单纯复形上随机信号的平稳性与谱特征

Madeline Navarro，电气与电子工程师学会学生会员，Andrei Buciulea，电气与电子工程师学会会员，
Santiago Segarra，电气与电子工程师学会高级会员，和 Antonio G. Marques，电气与电子工程师学会高级会员

arXiv:2602.03055v1 [eess.SP] 3 Feb 2026

摘要——**数据具有复杂结构的情况日益普遍，因此亟需新的模型和分析工具。图是一种重要的结构类型，可编码任意两个实体（节点）之间的关系。然而，图既无法表示非成对连接，也无法表示节点集合之间的关系。因此，我们转向单纯复形，以连接两个以上的节点，并对单形（如边和三角形）之间的关系建模。由此，我们的数据由定义在以单纯复形表示的拓扑空间上的信号构成。近期有大量研究探讨这类拓扑信号，但主要采用确定性表述。我们提出一个概率框架，用于描述定义在单纯复形上的随机信号。具体而言，我们将经典的平稳性概念加以推广。根据霍奇理论和狄拉克理论的谱对偶性，我们将平稳拓扑信号定义为拓扑滤波器对白噪声的输出。该定义自然延伸了平稳性的理想性质，这些性质适用于时间序列和图信号。尤为重要的是，我们通过明确的谱刻画，恰当地定义了拓扑功率谱密度（PSD）。随后，我们借助功率谱密度讨论拓扑平稳性在谱性质方面的优势。此外，我们还通过多项合成模拟和真实场景仿真，实证展示了这些优势的实用性。**

索引词—**拓扑信号的随机建模、单纯复形信号处理、平稳性、拓扑信号处理。**

I. 引言

随着数据科学日益复杂，现代任务越来越多地涉及位于非规则域上的数据。尽管图是复杂系统的常用模型，但其仅限于成对交互的局限往往不够充分，从而促使人们引入关联多个实体的高阶连接。当多元关系也与数据相关联时，我们可以利用单纯复形作为简洁而富有表达力的域，以承载节点、边、三角形等上的信号 [1]–[2]。近年来，针对拓扑（单纯）信号已发展出一套丰富的确定性工具 [3]–[4]。然而，从生物建模到交通、金融和网络推断，实际应用通常并不具备完美测量的数据 [5]。现有确定性工具可能无法考虑由噪声、缺失值或模型偏差引起的不确定性。因此，我们有必要发展易于处理的概率模型与推断工具，以便从高维数据中可靠地得出结论。

作为自然的切入点，我们聚焦于平稳性，它是时间序列和随机场中谱分析与估计的基础 [6], [7]。图域中的信号也早已建立了平稳性的理论 [8]–[11]。例如，先前的研究以类似于时间序列的方式定义了随机图信号的弱平稳性 [12], [13]。与时域中的情形一样，这些图平稳性的定义简化了谱行为，从而推广了信号处理（SP）中的重要定义，例如功率谱密度（PSD）或谱卷积。

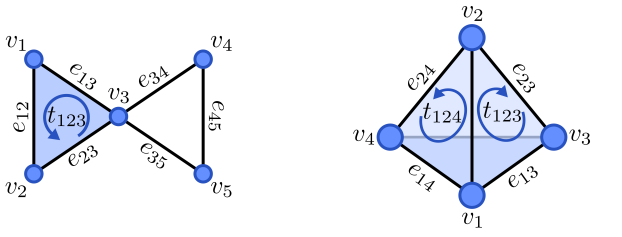
在此，我们研究定义在单纯复形上的信号。单纯复形涵盖图信号，同时还支持多方交互 [1], [2]。借助开创性的霍奇分解 [3]–[4], [14]–[16], ，任意给定阶数的单纯信号都可以分解为梯度分量、旋度分量和调和分量。这不仅在代数上简洁方便，也反映了真实数据中观察到的结构，例如交通网络中的流近乎无散，或无套利货币交易中边信号无旋 [15], [17]–[19]。此外，当数据存在于多个维度上（例如同时包含节点和边上的信号）时，我们无需分别处理它们；狄拉克算子能够耦合不同阶数的信号，从而实现多阶 谱处理 [20], [21]。霍奇算子和狄拉克算子都促成了许多经典概念的拓扑扩展，包括卷积架构、趋势滤波、神经网络、自回归（AR）建模、恢复方法、随机游走和高斯模型 [3]–[5], [15], [22]–[28]。

尽管已有这些进展，针对随机单纯形信号的概率建模仍处于起步阶段。一些著名的概率模型已被适配到单纯结构，例如扩散模型和高斯过程 [29], [30] 然而，我们需要更一般的随机框架，使其在不确定性建模、谱分析以及跨单纯形阶的统计可靠推断方面具有计算可行性。具体而言，我们寻求这样的模型：(i) 遵循霍奇/狄拉克谱几何，(ii) 提供清晰的谱刻画，以及 (iii) 导出适用于实际设置（如数据稀缺情形）的有原则的估计器与滤波器。

我们的方法。受图设定以及基于霍奇或狄拉克的拓扑谱性质的启发，我们为平稳拓扑随机信号建立了概率框架。在第三节中，我们将平稳拓扑过程定义为给定白噪声时拓扑滤波器的输出。我们还表明，可借助单纯复形的谱性质方便地刻画拓扑信号的行为，无论是针对单一给定阶，还是在多阶情形下。我们的定义在数学上简明直接，与图或时间序列上的平稳性相一致，同时又通过单纯复形所允许的多路关系提供了更丰富的建模能力。

具体而言，我们作出以下贡献：1）我们为随机单纯信号提出了拓扑平稳性 的定义，并推导出相应的功率谱密度。

Madeline Navarro 和 Santiago Segarra 任职于莱斯大学电子与计算机工程系。Andrei Buciulea 和 Antonio G. Marques 任职于胡安·卡洛斯国王大学信号理论与通信系。本论文得到西班牙国家研究署（AEI/10.13039/501100011033）项目 PID2023-149457OB-I00、马德里自治区的 IDEA-CM（TEC-2024/COM-89）项目以及 ELLIS马德里分部的支持。



(a) 二阶示例。 (b) 三阶示例。

图 1：不同阶数的单纯复形示例。(a) 一个包含一个三角形 $\{1,2,3\}$ 的二阶单纯复形。边 $\{1,2\}$ 与 $\{1,2,3\}$ 对齐，但 $\{1,3\}$ 与 $\{1,2,3\}$ 反对齐。(b) 一个包含一个四面体 $\{1,2,3,4\}$ 的二阶单纯复形。三角形 $\{1,2,4\}$ 与 $\{1,2,3,4\}$ 对齐，但 $\{1,2,3\}$ 与 $\{1,2,3,4\}$ 反对齐。

2）我们开发并分析了非参数方法和参数方法，用于估计平稳随机拓扑信号的功率谱密度或协方差。

3) 我们将经典统计定义和工具（包括维纳滤波器）推广到拓扑域，并通过信号去噪等典型信号处理任务的实例，展示这些工具的用途。

组织结构。第二节首先回顾单纯复形、霍奇算子和狄拉克算子以及谱分解的预备知识。随后，第三节对拓扑信号的平稳性进行形式化定义，并将其与先前的表述联系起来。第四节介绍多种用于估计平稳拓扑信号的协方差矩阵和功率谱密度的估计器；第五节则展示如何将维纳滤波器等经典工具推广到拓扑域。第六节报告数值实验，以验证这些工具在平稳性假设下估计拓扑信号协方差和功率谱密度的效用。最后，我们在第七节总结论文。

II. 拓扑信号处理基础

我们首先简要回顾拓扑信号处理的基本概念，为下文定义的平稳单纯信号奠定基础。有关这部分内容的更详细介绍见 [31]。

A. 单纯复形与单纯信号

单纯复形由一组 $\mathcal{V}$ 顶点之间的多元关系构成。具体而言，$\mathcal{P}^K$ 的 $K \leq N_0$ 阶单纯复形包含单形，即 $\mathcal{V}$ 的子集，其中最大单形的基数为 $K+1$。此外，如果 $\mathcal{W} \in \mathcal{P}^K$ 包含 $k+1$ 个顶点，那么 $\mathcal{W}$ 就是一个 k-单形。我们可以将 $\mathcal{P}^K$ 中 k-单形的集合记作 $\{\mathcal{W}_n^k\}_{n=1}^{N_k}$，其中 N_k 是 k-单形的数量，单形总数为 $N = \sum_{k=1}^{K} N_k$。因此，我们可以将 $\mathcal{P}^K$ 定义为集合 $\{\mathcal{W}_n^0\}_{n=1}^{N_0}$、$\{\mathcal{W}_n^1\}_{n=1}^{N_1}$、……、$\{\mathcal{W}_n^K\}_{n=1}^{N_K}$，并满足包含性质：对于每个单形 $\mathcal{W} \in \mathcal{P}^K$，它的所有子集也都属于 $\mathcal{P}^K$，即对于所有 $\mathcal{W}' \subset \mathcal{W}$，$\mathcal{W}' \in \mathcal{P}^K$。对于给定的 k-单形 $\mathcal{W}^k$，如果 $\mathcal{W}^{k-1} \subset \mathcal{W}^k$，则 $(k-1)$-单形 $\mathcal{W}^{k-1}$ 是 $\mathcal{W}^k$ 的面；如果 $\mathcal{W}^{k+1} \supset \mathcal{W}^k$，则 $(k+1)$-单形 $\mathcal{W}^{k+1}$ 是 $\mathcal{W}^k$ 的余面。尽管这个定义看似抽象，单纯复形却能直观地表示欧氏空间中的连通数据。图 1a 展示了一个 2 阶单纯复形，其中 0-单形表示顶点，1-单形表示边，2-单形表示三角形而图 1b 展示了一个 3 阶单纯复形，其中 3-单形是一个四面体。

接下来，我们介绍用于关联相邻维度信号的算子。为此，我们需要为 $\mathcal{P}^K$ 中的任意单形指定参考定向。我们将 $\mathcal{P}^k$ 中任意单形 $\mathcal{W} \subseteq \mathcal{V}$ 的规范表示定义为由顶点索引排序的 $\mathcal{V}$ 子集。例如，$\{1,2,4\}$、$\{2,1,4\}$ 和 $\{2,4,1\}$ 都对应同一个 2-单纯形，其规范表示为 $\{1,2,4\}$。注意，$\{2,1,4\}$ 与 $\{1,2,4\}$ 反对齐，因为有向边流 $2\to1\to4\to2$ 与规范 $1\to2\to4\to1$ 方向相反。随后，我们可以引入所有 $k=1,\ldots,K$ 对应的关联矩阵 $\mathbf{B}_k \in \{-1,0,1\}^{N_{k-1}\times N_k}$，它们连接相邻阶的单形 $k-1$ 和 k [3]。若 $[\mathbf{B}_k]_{nm} \neq 0$，则 $\mathcal{W}_m^k \in \mathcal{P}^K$ 存在且 $\mathcal{W}_n^{k-1} \subset \mathcal{W}_m^k$；$[\mathbf{B}_k]_{nm}$ 的符号表示 $\mathcal{W}_n^{k-1}$ 与 $\mathcal{W}_m^k$ 相对于各自参考定向是对齐还是反对齐，如图 1 所示。利用这些矩阵，我们为每个维度 [14] 引入霍奇拉普拉斯算子

$$\begin{aligned}&\mathbf{L}_0 = \mathbf{B}_1\mathbf{B}_1^\top,\ \mathbf{L}_K = \mathbf{B}_K^\top\mathbf{B}_K,\ \text{and}\\&\mathbf{L}_k = \mathbf{L}_{k,\ell} + \mathbf{L}_{k,u}\ \text{for}\ k=1,\ldots,K-1,\\&\qquad\text{with}\ \mathbf{L}_{k,\ell} = \mathbf{B}_k^\top\mathbf{B}_k,\ \text{and}\ \mathbf{L}_{k,u} = \mathbf{B}_{k+1}\mathbf{B}_{k+1}^\top.\end{aligned} \tag{1}$$

每个维度 $1 \leq k \leq K-1$ 都关联一个下拉普拉斯算子 $\mathbf{L}_{k,\ell}:=\mathbf{B}_k^\top\mathbf{B}_k$ 和一个上拉普拉斯算子 $\mathbf{L}_{k,u}:=\mathbf{B}_{k+1}\mathbf{B}_{k+1}^\top$，它们分别通过共享的面和余面连接 k-单形。(1) 中的霍奇拉普拉斯算子描述了给定阶数的单形如何连接，但我们也会将它们用作拓扑信号算子。

为此，我们必须先介绍单纯形信号，即定义在单纯复形上的实值数据。首先，我们考虑单一阶数上的数据。一个 k-阶单纯信号（简称为 k-信号）由向量 $\mathbf{s}_k = [s_1^k,\ldots,s_{N_k}^k]^\top \in \mathbb{R}^{N_k}$ 表示，其中条目 s_n^k 表示第 n 个 k-单形 [3] 的值。s_n^k 的符号表示信号的方向与第 n 个单形的参考方向对齐还是相反，图 1 对此作了示例。接下来，我们将所有阶数上的信号拼接起来，定义一个完整的单纯复形信号

$$\mathbf{s} = \left[\mathbf{s}^{0\top},\ldots,\mathbf{s}^{K\top}\right]^\top \in \mathbb{R}^N. \tag{2}$$

B. 霍奇分解

关联矩阵的一个关键性质是$\mathbf{B}_k\mathbf{B}_{k+1}=\mathbf{0}$ [32]。因此，根据式 (1) 中霍奇拉普拉斯算子的定义，我们得到霍奇分解，它可以将单纯信号空间分解为三个互相正交且各具意义的子空间。由于 $\text{span}(\mathbf{B}_k^\top)$ 与 $\text{span}(\mathbf{B}_{k+1})$ 正交，我们可以将任意 k-信号$\mathbf{s}^k$ 唯一分解为

$$\mathbf{s}^k = \mathbf{s}_{\text{G}}^k + \mathbf{s}_{\text{C}}^k + \mathbf{s}_{\text{H}}^k, \tag{3}$$

其中，$\mathbf{s}_{\text{G}}^k \in \text{span}(\mathbf{B}_k^\top)$、$\mathbf{s}_{\text{C}}^k \in \text{span}(\mathbf{B}_{k+1})$ 和 $\mathbf{s}_{\text{H}}^k \in \text{kernel}(\mathbf{L}_k)$ 分别属于互相正交的子空间，对应于梯度、旋度和调和分量 [14]。直观来说，给定一个 k-信号$\mathbf{s}$，我们对各个分量的解释如下 [15]。

- **梯度：** 变换$\mathbf{B}_{k+1}^\top\mathbf{s}$ 将$\mathbf{s}$ 中的值映射到所有 k-单形的 $(k+1)$-余面，表示 k-信号值围绕 $(k+1)$-单形的环流，从而产生离散旋度运算。
- **旋度：** 我们将$\mathbf{s}$ 映射到每个 k-单纯形的$(k-1)$-面，经由$\mathbf{B}_k\mathbf{s}$，这将对每个$(k-1)$-单纯形处的 k-信号值求和，

表示净流，从而诱导离散散度运算。

• **调和：** 任何信号$\mathbf{s} \in$ 核$(\mathbf{B}_{k+1}^{\top}) \cap$核$(\mathbf{B}_k)$均为无旋且无散，因此既不会对绕$(k+1)$-单纯形的环流产生贡献，也不会对穿过$(k-1)$-单纯形的净流产生贡献。

因此，我们可以将梯度分量和旋度分量写为

$$\mathbf{s}_{\mathrm{G}}^{k} = \mathbf{B}_{k}^{\top}\mathbf{x}^{k-1}, \qquad \mathbf{s}_{\mathrm{C}}^{k} = \mathbf{B}_{k+1}\mathbf{x}^{k+1}$$

对于某个$(k-1)$-信号$\mathbf{x}^{k-1}$和$(k+1)$-信号$\mathbf{x}^{k+1}$。因此，$\mathbf{s}_{\mathrm{G}}^{k}$由下层连接驱动，也就是说，由$(k-1)$-面驱动；而$\mathbf{s}_{\mathrm{C}}^{k}$由上层连接驱动，也就是说，由$(k+1)$-余面驱动；调和分量$\mathbf{s}_{\mathrm{H}}^{k}$则表示$\mathbf{s}^{k}$中与这两者均正交的剩余部分。

由于式(3)中的三个分量分别对应互相正交子空间，我们还可以基于霍奇拉普拉斯算子$\mathbf{L}_k$的特征分解，对信号进行谱刻画。令$\mathbf{U}_k$表示$\mathbf{L}_k$的特征向量，并将其进一步分解为$\mathbf{U}_k = [\mathbf{U}_{k,\mathrm{G}}、\mathbf{U}_{k,\mathrm{C}}、\mathbf{U}_{k,\mathrm{H}}]$，其中$\mathbf{U}_{k,\mathrm{G}}$和$\mathbf{U}_{k,\mathrm{C}}$分别表示与$\mathbf{L}_k$，$\ell$和$\mathbf{L}_{k,u}$的非零特征值相关联的特征向量，而$\mathbf{U}_{k,\mathrm{H}}$张成$\mathbf{L}_k$的零空间。因此，我们也可以等价地将 k-信号$\mathbf{s}^{k}$ 写为

$$\mathbf{s}^{k} = \mathbf{U}_{k,\mathrm{G}}\tilde{\mathbf{s}}_{\mathrm{G}}^{k} + \mathbf{U}_{k,\mathrm{C}}\tilde{\mathbf{s}}_{\mathrm{C}}^{k} + \mathbf{U}_{k,\mathrm{H}}\tilde{\mathbf{s}}_{\mathrm{H}}^{k}, \tag{4}$$

其中，谱分量 $\tilde{\mathbf{s}}_{\mathrm{G}}^{k} := \mathbf{U}_{k,\mathrm{G}}^{\top}\mathbf{s}^{k}$, $\tilde{\mathbf{s}}_{\mathrm{C}}^{k} := \mathbf{U}_{k,\mathrm{C}}^{\top}\mathbf{s}^{k}$ 和 $\tilde{\mathbf{s}}_{\mathrm{H}}^{k} := \mathbf{U}_{k,\mathrm{H}}^{\top}\mathbf{s}^{k}$ 构成了 $\tilde{\mathbf{s}}^{k} := \mathbf{U}_{k}^{\top}\mathbf{s}^{k}$，其中包含拓扑频率系数。更具体地说，$\mathbf{U}_{k}^{\top}\mathbf{s}^{k}$ 给出了 $\mathbf{s}^{k}$ 的拓扑傅里叶变换（TFT），使我们能够将信号在单纯复形上的频率分解加以推广 [3], [4]。

C. 狄拉克分解

霍奇分解是一种经典且理论基础扎实的方法，用于刻画给定单一阶数下单纯形信号的谱结构。我们还可以进一步运用霍奇拉普拉斯算子构建算子，以刻画不同阶数之间的信号交互。当我们关注多阶信号，即同时在所有阶数上观测到的单纯形信号时，就会采用狄拉克算子$\mathbf{D}$ [21], [33], , 其尺寸必须为 $N \times N$，以涵盖单纯复形 $\mathcal{P}^{K}$ 中所有阶数的 $N = \sum_{k=1}^{K} N_k$ 单形。此外，$\mathbf{D}$ 具有$K \times K$ 分块结构，其中第 (k, k') 个分块记为$[\![\mathbf{D}]\!]_{k,k'}$，关联阶数 k 和 k'。形式上，每个分块定义为 [21], [33]

$$[\![\mathbf{D}]\!]_{k,k'} = \begin{cases} \mathbf{B}_k & \text{if } k' = k+1 \\ \mathbf{B}_{k'}^{\top} & \text{if } k' = k-1 \\ \mathbf{0} & \text{otherwise} \end{cases}. \tag{5}$$

请注意，由于我们只考虑相邻阶之间的交互，$\mathbf{D}$ 的大多数块均为零值，只有上对角块和下对角块例外。将$\mathbf{D}$ 作为算子$\mathbf{y} = \mathbf{D}\mathbf{x}$ 应用时，它将单纯复形信号空间映射到自身。不过，若将$\mathbf{x}$ 和$\mathbf{y}$ 组织成与$\mathbf{D}$相符的分块向量，则可将$\mathbf{y}$ 的第 k个子向量写为 $[\![\mathbf{y}]\!]_k = \mathbf{B}_k[\![\mathbf{x}]\!]_{k-1} + \mathbf{B}_{k+1}^{\top}[\![\mathbf{x}]\!]_{k+1}$，也就是说，$\mathbf{D}$ 将信号值映射为来自相邻维度 $k \pm 1$的k信号 $[\![\mathbf{y}]\!]_k$。

同样重要的是$\mathbf{D}$的谱特性。狄拉克算子的平方得到一个块对角矩阵 $\mathcal{L} = \mathbf{D}^2$ ，其对角线上包含霍奇拉普拉斯算子，即 $\mathcal{L} := \mathrm{blkdiag}(\{\mathbf{L}_k\}_{k=0}^{K})$。事实上，算子$\mathbf{D}$ 将 $k \pm 1$ 阶信号映射为 k阶信号，因此，连续应用两次$\mathbf{D}^2$ 应通过上层和下层连接，将 k阶信号映射为 k阶信号，也就是式（1）中的上拉普拉斯算子和下拉普拉斯算子。回顾 $\mathrm{span}(\mathbf{B}_k^{\top})$ 与 $\mathrm{span}(\mathbf{B}_{k+1})$ 的正交性以及式（1）的定义，我们可以将 $\mathbf{D}$ 的特征向量与关联矩阵的奇异向量联系起来。此外，利用$\mathbf{D}$的任意偶次幂，我们可以得到第 II-B 节中霍奇分解的谱表示。因此，我们可以通过$\mathbf{D}$的特征向量对多阶信号应用拓扑傅里叶变换，这与式（4）中基于霍奇算子的拓扑傅里叶变换类似。

III. 拓扑弱平稳性

现在我们来讨论本文的核心内容：平稳随机拓扑信号的定义。在经典时间序列分析中，如果一个随机过程的相关矩阵在时间平移 [6], [7]下保持不变，则称其为弱平稳。这意味着时间平稳过程具有两个基本性质。首先，当且仅当信号可表示为白噪声经过线性时不变滤波器后的输出时，该信号才是弱时间平稳的。其次，任何时间平稳过程的相关矩阵的谱性质都能完整描述其二阶统计特性。为了将这一概念推广到拓扑域，我们采用这两个性质来定义单纯复形信号的弱平稳性。

然而，与图域上的平稳性不同， [10], [11],我们还必须考虑该域的多阶特性。事实上，我们可能关注在所有阶上观测到的多阶信号。即使我们只要求信号定义在单阶上，单纯复形域也会体现节点、边、三角形及更高维单形之间的高阶交互。因此，我们希望定义一种弱平稳性，使其与第 II-B 节中由霍奇拉普拉斯算子诱导的谱几何或第 II-C 节中由狄拉克算子诱导的谱几何相一致。具体而言，我们将平稳拓扑随机信号定义为给定白噪声时某个拓扑滤波器 的输出，该滤波器由$\mathbf{L}_k$ 或$\mathbf{D}$表征，具体取决于我们的研究关注。

定义 1。对于某个单纯复形 $\mathcal{P}^{K}$，设$\mathbf{T} \in \{\mathbf{L}_k, \mathbf{D}\}$。若零均值随机拓扑信号$\mathbf{s}$ 可以表示为如下形式，则称其相对于$\mathbf{T}$ 是弱平稳的：

$$\mathbf{s} = \mathbf{H}\mathbf{w}, \tag{6}$$

其中 $\mathbf{w}$ 是满足 $\mathbb{E}[\mathbf{w}] = \mathbf{0}$ 且$\mathbf{E}[\mathbf{w}\mathbf{w}^{\top}] = \mathbf{I}$的白单纯信号，并且拓扑滤波器 $\mathbf{H}$ 与$\mathbf{T}$可交换。

因此，我们对单纯形信号弱平稳性的定义是通过线性移不变滤波器来刻画的。不过，与时间域或图域相比，我们对拓扑滤波器的描述更加灵活。我们考虑在线性滤波器中，将输入信号在底层域上经过移位的版本进行聚合，其中移位对应于上层或下层连接，即面或余面。我们还可以选择对单阶的 k-信号进行滤波，或对联合单纯形空间上的多阶信号 (2) 进行滤波。更具体地说，令 $\mathbf{T} \in \{\mathbf{L}_k, \mathbf{D}\}$ 表示一个

我们感兴趣的拓扑算子。我们将拓扑滤波器定义为 $\mathbf{T}$ 的多项式[1]，

$$\mathbf{H} = \sum_{r=0}^{R-1} h_r \mathbf{T}^r \tag{7}$$

对于实值系数 $\mathbf{h} := [h_0, \ldots, h_{R-1}]^\top \in \mathbb{R}^R$。若$\mathbf{T} = \mathbf{L}_k$，则对于给定阶数 k，我们得到一个霍奇域滤波器，通过上层和下层连接处理 k-信号。或者，对于$\mathbf{T} = \mathbf{D}$，我们处理同时支撑于所有单纯形阶上的信号，从而得到一个狄拉克域滤波器。此外，注意到$\mathbf{L}_k$ 和$\mathbf{D}$均为对称矩阵。因此，$\mathbf{T}$ 是对称矩阵，故可通过特征分解对角化为$\mathbf{T} = \mathbf{U}\mathrm{diag}(\boldsymbol{\lambda})\mathbf{U}^\top$。于是，滤波器$\mathbf{H}$ 具有如下谱表示

$$\mathbf{H} = \mathbf{U}\mathrm{diag}\big(h(\boldsymbol{\lambda})\big)\mathbf{U}^\top, \quad h(\lambda_i) = \sum_{r=0}^{R-1} h_r \lambda_i^r, \tag{8}$$

其中，$h(\cdot)$是该滤波器的拓扑频率响应，它逐元素作用于特征值 $\boldsymbol{\lambda}$。

回顾一下，弱平稳性与两个性质相关：作为白噪声经过滤波器后的输出，以及信号相关性的谱性质。我们的定义 1 显然受到第一个性质的启发。因此，接下来我们引入平稳拓扑信号的协方差矩阵，以考察第二个性质。根据定义 1，我们可以将平稳拓扑信号$\mathbf{s}$ 的协方差表示为

$$\begin{aligned}\mathbf{C} &:= \mathbb{E}[\mathbf{s}\mathbf{s}^\top] = \mathbb{E}[(\mathbf{H}\mathbf{w})(\mathbf{H}\mathbf{w})^\top] = \mathbf{H}\mathbb{E}[\mathbf{w}\mathbf{w}^\top]\mathbf{H}^\top \\ &= \mathbf{H}\mathbf{H}^\top = \mathbf{H}^2,\end{aligned} \tag{9}$$

其中，最后一个不等式由$\mathbf{H}$的对称性得出，因为$\mathbf{T}$ 是对称的。因此，协方差完全由滤波器$\mathbf{H}$决定，而$\mathbf{H}$本身取决于算子$\mathbf{T}$的选择以及滤波器系数$\mathbf{h}$。注意，由于$\mathbf{C} = \mathbf{H}^2 = \mathbf{U}\mathrm{diag}^2\big(h(\boldsymbol{\lambda})\big)\mathbf{U}^\top$，$\mathbf{C}$ 的秩与$\mathbf{H}$的秩相同。

对$\mathbf{C}$的这一刻画引出了一个几乎等价且通常更易解释的弱平稳性定义，该定义由协方差$\mathbf{C}$的谱推导而来。在时域中，平稳过程的自相关可由傅里叶变换对角化。类似地，平稳拓扑过程的协方差可通过$\mathbf{T} \in \{\mathbf{L}_k, \mathbf{D}\}$的特征基，经由 TFT 对角化。以下定义将对此进行形式化表述。

定义 2. 若零均值随机拓扑信号$\mathbf{s}$ 相对于$\mathbf{T} \in \{\mathbf{L}_k, \mathbf{D}\}$ 是弱平稳的，则其协方差$\mathbf{C}$与$\mathbf{T}$ 可同时对角化，即，

$$\mathbf{C} = \mathbf{U}\mathrm{diag}(\mathbf{p})\mathbf{U}^\top, \tag{10}$$

其中，$\mathbf{U}$ 表示$\mathbf{C}$ 和$\mathbf{T}$的共享特征向量，$\mathbf{p} \geq \mathbf{0}$包含$\mathbf{C}$的特征值，我们将其称为功率谱密度。

请注意，定义 2 比定义 1 稍微更一般。要理解这一点，回想一下，对于式 (6) 中的任意平稳信号$\mathbf{s}$，其协方差矩阵$\mathbf{C}$ 具有式 (9) 的形式，与$\mathbf{T}$ 共享特征向量$\mathbf{U}$，且特征值为 $\mathrm{diag}^2\big(h(\boldsymbol{\lambda})\big)$。因此，$\mathbf{s}$也满足定义 2。然而，特征值并不受约束

[1]由于篇幅所限，本文未予介绍；此外，也可以考虑分别在下拉普拉斯算子和上拉普拉斯算子上独立运作的生成滤波器 [34]。

式 (10) 中$\mathbf{C}$ 的特征值，而定义 1 则隐含地要求：对于$\mathbf{T}$的任意两个特征值 λ_i 和 λ_j，如果 $\lambda_i = \lambda_j$，那么 $p_i = p_j$。重要的是，定义 2 强调了功率谱密度$\mathbf{p}$在刻画平稳拓扑信号的二阶统计量方面的重要性。给定$\mathbf{T}$的特征向量，功率谱密度即可完整描述协方差矩阵$\mathbf{C}$。我们将$\mathbf{p}$ 的定义形式化如下。

定义3. 对于一个相对于$\mathbf{T} = \mathbf{U}\mathrm{diag}(\boldsymbol{\lambda})\mathbf{U}^\top \in \{\mathbf{L}_k, \mathbf{D}\}$平稳的随机拓扑信号$\mathbf{s}$，其功率谱密度（*PSD*）定义为

$$\mathbf{p} := \mathrm{diag}(\mathbf{U}^\top \mathbf{C}\mathbf{U}). \tag{11}$$

给定特征基$\mathbf{U}$，向量$\mathbf{p}$ 不仅能够完整表征$\mathbf{C}$，而且也是经典功率谱密度的直接对应物，后者用于平稳时间序列过程。事实上，平稳拓扑信号的功率谱密度继承了其时域对应物的相同关键性质，具体表述如下。

推论1. 设$\mathbf{U}$表示底层拓扑算子$\mathbf{T} \in \{\mathbf{L}_k, \mathbf{D}\}$（对于 $k \in \{0, 1, \cdots, K\}$）的特征基，并考虑$\mathbf{s}$的拓扑傅里叶变换，即其谱表示

$$\tilde{\mathbf{s}} = \mathbf{U}^\top \mathbf{s}. \tag{12}$$

那么，$\tilde{\mathbf{s}}$ 的协方差是对角的，且可表示为

$$\tilde{\mathbf{C}} = \mathbb{E}[\tilde{\mathbf{s}}\tilde{\mathbf{s}}^\top] = \mathrm{diag}(\mathbf{p}). \tag{13}$$

证明。结果可直接得出，因为

$$\mathbb{E}[\tilde{\mathbf{s}}\tilde{\mathbf{s}}^\top] = \mathbf{U}^\top \mathbb{E}[\tilde{\mathbf{s}}\tilde{\mathbf{s}}^\top]\mathbf{U} = \mathbf{U}^\top \mathbf{C}\mathbf{U} = \mathrm{diag}(\mathbf{p}). \qquad \square$$

该推论正式阐明了我们的主张：定义 3 类似于时间平稳过程的功率谱密度。具体而言，请注意，协方差$\mathbf{C}$是对角的，这表明平稳拓扑信号在谱域中不相关，其中信号的谱表示由拓扑算子$\mathbf{T}$的特征向量定义。这种简化本身已足以证明将谱域用于信号分析的合理性，同时也意味着平稳性的另一项一般性质。由于$\mathbf{p}$的各项彼此不相关，我们只需考虑第i项 p_i，便可完整表示该平稳过程在第 i个拓扑频率上的功率。因此，式 (10) 中的功率谱密度$\mathbf{p}$可视为随机信号$\mathbf{s}$ 在由$\mathbf{T}$定义的各个拓扑频率上的能量分布。

此外，在满足定义 1 的条件下，我们可以将时间域和图域中线性移位不变滤波器与平稳过程功率谱密度之间的经典关系推广至单纯复形域。具体而言，给定式 (6) 中的单纯信号$\mathbf{s} = \mathbf{H}\mathbf{w}$ 以及式 (9) 中的协方差，我们可以将$\mathbf{p}$写为

$$p_i = h(\lambda_i)^2 \quad \text{for all } i \in [N] := \{1, \ldots, N\}, \tag{14}$$

因此，功率谱密度对应于滤波器频率响应的模平方，这与预期一致。

根据 (11) 和 (14)，我们可以看到，平稳性在概率建模方面还有另一项优势。给定拓扑算子 $\mathbf{T} = \mathbf{U}\mathrm{diag}(\boldsymbol{\lambda})\mathbf{U}^\top$，描述平稳信号 $\mathbf{s}$的协方差 $\mathbf{C}$仅需 N 个自由度，而非 $N(N+1)/2$。此外，若 $R < N$ 适用于滤波器系数$\mathbf{h} \in \mathbb{R}^R$，我们还可以按照 (14) 对功率谱密度进行建模，进一步减少自由度。事实上，这种简化至关重要，尤其是因为对于高维拓扑信号，$N(N+1)/2$ 可能会变得非常大；我们将在下一节利用这一点进行协方差估计。

IV. 协方差估计

为了描述单形之间的统计关系，我们考虑从一组 M个实现$\mathbf{S} = [\mathbf{s}_1, \ldots, \mathbf{s}_M] \in \mathbb{R}^{N\times M}$中估计平稳拓扑信号$\mathbf{s}$ 的协方差矩阵$\mathbf{C} = \mathbb{E}[\mathbf{s}\mathbf{s}^\top]$ 这一任务。恢复$\mathbf{C}$的典型方法是采用样本估计

$$\hat{\mathbf{C}} := \frac{1}{M}\mathbf{S}\mathbf{S}^\top = \frac{1}{M}\sum_{m=1}^{M}\mathbf{s}_m\mathbf{s}_m^\top, \tag{15}$$

计算需要 $\mathcal{O}(MN^2)$ 次运算。虽然式 (15) 很直接，但它没有利用第 III 节中针对平稳信号所介绍的 $\mathbf{C}$ 的已知结构。因此，我们旨在利用式 (11) 和 (14) 中的特征，开发更高效的功率谱密度 $\mathbf{p}$ 和协方差 $\mathbf{C}$ 估计器。回顾一下，我们可以通过霍奇拉普拉斯算子 $\mathbf{L}_k$ 考虑单阶信号 k ，也可以通过狄拉克算子 $\mathbf{D}$ 考虑涵盖所有阶的多阶信号 $\in \{$。因此，对于以下估计器，我们用 $\mathbf{T} \in \{\mathbf{L}_k, \mathbf{D}\}$ 表示这两种情形，其特征分解为 $\mathbf{T} = \mathbf{U}\mathrm{diag}(\boldsymbol{\lambda})\mathbf{U}^\top$，其中 $\mathbf{U}^\top$ 还进一步诱导出拓扑傅里叶变换。

A. 非参数估计量

我们介绍两种非参数方法：一种是在空间域中应用信号的方法（基于相关图），另一种是考虑信号谱表示的方法（基于周期图）。

1）基于相关图的估计量：我们的方法分为两个步骤。首先，将样本协方差$\mathbf{C}$ 变换到谱域，以估计功率谱密度（PSD），

$$\hat{\mathbf{p}}_{\mathrm{cg}} := \mathrm{diag}(\mathbf{U}^\top\hat{\mathbf{C}}\mathbf{U}) \in \mathbb{R}_+^N. \tag{16}$$

然后根据$\hat{\mathbf{p}}^{\mathrm{cg}}$ 重构协方差估计值，形式如下：

$$\hat{\mathbf{C}}_{\mathrm{cg}} := \mathbf{U}\mathrm{diag}(\hat{\mathbf{p}}_{\mathrm{cg}})\mathbf{U}^\top. \tag{17}$$

因此，$\mathbf{C}^{\mathrm{cg}}$ 可以看作$\mathbf{C}$ 在由$\mathbf{U}$对角化的对称矩阵子空间上的投影。就计算复杂度而言，注意我们必须先获得$\mathbf{C}$ ，其计算复杂度为$\mathcal{O}(MN^2)$，随后进行两次谱域变换，每次的计算复杂度均为 $\mathcal{O}(N^3)$。

2）基于周期图的估计量：或者，我们采用拓扑傅里叶变换（TFT），以降低计算量。首先将信号变换到谱域，

$$\tilde{\mathbf{S}} = \mathbf{U}^\top\mathbf{S} = [\tilde{\mathbf{s}}_1, \ldots, \tilde{\mathbf{s}}_M].$$

然后通过对变换后信号的平方幅值取平均来估计功率谱密度（PSD），

$$[\hat{\mathbf{p}}_{\mathrm{pg}}]_i := \frac{1}{M}\sum_{m=1}^{M}[\tilde{\mathbf{s}}_m]_i^2 \quad \text{for all } m \in [M], \tag{18}$$

或者等价地，$\hat{\mathbf{p}}_{\mathrm{pg}} = \frac{1}{M}(\tilde{\mathbf{S}} \circ \tilde{\mathbf{S}})\mathbf{1}$，其中 $\circ$ 表示哈达玛积（逐元素乘积）。协方差估计值的求法与前文相同，

$$\hat{\mathbf{C}}_{\mathrm{pg}} := \mathbf{U}\mathrm{diag}(\hat{\mathbf{p}}_{\mathrm{pg}})\mathbf{U}^\top. \tag{19}$$

注意到，采用式 (16) 得到 $\hat{\mathbf{p}}^{\mathrm{cg}}$需要 $\mathcal{O}(\max(M,N)N^2)$，而采用式 (18) 得到 $\hat{\mathbf{p}}^{\mathrm{pg}}$ 需要 $\mathcal{O}(MN^2)$。不过，方法 (16) 和 (18) 得到的估计值等价，这一点很容易验证。因此，在低样本情形下，$M < N$，通过式 (18) 估计功率谱密度在计算上更有优势。

接下来，我们通过分析其均值和方差 [11]，验证式 (16) 和 (18) 中所提出的、数值等价的方法的价值。由于二者等价，我们将分析 $\hat{\mathbf{p}}^{\mathrm{pg}}$，但以下讨论同样适用于 $\hat{\mathbf{p}}^{\mathrm{cg}}$。特别地，下面的命题表明，$\hat{\mathbf{p}}^{\mathrm{pg}}$ 是一个无偏估计值，其各个条目彼此不相关，且方差取决于$\mathbf{p}$中条目的幅度。

命题 1。考虑一个相对于$\mathbf{T} \in \{\mathbf{L}_k, \mathbf{D}\}$ 平稳的拓扑信号$\mathbf{s}$，其功率谱密度为$\mathbf{p}$。给定由 M 个$\mathbf{s}$的独立实现构成的矩阵$\mathbf{S} \in \mathbb{R}^{N\times M}$，令 $\hat{\mathbf{p}}^{\mathrm{pg}}$ 为式 (18) 中基于周期图的功率谱密度估计。则，

(i) $\hat{\mathbf{p}}^{\mathrm{pg}}$ 的偏差$\mathbf{b}^{\mathrm{pg}}$ 为

$$\mathbf{b}_{\mathrm{pg}} := \mathbb{E}[\hat{\mathbf{p}}_{\mathrm{pg}}] - \mathbf{p} = \mathbf{0}; \tag{20}$$

(ii) 如果$\mathbf{s}$ 为高斯信号，则 $\hat{\mathbf{p}}^{\mathrm{pg}}$ 的协方差 $\boldsymbol{\Sigma}_{\mathrm{pg}}$ 为

$$\boldsymbol{\Sigma}_{\mathrm{pg}} := \mathbb{E}\big[(\hat{\mathbf{p}}_{\mathrm{pg}} - \mathbf{p})(\hat{\mathbf{p}}_{\mathrm{pg}} - \mathbf{p})^\top\big] = \frac{2}{M}\mathrm{diag}^2(\mathbf{p}). \tag{21}$$

证明。证明过程与 [11, 命题 2],中的步骤相同，只需将图傅里叶变换替换为拓扑傅里叶变换。 □

因此，我们证明了 $\hat{\mathbf{p}}^{\mathrm{pg}}$ 是无偏估计值，即$\mathbf{b}_{\mathrm{pg}} = \mathbf{0}$。此外，如果信号$\mathbf{s}$ 服从高斯分布，我们可以保证估计值 $\hat{\mathbf{p}}^{\mathrm{pg}}$ 的各项互不相关。对于 $\hat{\mathbf{p}}^{\mathrm{pg}}$ 的协方差而言，高斯性的限制与时间域和图域中的平稳性类似，这是由于需要计算$\mathbf{s}$ [6, 的四阶矩（第 8 节2]）。除了各项互不相关之外，我们还注意到，$[\hat{\mathbf{p}}_{\mathrm{pg}}]_i$ 的方差与 p_i^2成正比。因此，由于 $\hat{\mathbf{p}}^{\mathrm{pg}}$ 是无偏估计值，其期望误差由其方差决定，进而取决于$\mathbf{p}$的大小。更具体地，根据 $\boldsymbol{\Sigma}_{\mathrm{pg}}$的定义，我们有

$$\begin{aligned}\mathrm{MSE}_{\mathrm{pg}} &= \mathbb{E}\big[\|\hat{\mathbf{p}}_{\mathrm{pg}} - \mathbf{p}\|_2^2\big] = \mathbb{E}\big[\mathrm{tr}\big((\hat{\mathbf{p}}_{\mathrm{pg}} - \mathbf{p})(\hat{\mathbf{p}}_{\mathrm{pg}} - \mathbf{p})^\top\big)\big] \\ &= \mathrm{tr}(\hat{\mathbf{C}}_{\mathrm{pg}}) = \frac{2}{M}\|\mathbf{p}\|_2^2.\end{aligned} \tag{22}$$

不出所料，$\hat{\mathbf{p}}^{\mathrm{pg}}$ 的均方误差与观测值数量 M成反比。在 M 较低的情况下，我们可以考虑采用其他方法来降低均方误差^{pg} ，但这可能会引入偏差；相关探索留待未来进行。注意，根据式（10）和式（19），真实协方差$\mathbf{C}$ 和估计值$\mathbf{C}^{\mathrm{pg}}$分别由 $\mathrm{diag}(\mathbf{p})$ 和 $\mathrm{diag}(\hat{\mathbf{p}}^{\mathrm{pg}})$ 经酉变换得到，因此$\mathbf{C}^{\mathrm{pg}}$ 同样是 $\mathbf{C}$ 的无偏估计值，其均方误差为$\frac{2}{M}\|\mathbf{p}\|_2^2 = \frac{2}{M}\|\mathbf{C}\|_F^2$。

3）子空间感知估计：现实世界中的拓扑信号通常受限于特定子空间。例如，在（3）中，表示通信流量的边流信号通常由调和分量主导，主要位于$\mathbf{U}_{k,\,\mathrm{H}}$ 子空间中。在这种情况下，估计协方差$\mathbf{C}$时，我们可以利用信号在谱域中的受限支撑作为额外结构。

令$\mathbf{U}_{\mathcal{S}}$ 表示与感兴趣子空间相对应的特征向量$\mathbf{U}$的列，其索引为 $\mathcal{S} \subseteq [N]$。当$\mathbf{U}_{\mathcal{S}}$ 已知时，我们可以将协方差改写为$\mathbf{C} = \mathbf{U}\mathrm{diag}(\mathbf{p})\mathbf{U}^\top = \mathbf{U}_{\mathcal{S}}\mathrm{diag}(\mathbf{p}_{\mathcal{S}})\mathbf{U}_{\mathcal{S}}^\top$，其中$\mathbf{p}_{\mathcal{S}}$ 包含功率谱密度的非零值。于是，估计器 (16) 和 (18) 中的非零项变为

$$[\hat{\mathbf{p}}_{\mathrm{cg}}]_{\mathcal{S}} = \mathrm{diag}(\mathbf{U}_{\mathcal{S}}\hat{\mathbf{C}}\mathbf{U}_{\mathcal{S}}^\top), \quad [\hat{\mathbf{p}}_{\mathrm{pg}}]_{\mathcal{S}} = \frac{1}{M}\big((\mathbf{U}_{\mathcal{S}}^\top\mathbf{S}) \circ (\mathbf{U}_{\mathcal{S}}^\top\mathbf{S})\big)\mathbf{1},$$

随后按 (17) 和 (19) 进行协方差估计，其中用$\mathbf{U}_{\mathcal{S}}$ 替代$\mathbf{U}$。

在某些情况下，我们可能知道信号 $\mathbf{s}$ 局限于一个未知子空间。例如，考虑一个 k-信号 $\mathbf{s}$ 和 $\mathbf{T} = \mathbf{L}_k = \mathbf{U}_k\mathrm{diag}(\boldsymbol{\lambda}_k)\mathbf{U}_k^\top$，它被限制在梯度、旋度或调和子空间中，也就是说，$\mathbf{U}_{\mathcal{S}} \in \{\mathbf{U}_{k,\mathrm{G}}, \mathbf{U}_{k,\mathrm{C}}, \mathbf{U}_{k,\mathrm{H}}\}$，如式（3）所定义，但我们不知道具体是哪一个。在估计功率谱密度$\mathbf{p}$ 或协方差 $\mathbf{C}$之前，我们可以推断哪个子空间处于激活状态。一种简单而有效的策略是将 $\mathbf{s}$ 投影到三个互相正交子空间中的每一个，并计算各自的能量，

$$E_{\mathrm{G}} = \|\mathbf{U}_{k,\mathrm{G}}^\top \mathbf{s}\|_2^2,\ E_{\mathrm{C}} = \|\mathbf{U}_{k,\mathrm{C}}^\top \mathbf{s}\|_2^2,\ E_{\mathrm{H}} = \|\mathbf{U}_{k,\mathrm{H}}^\top \mathbf{s}\|_2^2. \quad (23)$$

然后，我们可以通过能量阈值或基于似然的检验 [31],识别最可能的分量，再像之前一样进行功率谱密度和协方差估计。

B. 参数估计器

接下来，我们介绍用于单纯信号的拓扑滤波器模型，它支持参数化方法来估计$\mathbf{p}$和$\mathbf{C}$。具体而言，我们为单纯形域定义移动平均（MA）和 AR 随机过程的推广形式。这些模型对应于平稳过程；正如我们将要展示的，这使协方差$\mathbf{C}$ 建模的复杂度从 $\mathcal{O}(N^2)$ 降至 $\mathcal{O}(R)$，即滤波器参数的数量。

1) MA 拓扑过程：与定义 1 中一样，我们假设平稳拓扑信号是由白噪声驱动的式（7）中的有限阶拓扑滤波器的输出生成的。如第 III 节所述，由于我们只需恢复滤波器参数，因此可通过式（14）中的模型得到参数协方差估计器。

这使我们能够定义*MA*拓扑过程，即遵循式（6）的信号。为形式化 MA 过程的参数模型，令 $\boldsymbol{\beta} = [\beta_0, \ldots, \beta_{R-1}]^\top$ 表示滤波器系数，并令 $\mathbf{T} \in \{\mathbf{L}_k, \mathbf{D}\}$ 表示我们关注的对称拓扑算子。于是，我们将有限脉冲响应（FIR）拓扑滤波器定义为

$$\mathbf{H}(\boldsymbol{\beta}) = \sum_{r=0}^{R-1} \beta_r \mathbf{T}^r, \quad (24)$$

其中 $R < N$，且在实际应用中，$R \ll N$。由于我们的平稳信号$\mathbf{s}$由白噪声$\mathbf{w}$生成，因此有

$$\mathbf{s} = \mathbf{H}(\boldsymbol{\beta})\mathbf{w}, \quad \mathbb{E}[\mathbf{w}\mathbf{w}^\top] = \mathbf{I}, \quad (25)$$

其协方差为

$$\mathbf{C}^\beta(\boldsymbol{\beta}) = \mathbf{H}(\boldsymbol{\beta})\mathbf{H}(\boldsymbol{\beta})^\top = \sum_{r=0}^{R-1}\sum_{r'=0}^{R-1} \beta_r \beta_{r'} \mathbf{T}^{r+r'}, \quad (26)$$

它本身是 $\mathbf{T}$ 的一个次数为 $2(R-1)$ 的多项式

$$\mathbf{C}^\gamma(\boldsymbol{\gamma}) = \sum_{r=0}^{2(R-1)} \gamma_r \mathbf{T}^r, \qquad \gamma_r := \sum_{r'+r''=r} \beta_{r'}\beta_{r''}, \quad (27)$$

系数 $\boldsymbol{\gamma} \in \mathbb{R}^{2(R-1)}$ 通过卷积获得。

然后，根据特征分解 $\mathbf{T} = \mathbf{U}\mathrm{diag}(\boldsymbol{\lambda})\mathbf{U}^\top$ 以及公式 (11) 中的定义，与 $\mathbf{C}^\beta(\boldsymbol{\beta})$ 对应的功率谱密度为

$$\mathbf{p}(\boldsymbol{\beta}) = h^\beta(\boldsymbol{\lambda}) \circ h^\beta(\boldsymbol{\lambda}) = (\boldsymbol{\Psi}^\beta\boldsymbol{\beta}) \circ (\boldsymbol{\Psi}^\beta\boldsymbol{\beta}), \quad (28)$$

其中，$h^\beta(\lambda) = \sum_{r=0}^{R-1} \beta_r \lambda^r$ 表示 $\mathbf{H}^\beta(\boldsymbol{\beta})$ 的拓扑频率响应，$\boldsymbol{\Psi}^\beta \in \mathbb{R}^{N\times R}$是一个范德蒙德矩阵，使得 $\Psi^\beta_{ir} = \lambda_i^{r-1}$对所有 $i \in [N], r \in [R]$均成立。因此，我们对 MA拓扑过程的定义与时间域和图域中的定义相似，在后两者中，功率谱密度等于滤波器频率响应的平方。

若$\mathbf{s}$ 是一个MA过程，我们可以通过ˆ恢复 $\boldsymbol{\beta}$，利用式（15）中的样本协方差$\mathbf{C}$来估计协方差$\mathbf{C}$，

$$\hat{\boldsymbol{\beta}}_{\mathbf{C}} = \mathrm{argmin}_{\boldsymbol{\beta}}\, \mathcal{D}_{\mathbf{C}}\Big(\hat{\mathbf{C}}, \mathbf{C}^\beta(\boldsymbol{\beta})\Big), \quad (29)$$

其中，式（26）中的$\mathbf{C}^\beta(\boldsymbol{\beta})$以及作为量化协方差误差的距离度量 $\mathcal{D}_{\mathbf{C}}(\cdot,\cdot)$。作为一种可能更高效的替代方法，我们可以在频域中估计 $\boldsymbol{\beta}$，给定功率谱密度的估计值，例如式（18）中基于周期图的ˆ$\mathbf{p}_{\mathrm{pg}}$，

$$\hat{\boldsymbol{\beta}}_{\mathbf{p}} = \mathrm{argmin}_{\boldsymbol{\beta}}\, \mathcal{D}_{\mathbf{p}}\big(\hat{\mathbf{p}}_{\mathrm{pg}}, (\boldsymbol{\Psi}^\beta\boldsymbol{\beta} \circ \boldsymbol{\Psi}^\beta\boldsymbol{\beta})\big), \quad (30)$$

其中，度量 $\mathcal{D}_{\mathbf{p}}(\cdot,\cdot)$用于衡量功率谱密度误差。

无论选择式（29）还是式（30），两个目标函数关于 $\boldsymbol{\beta}$ 都是二次函数，因此通常是非凸的。然而，在对$\mathbf{T}$或 $\boldsymbol{\beta}$施加额外假设的情况下，我们可以将式（30）松弛为凸问题或易于求解的问题，例如$\mathbf{T}$ 为半正定且 $\boldsymbol{\beta}$为非负的情况。此外，在 $\mathcal{D}_{\mathbf{p}}(\mathbf{p}_1,\mathbf{p}_2) = \|\mathbf{p}_1 - \mathbf{p}_2\|_2^2$这一特定情形下，我们可以采用高效的相位恢复算法（如Wirtinger flow [35], [36]）求解式（30）。下面介绍式（29）的一种凸松弛方法，即忽略式（27）中 γ 的卷积结构。具体而言，我们估计$\mathbf{C}^\gamma(\boldsymbol{\gamma})$，而不是$\mathbf{C}^\beta(\boldsymbol{\beta})$，

$$\hat{\boldsymbol{\gamma}}_{\mathbf{C}} = \mathrm{argmin}_{\boldsymbol{\gamma}}\, \mathcal{D}_{\mathbf{C}}\Big(\hat{\mathbf{C}}, \mathbf{C}^\gamma(\boldsymbol{\gamma})\Big), \quad (31)$$

当度量 $\mathcal{D}_{\mathbf{C}}(\cdot,\cdot)$ 为凸时，其为凸函数。同样，对于满足 $\Psi^\gamma_{ir} = \lambda_i^{r-1}$ 的 $\boldsymbol{\Psi}^\gamma \in \mathbb{R}^{N\times 2R}$ $(i \in [N], r \in [2R-1])$，我们可以通过以下方式在频域估计 $\boldsymbol{\gamma}$：

$$\hat{\boldsymbol{\gamma}}_{\mathbf{p}} = \mathrm{argmin}_{\boldsymbol{\gamma}}\, \mathcal{D}_{\mathbf{p}}\big(\hat{\mathbf{p}}_{\mathrm{pg}}, \boldsymbol{\Psi}^\gamma\boldsymbol{\gamma}\big). \quad (32)$$

注意，如果 $\mathcal{D}_{\mathbf{C}}(\mathbf{C}_{1,2}, \mathbf{C}_{1-1,2}) = \|\mathbf{C}_1 - \mathbf{C}_2\|_2^2$ 且 $\mathcal{D}_{\mathbf{p}}(\mathbf{pp})$ $\mathbf{pp}$，那么目标函数以及式（31）和式（32）的解是等价的。然而，求得闭式解的复杂程度有所不同，我们将在第VI节通过实证展示。因此，我们对单纯复形的经典参数化MA过程所作的推广，将协方差矩阵 $\mathbf{C}$ 的表示从$\mathcal{O}(N^2)$个自由度简化为 $\mathcal{O}(R)$ 或 $\mathcal{O}(R^2)$，从而降低了恢复 $\mathbf{C}$ 的复杂度。

2) 自回归拓扑过程：接下来，我们介绍用于拓扑信号的自回归滤波器模型。该模型描述了在给定白噪声的情况下，通过递归滤波得到信号 $\mathbf{s}$，

$$\mathbf{s} = \sum_{r=1}^{R} \alpha_r \mathbf{T}^r \mathbf{s} + \mathbf{w}, \quad (33)$$

其中$\mathbf{w}$ 是白单纯信号，满足 $\mathbb{E}[\mathbf{w}] = \mathbf{0}$ 和$\mathbb{E}[\mathbf{w}\mathbf{w}^\top] = \mathbf{I}$，且 $\boldsymbol{\alpha} = [\alpha_1, \ldots, \alpha_R]^\top \in \mathbb{R}^R$包含 AR系数。我们也可以将式（33）中的递归过程等价地写成一种滤波操作，其中滤波器为

$$\mathbf{H}(\boldsymbol{\alpha})\mathbf{s} = \mathbf{w}, \quad \mathbf{H}(\boldsymbol{\alpha}) := \mathbf{I} - \sum_{r=1}^{R} \alpha_r \mathbf{T}^r, \quad (34)$$

从而得到s的生成形式

$$\mathbf{s} = \mathbf{H}^{-1}(\boldsymbol{\alpha})\mathbf{w}. \tag{35}$$

那么，s的协方差矩阵为

$$\mathbf{C}^{\alpha}(\boldsymbol{\alpha}) = \mathbf{H}^{-2}(\boldsymbol{\alpha}), \tag{36}$$

而功率谱密度可通过$\mathbf{H}(\boldsymbol{\alpha})$的频率响应定义，即

$$[\mathbf{p}(\boldsymbol{\alpha})]_i := \frac{1}{\left(h^{\alpha}(\lambda_i)\right)^2} = \frac{1}{[\mathbf{1} - \boldsymbol{\Psi}^{\alpha}\boldsymbol{\alpha}]_i^2} \text{ for all } i \in [N] \tag{37}$$

对于 $h^{\alpha}\ \lambda = 1 - \sum_{r=1}^{R} \alpha_r \lambda^r$ $\boldsymbol{\Psi}^{\alpha} \in \mathbb{R}^{N\times R}$，以及范德蒙德矩阵，其中所有 $i \in [N]$ 均满足 $\Psi_{ir}^{\alpha} = \lambda_i^r, r \in [R]$，其中拓扑算子再次满足 $\mathbf{T} = \mathbf{U}\mathrm{diag}(\boldsymbol{\lambda})\mathbf{U}^{\top}$。

为了估计自回归拓扑过程的协方差矩阵，我们通过最小化 $\hat{}\mathbf{C}^{\alpha}(\boldsymbol{\alpha})$ 与式（15）中的样本协方差 $\mathbf{C}$ 之间的差异，来恢复滤波器参数 $\boldsymbol{\alpha}$ 。然而，为避免在式（36）中需要对 $\mathbf{H}(\boldsymbol{\alpha})$求逆，我们不直接比较两个矩阵$\mathbf{C}^{\alpha}(\boldsymbol{\alpha})$和$\hat{}\mathbf{C}$ ，而是将$\hat{}\mathbf{C}\big(\mathbf{C}^{\alpha}(\boldsymbol{\alpha})\big)^{-1}$ 与单位矩阵进行比较，从而得到以下优化问题

$$\hat{\boldsymbol{\alpha}}_{\mathbf{C}} = \operatorname{argmin}_{\boldsymbol{\alpha}} \mathcal{D}_{\mathbf{C}}\Big(\hat{\mathbf{C}}\big(\mathbf{C}^{\alpha}(\boldsymbol{\alpha})\big)^{-1}, \mathbf{I}\Big). \tag{38}$$

与移动平均情形类似，我们也可以通过功率谱密度(37)在频域中估计滤波器参数 $\boldsymbol{\alpha}$,

$$\hat{\boldsymbol{\alpha}}_{\mathbf{p}} = \operatorname{argmin}_{\boldsymbol{\alpha}} \mathcal{D}_{\mathbf{p}}\Big(\mathrm{diag}^{-1}\big(\mathbf{p}(\boldsymbol{\alpha})\big)\hat{\mathbf{p}}_{\mathrm{pg}}, \mathbf{1}\Big), \tag{39}$$

其中$\hat{}\mathbf{p}_{\mathrm{pg}}$ 表示(18)中基于周期图的功率谱密度估计值。如前所述，问题(38)和(39)关于 $\boldsymbol{\alpha}$ 是二次的，因此对于 $\mathcal{D}_{\mathbf{C}}$ 和 $\mathcal{D}_{\mathbf{p}}$的常见选择可能是非凸的。因此，我们可以再次考虑估计由 $\boldsymbol{\alpha}$ 与其自身的卷积所定义的另一组参数 $\boldsymbol{\eta} \in \mathbb{R}^{2R}$ ，其中

$$\mathbf{C}^{\eta}(\boldsymbol{\eta}) = \left(\mathbf{I} - \sum_{r=1}^{2R} \eta_r \mathbf{T}^r\right)^{-1}, \quad \eta_r := \sum_{r'+r''=r} \alpha_{r'}\alpha_{r''}. \tag{40}$$

这导出了简化形式 [参见(38)和(39)]

$$\hat{\boldsymbol{\eta}}_{\mathbf{C}} = \operatorname{argmin}_{\boldsymbol{\eta}} \mathcal{D}_{\mathbf{C}}\Big(\hat{\mathbf{C}}\big(\mathbf{I} - \textstyle\sum_{r=1}^{2R}\eta_r\mathbf{T}^r\big), \mathbf{I}\Big) \tag{41}$$

$$\hat{\boldsymbol{\eta}}_{\mathbf{p}} = \operatorname{argmin}_{\boldsymbol{\eta}} \mathcal{D}_{\mathbf{p}}\Big(\hat{\mathbf{p}}_{\mathrm{pg}} \circ \big(\mathbf{1} - \boldsymbol{\Psi}^{\eta}\boldsymbol{\eta}\big), \mathbf{1}\Big) \tag{42}$$

其中 $\boldsymbol{\Psi}^{\eta} \in \mathbb{R}^{N\times 2R}$ 是对于 $i \in [N], r \in [2R]$具有 $\Psi_{ir}^{\eta} = \lambda_i^r$的范德蒙德矩阵。虽然与MA模型一样，(42)在计算上比(41)更简单，但请注意， $\mathcal{D}_{\mathbf{C}}(\mathbf{C}_1, \mathbf{C}_2) = \|\mathbf{C}_1 - \mathbf{C}_2\|_F^2$与 $\mathcal{D}_{\mathbf{p}}(\mathbf{p}_1, \mathbf{p}_2) = \|\mathbf{p}_1 - \mathbf{p}_2\|_2^2$ 并不会产生等价的目标函数，这是由于(41)中来自非对角矩阵元素的附加项。由于(41)需要优化 $N(N+1)/2$ 项，而不是仅优化(42)的 N 项，我们需要更多信号 M 才能从(41)得到有竞争力的解。因此，在$\hat{}$我们对$\hat{}\mathbf{p}_{\mathrm{pg}}$ 的估计值比$\mathbf{C}$更可靠的情形中，估计$\hat{}\boldsymbol{\eta}_{\mathbf{p}}$ 在计算和样本复杂度两方面都可能优于$\hat{}\boldsymbol{\eta}_{\mathbf{C}}$。

一个有趣的情形是一阶自回归过程（$R = 1$），

$$\mathbf{C} = (\mathbf{I} - \alpha\mathbf{T})^{-2},$$

其在估计单纯复形性质方面具有若干优势。首先，由于（34）中的滤波器仅需要一个标量参数 $\alpha \in \mathbb{R}$，我们对（38）中 $\mathbf{C}$ 以及（39）中 $\mathbf{p}$ 的估计在已知拓扑算子 $\mathbf{T} \in \{\mathbf{L}_k, \mathbf{D}\}$ 的情况下，在计算上变得更为简单。其次，$\mathbf{T}$的稀疏模式直接影响精度矩阵 $\mathbf{C}^{-1} = \mathbf{I} - 2\alpha\mathbf{T} + \alpha^2\mathbf{T}^2$ 的支撑。因此，我们可以通过首先估计一个稀疏的$\mathbf{C}^{-1}$来恢复 $\mathbf{C}$。第三，我们同样可以利用这一关系，通过联合估计 $\mathbf{C}^{-1}$ 与$\mathbf{T}$来估计未知的 $\mathbf{T}$，其实际优势与理论优势将在未来工作中加以探讨。

最后，关于这两种信号模型，我们可以结合移动平均和自回归拓扑过程的定义，得到自回归–移动平均（ARMA）信号，其形式为 $\mathbf{H}_1(\boldsymbol{\alpha})\mathbf{s} = \mathbf{H}_2(\boldsymbol{\beta})\mathbf{w}$，其中多项式算子 $\mathbf{H}_1(\boldsymbol{\alpha})$ 如式（24）所示，$\mathbf{H}_2(\boldsymbol{\beta})$ 如式（34）所示。尽管这些 ARMA 模型很有趣，我们仍将其研究留待未来探索。

3）非多项式滤波：到目前为止，我们考虑的拓扑信号模型均基于算子 $\mathbf{T} \in \{\mathbf{L}_k, \mathbf{D}\}$ 的多项式滤波器，这些模型已被广泛用于描述单纯复形上的扩散信号。这里，我们将考虑更一般的非多项式滤波器，作为平稳拓扑过程的额外参数化表示。

谱域滤波器参数化。 我们不在空间域中定义拓扑滤波器，而是考虑由低维向量 $\boldsymbol{\theta}$ 参数化的频率响应 $h^{\theta}(\lambda)$。与 MA 和 AR 情形类似，我们可以通过滤波器频率响应估计协方差 $\mathbf{C}$，这要求我们能够通过以下方式恢复 $\boldsymbol{\theta}$

$$\hat{\boldsymbol{\theta}}_{\mathrm{spec}} = \operatorname{argmin}_{\boldsymbol{\theta}} \mathcal{D}_{\mathbf{p}}\big(\hat{\mathbf{p}}_{\mathrm{pg}}, h^{\theta}(\boldsymbol{\lambda}) \circ h^{\theta}(\boldsymbol{\lambda})\big). \tag{43}$$

这种表述相当一般，但此类参数化模型的经典示例包括指数型或有理型响应，例如

$$h^{\theta}(\lambda) = e^{-\theta_1\lambda}, \quad h^{\theta}(\lambda) = \mathrm{sigmoid}(\theta_1\lambda + \theta_2),$$

它们分别对单纯复形上的信号起到经典热扩散滤波器或带通滤波器的作用。

核参数化。 或者，我们也可以直接对功率谱密度 $\mathbf{p}_{\theta} = \kappa^{\theta}(\boldsymbol{\lambda})$ 进行参数化，而不是对滤波器进行参数化，即学习一个谱核——它是单纯复形谱的函数。随后，我们通过其参数 $\boldsymbol{\theta}$ 将功率谱密度估计为

$$\boldsymbol{\theta}_{\mathrm{ker}} = \operatorname{argmin}_{\boldsymbol{\theta}} \mathcal{D}_{\mathbf{p}}\big(\hat{\mathbf{p}}_{\mathrm{pg}}, \kappa^{\theta}(\boldsymbol{\lambda})\big), \tag{44}$$

示例包括高斯核或拉普拉斯核,

$$\kappa^{\theta}(\lambda) = e^{-\theta_1\lambda^2}, \quad \kappa^{\theta}(\lambda) = \frac{1}{(1+\theta_1\lambda)^{\theta_2}}.$$

这些特定的核可以对平滑功率谱密度或幂律行为进行建模，而这些现象在许多真实拓扑数据集 [1], [2], [37] 中都可以观察到。请注意，式（43）中的基于谱的方法与式（44）中的基于核的方法之间的差异，主要体现在诠释上。二者都使 $\hat{}\mathbf{p}_{\mathrm{pg}}$ 与某个参数化函数之间的差异最小化，而该函数的含义决定了我们的关注重点。

基于谱和核的滤波器模型 $h^{\theta}(\lambda)$ 和 $\kappa^{\theta}(\lambda)$ 不仅能让我们采用更灵活的低维参数化方式，还能将它们与拓扑机器学习联系起来。例如，我们可以将 $\kappa^{\theta}(\lambda)$ 视为可学习核，在拓扑频域上定义一个高斯过程，并根据其关联拓扑算子 $\mathbf{T} \in \{\mathbf{L}_k, \mathbf{D}\}$ 的谱 $\boldsymbol{\lambda}$，预测任意单纯复形的功率谱密度（PSD）。更具体地说，根据（13），我们可以构造协方差

任意平稳拓扑过程 $\mathbf{s}$ 的频率响应 $\tilde{\mathbf{s}} = \mathbf{U}^\top \mathbf{s}$ 的矩阵为 $\mathrm{diag}(\kappa^\theta(\boldsymbol{\lambda}))$。此外，我们还可以采用数据驱动的方法来学习滤波器频率响应：将 $h^\theta(\lambda)$ 建模为神经网络，并根据观测到的拓扑信号拟合滤波器参数 $\boldsymbol{\theta}$ 。对于这些示例，我们仍然可以利用拓扑平稳性，在频域中学习滤波器参数；此时模型复杂度为 $\mathcal{O}(N)$，而非之前的 $\mathcal{O}(N^2)$。

注释 1 (单阶与多阶模型)。 我们关注拓扑算子 $\mathbf{T}$ ，它可以是霍奇拉普拉斯算子 $\mathbf{L}_k$ ，作用于k位于单一阶上的信号；也可以是狄拉克算子 $\mathbf{D}$，用于多阶信号，此时数据在所有维度上同时观测。对于前者，我们仍可通过考虑滤波器及其系数 $\{\mathbf{h}_k\}_{k=0}^K$，为每个阶分别设置一组，来对所有阶上的信号建模，尽管不同维度的滤波器之间没有任何关系。对于后者，我们假定各阶之间的扩散动力学相同，也就是说，所有维度共享相同的滤波器系数$\mathbf{h}$ 。不过，我们也可以考虑介于这两个极端之间的情况。例如，我们可以对系数集 $\{\mathbf{h}_k\}_{k=0}^K$施加结构关系，例如要求其随阶平滑变化，或在不同阶之间具有层级依赖。尽管这能提高建模灵活性，我们仍可使用本节提出的估计器，只需对其进行相应修改，以纳入对滤波器系数的约束。

V. 基于平稳性的模型与工具

接下来，我们将展示平稳拓扑过程在经典信号处理任务中的价值。我们研究拓扑去噪（对维纳滤波器的推广）和插值（表明无需直接观测 k 阶信息，也能恢复单纯形 k 信号），并考察高斯拓扑平稳性，最后讨论尚待解决的问题。

A. 拓扑信号的维纳滤波

我们首先研究如何恢复一个受加性噪声污染的未知拓扑信号 $\mathbf{s}$ ，并将经典维纳滤波器推广到单纯复形域。与协方差估计类似，若$\mathbf{s}$ 是平稳的，则在频域中进行处理可大幅提高任务效率；我们将说明，这相当于一种简单的逐元素缩放操作。

我们希望恢复一个相对于$\mathbf{T} \in \{\mathbf{L}_k, \mathbf{D}\}$ 平稳且协方差为 $\mathbf{C} = \mathbb{E}[\mathbf{s}\mathbf{s}^\top]$的零均值拓扑信号 $\mathbf{s}$ 。特别地，给定受污染的信号

$$\mathbf{y} = \mathbf{s} + \mathbf{v}, \tag{45}$$

其中$\mathbf{v}$ 是一个均值为零的白噪声信号，与$\mathbf{s}$ 相互独立，其协方差为 $\mathbf{C}_v = \sigma_v^2\mathbf{I}$ 对于 $\sigma_v^2 > 0$，我们寻求根据$\mathbf{y}$对$\mathbf{s}$ 的最小均方误差（MMSE）估计，即

$$\hat{\mathbf{s}} = \mathrm{argmin}_{\mathbf{z}}\, \mathbb{E}\big[\|\mathbf{s}-\mathbf{z}\|_2^2|\mathbf{y}\big] = \mathbf{C}(\mathbf{C}+\mathbf{C}_v)^{-1}\mathbf{y} = \mathbf{G}\mathbf{y}, \tag{46}$$

其中，最优解由$\mathbf{s}$、$\mathbf{y}$的定义以及拓扑平稳性得出，$\mathbf{G}$ 是维纳滤波矩阵 [6], [38]。虽然计算过程简单直接，但对于中等规模的单纯复形，计算$\mathbf{G}$ 可能十分繁琐，甚至不可行，因为这需要求一个 $N \times N$矩阵的逆。然而，利用拓扑平稳性，我们可以大幅简化 $\hat{\mathbf{s}}$ 的求取过程。回想一下，$\mathbf{C}$ 与 $\mathbf{U}$ 具有相同的特征向量$\mathbf{T}$ ，因为$\mathbf{s}$ 是平稳的。然后，根据定义3以及$\mathbf{v}$的白噪声性质，我们有

$$\mathbf{C} = \mathbf{U}\mathrm{diag}(\mathbf{p})\mathbf{U}^\top, \quad \mathbf{C}_v = \mathbf{U}\mathrm{diag}(\sigma_v^2\mathbf{1})\mathbf{U}^\top.$$

然后，利用信号的拓扑傅里叶变换 $\tilde{\hat{\mathbf{s}}} = \mathbf{U}^\top\mathbf{s}$ 和 $\tilde{\mathbf{y}} = \mathbf{U}^\top\mathbf{y}$，我们可以在频域中将（46）重写为

$$\tilde{\hat{\mathbf{s}}} = \mathbf{U}^\top\mathbf{G}\mathbf{U}\tilde{\mathbf{y}} = \mathrm{diag}(\mathbf{p})\mathrm{diag}^{-1}\big(\mathbf{p}+\sigma_v^2\mathbf{1}\big)\tilde{\mathbf{y}} = \tilde{\mathbf{g}}\circ\tilde{\mathbf{y}} \tag{47}$$

其中 $\tilde{\mathbf{g}}$ 表示拓扑维纳滤波器$\mathbf{G}$的频率响应，其中

$$\tilde{g}_i := \frac{p_i}{p_i+\sigma_v^2} \quad \text{for all } i \in [N]. \tag{48}$$

因此，即使对于大型单纯复形，获得$\mathbf{s}$ 的最小均方误差也具有合理的可扩展性，因为我们只需对拓扑傅里叶变换 $\tilde{\mathbf{y}}$与 $\tilde{\mathbf{g}}$进行逐元素缩放，适用于 $\mathcal{O}(N)$，然后将其变换回空间域 $\hat{\mathbf{s}} = \mathbf{U}\tilde{\hat{\mathbf{s}}}$，适用于 $\mathcal{O}(N^2)$，而不必执行（46）的 $\mathcal{O}(N^3)$。根据（48），解 $\hat{\mathbf{s}}$ 会保留$\mathbf{y}$ $\tilde{\mathbf{}}$中具有较高信噪比的元素，即 $i \in [N]$ 满足 $p_i \gg \sigma_v^2$ ($g_i \approx 1$)的元素，同时衰减信噪比较低的元素，其中is相对于 σ_v^2 ($g_i \approx 0$)较小。最后请注意，当$\mathbf{p}$未知时，我们可以用估计值（例如（18）中的 $\hat{\mathbf{p}}^{\mathrm{pg}}$ ）替换（48）中的功率谱密度$\mathbf{p}$。

B. 拓扑信号重建与插值

接下来，我们考虑平稳性假设下的拓扑信号重建。在这种情况下，我们给定部分观测信号，例如仅测量一部分节点、边或高阶单纯形，或缺失整个单纯形阶，并希望恢复完整的原始信号。与前文一样，我们利用平稳性所带来的额外协方差结构，实现统计最优的信号插值。

与第 V-A 节一样，我们寻求零均值拓扑信号$\mathbf{s}$ ，该信号相对于 $\mathbf{T} \in \{\mathbf{L}_k,\mathbf{D}\}$ 平稳，且协方差为 $\mathbf{C}$；但在此处，我们仅观测其部分条目。更正式地说，我们将部分观测建模为

$$\bar{\mathbf{s}} = \boldsymbol{\Theta}\mathbf{s} + \mathbf{v}, \tag{49}$$

其中， $\boldsymbol{\Theta} \in \{0,1\}^{P\times N}$选取$\mathbf{s}$的 $P < N$观测条目，而$\mathbf{v}$ 仍表示与$\mathbf{s}$ 相互独立、协方差为$\mathbf{C}_v = \sigma_v^2\mathbf{I}$的零均值白噪声。与前文一样，我们首先考虑式（49）中$\mathbf{s}$ 的最优线性最小均方误差估计，其表达式为

$$\hat{\mathbf{s}} = \mathbf{C}\boldsymbol{\Theta}^\top\big(\boldsymbol{\Theta}\mathbf{C}\boldsymbol{\Theta}^\top+\sigma_v^2\mathbf{I}\big)^{-1}\bar{\mathbf{s}} = \big(\boldsymbol{\Theta}^\top\boldsymbol{\Theta}+\sigma_v^2\mathbf{C}^{-1}\big)^{-1}\boldsymbol{\Theta}^\top\bar{\mathbf{s}}, \tag{50}$$

其中 $\boldsymbol{\Theta}^\top\boldsymbol{\Theta}$是一个二值对角矩阵，用于鼓励保留已观测单形 $\bar{}$$\mathbf{s}$ ，而$\boldsymbol{\Theta}^\top\bar{}\mathbf{s}$ 是部分观测信号 $\bar{}\mathbf{s}$的零填充插值。不过请注意，(50)也是以下二次问题的解

$$\hat{\mathbf{s}} = \mathrm{argmin}_{\mathbf{z}}\, \frac{1}{\sigma_v^2}\|\bar{\mathbf{s}}-\boldsymbol{\Theta}\mathbf{z}\|_2^2 + \mathbf{z}^\top\mathbf{C}^{-1}\mathbf{z}, \tag{51}$$

因此，最优 MMSE 解 $\hat{\mathbf{s}}$ 也可以看作是通过一个正则化优化问题获得的，其中的项 $\mathbf{z}^\top\mathbf{C}^{-1}\mathbf{z}$ 促使估计值 $\hat{\mathbf{s}}$呈现平稳行为。此外，如果 $\mathbf{s}$ 服从高斯分布，那么式（51）中的 $\mathbf{z}^\top\mathbf{C}^{-1}\mathbf{z}$ 可作为先验，而式（50）则是最大后验估计（MAP）估计值。因此，$\mathbf{C}^{-1}$ 的结构会影响我们对 $\mathbf{s}$的估计。借助出现在 $\mathbf{C}$中的平稳信号 $\mathbf{s}$ 的已知性质，我们可以确定正则化项的表达式，甚至还可能

施加额外的结构以改进估计。下面列出几种尤其相关的选择。

- **平滑性：** 我们通常关注平滑信号，即在连接上变化不大的信号。对于平稳拓扑信号，若$\mathbf{C}^{-1}=\mathbf{L}_k$，则$\mathbf{s}$是一个关于上层和下层连接平滑的$k$-信号；若$\mathbf{C}^{-1}=\mathbf{D}^2=\mathcal{L}$，则$\mathbf{s}$是一个在所有阶上都平滑的多阶信号。(51)中的正则化项$\mathbf{z}^\top\mathbf{C}^{-1}\mathbf{z}$会惩罚单纯形连接上的高信号方差，从而优先选择更平滑的估计值$\hat{\mathbf{s}}$。对于多阶设置，(51)变为

$$\begin{aligned}\hat{\mathbf{s}} &= \operatorname{argmin}_{\mathbf{z}} \frac{1}{\sigma_v^2}\|\bar{\mathbf{s}}-\mathbf{\Theta}\mathbf{z}\|_2^2+\mathbf{z}^\top\mathbf{D}^2\mathbf{z}\\ &= \operatorname{argmin}_{\mathbf{z}} \frac{1}{\sigma_v^2}\|\bar{\mathbf{s}}-\mathbf{\Theta}\mathbf{z}\|_2^2+\sum_{k=0}^{K}\mathbf{z}^{k\top}\mathbf{L}_k\mathbf{z}^k, \end{aligned} \tag{52}$$

其中$\mathbf{z}$是一个多阶信号，$\mathbf{z}^k$表示k-单形上的信号值。

- **一阶自回归：** 我们回到第IV-B2节讨论的$\mathbf{s}$为一阶自回归过程的情形，其精度矩阵为$\mathbf{C}^{-1}=\mathbf{I}+\alpha^2\mathbf{T}^2-2\alpha\mathbf{T}$，其中$\mathbf{T}\in\{\mathbf{L}_k,\mathbf{D}\}$。在这种设置下，$\mathbf{s}$遵循单纯复形上的结构方程模型（SEM），其中每个单纯形上的信号都被建模为相邻单纯形上信号的含噪线性组合。我们将(51)重写为

$$\hat{\mathbf{s}}=\operatorname{argmin}_{\mathbf{z}} \frac{1}{\sigma_v^2}\|\bar{\mathbf{s}}-\mathbf{\Theta}\mathbf{z}\|_2^2+\mathbf{z}^\top\big(\mathbf{I}+\alpha^2\mathbf{T}^2-2\alpha\mathbf{T}\big)\mathbf{z}. \tag{53}$$

当$\mathbf{T}=\mathbf{D}$时，项$\alpha^2\mathbf{D}^2$再次促进所有阶上的信号平滑性。项$-2\alpha\mathbf{D}$鼓励相邻阶之间的信号行为保持一致，即$(k-1)$-信号的旋度方向不会与k-信号的梯度方向严重相反。

- **子空间感知：** 我们回到第IV-A3节中的设定，其中我们知道$\mathbf{s}$主要受限于某个子空间，该子空间记为$\mathbf{U}_{\mathcal{S}}$；这在某些领域中很常见。回忆一下，这对应于

$$\mathbf{s}\approx\mathbf{U}_{\mathcal{S}}\tilde{\mathbf{s}}_{\mathcal{S}},\quad \mathbf{C}\approx\mathbf{U}_{\mathcal{S}}\operatorname{diag}(\mathbf{p}_{\mathcal{S}})\mathbf{U}_{\mathcal{S}}^\top, \tag{54}$$

其中，$\mathcal{S}$不仅对$\mathbf{T}$的特征向量$\mathbf{U}$中的活跃列进行索引，还选出信号的拓扑傅里叶变换$\mathbf{s}$和功率谱密度$\mathbf{p}$中的非零项。于是，(51)中的$\mathbf{C}^{-1}=\mathbf{U}_{\mathcal{S}}\operatorname{diag}^{-1}(\mathbf{p}_{\mathcal{S}})\mathbf{U}_{\mathcal{S}}^\top$。然而请注意，现在我们可以将(49)重写为

$$\bar{\mathbf{s}}=\mathbf{\Theta}\mathbf{U}_{\mathcal{S}}\tilde{\mathbf{s}}_{\mathcal{S}}+\mathbf{v},$$

这启发我们像第V-A节那样在频域中求解$\mathbf{s}$。由此可得估计值

$$\begin{aligned}\hat{\tilde{\mathbf{s}}}_{\mathcal{S}} &= \operatorname{argmin}_{\tilde{\mathbf{z}}} \frac{1}{\sigma_v^2}\|\bar{\mathbf{s}}-\mathbf{\Theta}\mathbf{U}_{\mathcal{S}}\tilde{\mathbf{z}}\|_2^2+\tilde{\mathbf{z}}^\top\operatorname{diag}^{-1}(\mathbf{p}_{\mathcal{S}})\tilde{\mathbf{z}}\\ &= \Big(\mathbf{U}_{\mathcal{S}}^\top\mathbf{\Theta}^\top\mathbf{\Theta}\mathbf{U}_{\mathcal{S}}+\sigma_v^2\operatorname{diag}^{-1}(\mathbf{p}_{\mathcal{S}})\Big)^{-1}\mathbf{U}_{\mathcal{S}}^\top\mathbf{\Theta}^\top\bar{\mathbf{s}}, \end{aligned} \tag{55}$$

$\hat{}$，借此可以得到信号估计值$\hat{\mathbf{s}}=\mathbf{U}_{\mathcal{S}}\tilde{\mathbf{s}}_{\mathcal{S}}$。

在经典的确定性方法中，我们不允许随机性，因此（55）中的$\sigma_v^2\operatorname{diag}(\mathbf{p}_{\mathcal{S}})=\mathbf{0}$为零；此时我们重新获得估计值$\tilde{\mathbf{s}}$，同时舍弃部分观测信号$\bar{\mathbf{s}}$中那些未明确贡献于活跃子空间$\mathbf{U}_{\mathcal{S}}$的元素。然而，由于（55）中既考虑了加性噪声$\mathbf{v}$，也考虑了平稳信号$\mathbf{s}$的方差所带来的随机性，我们可以将不确定性纳入对$\bar{\mathbf{s}}$的考量，并根据观测值中的噪声水平$\sigma_v^2$以及感兴趣信号的方差$\mathbf{p}_{\mathcal{S}}$，得到最可能的信号估计值$\hat{\mathbf{s}}$。此外，我们的随机方法还具有更强的实用性，因为它能够灵活适用于信号$\mathbf{s}$几乎但并非完全受限于$\mathbf{U}_{\mathcal{S}}$的情形。

以上是从任意一组已观测单形中重构信号的实用选择，适用于单阶情形或跨所有阶的情形。然而，对于单纯复形而言，多阶设置下的信号重构$\mathbf{T}=\mathbf{D}$尤为重要。如上所述，我们可能只能测量某些阶上的信号。例如，假设我们在一个2阶单纯复形上观测节点信号和三角形信号，但无法测量边上的值，也就是说，$\mathbf{\Theta}$仅选择$\mathbf{s}$中对应于0阶和2阶单形的条目。由于没有观测到边信号，我们无法利用面和余面之间的边连通性来近似未知值，因此只能依赖相邻阶的信号来确定$\mathbf{s}$中的边信号值。这只有在狄拉克算子$\mathbf{D}$规定了各阶信号行为之间的关系时才有可能。

C. 高斯性下的平稳拓扑信号

随机拓扑信号的一个重要类别是平稳信号$\mathbf{s}$为高斯分布的情形。我们特别关注这样的情况：赋予拓扑滤波器$\mathbf{H}$（式（6））的激励$\mathbf{w}$是高斯的，但也可以定义其他高斯拓扑过程模型，例如[28]。我们的情形类似于其他领域中平稳过程具有高斯性的经典情形，同时我们探讨底层域为单纯复形时的情况。

若输入白噪声$\mathbf{w}\sim\mathcal{N}(\mathbf{0},\mathbf{I})$为高斯，则由(6)中的线性可知，我们的信号$\mathbf{s}=\mathbf{H}\mathbf{w}$也为高斯，其中

$$\mathbf{s}\sim\mathcal{N}(\mathbf{0},\mathbf{C}),\quad \mathbf{C}=\mathbf{H}^2, \tag{56}$$

并且我们回顾$\mathbf{H}=\sum_{r=0}^{R-1}h_r\mathbf{T}^r$对应于算子$\mathbf{T}\in\{\mathbf{L}_k,\mathbf{D}\}$，这决定了我们关注的是单阶信号$k$还是多阶数据。于是，$\mathbf{s}$的概率密度函数为

$$f_{\mathbf{s}}(\mathbf{x})=\frac{(2\pi)^{2/N}}{\det(\mathbf{H})}\exp\Big\{-\frac{1}{2}\mathbf{x}^\top\mathbf{H}^{-2}\mathbf{x}\Big\}, \tag{57}$$

因为$\mathbf{C}=\mathbf{H}^2$。此外，根据定义1和定义3，我们回顾$\mathbf{C}=\mathbf{U}\operatorname{diag}(\mathbf{p})\mathbf{U}^\top=\mathbf{U}\operatorname{diag}^2(h(\boldsymbol{\lambda}))\mathbf{U}^\top$，其中滤波器频率响应为$h(\lambda)=\sum_{r=0}^{R-1}h_r\lambda^r$，且$\mathbf{U}$为$\mathbf{T}$的特征向量。因此，与平稳拓扑信号的通常情形一样，$\mathbf{s}$的二阶统计行为完全由功率谱密度$\mathbf{p}$描述，而单纯复形结构则通过$\mathbf{U}$纳入。然而，更重要的是，若$\mathbf{s}$为高斯，则$\mathbf{s}$的频率分量，即$\tilde{\mathbf{s}}=\mathbf{U}^\top\mathbf{s}$的各分量，不仅不相关而且相互独立。为看出这一点，回顾推论1表明$\tilde{\mathbf{s}}$的协方差为$\tilde{\mathbf{C}}=\operatorname{diag}(\mathbf{p})$。此外，拓扑傅里叶变换是线性运算，因此$\tilde{\mathbf{s}}\sim\mathcal{N}\big(\mathbf{0},\operatorname{diag}(\mathbf{p})\big)$也由高斯变量组成，其概率密度函数为

$$f_{\tilde{\mathbf{s}}}(\tilde{\mathbf{x}})=\prod_{i=1}^{N}\frac{1}{\sqrt{2\pi p_i}}\exp\Big\{-\frac{\tilde{x}_i^2}{2p_i}\Big\}. \tag{58}$$

因此，单纯结构通过$\mathbf{U}$影响$\mathbf{s}$在空间域中的协方差，但频率分量$\tilde{\mathbf{s}}$彼此独立，其方差由$\mathbf{p}$决定。

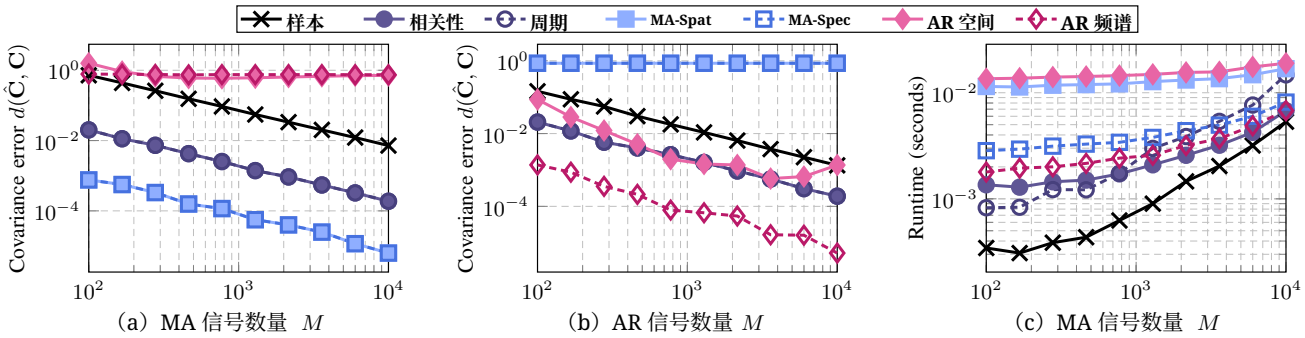


图 2：对于 (a) MA拓扑信号和 (b) AR拓扑信号，随着观测信号数量 M 增加，协方差估计误差的变化。(c) 随着观测信号数量 M 增加，运行时间（秒）的对比。

当我们想要根据独立实现$\mathbf{S} \in \mathbb{R}^{N\times M}$估计协方差$\mathbf{C}$ 时，已知式（15）中的样本协方差 $\hat{}\mathbf{C}$ 是$\mathbf{C}$的最大似然估计（MLE）。不过，我们也可以采用第 IV 节中的方法，在施加先验结构信息的同时，求得$\mathbf{C}$ 的最可能估计值。例如，对于参数化方法，我们可以考虑服从式（25）或式（35）中 MA 或 AR 模型的高斯拓扑信号。如果$\mathbf{C}^{-1} \in \{\mathbf{L}_k, \mathbf{D}^2\}$，我们可以对$\mathbf{C}^{-1}$施加稀疏性，以简洁地刻画单纯形相互依赖关系；也可以鼓励$\mathbf{C}^{-1}$ 具有低秩特性，以体现其内在的简约性。举例来说，设$\mathbf{s}$为一个自回归过程，其中 $\left(\mathbf{C}^{\alpha}(\boldsymbol{\alpha})\right)^{-1} = \mathbf{H}^2(\boldsymbol{\alpha})$，如式（34）所示。那么，我们不再将样本协方差$\mathbf{C}$作为朴素的最大似然估计 $\hat{}$，而是可以考虑以下正则化优化问题

$$\begin{aligned}\hat{\boldsymbol{\alpha}} &= \operatorname{argmin}_{\boldsymbol{\alpha}} \quad -\log\det(\boldsymbol{\Theta}_{\boldsymbol{\alpha}}) + \operatorname{tr}(\hat{\mathbf{C}}\boldsymbol{\Theta}_{\boldsymbol{\alpha}}) + \lambda\mathcal{R}(\boldsymbol{\Theta}_{\boldsymbol{\alpha}}) \\ &\quad \text{s.t.} \quad \boldsymbol{\Theta}_{\boldsymbol{\alpha}} = \mathbf{H}^2(\boldsymbol{\alpha}) \succ 0, \quad \boldsymbol{\Theta}_{\boldsymbol{\alpha}} \in \mathcal{D}, \end{aligned} \tag{59}$$

其中，$\lambda > 0$ 是某个惩罚权重，$\mathcal{R}(\boldsymbol{\Theta}_{\boldsymbol{\alpha}})$包含用于促进 $\boldsymbol{\Theta}_{\boldsymbol{\alpha}}$ 理想结构属性（如稀疏性）的正则化项，而 $\mathcal{D}$ 则确保 $\boldsymbol{\Theta}_{\boldsymbol{\alpha}}$ 符合已知特征，例如，当 $R = 1$时，其非对角元素仅限于 $\mathbf{T}$ 和$\mathbf{T}^2$ 的支撑集，此时 $\boldsymbol{\Theta}_{\boldsymbol{\alpha}} = \mathbf{H}^2(\boldsymbol{\alpha}) = \mathbf{I} - 2\alpha\mathbf{T} + \alpha^2\mathbf{T}^2$。

D. 其他研究方向

本节讨论的示例仅涉及一部分适用于拓扑平稳性或值得研究的问题。我们在此提出一些值得未来探索的其他研究方向。

时变拓扑信号。尽管许多问题关注静态数据的分析与应用，但数据随时间变化的情况极为常见。因此，一个自然的拓展方向是将现有的基于平稳性的框架应用于时变随机拓扑信号。针对这一场景，潜在的改进包括定义状态空间表示，以及推广适用于单纯复形上信号的动态滤波方法。

稳健协方差估计。另一种颇具挑战性但又切实存在的情形是：从不完整或受噪声污染的观测值中估计平稳过程的协方差。在这项工作中，我们提出了对拓扑信号进行去噪或插值的方法，这意味着采用一个简单的两步流程：首先估计协方差矩阵，然后利用它改进估计值。然而，这两个步骤完全独立，可能会忽略一些对估计有益的关键关系。这促使我们研究稳健协方差估计器，尤其是在确保计算仍然可行，或能够处理常见但具有挑战性的情形（例如存在离群值）的情况下。

超越单纯复形。虽然单纯复形确实能构成丰富的领域，我们也可以考虑将这一框架推广到其他高阶结构，例如超图、胞腔复形或非单纯形相互作用。近期在这方面的进展已将线性滤波的概念拓展至这些情形 [39], [40], ，后续自然可以进一步推广平稳性的概念。然而，缺少包含性质所施加的结构时，针对更一般对象的问题往往会变得不适定，因此，扩展工作可能需要更加审慎的分析。

单纯复形估计。图信号处理中的一项基本任务是根据一组观测信号估计图结构，其中包括假设信号平稳的方法 [41], [42]。因此，另一个自然的研究方向是从平稳拓扑信号中恢复单纯复形。例如，我们可以估计关联矩阵 $\{\mathbf{B}_k\}_{k=1}^K$，据此构造狄拉克算子 $\mathbf{D}$ 或霍奇拉普拉斯算子 $\{\mathbf{L}_k\}_{k=0}^K$；也可以直接估计 $\mathbf{D}$或$\mathbf{L}_k$。虽然从概念上说，这一推广很直接，但恢复底层单纯结构所需的考量比图情形复杂得多，还必须确保估计值有效，例如确保$\mathbf{B}_k\mathbf{B}_{k+1} = \mathbf{0}$。

随机单纯复形。最后，即使我们根据数据估计或构建单纯复形，也未必能确定它们是否反映了节点之间的潜在关系。除了能够构建单纯复形之外，我们通常还需要能够对信号以及连接节点集合的单形中的不确定性进行建模。因此，拓扑信号处理的另一个重要方向是针对单纯复形的概率建模。

VI. 数值结果

为完成对拓扑平稳性框架的介绍，并使所提出的概念具体化，我们通过数值实验展示协方差估计和信号恢复。对于以下实验，我们通过 $d(\hat{}\mathbf{X}, \mathbf{X}) = \|\hat{}\mathbf{X} - \mathbf{X}\|_F^2 / \|\mathbf{X}\|_F^2$ 测量误差，其中 $\hat{}\mathbf{X}$ 表示估计值，$\mathbf{X}$ 为目标值。所有结果均为 50 次独立试验测量值的中位数。

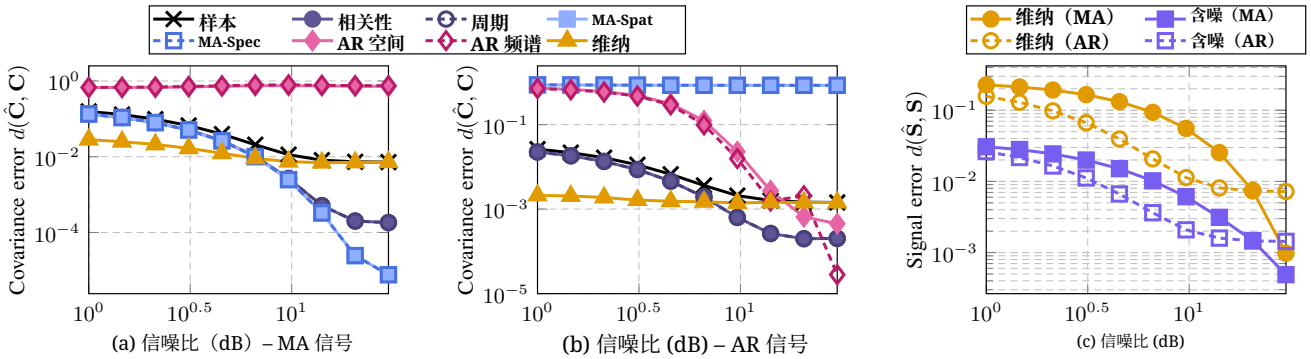


图 3：随着信噪比变化，从含噪或滤波后的信号估计协方差所产生的误差：(a) MA拓扑信号和 (b) AR拓扑信号。(c) MA 和 AR 拓扑信号的去噪误差随信噪比变化的情况。

A. 协方差估计

我们首先考虑第 IV 节所述的平稳拓扑信号的协方差估计，其中我们观测到M 信号 $\mathbf{S} \in \mathbb{R}^{N\times M}$，并比较估计相应协方差矩阵$\mathbf{C}$的方法。对于非参数方法，我们考虑式 (15) 中的样本协方差（记为**样本**）、式 (17) 中的相关图估计器（记为 **Cor**）以及式 (19) 中的周期图估计器（记为 **Per**）。对于参数估计器，记为**移动平均** 或**自回归** 的方法，分别假设$\mathbf{s}$ 是 MA 过程或 AR 过程。对于 MA 过程，**MA-Spat** 表示在空间域中的估计（31），而**MA-Spec**表示在谱域中的估计（32）。类似地，**AR 空间** 和**AR 频谱**分别表示估计值（41）和（42）。

我们生成阶数为 $K = 2$ 、具有$N_0 = 50$ 个节点的单纯复形，其中对每一对节点以概率0.2采样一条边，并对每三条边组成的三元组以概率0.3采样一个三角形。对于每个合成单纯复形，我们由(25)中的MA模型或(35)中的AR模型生成多阶平稳样本，其中对每个 $r \in [R]$ 和 $R = 3$有$\mathbf{T} = \mathbf{D}$ 以及 $\beta_r = \alpha_r = 0.1$ 。随着观测值数量从 $M = 10^2$ 增加到 $M = 10^4$，我们估计信号的协方差矩阵$\mathbf{C}$，并度量每种方法所得的误差（距离），两种信号模型的结果分别如图2a和图2b所示。不出所料，假设正确信号模型的方法优于其他方法；**移动平均** 与**自回归** 方法分别在图2a和图2b中取得最低误差。同样，假设错误模型会带来更高的误差，其中**移动平均**方法对 AR 信号的误差最差，反之亦然。然而，尽管**MA-Spat**与**MA-Spec** 得到等价解，因为问题(31)和(32)的目标函数返回等价的值，但请注意(41)和(42)并非如此。因此，我们还发现**AR 频谱** 优于**AR 空间** ，因为(41)中的空间方法相对于(42)需要多得多的样本才能进行充分估计。此外，请注意**相关图**与**周期图** 得到等价的估计值，因此它们在图2a和图2b中的表现相同，但(19)中的周期图估计值在低样本情形下更为高效。为说明这一点，我们在图2c中绘制了针对MA 信号的各方法运行时间，因为无论信号模型如何，各方法之间的运行时间比较都是等价的。如第IV-A节所述，当 M较低时，**周期图**比**相关图** 更快，但当 M 足够大时，相关图估计器**相关图** 变得更为高效。此外，如图2c所示，请注意**MA-Spec** 和**AR 频谱** 这类频域参数估计更为高效比它们的空间对应方法更为高效，尽管**MA-Spat**与**MA-Spec**的解是等价的。

B. 拓扑信号去噪

我们考虑第 V-A 节中的信号去噪问题，即希望从含噪观测 $\mathbf{S} \in \mathbb{R}^{N\times M}$ 中恢复 $M = 10^4$ 多阶信号 $\mathbf{Y} = \mathbf{S}+\mathbf{V}$，其中含有不同 σ_v^2值下的独立白高斯噪声 $V_{im} \sim \mathcal{N}(0, \sigma_v^2)$。对于 MA 和 AR 模型，我们像之前一样生成单纯复形及其平稳数据，并使用第 VI-A 节中的所有方法，从含噪信号 $\mathbf{Y}$ 估计其底层协方差。我们还利用维纳滤波对式 (47) 中得到的滤波后ˆ信号 $\mathbf{S}$进行处理，通过式 (15) 近似协方差。图 3a 和图 3b 分别绘出了 MA 和 AR 信号的协方差估计误差随信噪比增大（即σ_v^2 减小）而变化的情况。我们发现，当信噪比较低、$\mathbf{Y}$ 噪声较大时，维纳滤波效果更好，因为我们可以通过式 (47) 衰减噪声尤其严重的频率分量；而即使是基于正确信号模型的方法，也可能对受污染的数据过拟合。

我们还在图 3c 中直接测量了信号重建误差，比较两种信号模型下$\mathbf{Y}$ 与$\mathbf{S}$ 之间的误差（记为 ˆ**含噪**），以及$\mathbf{S}$与$\mathbf{S}$ 之间的误差（记为**维纳**）。如预期，维纳滤波始终能得到更低的信号误差，从而改善含噪观测值。此外，我们还观察到，在较高信噪比下，自回归信号的**维纳** 与**含噪** 之间的差距大于移动平均信号，这是因为自回归模型更复杂。

C. 拓扑信号插值

接下来，我们比较第 V-B 节中用于拓扑信号插值的正则项。对于一组 $M = 10^4$ 多阶单纯信号 $\mathbf{S} \in \mathbb{R}^{N\times M}$，我们仅对¯ $P < N$ 单形的一个子集 $\mathbf{S} = \mathbf{\Theta}\mathbf{S}$ +V 观测含噪值，其中 $V_{im} \sim \mathcal{N}(0, \sigma_v^2)$ 对应于 $\sigma_v^2 = 0.01$。我们采样具有$R = 3$ 和 $\beta_r = 0.3$ 的 MA 信号（对所有 $r \in [R]$）、具有$R = 1$ 和 $\alpha = 0.3$的 AR 信号，以及低通平稳信号。更具体地说，我们按 (6) 中的滤波方式生成低通数据，对于 H =Udiag($h_{\text{lp}}(\boldsymbol{\lambda})$)U$^\top$，其中 $\boldsymbol{\lambda}$，U 分别为 D 的特征值和特征向量，并且 $h_{\text{lp}}(\lambda) = (\lambda^2 + 10^{-3})^{-1}$。随后，我们在给定部分¯观测值 S 的情况下进行拓扑信号恢复，以比较第 V-B 节中的方法，其中最大后验估计对应于给定真实协方差矩阵 C 时的 (50)，Smooth 求解 (52)，SEM 求解 (53)，而零值法假设 C =I

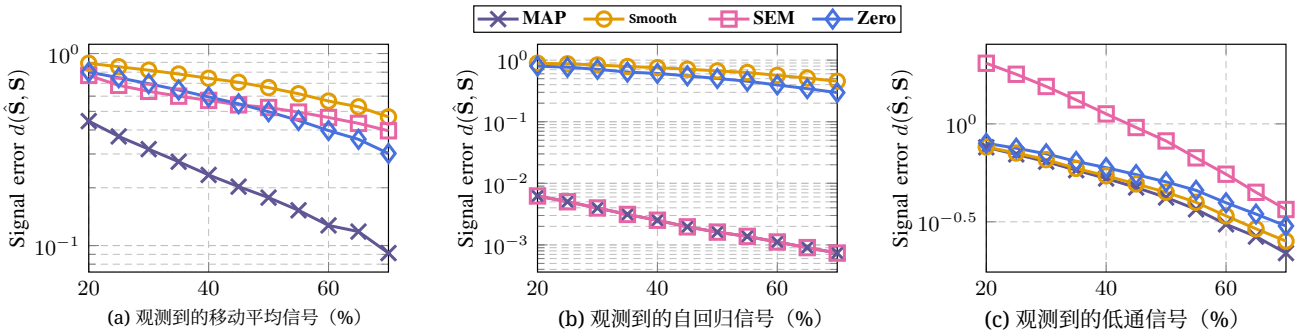


图 4：随着已观测单形的百分比增加，(a) MA拓扑信号、(b) AR拓扑信号和 (c) 低通拓扑信号的重构误差。

因此，它实现 $\boldsymbol{\Theta}^{\top}\bar{}\mathbf{S}$，用零填充未观测信号。在图 4 中，我们展示了全部三种信号模型在不同已观测单形比例 P/N 下的信号重建误差。我们观察到，对于图 4a 中的 MA 过程，带有先验结构假设的方法 **Smooth**、**SEM** 和 **零值法** 所取得的误差显著高于 **最大后验估计**（使用真实 $\mathbf{C}$），因为它们的假设未能刻画 (24) 中三阶拓扑滤波器的复杂性。然而，在图 4b 中，给定参数$\alpha \in \mathbb{R}$ 时，参数化方法**SEM** 等价于**最大后验估计**，因为标量 α 完全描述了 AR 信号协方差，但我们无需像 (50) 中那样对$\mathbf{C}$ 求逆。最后，对于图 4c 中的低通信号，**SEM** 取得最高误差，因为它假设 (35) 中为高通滤波器，但 **Smooth**的表现最接近最优的 **最大后验估计**，因为 **Smooth**会在估计信号中鼓励更强的低通特性。

D. 合著数据

最后，我们考察一个现实场景，其中一个二阶单纯复形根据学术作者之间的合作关系将他们连接起来。这里，节点代表作者，边和三角形则连接至少共同发表过一篇论文的研究人员。我们将每位作者使用 $M = 1903$ 各种关键词的平均频率视为 0-信号，并通过分别对由边和三角形连接的作者的 0-信号求平均，构建 1-信号和 2-信号。基于这些合著数据，我们在图 5 中针对这些关键词信号进行信号去噪和插值的重复模拟，每项实验进行 50 次独立试验；每次试验对应于随机选取一个由$N_0 = 50$ 作者构成的诱导子单纯复形。

对于图 5a，我们按照第VI-B节所述，向关键词信号$\mathbf{V}$ $\mathbf{S}$ 中添加白高斯噪声，以得到受噪信号$\mathbf{Y} = \mathbf{S} + \mathbf{V}$，并考察不同噪声方差 σ_v^2 下的情况。随后，我们比较含噪信号$\mathbf{Y}$ 与真实信号$\mathbf{S}$ 之间的误差，以及通过维纳滤波（47）得到的去噪信号$\mathbf{S}$ 的误差。此外，我们还将其与基于平滑性假设的估计进行比较，即，

$$\hat{\mathbf{S}}_{\text{smooth}} = \text{argmin}_{\mathbf{S}} \|\mathbf{S} - \mathbf{Y}\|_F^2 + \gamma \text{tr}(\mathbf{S}^{\top}\mathbf{D}^2\mathbf{S}) \quad (60)$$

对于 $\gamma = 0.1$。我们观察到，平滑性可以改善噪声很大的观测值，但如果没有先验信息来规定 γ，额外施加的结构可能会损害估计。另一方面，维纳滤波后的信号 $\mathbf{S}$ （记为 **维纳**）在估计中纳入了随机性，因此在图 5a 的所有信噪比值下，其误差都低于 $\mathbf{Y}$ （记为 **含噪**）和$\mathbf{S}_{\text{smooth}}$ （记为 **Smooth**）。

我们还重新探讨了信号插值：仅根据子集$\bar{}$单形 $\mathbf{S} = \boldsymbol{\Theta}$ $\mathbf{S}+\mathbf{V}$ 上的含噪观测值，恢复完整信号值 $\mathbf{S}$ ，其中 $V_{im} \sim \mathcal{N}$（0，1），并研究观测单形的比例 P/N 变化时的情况。我们通过求解对（51）的修改形式来插值$\bar{}\mathbf{S}$,

$$\hat{\mathbf{S}} = \text{argmin}_{\mathbf{S}} \|\bar{\mathbf{S}} - \boldsymbol{\Theta}\mathbf{S}\|_F^2 + \text{tr}(\mathbf{S}^{\top}\hat{\mathbf{C}}_{\text{pg}}^{-1}\mathbf{S}) + \gamma \text{tr}(\mathbf{S}^{\top}\mathbf{D}^2\mathbf{S}) \quad (61)$$

其中，我们用式（19）中的 $\hat{}\mathbf{C}$ 替代未知的真实协方差矩阵$\mathbf{C}$ pg 。我们还加入平滑性惩罚项，并改变权重 γ ，以展示底层信号平滑性在何种情况下有助于估计。我们再次发现，在一定程度上鼓励平滑性可以改善估计：当观测到的单形比例较低时，如图 5b 所示，$\gamma = 0.1$ 的误差低于 $\gamma \in \{0, 1\}$ （当 $P/N \leq 40$ 时）；但仅施加

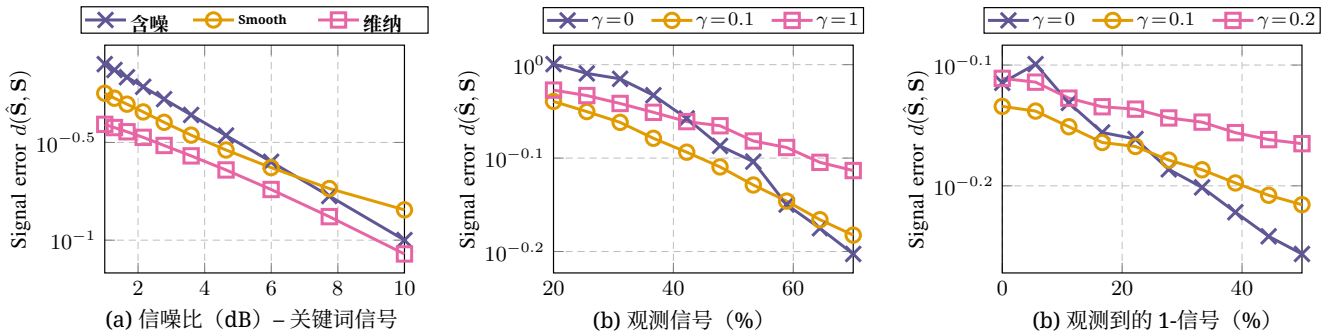


图 5：(a) 关键词信号的去噪误差随信噪比的变化。(b) 随着观测到的单形所占百分比增加，关键词信号的重构误差。(c) 随着观测到的 1-单形所占百分比增加，关键词信号的重构误差。

平稳性 $\gamma = 0$ 在观测到足够多单形时更可取。

我们也举例说明了缺失1-单形数据的情形。对于一组 $\bar{\mathbf{S}} \in \mathbb{R}^{P\times M}$ 部分观测，我们随机丢弃数量不等的1-信号，其中 $P = N_0 + P_1 + N_2$ 和 $P_1 < N_1$。图 5c绘制了针对 $\gamma \in \{0, 0.1, 0.2\}$求解(61)所得的信号重建误差，其中 $P_1/N_1 = 0$对应于在不观测任何边的情况下恢复1-信号的情形。我们观察到与图 5b类似的趋势：当P_1/N_1 较低时，通过 $\gamma = 0.1$施加的平滑性可以改善估计，但过强的惩罚 $\gamma = 0.2$ 会增大误差。因此，当数据稀缺时，先验结构知识有助于估计；但当观测到足够多的单形时，由我们的平稳性假设所带来的灵活性已足够。

VII. 结论

我们提出了单纯复形上随机信号的概率框架。具体而言，我们定义了平稳拓扑过程，以将经典平稳性扩展到拓扑域。我们通过对Hodge算子和Dirac算子的谱分析，探讨了平稳性对拓扑信号的影响，并由此定义了单纯形信号的功率谱密度。因此，尽管我们的工作遵循传统信号处理模型与技术的直觉，我们仍能够为复杂高阶连接上的随机过程对这类数据进行建模。此外，平稳拓扑过程在代数上易于处理的谱性质使我们能够开发信号协方差矩阵和功率谱密度的实用估计器，这在数据稀缺情形下尤为有用。我们还将信号处理中的经典任务（如信号去噪和插值）推广到单纯复形域。最后，我们讨论了未来方向，包括向更一般拓扑结构的扩展，以及将我们的框架应用于生成模型或贝叶斯模型。

参考文献

[1] C. Bick、E. Gross、H. A. Harrington 和 M. T. Schaub，“什么是高阶网络？”*SIAM* 综述，第 65 卷，第 3 期，第 686–731 页，2023 年。[2] V. Salnikov、D. Cassese 和 R. Lambiotte，“单纯复形与复杂系统”，《欧洲物理学杂志》，第 40 卷，第 1 期，文章编号 014001，2018 年。[3] S. Barbarossa 和 S. Sardellitti，“单纯复形上的拓扑信号处理”，*IEEE* 信号处理汇刊，第 68 卷，第 2992–3007 页，2020 年。[4] M. T. Schaub、Y. Zhu、J.-B. Seby、T. M. Roddenberry 和 S. Segarra，“高阶网络上的信号处理：立足边上……以及更远之处”，信号处理，第 187 卷，文章编号 108149，2021 年。[5] E. Isufi、G. Leus、B. Beferull-Lozano、S. Barbarossa 和 P. Di Lorenzo，“拓扑信号处理与学习：最新进展和未来挑战”，信号处理，第 233 卷，文章编号 109930，2025 年。[6] M. H. Hayes，《统计数字信号处理与建模》。John Wiley & Sons，2009 年。[7] P. Stoica 和 R. L. Moses，《信号谱分析》。普伦蒂斯霍尔，上萨德尔河，新泽西州，2005 年。[8] B. Girault，“利用等距图平移实现平稳图信号”，载于欧洲信号处理会议（*EUSIPCO*），2015 年，第 1516–1520 页。[9] B. Girault、P. Gonc̸alves 和 E. Fleury，“图上的平移：一种等距移位算子”，*IEEE* 信号处理快报，第 22 卷，第 12 期，第 2416–2420 页，2015 年。[10] N. Perraudin 和 P. Vandergheynst，“图上的平稳信号处理”，*IEEE* 信号处理汇刊，第 65 卷，第 3477 页，2017 年 13，3462，第 – 页。[11] A. G. Marques 、 S. Segarra 、 G. Leus 和 A. Ribeiro，“平稳图过程与谱估计”，*IEEE* 信号处理汇刊，第 65 卷，第 22 期，第 5911–5926 页，2017 年。[12] A. Sandryhaila 和 J. M. F. Moura，“图上的离散信号处理：频率分析”，*IEEE* 信号处理汇刊，第 62 卷，第 12 期，第 3042–3054 页，2014 年。[13] S. Segarra、A. G. Marques 和 A. Ribeiro，“最优图滤波器设计及其在分布式线性网络算子中的应用”，IEEE 信号处理汇刊，第 65 卷，第 15 期，第 4117–4131 页，2017 年。[14] L.-H. Lim，“图上的霍奇拉普拉斯算子”，SIAM 综述，第 62 卷，第 3 期，第 685–715 页，2020 年。[15] M. Yang、E. Isufi、M. T. Schaub 和 G. Leus，“单纯复形卷积滤波器”，IEEE 信号处理汇刊，第 70 卷，第 4633–4648 页，2022 年。[16] E. Isufi 和 M. Yang，“单纯复形上的卷积滤波”，载于 IEEE 国际声学、语音与信号处理会议（ICASSP），2022 年，第 5578–5582 页。[17] M. T. Schaub 和 S. Segarra，“流平滑与去噪：边空间中的图信号处理”，载于 IEEE 全球信号与信息处理会议（GlobalSIP），2018 年，第 735–739 页。[18] J. Jia、M. T. Schaub、S. Segarra 和 A. R. Benson，“基于图的边流半监督与主动学习”，载于 ACM SIGKDD 国际知识发现与数据挖掘会议（KDD），2019 年，第 761–771 页。[19] C. Liu、G. Leus 和 E. Isufi，“展开单纯复形弹性网络以重构边流信号”，IEEE 开放信号处理期刊，2023 年。[20] G. Bianconi，“网络与单纯复形的拓扑狄拉克方程”，《物理学：复杂性》，第 2 卷，第 3 期，文章编号 035022，2021 年。[21] F. Baccini、F. Geraci 和 G. Bianconi，“加权单纯复形及其表示高阶网络数据和拓扑的能力”，《物理评论E》，第 106 卷，第 3 期，文章编号 034319，2022 年。[22] M. Yang 和 E. Isufi，“单纯复形趋势滤波”，载于阿西洛马信号、系统与计算会议。电气与电子工程师学会，2022 年，第 930–934 页。[23] M. Yang、E. Isufi 和 G. Leus，“单纯复形卷积神经网络”，载于 IEEE 国际声学、语音与信号处理会议（ICASSP），2022 年，第 8847–8851 页。[24] C. Battiloro、L. Testa、L. Giusti、S. Sardellitti、P. Di Lorenzo 和 S. Barbarossa，“广义单纯复形注意力神经网络”，IEEE 信号与信息处理网络汇刊，2024 年。[25] J. Krishnan、R. Money、B. Beferull-Lozano 和 E. Isufi，“用于流式边流的单纯复形向量自回归模型”，载于 IEEE 国际声学、语音与信号处理会议（ICASSP），2023 年，第 1–5 页。[26] S. Reddy 和 S. P. Chepuri，“通过子采样邻域聚合恢复单纯复形上的信号”，IEEE 信号处理快报，2024 年。[27] M. T. Schaub、A. R. Benson、P. Horn、G. Lippner 和 A. Jadbabaie，“单纯复形上的随机游走与归一化霍奇 1-拉普拉斯算子”，SIAM 综述，第 62 卷，第 2 期，第 353–391 页，2020 年。[28] L. Marinucci、G. D’Acunto、P. Di Lorenzo 和 S. Barbarossa，“单纯复形高斯模型：表示与推断”，arXiv 预印本 arXiv:2510.12983，2025 年。[29] M. Yang，“拓扑薛定谔桥匹配”，载于国际学习表征会议（ICLR），2025 年。[30] M. Yang、V. Borovitskiy 和 E. Isufi，“霍奇复合边高斯过程”，载于国际人工智能与统计会议，S. Dasgupta、S. Mandt 和 Y. Li 编，第 238 卷。PMLR，2024 年 5 月 2–4 日，第 3754–3762 页。[31] C. Liu、V. M. Tenorio、A. G. Marques 和 E. Isufi，“匹配拓扑子空间检测器”，IEEE 信号处理汇刊，2026 年。[32] A. Hatcher，《代数拓扑》。剑桥大学出版社，2002 年。[33] L. Calmon、M. T. Schaub 和 G. Bianconi，“利用狄拉克算子进行高阶信号处理”，载于阿西洛马信号、系统与计算会议，2022 年，第 925–929 页。[34] T. M. Roddenberry、F. Frantzen、M. T. Schaub 和 S. Segarra，“霍奇小波：单纯复形上流的局部化谱表示”，载于 IEEE 国际声学、语音与信号处理会议（ICASSP）。电气与电子工程师学会，2022 年，第 5922–5926 页。[35] E. J. Candes、X. Li 和 M. Soltanolkotabi，“通过 Wirtinger flow 进行相位恢复：理论与算法”，IEEE 信息论汇刊，第 61 卷，第 4 期，第 1985–2007 页，2015 年。[36] Y. Shechtman、Y. C. Eldar、O. Cohen、H. N. Chapman、J. Miao 和 M. Segev，“应用于光学成像的相位恢复：当代综述”，IEEE 信号处理杂志，第 32 卷，第 109 页，2015 年第 3 期，第 87 页。[37] A.-L. Barab´asi，“网络科学”，《皇家学会哲学汇刊A辑》，第 371 卷，第 1987 期，文章编号 20120375，2013 年。[38] S. Kay，《统计信号处理基础：检测理论》。普伦蒂斯霍尔，1998 年。[39] T. M. Roddenberry、M. T. Schaub 和 M. Hajij，“胞腔复形上的信号处理”，载于 IEEE 国际声学、语音与信号处理会议（ICASSP）。电气与电子工程师学会，2022 年，第 8852–8856 页。[40] S. Sardellitti 和 S. Barbarossa，“广义胞腔复形上的拓扑信号处理”，IEEE 信号处理汇刊，2024 年。[41] X. Dong、D. Thanou、P. Frossard 和 P. Vandergheynst，“在平滑图信号表示中学习拉普拉斯矩阵”，IEEE 信号处理汇刊，第 64 卷，第 23 期，第 6160–6173 页，2016 年。[42] S. Segarra、A. G. Marques、G. Mateos 和 A. Ribeiro，“从谱模板推断网络拓扑”，IEEE 信号与信息处理网络汇刊，第 3 卷，第 3 期，第 467–483 页，2017 年。