots, d\}$，并在给定 $\widehat{\Sigma}$ 的条件下，令 $v \sim \mathcal{N}(0, (\widehat{\Sigma}^{-1})_{\Omega,\Omega})$。那么，以至少 $1-\delta$ 的概率，对于$(X_1, \ldots, X_N, v)$ 联合成立，

$$\|v\|_{\max} \lesssim \sqrt{\frac{\log(|\Omega|/\delta)}{c}}.$$

引理的证明 *C.5.* 根据高斯样本协方差矩阵的标准算子范数偏差界 [82],

$$\mathbb{P}\left[\|\widehat{\Sigma}-\Sigma\| \lesssim \|\Sigma\|\left(\sqrt{\tfrac{d+\log(2/\delta)}{N}}+\tfrac{d+\log(2/\delta)}{N}\right)\right] \geq 1-\tfrac{\delta}{2}.$$

若 $N \gtrsim (\frac{C}{c})^2(d+\log(1/\delta))$，则偏差项至多为 $\frac{1}{2}\sigma_{\min}(\Sigma)$。根据 Weyl 不等式,

$$\sigma_{\min}(\widehat{\Sigma}) \geq \sigma_{\min}(\Sigma) - \|\widehat{\Sigma}-\Sigma\| \geq \frac{1}{2}\sigma_{\min}(\Sigma) \geq c/2$$

$\widehat{\Sigma}$的概率至少为 $1-\delta/2$。将此事件记为 E；特别地， Σ 在 E 上可逆。

给定 $\widehat{\Sigma}$，并令 $v \sim \mathcal{N}(0, (\widehat{\Sigma}^{-1})_{\Omega,\,\Omega})$。由于 $v=(v_j)_{j\in\Omega}$ 的坐标为 $v_j \sim \mathcal{N}(0, (\widehat{\Sigma}^{-1})_{jj})$，其中 $j\in\Omega$，因此我们有

$$\mathbb{P}\Big[\|v\|_{\max} > t \,\Big|\, \widehat{\Sigma}\Big] \leq \sum_{j\in\Omega} 2\exp\Big(-\frac{t^2}{2(\widehat{\Sigma}^{-1})_{jj}}\Big) \leq 2|\Omega|\exp\Big(-\frac{t^2\,\sigma_{\min}(\widehat{\Sigma})}{2}\Big).$$

取 $t=\sqrt{2\log(4|\Omega|/\delta)\,\sigma_{\min}(\widehat{\Sigma})}$，得到条件概率至多为 $\delta/2$。在 E 上， $\sigma_{\min}(\widehat{\Sigma}) \geq c/2$，因此

$$\|v\|_{\max} \leq \sqrt{\frac{4\log(4|\Omega|/\delta)}{c}} \lesssim \sqrt{\frac{\log(|\Omega|/\delta)}{c}}$$

以至少为 $1-\delta/2$ 的条件概率。

将这两个事件相交，可得结论以至少为 $1-\delta$ 的概率同时关于 $(X_1, \ldots\ldots, X_N, v)$ 成立。 □

D 辅助材料第 6

我们解释为何通过更新后的支撑集定义的修正估计量 支撑集$^{\text{new}}(J, t, t')$ 在（6.4）中可以得到增强界（6.1）。

直接修改命题 5.10 的证明可得

$$\text{Error-Compression} \lesssim \varepsilon 2^{-J(t+t'-r)}.$$

事实上，回顾

$$\text{Error-Compression} = \|\mathbf{D}_{\Lambda_J}^{-t'}(\mathbf{A}^{\varepsilon}_{\Lambda_J}-\mathbf{A}_{\Lambda_J})\mathbf{D}_{\Lambda_J}^{-t}\| \leq \Bigg(\sum_{j,j'=j_0}^{J} 2^{-2jt'-2j't}\|(\mathbf{A}^{\varepsilon}_{\Lambda_J})_{j,j'}-(\mathbf{A}_{\Lambda_J})_{j,j'}\|^2\Bigg)^{1/2}.$$

对于执行新阈值化规则的指标 (j,j')，我们有

$$\|(\mathbf{A}^{\varepsilon}_{\Lambda_J})_{j,j'}-(\mathbf{A}_{\Lambda_J})_{j,j'}\| \lesssim 2^{-(j+j')\widetilde{d}}\cdot(\tau^{\text{new}}_{jj'})^{-(2\widetilde{d}+r)} \leq \frac{\varepsilon}{J}2^{-J(t+t'-r)+jt'+j't},$$

对于被丢弃的指标

$$\|(\mathbf{A}^{\varepsilon}_{\Lambda_J})_{j,j'}-(\mathbf{A}_{\Lambda_J})_{j,j'}\| = \|(\mathbf{A}_{\Lambda_J})_{j,j'}\| \lesssim 2^{(j+j')r/2-|j-j'|(\sigma-n/2)} \leq \frac{\varepsilon}{J}2^{-J(t+t'-r)+jt'+j't}.$$

将上述界代入，即可得到所声称的压缩误差估计。

接下来注意，新的阈值参数和斜率条件与（4.2）中的条件仅相差关于 J 的多项式因子，以及至多 $O(\log J)$ 个尺度偏移。因此，命题 5.11 和命题5.12 中的论证稍作修改后仍然成立。

对于命题 5.11，所需修改包括：在新的阈值化规则和斜率条件下验证支撑包含关系（如引理B.1中所述），并选取一个~经过适当调整的 J。由此得到

$$\text{Error-OVB} \lesssim \varepsilon 2^{-J(t+t'-r)}.$$

对于方差项，可将命题 C.2 中的支撑模式分析推广到新的支撑集 支撑集$^{\text{new}}(J,\ t,\ t')$。据此，命题 5.12 的证明可直接推广，从而得到

$$\text{Error-Var} \lesssim \sqrt{\frac{\log(1/\delta)}{N}} \cdot (J/\varepsilon)^{\kappa} \cdot 2^{\rho J(t+t'-r)/2},$$

对于某个常数 $\kappa > 0$，其中 ρ 与（5.2）中的指数相同。为简洁起见，我们省略技术细节。

最后，合并 误差-压缩、误差-OVB和 误差-方差的界可得估计式（6.1）。