arXiv:2305.01379v1 [stat.ML] 2 May 2023

LogSpecT：基于平稳信号并具有恢复保证的可行图学习模型

Shangyuan Liu* Linglingzhi Zhu† Anthony Man-Cho So‡

2023年5月3日

摘要

从信号中进行图学习是图信号处理（GSP）中的一项核心任务。从平稳信号中学习图最常用的模型之一是$SpecT$ [SMMR17]。然而，其实际形式$rSpecT$ 已知对超参数选择十分敏感，更糟的是，还可能出现不可行性。本文首次给出了保证 rSpecT 不可行的条件，并设计了一种新模型（$LogSpecT$）及其实际形式（$rLogSpecT$）以解决这一问题。与 rSpecT 不同，新型实际模型 rLogSpecT 始终可行。此外，我们给出了 rLogSpecT 的恢复保证，这些保证源自与上半连续收敛相关的现代优化工具。这些工具本身也可能具有独立价值，并对各种学习问题意义重大。为展示 rLogSpecT 在实践中的优势，我们提出了一种基于线性化交替方向乘子法（L-ADMM）的高效算法。L-ADMM 的子问题均有闭式解，且其收敛性得到保证。针对合成网络和真实网络的大量数值结果验证了所提方法的稳定性和优越性，彰显了其在各种图学习应用中的潜力。

1 引言

基于图的学习已在许多实际领域证明了其价值，例如生命科学[SMSK+11、STM15]，信号处理 [JHIM+19、JT19、TE20、TEOC20]，以及金融工程[AOTS15、MNBD21]，等等。然而，许多情况下图并未预先提供，我们只能观测到与图密切相关的数据。因此，图信号处理（GSP）的核心任务之一，是根据数据与图之间的相互作用学习潜在的图拓扑 [MSMR19]。

数据的一种常用属性是图信号平稳性 [Gir15、MSLR17、PV17]。这一属性将定义在空间/时间定义域上的信号平稳性概念扩展到

*香港中文大学系统工程与工程管理学系。电子邮箱：shangyuanliu@link.cuhk.edu.hk

†香港中文大学系统工程与工程管理学系。电子邮箱：llzzhu@se.cuhk.edu.hk

‡香港中文大学系统工程与工程管理学系。电子邮箱：manchoso@se.cuhk.edu.hk

具有不规则结构（即图）的信号 [SCML18]。尽管这一概念更多是从理论角度提出的，但多项研究表明，一些真实数据集可以近似视为平稳，或部分由平稳性假设解释。例如， [PV17]指出，广为人知的 USPS 数据集以及由裁剪人脸构成的 CMUPIE 数据集都表现出近似平稳性。 [Gir15] 发现，某些天气数据可以用平稳图信号来解释。此外，[SMMR17] 强调了平稳性假设在学习蛋白质结构中的重要性；这是一项生物信息学中的关键任务。

在平稳性假设下处理平稳图信号并学习拓扑的主流方法是基于谱模板的模型。这一研究方向始于 [SMMR17], ，其提出了一种称为$SpecT$ [SMMR17] 的基础模型，用于从平稳图信号中学习图。此后，人们对这一基础模型进行了许多扩展[SWUM17、BRM22]。在实际应用中，SpecT 需要未知的数据协方差矩阵。因此，有人提出了一种称为$rSpecT$ 的鲁棒形式[SMMR17], ，它引入了一个超参数，用于反映数据协方差矩阵的估计误差。因此，该模型对这个超参数十分敏感；若调参不当，可能会损害模型性能，甚至导致模型不可行。

目前，选择合适取值的方法计算开销很大，因为它需要对二阶锥规划求得高度精确的解。这个问题不仅会导致 rSpecT 的性能不稳定，还会使现有的恢复保证不完整。另一种较少受到关注的模型则试图通过将约束转化为惩罚项 [SM20]来规避模型不可行性问题。然而，这种方法引入了另一个既难以调参、也难以进行理论分析的超参数。此外，按照经验，随着收集到的样本越来越多，这个超参数应趋于无穷大，这可能会加剧模型的不稳定性。因此，我们自然会提出以下问题:

是否存在一种基于平稳信号的基础图学习模型，既能抵御超参数的影响，又具有可靠的理论保证?

本文通过提出一种替代性表述，来回答这一关键问题：基于平稳信号学习不含孤立节点的图，并提供严格的恢复保证。具体而言，我们的贡献如下。

1.1 贡献

首先，我们给出了一个能够保证基础模型 rSpecT 不可行的条件。为解决可行性问题，我们基于谱模板和对数障碍提出了一种名为 LogSpecT 的新表述，用于学习不含孤立节点的图。由于 LogSpecT 需要未知协方差矩阵的信息，我们又提出了一种名为 rLogSpecT 的实用表述，以处理有限样本情形。与 rSpecT 不同，rLogSpecT 始终保证可行。此外，我们提出的表述具有足够的通用性，能够继承 rSpecT 几乎所有现有的扩展。

其次，我们研究所提出模型的若干理论性质，并通过非渐近恢复分析将 rLogSpecT 与 LogSpecT 联系起来。利用与上半连续收敛相关的现代优化工具，可以证明恢复保证 [RW21]。不同于当前基于 ℓ_1 分析模型 [ZYY16], 、针对 rSpecT 构建的恢复保证，我们基于上半连续收敛的方法不受优化问题类型的限制（例如将

ℓ_1-范数与目标函数中的对数障碍相结合），因此能够适用于更广泛的应用场景，并具有独立的研究价值。

在算法方面，我们设计了一种线性化交替方向乘子法（L-ADMM）来求解 rLogSpecT。借助线性化技术，L-ADMM 的子问题具有闭式解，因而可以高效实现。此外，我们还证明了所提方法的收敛性。

最后，我们在合成数据和真实网络上进行了大量实验。在真实网络实验中，频繁观察到 rSpecT 的不可行性问题，这表明该问题并非仅仅是理论上的顾虑。与其他方法相比，我们观察到新提出的模型（即 LogSpecT 和 rLogSpecT）更加准确、稳定。对于经典模型失效的各种情形（例如 BA 图），我们的方法能够取得出色的性能。实验结果表明，我们的模型可作为基础模型，成为经典方法的更优替代方案。

1.2 相关工作

基于平稳信号的图学习是近年来备受关注的研究主题。 [ZJSL17] 采用了一阶近似，从平稳信号中学习图结构。 [SMMR17] 提出了首个基于谱模板的通用高效模型。为解决样本不足问题，他们还引入了鲁棒形式。此后，许多扩展方法相继提出。 [SWUM17] 从平稳信号中联合推断多个图。 [BRCM19] 处理了存在隐藏节点的情形。[SHMV19] 转向在线设置，并考虑了流信号。 [RBN+22] 解决了存在隐藏节点时的联合推断问题。 [SM20]考虑了存在隐藏节点时的在线设置。 [BRM22] 从平稳信号和平滑信号中学习含有隐变量的图。

不同于这些设置的扩展，我们重新审视基本模型，并对度施加正则化项。正如[dMCYP21]，所讨论的，将度限制为正值对于学习具有特殊结构的图、避免出现孤立节点至关重要。然而，直接采用约束甚至可能加剧我们问题中的可行性问题。我们采用对数障碍作为度正则化项。在图学习中，对数障碍最早见于 [Kal16] ，用于从平滑信号中学习图。后续研究局限于不同设置下从平滑信号中学习 [KLTF17、MF20、ZZL+22]。据我们所知，这是首项从理论和实证两方面证明对数障碍在基于平稳信号的图学习中的相关性的研究。

尽管鲁棒形式在各种基于平稳信号的图学习模型中无处不在，但其恢复保证却鲜有研究。 [PGM+17] 通过实证表明，随着样本数量增加，鲁棒形式的可行集会逼近原始可行集。 [NWM+22] 首次为 rSpecT 提供了恢复保证。他们的工作很大程度上依赖于目标函数为 ℓ_1-范数且约束为线性这一事实[ZYY16]。具体而言，他们的方法不适用于我们的模型。此外，他们工作所需的条件不仅限制严格，而且难以检验，因为这些条件取决于一个与随机样本相关的大规模矩阵的满秩性。

1.3 记号

本文采用的记号是标准的。我们用 $[m]$ 表示集合 $\{1,2,\dots,m\}$，其中 m为任意正整数。赋予全体实矩阵构成的欧几里得空间以内积 $\langle \boldsymbol{X},\boldsymbol{Y}\rangle \coloneqq \mathrm{tr}(\boldsymbol{X}^\top \boldsymbol{Y})$，其中 $\boldsymbol{X},\boldsymbol{Y}$为任意矩阵；由此诱导的 Frobenius 范数记为 $\|\cdot\|_F$（当自变量为向量时记为 $\|\cdot\|_2$）。对于任意 $\boldsymbol{X}\in\mathbb{R}^{m\times n}$，我们用 $\mathbb{B}(\boldsymbol{X},\rho)$ 表示以 $\boldsymbol{X}$ 为中心、半径为 ρ的闭欧几里得球。令 $\|\cdot\|$ 为算子范数，$\|\boldsymbol{X}\|_{1,1} \coloneqq \sum_{i,j}|\boldsymbol{X}_{ij}|$，$\|\boldsymbol{X}\|_{\infty,\infty}\coloneqq \max_{i,j}|\boldsymbol{X}_{ij}|$，并令 $\mathrm{diag}(\boldsymbol{X})$ 表示 $\boldsymbol{X}$的对角元素构成的向量。对于列向量 $\boldsymbol{x}$，令 $\mathrm{Diag}(\boldsymbol{x})$ 表示其对角元素由 $\boldsymbol{x}$给出的对角矩阵。给定闭凸集 $\mathcal{C}$，我们用 $\Pi_{\mathcal{C}}(\boldsymbol{w}) \coloneqq \mathrm{argmin}_{\boldsymbol{v}\in\mathcal{C}}\|\boldsymbol{v}-\boldsymbol{w}\|_2$ 表示将点 $\boldsymbol{w}$ 投影到集合 $\mathcal{C}$上的投影。我们用 $\mathbf{1}$（分别用 $\mathbf{0}$）表示全 1 向量（分别为全 0 向量），其维数由上下文确定。对于集合 $\mathcal{S}$ 和任意实数 α，令 $\alpha\mathcal{S} \coloneqq \{\alpha x \mid x\in\mathcal{S}\}$, $\mathcal{S}\mathbf{1} \coloneqq \{\boldsymbol{x}\mathbf{1} \mid \boldsymbol{x}\in\mathcal{S}\}$，并令 $\iota_{\mathcal{S}}(\cdot)$ 表示 $\mathcal{S}$的指示函数。对于两个非空紧集 $\mathcal{X}$ 和 $\mathcal{Y}$，它们之间的距离定义为 $\mathrm{dist}(\mathcal{X},\mathcal{Y}) \coloneqq \sup_{\boldsymbol{X}\in\mathcal{X}}\inf_{\boldsymbol{Y}\in\mathcal{Y}}\|\boldsymbol{X}-\boldsymbol{Y}\|_F$[1]。

2 预备知识

设 $\mathcal{G}=(\mathcal{V},\mathcal{E})$ 是一个图，其中 $\mathcal{V}=[m]$ 是节点集合，$\mathcal{E}\subset\mathcal{V}\times\mathcal{V}$ 是边集合。设 $\boldsymbol{S}$ 为与图 $\mathcal{G}$ 相关联的权重矩阵，其中 $\boldsymbol{S}_{ij}$ 表示节点 i 和 j之间边的权重。本文考虑不含自环的无向图。因此，有效邻接矩阵的集合为

$$\mathcal{S} \coloneqq \{\boldsymbol{S}\in\mathbb{R}^{m\times m} \mid \boldsymbol{S}=\boldsymbol{S}^\top,\ \mathrm{diag}(\boldsymbol{S})=\mathbf{0},\ \boldsymbol{S}\geq 0\}.$$

设邻接矩阵 $\boldsymbol{S}$ 可进行特征分解 $\boldsymbol{S}=\boldsymbol{U}\boldsymbol{\Lambda}\boldsymbol{U}^\top$，其中 $\boldsymbol{\Lambda}$ 是对角矩阵，$\boldsymbol{U}$ 是正交矩阵。图滤波器是一个线性算子 $h:\mathbb{R}^{m\times m}\to\mathbb{R}^{m\times m}$，定义为

$$h(\boldsymbol{S}) = \sum_{i=0}^{p} h_i \boldsymbol{S}^i = \boldsymbol{U}\left(\sum_{i=0}^{p} h_i \boldsymbol{\Lambda}^i\right)\boldsymbol{U}^\top,$$

其中 $p>0$ 是图滤波器的阶数，$\{h_i\}_{i=0}^p$ 是滤波器系数。按照惯例，我们有 $\boldsymbol{S}^0=\boldsymbol{I}$。

图信号可表示为向量 $\boldsymbol{x}\in\mathbb{R}^m$，其中第 i 个元素 x_i 是与节点 i 相关联的信号值。若信号 $\boldsymbol{x}$ 由以下方式生成，则称其为平稳信号

$$\boldsymbol{x} = h(\boldsymbol{S})\boldsymbol{w}, \tag{1}$$

其中，$\boldsymbol{w}$ 满足 $\mathbb{E}[\boldsymbol{w}]=\mathbf{0}$ 和 $\mathbb{E}[\boldsymbol{w}^\top\boldsymbol{w}]=\boldsymbol{I}$。简单计算可得，$\boldsymbol{x}$的协方差矩阵（记为 $\boldsymbol{C}_\infty$）与 $\boldsymbol{S}$具有相同的特征向量。因此，我们有如下约束

$$\boldsymbol{C}_\infty\boldsymbol{S} = \boldsymbol{S}\boldsymbol{C}_\infty. \tag{2}$$

[1]当集合 $\mathcal{Y}$（分别为 $\mathcal{X}$ 和 $\mathcal{Y}$）是单元素集合时，它可简化为经典的点到集合（分别为点到点）距离。

基于此，我们提出以下基础模型 SpecT，用于在不知道图滤波器的情况下从平稳信号中学习图 [SMMR17，SM20]:

$$\begin{aligned} \min_{\boldsymbol{S}} \ & \|\boldsymbol{S}\|_{1,1} \\ \text{s.t.} \ & \boldsymbol{C}_\infty \boldsymbol{S} = \boldsymbol{S}\boldsymbol{C}_\infty, \\ & \boldsymbol{S} \in \mathcal{S} \cap \left\{ \boldsymbol{S} \in \mathbb{R}^{m\times m} : \sum_{j=1}^{m} \boldsymbol{S}_{1j} = 1 \right\}, \end{aligned} \tag{SpecT}$$

其中约束条件 $\sum_{j=1}^{m} \boldsymbol{S}_{1j} = 1$ 用于排除平凡的最优解 $\boldsymbol{S}^* = \boldsymbol{0}$。当$\boldsymbol{C}_\infty$ 未知且只有 n 来自 $\{\boldsymbol{x}_i\}$ (1) 的独立同分布样本可用时，基于估计量 $\boldsymbol{C}_n$ 对 $\boldsymbol{C}_\infty$ 进行估计的鲁棒形式，称为 $rSpecT$，可表示为:

$$\begin{aligned} \min_{\boldsymbol{S}} \ & \|\boldsymbol{S}\|_{1,1} \\ \text{s.t.} \ & \|\boldsymbol{C}_n \boldsymbol{S} - \boldsymbol{S}\boldsymbol{C}_n\|_F \le \delta, \\ & \boldsymbol{S} \in \mathcal{S} \cap \left\{ \boldsymbol{S} \in \mathbb{R}^{m\times m} : \sum_{j=1}^{m} \boldsymbol{S}_{1j} = 1 \right\}. \end{aligned} \tag{rSpecT}$$

对于这种鲁棒形式，恢复保证在 [PGM+17] 中进行了实证研究，并在 [NWM+22] 中进行了理论研究，但所需条件具有较强限制性且难以检验。

3 rSpecT 的不可行性与新模型

尽管 rSpecT 已广受欢迎，但很少有研究讨论超参数 δ 的选择以及不可行性问题。本节将给出一个条件，在该条件下可以保证 rSpecT 不可行，并随后提出一种替代形式。为阐述我们的结果，先考虑以下双节点示例。

例 3.1。考虑一个包含 2 个节点的简单图。于是，rSpecT 的第二个约束所给出的集合为

$$\mathcal{S} \cap \left\{ \boldsymbol{S} \in \mathbb{R}^{m\times m} : \sum_{j=1}^{m} \boldsymbol{S}_{1j} = 1 \right\} = [0, 1; 1, 0],$$

这是一个单元素集合。假设样本协方差矩阵为 $\boldsymbol{C}_n = [h_{11}, h_{12}; h_{12}, h_{22}]$。那么，约束 $\|\boldsymbol{C}_n\boldsymbol{S} - \boldsymbol{S}\boldsymbol{C}_n\|_F \le \delta$ 可简化为 $2(h_{11} - h_{22})^2 \le \delta^2$。因此，当 $h_{11} \ne h_{22}$且 $\delta < \sqrt{2}\,|h_{11} - h_{22}|$时，rSpecT 不存在可行解。

在深入探讨一般情形之前，我们先介绍一个线性算子 $\boldsymbol{B} \in \mathbb{R}^{m^2 \times m\,(m-1)\,/2}$，它将向量 $\boldsymbol{x} \in \mathbb{R}_+^{(m-1)m/2}$ 映射为简单无向图的邻接矩阵 $\boldsymbol{S} \in \mathcal{S}$ 的向量化形式:

$$(\boldsymbol{B}\boldsymbol{x})_{m(i-1)+j} = \begin{cases} x_{i-j+\frac{i-1}{2}(2m-j)}, & \text{if } \ i > j, \\ 0, & \text{if } \ i = j, \\ x_{j-i+\frac{i-1}{2}(2m-i)}, & \text{if } \ i < j, \end{cases} \tag{3}$$

其中 i、$j \in [m]$。我们还定义 $\boldsymbol{A}_n := \boldsymbol{I} \otimes \boldsymbol{C}_n - \boldsymbol{C}_n \otimes \boldsymbol{I} \in \mathbb{R}^{m^2 \times m^2}$，这样一来，rSpecT 的第一个约束可以改写为

$$\|\boldsymbol{C}_n\boldsymbol{S} - \boldsymbol{S}\boldsymbol{C}_n\|_F \le \delta \iff \|\boldsymbol{A}_n \mathrm{vec}(\boldsymbol{S})\|_2 \le \delta,$$

其中，vec(·) 是向量化算子。现在给出一个保证 rSpecT 不可行的条件。

定理 3.2。考虑线性系统

$$\boldsymbol{A}_n\boldsymbol{B}\boldsymbol{y} = \boldsymbol{0}, \quad \boldsymbol{y} \le \boldsymbol{0}, \quad \sum_{i=1}^{m-1} y_i \ne \boldsymbol{0}, \tag{4}$$

其中 $\boldsymbol{y} \in \mathbb{R}^{\frac{m(m-1)}{2}}$。如果 (4) 没有可行解，则存在 $\bar{\delta}_n > 0$，使得对于所有 $\bar{}$，rSpecT $\delta_n \in [0, \delta_n)$均不可行。

备注 3.3。 由定理 *3.2*可知，对于任意固定的 n，当线性系统 (4) 中的 $\boldsymbol{A}_n\boldsymbol{B}$ 具有列满秩时（例如，例 *3.1*），该系统没有解。这将导致 $\bar{}$下的 rSpecT 不可行 $\delta_n \in [0, \delta_n)$。正如 *[SMMR 17]*, 中所述，应谨慎处理 rSpecT 的可行性问题。

rSpecT（SpecT）失效的原因在于约束（$\boldsymbol{S}\boldsymbol{1}$）$_1 = 1$的存在，该约束用于排除平凡解 $\boldsymbol{S}^* = \boldsymbol{0}$。因此，我们转而采用一种替代方法来规避平凡解。当假设图中不存在孤立节点时，通常会应用对数障碍 [Kal16]。在这些情况下，零解会自然地被排除。这一观察启发我们提出以下新颖的公式：将对数障碍与谱模板（2）相结合，从平稳信号中学习不含孤立节点的图：

$$\begin{aligned} \min_{\boldsymbol{S}} \ & \|\boldsymbol{S}\|_{1,1} - \alpha \boldsymbol{1}^\top \log(\boldsymbol{S}\boldsymbol{1}) \\ \text{s.t.} \ & \boldsymbol{S} \in \mathcal{S}, \quad \boldsymbol{S}\boldsymbol{C}_\infty = \boldsymbol{C}_\infty \boldsymbol{S}, \end{aligned} \tag{LogSpecT}$$

其中 $\|\cdot\|_{1,1}$ 是促进图稀疏性的 $\|\cdot\|_0$ 凸松弛， $\boldsymbol{1}^\top \log(\boldsymbol{S}\boldsymbol{1})$ 是用于保证不存在孤立节点的惩罚项，而 α 是调节参数。如下文命题所示，LogSpecT 中的超参数 α 只影响边权重的尺度，不影响图的连通结构。

命题 3.4. 设 Opt($\boldsymbol{C}, \alpha$) 为 LogSpecT 在输入协方差矩阵 $\boldsymbol{C}$ 和参数 $\alpha > 0$ 下的最优解集。那么，对任意 $\gamma > 0$，有

$$\mathrm{Opt}(\boldsymbol{C}, \alpha) = \gamma \, \mathrm{Opt}(\boldsymbol{C}, \alpha/\gamma) = \alpha \, \mathrm{Opt}(\boldsymbol{C}, 1).$$

注 3.5。 命题 *3.4* 的结果使我们在处理二值图时无需调整超参数 α。事实上，某些归一化操作可以消除不同取值 α 带来的影响，同时保留连通性信息。因此，在实现中，我们可以直接设为 $\alpha = 1$。

请注意，LogSpecT 中的真实协方差矩阵 $\boldsymbol{C}_\infty$ 通常未知，而我们可以从 n 独立同分布样本中获得估计值$\boldsymbol{C}_n$。为应对估计误差，我们引入以下鲁棒形式：

$$\begin{aligned} \min_{\boldsymbol{S}} \ & \|\boldsymbol{S}\|_{1,1} - \alpha \mathbf{1}^\top \log(\boldsymbol{S}\mathbf{1}) \\ \text{s.t.} \ & \boldsymbol{S} \in \mathcal{S}, \ \ \|\boldsymbol{S}\boldsymbol{C}_n - \boldsymbol{C}_n\boldsymbol{S}\|_F^2 \le \delta_n^2. \end{aligned} \tag{rLogSpecT}$$

该表述将 $\boldsymbol{C}_\infty$ 替换为 $\boldsymbol{C}_n$ ，并将等式约束放宽为带有容差阈值 δ_n的不等式约束。与rSpecT不同，我们证明rLogSpecT始终可行。

命题 3.6. 对于 $\delta_n > 0$ 和 $\boldsymbol{C}_n$ ，任意给定的 n下，rLogSpecT始终存在非平凡可行解。

接下来，我们讨论所提出模型的最优解/最优值的一些性质，这些性质有助于推导恢复保证。具体而言，我们得到最优解（可能不唯一）的一个上界，该上界与样本量 n 和不准确参数 δ_n无关。此外，还可得到最优值的一个下界。

命题 3.7. 以下结论成立：

- 对于LogSpecT的一个最优解 $\boldsymbol{S}^*$ （分别对于rLogSpecT在给定样本量 n下的一个最优解$\boldsymbol{S}_n^*$），有

$$\|\boldsymbol{S}^*\|_{1,1} = \alpha m \quad and \quad \|\boldsymbol{S}_n^*\|_{1,1} \le \alpha m, \quad \forall \delta_n > 0.$$

- 若 $\delta_n \ge 2\alpha m \|\boldsymbol{C}_n - \boldsymbol{C}_\infty\|$，则

$$\alpha m(1 - \log \alpha) \le f_n^* \le f^*, \quad \forall n,$$

其中， f^* （分别为 f_n^*）表示 LogSpecT（分别为 rLogSpecT）的最优值。

4 rLogSpecT 的恢复保证

本节研究在收集越来越多的独立同分布样本时，rLogSpecT 的非渐近行为。为简洁起见，记 $f := \|\cdot\|_{1,\ 1} - \alpha \mathbf{1}^\top \log(\cdot \mathbf{1})$。理论恢复保证如下：

定理 4.1. 若 $\delta_n \ge 2\rho\|\boldsymbol{C}_n - \boldsymbol{C}_\infty\|$ 在给定 $\rho := \max\{f^*, \alpha m, \alpha m(\log \alpha - 1)\}$ 的情况下，则存在常数 c_1、$\tilde{c}_1$、$\tilde{c}_2 > 0$，使得

$$\begin{aligned} &\text{(i)} \quad |f_n^* - f^*| \le c_1 \varepsilon_n, \quad \text{(ii)} \\ &\text{dist}\left(\mathcal{S}^{n,\ *}, \ \mathcal{S}_{2\varepsilon_n}^*\right) \le \varepsilon_n, \quad \text{(iii)} \\ &\text{dist}\left(\mathcal{S}^{n,\ *}\mathbf{1}, \ \mathcal{S}_0^*\mathbf{1}\right) \le \tilde{c}_1\varepsilon_n + \tilde{c}_2\sqrt{\varepsilon_n}, \end{aligned}$$

where $\varepsilon_n := \delta_n + 2\rho\|\boldsymbol{C}_n - \boldsymbol{C}_\infty\|$, $\mathcal{S}^{n,\ *}$ 是 rLogSpecT 的最优解集，并且

$$\mathcal{S}_\varepsilon^* := \{\boldsymbol{S} \in \mathcal{S} \mid \boldsymbol{S}\boldsymbol{C}_\infty = \boldsymbol{C}_\infty\boldsymbol{S}, \ f(\boldsymbol{S}) \le f^* + \varepsilon\}$$

是 LogSpecT 的 ε-次优解集。

注 4.2。 (i) 与定理 (ii) 中的结论 *4.1*相比，结论 (iii) 将 rLogSpecT 和 LogSpecT 的最优解集联系起来，而不是次优解。
(ii) 证明的一个副产品（详见附录 *D.2*）表明，最优节点度向量集合（$\mathcal{S}_0^*\mathbf{1}$）是一个单元素集合。然而，无法保证 $\boldsymbol{S}^*$ 本身的唯一性。我们将讨论这种非唯一性对模型性能的影响。相关讨论将在第 *6.4*节进行。

注 4.3。 定理 *4.1*的证明依赖于两个重要的优化概念：上半连续收敛和截断豪斯多夫距离。上半连续收敛与近似最小化问题的渐近解行为密切相关，而截断豪斯多夫距离则用于非渐近地刻画上半连续收敛。在 Kenmochi 条件的帮助下，我们能够显式计算截断豪斯多夫距离，从而研究模型最优值和最优解的非渐近行为。有关更多细节，请参阅附录 *D* 以及 *[RW09*的第 *7], [RW21*的第 6 章。*]*。

推论 4.4。 在定理 *4.1*的假设下，可得若 $\delta_n \to 0$，则

$$\lim_{n\to\infty} \operatorname{dist}(\mathcal{S}^{n,*}, \mathcal{S}_0^*) = 0. \tag{5}$$

注 4.5。 根据大数强定律，若选择 $\delta_n = \mathcal{O}(\|\boldsymbol{C}_n - \boldsymbol{C}_\infty\|)$，则几乎必然有 $\lim_{n\to\infty} \|\boldsymbol{C}_n - \boldsymbol{C}_\infty\| = 0$，从而 $\delta_n \to 0$。因此，推论 *4.4*表明 (5)几乎必然成立。

本节余下部分将研究在特定统计假设下 δ_n 的选择。大量分布（例如高斯分布、指数分布以及任意有界分布）都可归入次高斯分布（参见[Ver18，第 2 节。5]），其正式定义如下。

定义 4.6（次高斯分布）。若随机向量$\boldsymbol{w}$ 的概率分布满足存在正常数 C、v，使得对每个 $t > 0$，则称其为次高斯分布。

$$\mathbb{P}(\|\boldsymbol{w}\|_2 > t) \le Ce^{-vt^2}.$$

考虑如下情形： $\boldsymbol{w}$ 生成模型（1）服从次高斯分布。以下结果改编自 [Ver12，命题 2。1]。

引理 4.7。假设 $\boldsymbol{w}$ 在生成模型中（1）服从次高斯分布。那么， $\boldsymbol{x}$ 服从次高斯分布，且以大于 $1 - \mathcal{O}(\frac{1}{n})$ 的概率，

$$\|\boldsymbol{C}_n - \boldsymbol{C}_\infty\| \le \mathcal{O}\left(\sqrt{\frac{\log n}{n}}\right).$$

借助定理4.1中的非渐近结果，我们可以选择 δ_n 使其具有特定的衰减速率。

推论 4.8. 如果输入 $\boldsymbol{w}$ 在 (1) 服从次高斯分布，且 $\delta_n = \mathcal{O}(\sqrt{\log n/n})$，那么定理 *4.1*中的假设以大于 $1 - \mathcal{O}(\frac{1}{n})$的概率成立。

推论 4.8 与定理 4.1 一起说明，rLogSpect 的最优函数值／次优解集／最优节点度向量从 n 个独立同分布样本收敛到理想模型 LogSpect，其非渐近收敛速率为 $\mathcal{O}(\varepsilon_n)/\mathcal{O}(\varepsilon_n)/\mathcal{O}(\sqrt{\varepsilon_n})$，其中$\varepsilon_n = \mathcal{O}\left(\|\boldsymbol{C}_n - \boldsymbol{C}_\infty\|\right) \leq \mathcal{O}(\sqrt{\log n/n})$，适用于次高斯 ω。这与经典谱模板模型的收敛速率 $\mathcal{O}(\sqrt{\log n/n})$一致，例如，SpecT 模型的 [SMMR17] 中的命题 2 和 [NWM+22] 中的定理 2，这表明 LogSpecT 在恢复保证方面也具有竞争力。

5 用线性化交替方向乘子法求解 rLogSpecT

本节设计一种线性化交替方向乘子法算法来求解 rLogSpecT，该算法的子问题均具有闭式解。基于交替方向乘子法的算法已成功应用于解决各种图学习任务 [ZWKP19，WYLS21]。

首先，我们对 rLogSpecT 进行重新表述，使其符合交替方向乘子法框架：

$$\begin{aligned} \min_{\boldsymbol{S},\boldsymbol{Z},\boldsymbol{q}} \ & \langle \boldsymbol{S}, \mathbf{1}\mathbf{1}^\top \rangle - \alpha \mathbf{1}^\top \log \boldsymbol{q} + \iota_{\mathcal{S}}(\boldsymbol{S}) + \iota_{\mathbb{B}(\mathbf{0},\delta_n)}(\boldsymbol{Z}) \\ \text{s.t.} \ & \boldsymbol{C}_n\boldsymbol{S} - \boldsymbol{S}\boldsymbol{C}_n = \boldsymbol{Z}, \quad \boldsymbol{q} = \boldsymbol{S}\mathbf{1}. \end{aligned} \tag{6}$$

问题（6）的增广拉格朗日函数为

$$\begin{aligned} & L(\boldsymbol{S}, \boldsymbol{Z}, \boldsymbol{q}, \boldsymbol{\Lambda}, \boldsymbol{\lambda}_2) \\ = \ & \langle \mathbf{1}\mathbf{1}^\top, \boldsymbol{S} \rangle - \alpha \mathbf{1}^\top \log \boldsymbol{q} + \iota_{\mathcal{S}}(\boldsymbol{S}) + \iota_{\mathbb{B}(\mathbf{0},\delta_n)}(\boldsymbol{Z}) + \langle \boldsymbol{\Lambda}, \boldsymbol{C}_n\boldsymbol{S} - \boldsymbol{S}\boldsymbol{C}_n - \boldsymbol{Z} \rangle + \boldsymbol{\lambda}_2^\top (\boldsymbol{q} - \boldsymbol{S}\mathbf{1}) \\ & + \frac{\rho}{2}\|\boldsymbol{C}_n\boldsymbol{S} - \boldsymbol{S}\boldsymbol{C}_n - \boldsymbol{Z}\|_F^2 + \frac{\rho}{2}\|\boldsymbol{q} - \boldsymbol{S}\mathbf{1}\|_2^2. \end{aligned}$$

由于给定 $\boldsymbol{S}$时，$\boldsymbol{q}$ 和 $\boldsymbol{Z}$ 可分离，因此我们将原始变量分为两块进行更新：$\boldsymbol{S}$ 和（$\boldsymbol{Z}$，$\boldsymbol{q}$）。更具体地说，在第 k次迭代中，$(\boldsymbol{Z}, \boldsymbol{q})$ 的更新为

$$(\boldsymbol{Z}^{(k+1)}, \boldsymbol{q}^{(k+1)}) := \operatorname*{argmin}_{(\boldsymbol{Z},\boldsymbol{q})} L(\boldsymbol{S}^{(k)}, \boldsymbol{Z}, \boldsymbol{q}, \boldsymbol{\Lambda}^{(k)}, \boldsymbol{\lambda}_2^{(k)}), \tag{7}$$

其闭式解如下：

命题 5.1。 更新 (7)可以明确改写为

$$\begin{aligned} \boldsymbol{Z}^{(k+1)} &= \min\left\{1, \frac{\delta_n}{\|\tilde{\boldsymbol{Z}}\|_F}\right\} \tilde{\boldsymbol{Z}}, \\ \boldsymbol{q}^{(k+1)} &= \frac{\tilde{\boldsymbol{q}} + \sqrt{\tilde{\boldsymbol{q}}^2 + 4\alpha/\rho \mathbf{1}}}{2}, \end{aligned}$$

其中 $\tilde{\boldsymbol{Z}} = \boldsymbol{C}_n\boldsymbol{S}^{(k)} - \boldsymbol{S}^{(k)}\boldsymbol{C}_n + \frac{\boldsymbol{\Lambda}^{(k)}}{\rho}$、 $\tilde{\boldsymbol{q}} = \boldsymbol{S}^{(k)}\mathbf{1} - \frac{1}{\rho}\boldsymbol{\lambda}_2^{(k)}$。

$\boldsymbol{S}$ 的 ADMM 更新不具有闭式解。因此，当除 $\boldsymbol{S}$ 以外的所有变量固定时，我们求解增广拉格朗日函数 L 的线性化 ˆ版本 L_k，其形式为

$$\hat{L}_k(\boldsymbol{S}) = \langle \boldsymbol{C}^{(k)}, \boldsymbol{S} \rangle + \frac{\rho\tau}{2}\|\boldsymbol{S} - \boldsymbol{S}^{(k)}\|_F^2 + \iota_{\mathcal{S}}(\boldsymbol{S}),$$

其中 $\boldsymbol{C}^{(k)} := \mathbf{1}\mathbf{1}^\top + \boldsymbol{C}_n\boldsymbol{\Lambda}^{(k)} - \boldsymbol{\Lambda}^{(k)}\boldsymbol{C}_n - \boldsymbol{\lambda}_2^{(k)}\mathbf{1}^\top + \rho(\boldsymbol{S}^{(k)}\boldsymbol{C}_n^2 + \boldsymbol{C}_n^2\boldsymbol{S}^{(k)} - 2\boldsymbol{C}_n\boldsymbol{S}^{(k)}\boldsymbol{C}_n + \boldsymbol{Z}^{(k+1)}\boldsymbol{C}_n - \boldsymbol{C}_n\boldsymbol{Z}^{(k+1)} - \boldsymbol{q}^{(k+1)}\mathbf{1}^\top + \boldsymbol{S}^{(k)}\mathbf{1}\mathbf{1}^\top)$。因此，在第 k 次迭代中， $\boldsymbol{S}$ 的更新为

$$\boldsymbol{S}^{(k+1)} := \operatorname*{argmin}_{\boldsymbol{S}\in\mathcal{S}} \hat{L}_k(\boldsymbol{S}) = \Pi_{\mathcal{S}}(\boldsymbol{S}^{(k)} - \boldsymbol{C}^{(k)}/\rho\tau), \tag{8}$$

其中， $\Pi_{\mathcal{S}}(\boldsymbol{X})$ 是将 $\boldsymbol{X}$ 投影到闭凸集 $\mathcal{S}$ 上的结果，可以通过以下方式计算

$$(\Pi_{\mathcal{S}}(\boldsymbol{X}))_{ij} = \begin{cases} \frac{1}{2}\max\{0, X_{ij} + X_{ji}\}, & \text{if } i \neq j, \\ 0, & \text{if } i = j. \end{cases}$$

最后，我们通过对增广拉格朗日函数执行梯度上升步骤来更新对偶变量 $\boldsymbol{\Lambda}$ 和 $\boldsymbol{\lambda}_2$：

$$\boldsymbol{\Lambda}^{(k+1)} := \boldsymbol{\Lambda}^{(k)} + \rho(\boldsymbol{C}_n\boldsymbol{S}^{(k+1)} - \boldsymbol{S}^{(k+1)}\boldsymbol{C}_n - \boldsymbol{Z}^{(k+1)}), \tag{9}$$

$$\boldsymbol{\lambda}_2^{(k+1)} := \boldsymbol{\lambda}_2^{(k)} + \rho(\boldsymbol{q}^{(k+1)} - \boldsymbol{S}^{(k+1)}\mathbf{1})。 \tag{10}$$

然后，主算法总结如下，见算法 1。

算法1 rLogSpecT 的 L-ADMM

输入：初始化 $\boldsymbol{S}^{(0)}$、 $\boldsymbol{\Lambda}^{(0)}$、 $\boldsymbol{\lambda}_2^{(0)}$，步长 τ， ρ 的更新规则，以及 $k = 0$。输出： $\boldsymbol{S}^{(k)}$1: 当停止准则未满足时，执行以下操作2: 通过 (7) 更新 $\boldsymbol{Z}^{(k+1)}$ 和 $\boldsymbol{q}^{(k+1)}$； 3: 通过 (8) 更新 $\boldsymbol{S}^{(k+1)}$； 4: 通过 (9) 更新 $\boldsymbol{\Lambda}^{(k+1)}$； 5: 通过 (10) 更新 $\boldsymbol{\lambda}_2^{(k+1)}$； 6: 根据更新规则更新 ρ。7: $k = k + 1$; 8: 结束循环

对于 L-ADMM 的收敛性分析，我们将 $(\boldsymbol{Z}、\ \boldsymbol{q})$ 视为一个变量，并且约束 $\boldsymbol{C}_n\boldsymbol{S} - \boldsymbol{S}\boldsymbol{C}_n = \boldsymbol{Z}$ 和 $\boldsymbol{q} = \boldsymbol{S}\mathbf{1}$ 可以写成一个约束：

$$[\boldsymbol{I}\otimes\boldsymbol{C}_n - \boldsymbol{C}_n\otimes\boldsymbol{I}; \boldsymbol{I}\otimes\mathbf{1}^\top]\,\mathrm{vec}(\boldsymbol{S}) = [\mathrm{vec}(\boldsymbol{Z}); \boldsymbol{q}].$$

然后，我们可以通过交替更新 $\boldsymbol{S}$ （该部分的详细内容请参见第 E.4 节）和（$\boldsymbol{Z}、\ \boldsymbol{q}$），将双块近端 ADMM（即 L-ADMM）应用于该问题。因此，可以直接调用 [ZBO11 中的定理 4，2] 从而推导出以下定理。

定理5.2。如果 $\tau > m + \|\boldsymbol{C}_n\otimes\boldsymbol{I}_m - \boldsymbol{I}_m\otimes\boldsymbol{C}_n\|^2$，则

$$\lim_{k\to\infty} f(\boldsymbol{S}^{(k)}) = f_n^*.$$

6 实验

本节将介绍合成网络和真实网络上的实验设置与结果。

6.1 评估指标

为了评估学习得到的图的质量，我们采用以下标准指标（例如，参见[MSR08]）：

$$\text{F-measure} = \frac{2\mathsf{TP}}{2\mathsf{TP} + \mathsf{FN} + \mathsf{FP}},$$
$$\text{Precision} = \frac{\mathsf{TP}}{\mathsf{TP} + \mathsf{FP}}, \quad \text{Recall} = \frac{\mathsf{TP}}{\mathsf{TP} + \mathsf{FN}}.$$

这里，TP、FP 和 FN 分别表示真正例、假正例和假负例的数量。精确率和召回率这两个指标分别衡量：在学习得到的图的所有边中，以及在真实图的所有边中，正确检索出的边所占的比例。需要注意的是，仅其中一个指标取值较高，并不意味着图已被准确学习。这促使我们采用 F值这一指标，它是精确率和召回率的调和平均值。如果学习算法取得较高的 F值，则认为该算法表现良好。

6.2 数据生成

随机图。在实验中，我们考虑两种类型的合成图，即 Erd˝os-R´enyi（ER）图 [ER+60] 和 Barabasi-Albert模型图（BA） [BA99]。ER图通过以概率 $p = 0.2$ 在每对节点之间独立放置一条边来生成，并将每条边的权重设为 1。BA 图的生成方式是：初始时有两个相互连接的节点，然后逐个添加新节点，其中每个新节点恰好连接到一个先前的节点，该节点按照与其当时的度成比例的概率随机选取。

图滤波器。实验中使用了三种图滤波器。第一种是低通图滤波器（lowpass-EXP） $h(\boldsymbol{S}) = \exp(\frac{1}{2}\boldsymbol{S})$。第二种是高通图滤波器（highpass-EXP） $h(\boldsymbol{S}) = \exp(-\boldsymbol{S})$。最后一种是二次图滤波器（QUA） $h(\boldsymbol{S}) = \boldsymbol{S}^2 + \boldsymbol{S} + \boldsymbol{I}$。注意，二次图滤波器既不是低通滤波器，也不是高通滤波器。

图信号。为了生成平稳图信号，我们遵循 [SMMR17]中的生成模型：

$$\boldsymbol{x} = h(\boldsymbol{S})\boldsymbol{w},$$

其中， $\boldsymbol{S}$ 和 $h(\boldsymbol{S})$分别是随机图的邻接矩阵和我们之前介绍的图滤波器， $\boldsymbol{w} \sim \mathcal{N}(0, \boldsymbol{I}_m)$是服从正态分布的随机向量。如前所述， $\boldsymbol{x}$ 是平稳图信号，并且服从次高斯分布。

6.3 如何推断二值图

许多实际任务需要二值图，而非加权图。然而，LogSpect 和 rLogSpecT 得到的结果通常是加权图。本节将探讨如何将加权图转换为二值图。

首先，将图 $\boldsymbol{W}$ 中的每条边除以最大权重进行归一化，从而得到一组介于 0 和 1 之间的数值。接下来，我们需要选择一个阈值 ε，将这些数值舍入为 0 或 1，并据此构建二值图 $\boldsymbol{W}^*$。用数学公式表示，这一过程如下：

$$(W^*)_{ij} = \begin{cases} 1 & \text{if } W_{ij}/\max\{W_{i'j'}\} \geq \varepsilon, \\ 0 & \text{if } W_{ij}/\max\{W_{i'j'}\} < \varepsilon. \end{cases} \tag{11}$$

我们介绍两种选择阈值 ε 的方案：基于训练的策略和基于搜索的策略。

基于训练的。基于训练的策略使用训练数据集和k 图训练模型，并选择 ε^* 以获得最佳平均性能。对于给定的图，我们应用 ε^* 并按照公式（11）将其转换为二值图。

基于搜索的策略。这种策略更直接。对于一个图，我们搜索能够取得最佳性能的 ε^*。与基于训练的策略相比，基于搜索的策略仅适用于能够衡量学习得到的图的性能的情况。尽管存在这一局限，基于搜索的策略通常具有更好的性能。

6.4 合成网络上的实验

为评估 LogSpecT 和 rLogSpecT 的有效性，我们在合成数据上将它们与 SpecT 和 rSpecT 进行比较。我们注意到，实践中通过对最小度施加限制，纳入了不存在孤立节点这一先验。这种方法引入了一个额外的超参数，还可能加剧鲁棒形式中的可行性问题。此外，这一技巧并未提供任何恢复保证。基于这些原因，我们仍将 LogSpecT 和 rLogSpecT 与 SpecT 和 rSpecT 进行比较。

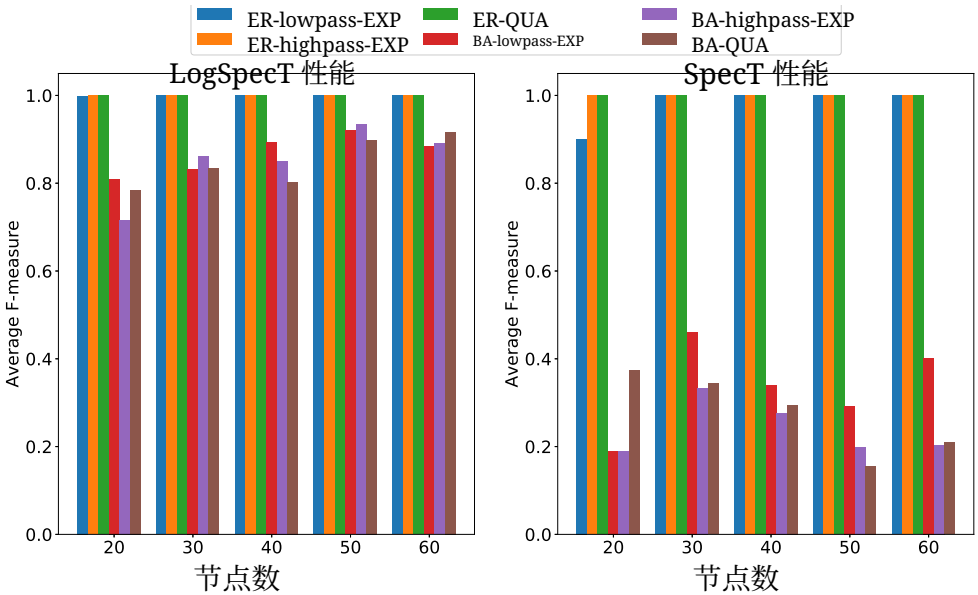

图 1：LogSpecT 与 SpecT 在合成数据上的比较

LogSpecT 的性能。我们首先在两类随机图上，使用三种图滤波器，比较理想模型 LogSpecT 和 SpecT 的性能。我们在节点数为 20 至 60 的图上进行实验，并采用基于搜索的策略来设置阈值ε^*。经过 10 次重复实验后，我们在图 1中报告平均结果。左列展示 LogSpecT 的性能，右列展示 SpecT 的性能。首先，我们观察到，在 ER图 上，两种模型都取得了良好的性能，几乎能够完美恢复图。然而，它们在 BA 图上的表现有所不同。在这种情况下，LogSpecT 能够有效运行。LogSpecT 学习得到的图的 F值远高于 SpecT 学习得到的图。其次，对不同图滤波器的比较表明，对于 LogSpecT 和 SpecT 而言，不同的图滤波器对恢复性能几乎没有影响。最后，我们观察到，随着节点数增加，LogSpecT 在 BA 图上的输出往往具有更高的 F值。这表明 LogSpecT 在较大的网络上可能表现更好。然而，从 SpecT 的结果中观察不到这种现象。

rLogSpecT 的性能。由于 LogSpecT 的表现显著优于 SpecT，因此我们只展示随着样本不断增多，rLogSpecT 的性能。首先，我们通过实验研究 rLogSpecT 的解如何趋近于 LogSpecT。为此，我们利用 QUA 图滤波器，在含有 20 个节点的 ER图上生成平稳图信号。样本量范围从 10^5到 $3*10^5$，且 $\delta_n = \|\boldsymbol{C}_n - \boldsymbol{C}_\infty\|_\infty$。实验结果见图2。图中显示，估计误差 $\|\boldsymbol{C}_n - \boldsymbol{C}\|_\infty$、函数值的相对误差$|f_n - f^*|/f^*$以及度向量的相对误差 $\|\boldsymbol{d} - \boldsymbol{d}^*\|_2/\|\boldsymbol{d}^*\|_2$高度相关。这表明，就函数值和度向量而言，rLogSpecT 与 LogSpecT 之间的相对误差可以通过 $\boldsymbol{C}_\infty$ 的估计误差得到有效控制。

如前所述，LogSpecT 的最优解不一定唯一。因此，我们无法证明 rLogSpecT 的最优解会收敛到 LogSpecT 的最优解。此外，直观而言，非唯一性可能会影响 rLogSpecT 的性能。接下来的实验表明，在实际中，当收集到足够多的样本时，rLogSpecT 学习得到的图往往具有较高的 F值。

在 BA 图上，使用来自全部三种图滤波器的图信号对 rLogSpecT 进行测试。样本量 n 取值范围为 10 至 10^6，并且 δ_n 设为 $0.2\sqrt{\log n/n}$。我们采用基于训练的策略，从 10 个随机选取的训练图中确定最佳阈值 ε^*。随后，我们计算测试集中学习得到的二值图的 F值。结果如图3所示。结果表明，rLogSpecT 适用于 BA 图上的全部三种图滤波器，并且随着收集到的信号越来越多，其性能往往会有所提升。这说明与 LogSpecT 类似，rLogSpecT 对 BA 图上的不同类型图滤波器均有效，包括高通滤波器。更多实验和讨论请参阅附录F。

6.5 真实网络上的实验

在这组实验中，我们将 LogSpecT（分别为 rLogSpecT）与来自统计学和图信号处理领域的 SpecT（分别为 rSpecT）及其他方法进行比较，测试数据集为一个生物图数据库$Protein$ 和一个社交图数据库$Reddit$，来自 [CN21]。我们从数据库中选取节点数少于 50 的图。这最终得到 Protein 数据库中的 871 个测试图和 Reddit 数据库中的 309 个测试图。

rSpecT 的不可行性。对于这些真实网络，我们首先检查 rSpecT 遇到的不可行性问题是否严重。为此，我们采用随机图滤波器

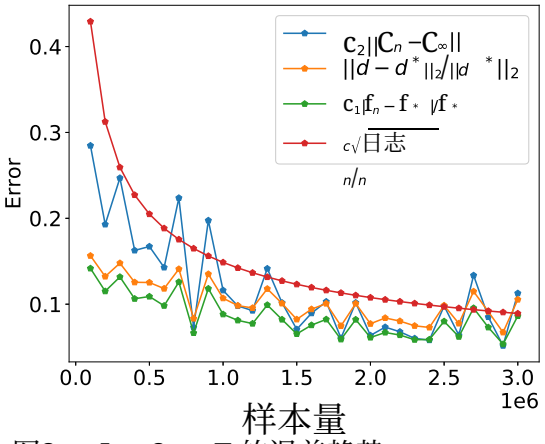

图2：rLogSpecT 的误差趋势。

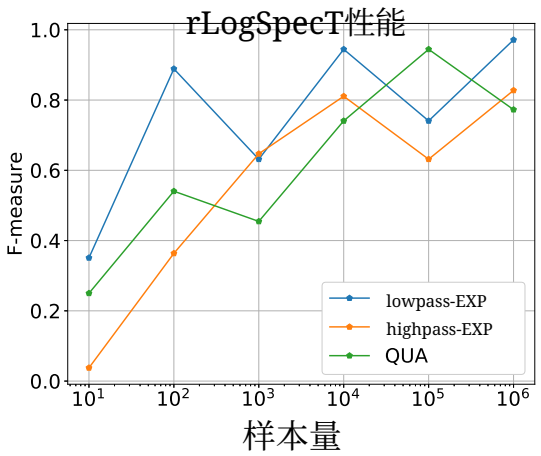

图 3：使用 $\delta_n = 0.2\sqrt{\log n/n}$ 的 rLogSpecT。

$t_1\boldsymbol{S}^2 + t_2\boldsymbol{S} + t_3\boldsymbol{I}$，其中 t_i，$i = 1$，2，3 是服从高斯分布的随机变量，其参数为$\mu = 0$ 和 $\sigma = 2$。随后，我们计算使 rSpecT 可行的最小 δ_n。如果最小的$\delta_n > 0$，则 rSpecT 可能会遇到不可行性问题。观测到的图信号数量不同时，结果如表 1所示。

表 1：不可行性的可能性

Protein			
样本量	10	100	1000
频率	0.980	0.999	0.994
$\delta_{\min}$的均值	38.514	26.012	14.813
Reddit			
样本量	10	100	1000
频率	0.984	1	1
$\delta_{\min}$的均值	1094.788	830.642	531.006

实验结果表明，在我们使用的真实数据集中，很有可能会遇到不可行性问题，并且增加样本量并不能降低这一概率。然而，我们确实观察到 $\delta_{\min}$的均值有所下降；这是合理的，因为当样本量趋于无穷时，SpecT 应当是可行的。不可行情形的频繁出现，使得在为 rSpecT 选择 δ 时必须予以谨慎考虑。一种常用方法是将$\delta = \delta_{\min}$ [SMMR17]设定为该值。然而，这种方法的计算成本较高，因为它需要在第一阶段精确求解二阶锥规划（SOCP），以获得 $\delta_{\min}$，并且缺乏非渐近分析。

不同图学习模型的性能。平稳图信号由低通滤波器 $h(\boldsymbol{S}) = \exp(\boldsymbol{S})$在这些图上生成。我们选择著名的统计方法（阈值化）相关性 [Kol09] 以及第一个将对数障碍应用于图推断的图信号处理方法 [Kal16] 作为基线。相关性方法的最优阈值从0.1到0.6中选取，并通过搜索 $\{0.01, 0.1, 1, 10, 100, 1000\}$ 以获得Kalofolias模型中的最佳超参数。rLogSpecT中的参数 δ_n 设为 $10\sqrt{\log n/n}$ 和

在rSpecT中，该参数设为能够得到可行解的最小值 [SMMR17]。我们还采用基于搜索的策略，将LogSpecT和SpecT学习得到的图转换为二值图。

真实协方差矩阵下理想模型的结果 C_∞ 汇总于图4中。我们观察到，在真实图上，LogSpecT 的平均表现最佳（红线表示中位数）。此外，与 SpecT 相比，LogSpecT 的表现更加稳定。需要指出的是，由于图信号不一定平滑，Kalofolias模型无法保证性能，尤其是在Reddit网络上。

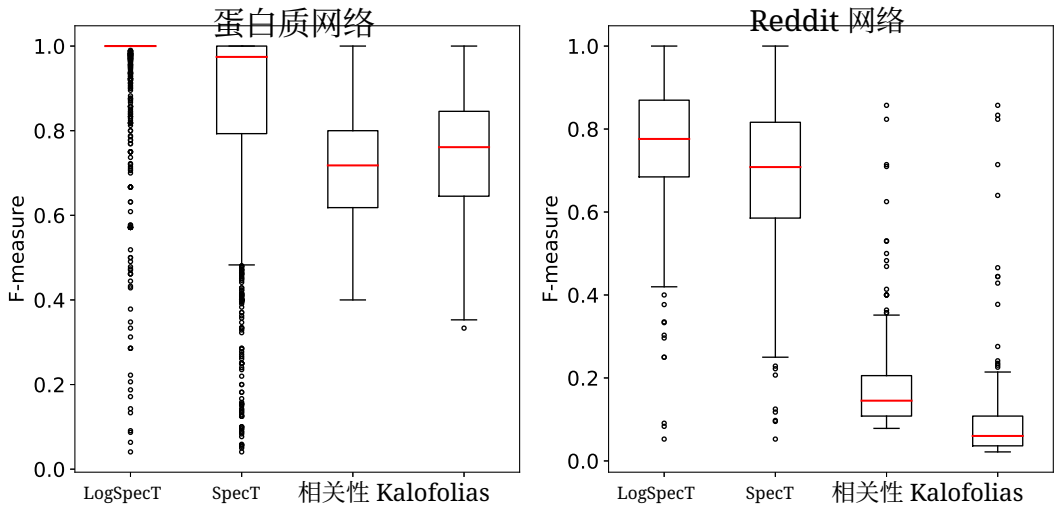

图4：LogSpecT 和 SpecT 的性能。

图5 比较了在观测到不同数量的信号时四种不同方法的性能。[2] 随着样本量增加，侧重于平稳性的模型能够更准确地恢复图，而相关性方法和 Kalofolias 的方法则无法利用信号信息。从图4 中的实验结果也可以推断出这一点，因为这些模型在拥有完整信息时都无法取得良好性能，更不用说只有部分信息了。实验还表明，在观测到固定数量的样本时，rLogSpecT 学习得到的图比 rSpecT 更接近真实值。这进一步证实了 rLogSpecT 在利用平稳信号进行图学习方面的优越性。

7 结论

本文针对基于平稳信号进行图学习的最基础模型，首次提出了不可行性条件 [SMMR17] ，并提出了一种高效的替代方案。随后，我们借助先进的优化工具分析了有限样本情形下鲁棒形式的恢复保证，这些工具或可在更广泛的学习任务中得到应用。与现有文献 [NWM+22]相比，这些理论结果所需的条件更宽松。我们还设计了

[2]中的模型[SM20] 是 rSpecT 的替代模型，用于近似 SpecT，且其超参数难以确定。因此，我们省略了该模型的性能。

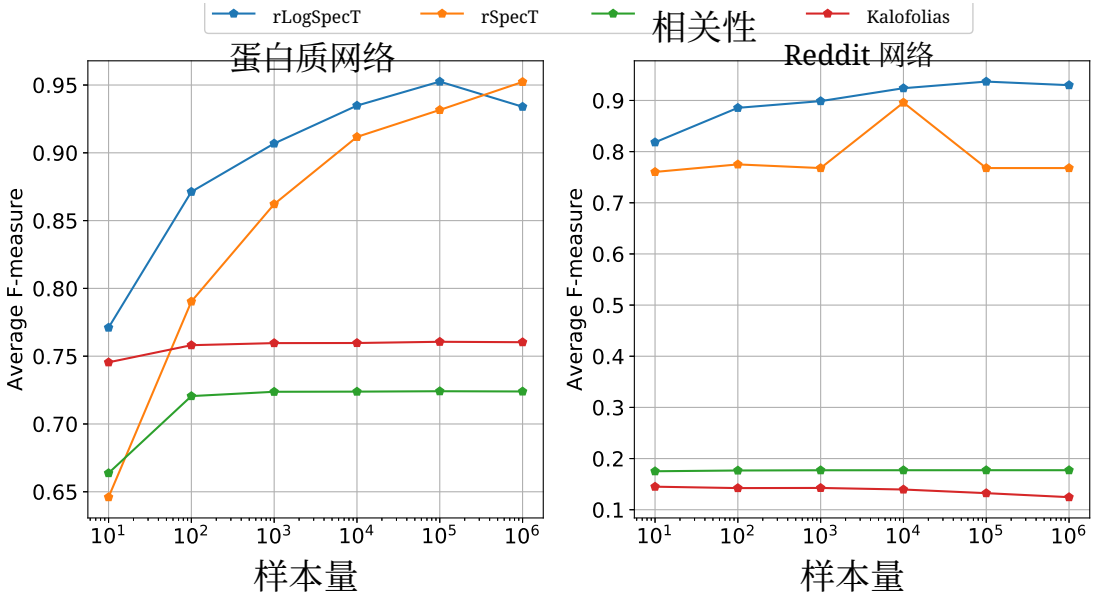

图5：rLogSpecT 和 rSpecT 的渐近性能。

一种 L-ADMM 算法，可实现高效运行并保证理论收敛。本文针对合成数据和真实数据进行了大量实验。结果表明，我们提出的模型显著优于现有模型（例如，在 BA 图上的实验）。

我们认为，这项工作代表着在基础模型之上的重要进展。其通用形式使当前针对 SpecT 的扩展能够得到迁移。使用这些扩展测试我们提出的模型，是未来的一个研究方向。此外，我们注意到，尽管鲁棒形式的恢复保证已很明确，但理想情况模型（即 SpecT 和 LogSpecT）的估计性能分析仍不完整。探究精确恢复条件是另一个未来的研究方向。

参考文献

[AOTS15] Daron Acemoglu、Asuman Ozdaglar 和 Alireza Tahbaz-Salehi。金融网络中的系统性风险与稳定性。《美国经济评论》，105(2):564–608，2015年。[BA99] Albert-L´aszl´o Barab´asi 和 R´eka Albert。随机网络中尺度律的涌现。《科学》，286(5439):509–512，1999年。[BPC+11] Stephen Boyd、Neal Parikh、Eric Chu、Borja Peleato、Jonathan Eckstein 等。通过交替方向乘子法实现分布式优化与统计学习。《机器学习基础与趋势》®，3(1):1–122，2011年。[BRCM19] Andrei Buciulea、Samuel Rey、Cristobal Cabrera 和 Antonio G Marques。利用含隐变量的图平稳信号重建网络。载于第53届阿西洛马会议：信号、系统与计算论文集，第56–60页。IEEE，2019年。

[BRM22] Andrei Buciulea、Samuel Rey 和 Antonio G Marques。利用含隐变量的平滑信号和平稳图信号学习图。IEEE网络信号与信息处理汇刊，8:273–287，2022年。[CN21] Samir Chowdhury 和 Tom Needham。通过 Gromov-Wasserstein 学习实现广义谱聚类。载于第24届国际人工智能与统计学会议论文集（AISTATS 2021），第712–720页。PMLR，2021年。[dMCYP21] Jose Vinicius de Miranda Cardoso、Jiaxi Ying 和 Daniel Palomar。厚尾市场中的图模型。《神经信息处理系统进展》，34:19989–20001，2021年。[ER+60] Paul Erd˝os、Alfr´ed R´enyi 等。随机图的演化。匈牙利科学院数学研究所出版物，5(1):17–60，1960年。[Gir15] Benjamin Girault。利用等距图平移构造平稳图信号。载于第23届欧洲信号处理会议论文集（EUSIPCO 2015），第1516–1520页。IEEE，2015年。[Hof52] Alan J Hoffman。关于线性不等式组的近似解。《美国国家标准局研究杂志》，49(4):263–265，1952年。[HW22] Yiran He 和 Hoi-To Wai。通过结构化因子分析，从低秩激励图信号中检测中心节点。IEEE信号处理汇刊，70:2416–2430，2022年。[JHIM+19] Alexander Jung、Alfred O Hero III、Alexandru Cristian Mara、Saeed Jahromi、Ayelet Heimowitz 和 Yonina C Eldar。通过总变差最小化对网络结构数据进行半监督学习。IEEE信号处理汇刊，67(24):6256–6269，2019年。[JT19] Alexander Jung 和 Nguyen Tran。网络数据中的局部线性回归。IEEE信号处理快报，26(7):1090–1094，2019年。[Kal16] Vassilis Kalofolias。如何从平滑信号中学习图。载于第19届国际人工智能与统计学会议论文集（AISTATS 2016），第920–929页。PMLR，2016年。[KLTF17] Vassilis Kalofolias、Andreas Loukas、Dorina Thanou 和 Pascal Frossard。学习时变图。载于2017年IEEE国际声学、语音与信号处理会议论文集（ICASSP 2017），第2826–2830页。IEEE，2017年。[Kol09] Eric D. Kolaczyk。《网络数据的统计分析：方法与模型》。施普林格出版社，纽约州纽约市，2009年。[MF20] Hermina Petric Maretic 和 Pascal Frossard。图拉普拉斯混合模型。IEEE网络信号与信息处理汇刊，6:261–270，2020年。

[MNBD21] Gautier Marti、Frank Nielsen、Miko laj Bi´nkowski 和 Philippe Donnat。金融市场中相关性、层级结构、网络与聚类二十年发展的综述。《信息几何进展》，第245–274页，2021年。[MSLR17] Antonio G Marques、Santiago Segarra、Geert Leus 和 Alejandro Ribeiro。平稳图过程与谱估计。《IEEE信号处理汇刊》，65(22):5911–5926，2017年。[MSMR19] Gonzalo Mateos、Santiago Segarra、Antonio G Marques 和 Alejandro Ribeiro。连点成网：通过图信号处理识别网络结构。《IEEE信号处理杂志》，36(3):16–43，2019年。[MSR08] D. Christopher Manning、Hinrich Sch¨utze 和 Prabhakar Raghavan。《信息检索导论》。剑桥大学出版社，2008年。[NWM+22] Madeline Navarro、Yuhao Wang、Antonio G Marques、Caroline Uhler 和 Santiago Segarra。基于矩阵多项式的多个图联合推断。《机器学习研究杂志》，23(76):1–35，2022年。[PGM+17] Bastien Pasdeloup、Vincent Gripon、Gr´egoire Mercier、Dominique Pastor 和 Michael G Rabbat。基于平稳信号观测的图扩散过程刻画与推断。《IEEE网络信号与信息处理汇刊》，4(3):481–496，2017年。[PV17] Nathana¨el Perraudin 和 Pierre Vandergheynst。图上的平稳信号处理。《IEEE信号处理汇刊》，65(13):3462–3477，2017年。[RBN+22] Samuel Rey、Andrei Buciulea、Madeline Navarro、Santiago Segarra 和 Antonio G Marques。基于平稳图信号对含隐变量的多个图进行联合推断。载于《2022年IEEE声学、语音与信号处理国际会议（ICASSP 2022）论文集》，第5817–5821页。IEEE，2022年。[RW09] R Tyrrell Rockafellar 和 Roger J-B Wets。《变分分析》，第317卷。施普林格科学与商业媒体，2009年。[RW21] Johannes O Royset 和 Roger JB Wets。《优化入门》。施普林格，2021年。[RWS20] Raksha Ramakrishna、Hoi-To Wai 和 Anna Scaglione。低通图信号处理及其应用用户指南：工具与应用。《IEEE信号处理杂志》，37(6):74–85，2020年。[SCML18] Santiago Segarra、Sundeep Prabhakar Chepuri、Antonio G Marques 和 Geert Leus。统计图信号处理：平稳性与谱估计。《协同与图信号处理》，第325–347页，2018年。[SHMV19] Rasoul Shafipour、Abolfazl Hashemi、Gonzalo Mateos 和 Haris Vikalo。基于流式平稳图信号的在线拓扑推断。载于《2019年IEEE数据科学研讨会（DSW）》，第140–144页。IEEE，2019年。

[SM20] Rasoul Shafipour 和 Gonzalo Mateos。基于具有部分连通性信息的流式平稳图信号进行在线拓扑推断。《算法》，13(9):228，2020年。[SMMR17] Santiago Segarra、Antonio G Marques、Gonzalo Mateos 和 Alejandro Ribeiro。基于谱模板的网络拓扑推断。《IEEE网络信号与信息处理汇刊》，3(3):467–483，2017年。[SMSK+11] Stephen M Smith、Karla L Miller、Gholamreza Salimi-Khorshidi、Matthew Web-ster、Christian F Beckmann、Thomas E Nichols、Joseph D Ramsey 和 Mark W Woolrich。功能磁共振成像的网络建模方法。《神经影像》，54(2):875–891，2011年。[STM15] Oliver Stegle、Sarah A Teichmann 和 John C Marioni。单细胞转录组学中的计算与分析挑战。《自然评论：遗传学》，16(3):133–145，2015年。[SWUM17] Santiago Segarra、Yuhao Wang、Caroline Uhler 和 Antonio G Marques。基于平稳图信号的网络联合推断。载于《第51届阿西洛马会议：信号、系统与计算论文集》，第975–979页。IEEE，2017年。[TE20] Yuichi Tanaka 和 Yonina C Eldar。具有子空间与平滑先验的图上广义采样。《IEEE信号处理汇刊》，68:2272–2286，2020年。[TEOC20] Yuichi Tanaka、Yonina C Eldar、Antonio Ortega 和 Gene Cheung。图上信号采样：从理论到应用。《IEEE信号处理杂志》，37(6):14–30，2020年。[Ver12] Roman Vershynin。样本协方差矩阵与实际协方差矩阵有多接近？《理论概率学杂志》，25(3):655–686，2012年。[Ver18] Roman Vershynin。《高维概率：简介及其在数据科学中的应用》，第47卷。剑桥大学出版社，2018年。[WYLS21] Xiaolu Wang、Chaorui Yao、Haoyu Lei 和 Anthony Man-Cho So。基于平滑信号进行图学习的高效交替方向方法。载于《2021年IEEE声学、语音与信号处理国际会议（ICASSP 2021）论文集》，第5380–5384页。IEEE，2021年。[ZBO11] Xiaoqun Zhang、Martin Burger 和 Stanley Osher。基于Bregman迭代的统一原始–对偶算法框架。《科学计算杂志》，46(1):20–46，2011年。[ZJSL17] Zhaowei Zhu、Shengda Jin、Xuming Song 和 Xiliang Luo。通过一阶近似利用平稳图信号学习图结构。载于《第22届数字信号处理国际会议（DSP 2017）论文集》，第1–5页。IEEE，2017年。

[ZWKP19] Licheng Zhao、Yiwei Wang、Sandeep Kumar 和 Daniel P Palomar。通过交替方向乘子法和最大化最小化进行图拉普拉斯矩阵估计的优化算法。《IEEE信号处理汇刊》，67(16):4231–4244，2019年。[ZYY16] Hui Zhang、Ming Yan 和 Wotao Yin。l1合成最小化和l1分析最小化解的唯一性与鲁棒性的一项条件。《计算数学进展》，42(6):1381–1399，2016年。[ZZL+22] Shuai Zheng、Zhenfeng Zhu、Zhizhe Liu、Zhenyu Guo、Yang Liu、Yuchen Yang 和 Yao Zhao。用于疾病预测的多模态图学习。《IEEE医学影像汇刊》，41(9):2207–2216，2022年。

附录

附录包含缺失的证明、对正文中某些论证的详细讨论，以及更多数值实验。

定理3.2 的证明

由于线性系统（4）无解，根据法卡斯引理，我们知道以下系统有解：

$$\begin{cases} \begin{bmatrix} \boldsymbol{I}_{m-1} & \boldsymbol{0}_{\frac{(m-1)(m-2)}{2}} \end{bmatrix} \boldsymbol{B}^\top \boldsymbol{A}_n^\top \boldsymbol{x} < \boldsymbol{0}_{(m-1)\times 1}, \\ \begin{bmatrix} \boldsymbol{0}_{\frac{(m-1)(m-2)}{2}\times(m-1)} & \boldsymbol{I}_{\frac{(m-1)(m-2)}{2}} \end{bmatrix} \boldsymbol{B}^\top \boldsymbol{A}_n^\top \boldsymbol{x} \leq \boldsymbol{0}_{\frac{(m-1)(m-2)}{2}\times 1}. \end{cases} \tag{12}$$

令 $\boldsymbol{x}^* \in \mathbb{R}^{m^2}$ 为（12）的一个解。记 $\boldsymbol{x}_+ :=$ 最大值$\{\boldsymbol{x}^*, \boldsymbol{0}\}$，$\boldsymbol{x}_- :=$ 最大值$\{-\boldsymbol{x}^*, \boldsymbol{0}\}$。那么，存在 $c \in (0, 1]$ 使得

$$\boldsymbol{B}^\top \boldsymbol{A}_n^\top (\boldsymbol{x}_+ - \boldsymbol{x}_-) + c\boldsymbol{1}_{m^2}^\top (\boldsymbol{x}_+ + \boldsymbol{x}_-)[\boldsymbol{1}_{m-1}; \boldsymbol{0}_{\frac{(m-1)(m-2)}{2}}] \leq \boldsymbol{0}.$$

定义 $y := -\boldsymbol{1}_{m^2}^\top (\boldsymbol{x}_+ + \boldsymbol{x}_-)$，$z := c\boldsymbol{1}_{m^2}^\top (\boldsymbol{x}_+ + \boldsymbol{x}_-)$，并令 $\bar{\delta} = c$。对于所有 $\delta \in [0, \bar{\delta})$，$(\boldsymbol{x}_+, \boldsymbol{x}_-, y, z)$ 都是以下线性系统的一个解：

$$\begin{cases} \boldsymbol{B}^\top \boldsymbol{A}_n^\top (\boldsymbol{x}_+ - \boldsymbol{x}_-) + z[\boldsymbol{1}_{m-1}; \boldsymbol{0}_{\frac{(m-1)(m-2)}{2}}] \leq \boldsymbol{0}, \\ \boldsymbol{1}_{m^2}^\top (\boldsymbol{x}_+ + \boldsymbol{x}_-) + y \leq 0, \\ \delta y + z > 0, \\ \boldsymbol{x}_+, \boldsymbol{x}_-, -y \geq \boldsymbol{0}. \end{cases}$$

再次，根据法卡斯引理 根据 rkas 引理，这意味着以下线性系统无解：

$$\begin{cases} \boldsymbol{A}_n \boldsymbol{B} \boldsymbol{s} + t\boldsymbol{1}_{m^2} \geq \boldsymbol{0}, \\ \boldsymbol{A}_n \boldsymbol{B} \boldsymbol{s} - t\boldsymbol{1}_{m^2} \leq \boldsymbol{0}, \\ t \leq \delta, \\ \begin{bmatrix} \boldsymbol{1}_{m-1} & \boldsymbol{0}_{\frac{(m-1)(m-2)}{2}} \end{bmatrix} \boldsymbol{s} = 1, \end{cases} \tag{13}$$

其中 $\boldsymbol{s} \in \mathbb{R}^{m(m-1)/2}$ 且 $t \in \mathbb{R}$。由于（13）等价于：

$$\begin{cases} \|\boldsymbol{C}_n \boldsymbol{S} - \boldsymbol{S}\boldsymbol{C}_n\|_{\infty,\infty} \leq \delta, \\ (\boldsymbol{S}\boldsymbol{1})_1 = 1, \\ \boldsymbol{S} \in \mathcal{S}, \end{cases} \tag{14}$$

上述论证表明（14）无解。假设 rSpecT 存在一个可行解 $\boldsymbol{S}'$，则

$$\|\boldsymbol{C}_n \boldsymbol{S}' - \boldsymbol{S}' \boldsymbol{C}_n\|_{\infty,\infty} \leq \|\boldsymbol{C}_n \boldsymbol{S}' - \boldsymbol{S}' \boldsymbol{C}_n\|_F \leq \delta.$$

因此，$\boldsymbol{S}'$ 也是（14）的一个解。然而，（14）没有解。由此可得，在这种情况下，rSpecT 不可行。

B 定理中充分条件的解释 3.2

我们进一步阐述$\boldsymbol{A}_n\boldsymbol{B}$列满秩时的不可行性条件。该条件的一个应用是例3.1。具体而言，我们知道在这种情况下,

$$\boldsymbol{B}=\begin{pmatrix}0\\1\\1\\0\end{pmatrix}\quad\text{and}\quad\boldsymbol{A}_n=\begin{pmatrix}0&h_{12}&-h_{12}&0\\h_{12}&h_{22}-h_{11}&0&-h_{12}\\-h_{12}&0&h_{11}-h_{22}&h_{12}\\0&-h_{12}&h_{12}&0\end{pmatrix}.$$

这意味着

$$\boldsymbol{A}_n\boldsymbol{B}=\begin{pmatrix}0\\h_{22}-h_{11}\\h_{11}-h_{22}\\0\end{pmatrix}.$$

因此，当 $h_{11}\neq h_{22}$， $\boldsymbol{A}_n\boldsymbol{B}$ 列满秩。这意味着，当 δ 足够小时（由示例 3.1 可知 $\tilde{\delta}=\sqrt{2}|h_{11}-h_{22}|$），模型 rSpecT 不可行。

C (r)LogSpecT 的性质证明

C.1 命题 3.4 的证明

由于常数　　由于约束集 $\mathcal{S}$ 是一个锥，因此对所有 $\gamma>0$、$\gamma\mathcal{S}=\mathcal{S}$ 都成立。于是我们知道

$$\begin{aligned}\mathrm{Opt}(\boldsymbol{C},\alpha)&=\operatorname*{argmin}_{\boldsymbol{S}\in\mathcal{S},\boldsymbol{C}\boldsymbol{S}=\boldsymbol{S}\boldsymbol{C}}\|\boldsymbol{S}\|_{1,1}-\alpha\mathbf{1}^\top\log(\boldsymbol{S}\mathbf{1})\\&=\gamma\cdot\operatorname*{argmin}_{\gamma\boldsymbol{S}\in\mathcal{S},\boldsymbol{C}\gamma\boldsymbol{S}=\gamma\boldsymbol{S}\boldsymbol{C}}\|\gamma\boldsymbol{S}\|_{1,1}-\alpha\mathbf{1}^\top\log(\gamma\boldsymbol{S}\mathbf{1})\\&=\gamma\cdot\operatorname*{argmin}_{\boldsymbol{S}\in\frac{1}{\gamma}\mathcal{S},\boldsymbol{C}\boldsymbol{S}=\boldsymbol{S}\boldsymbol{C}}\gamma\|\boldsymbol{S}\|_{1,1}-\alpha\mathbf{1}^\top\log(\boldsymbol{S}\mathbf{1})\\&=\gamma\cdot\operatorname*{argmin}_{\boldsymbol{S}\in\mathcal{S},\boldsymbol{C}\boldsymbol{S}=\boldsymbol{S}\boldsymbol{C}}\|\boldsymbol{S}\|_{1,1}-\frac{\alpha}{\gamma}\mathbf{1}^\top\log(\boldsymbol{S}\mathbf{1})\\&=\gamma\,\mathrm{Opt}(\boldsymbol{C},\alpha/\gamma),\end{aligned}$$

其中第三个等式来自对数函数的基本微积分法则。令 $\gamma=\alpha$，则 $\mathrm{Opt}(\boldsymbol{C},\alpha)=\alpha\,\mathrm{Opt}(\boldsymbol{C},1)$，证毕。

C.2 命题证明 3.6

证明将通过为 rLogSpecT 构造一个可行解来进行。回顾一下，$\boldsymbol{A}_n=\boldsymbol{I}\otimes\boldsymbol{C}_n-\boldsymbol{C}_n\otimes\boldsymbol{I}$ 以及矩阵 $\boldsymbol{B}\in\mathbb{R}^{m^2\times m_{(m-1)/2}}$ 可将非负向量映射为有效邻接矩阵的向量化形式。令 $\boldsymbol{S}=\min\{\frac{\delta}{\|\boldsymbol{A}_n\boldsymbol{B}\boldsymbol{s}\|_2},1\}\cdot\mathrm{mat}(\boldsymbol{B}\boldsymbol{s})$，其中$\boldsymbol{s}\in\mathbb{R}^{(m-1)m/2}$ 为非负向量，mat(·) 为矩阵化算子。注意,

$$\mathrm{vec}(\boldsymbol{C}_n\boldsymbol{S}-\boldsymbol{S}\boldsymbol{C}_n)=(\boldsymbol{I}\otimes\boldsymbol{C}_n-\boldsymbol{C}_n\otimes\boldsymbol{I})\,\mathrm{vec}(\boldsymbol{S})=\boldsymbol{A}_n\mathrm{vec}(\boldsymbol{S}).$$

然后，我们知道，

$$\|\boldsymbol{C}_n\boldsymbol{S}-\boldsymbol{S}\boldsymbol{C}_n\|_F=\|\mathrm{vec}(\boldsymbol{C}_n\boldsymbol{S}-\boldsymbol{S}\boldsymbol{C}_n)\|_2=\min\left\{\frac{\delta}{\|\boldsymbol{A}_n\boldsymbol{B}\boldsymbol{s}\|_2},1\right\}\cdot\|\boldsymbol{A}_n\boldsymbol{B}\boldsymbol{s}\|_2\le\delta.$$

T因此，给定的 $\boldsymbol{S}$ 是 rLogSpecT 的可行解，由此完成证明 .

C.3 命题 3.7 的证明

对于第一个结论，我们来考察 LogSpecT 和 rLogSpecT 的 Karush-Kuhn-Tucker（KKT）条件。由于 LogSpecT 是一个凸问题，且满足 Slater 条件，因此 KKT 条件是最优性的充要条件，即存在 $(\boldsymbol{\Lambda}_1,\boldsymbol{\Lambda}_2)\in\mathbb{R}^{m\times m}\times\mathcal{N}_{\mathcal{S}}(\boldsymbol{S}^*)$，使得

$$\begin{cases}\nabla_{\boldsymbol{S}}(\|\boldsymbol{S}^*\|_{1,1}-\alpha\mathbf{1}^\top\log(\boldsymbol{S}^*\mathbf{1}))+\boldsymbol{C}_\infty\boldsymbol{\Lambda}_1-\boldsymbol{\Lambda}_1\boldsymbol{C}_\infty+\boldsymbol{\Lambda}_2=\mathbf{0},\\ \boldsymbol{C}_\infty\boldsymbol{S}^*=\boldsymbol{S}^*\boldsymbol{C}_\infty,\\ \boldsymbol{S}^*\in\mathcal{S},\end{cases}\tag{15}$$

其中 $\mathcal{N}_{\mathcal{S}}(\boldsymbol{S}^*):=\{\boldsymbol{N}\in\mathbb{R}^{m\times m}:\sup_{\boldsymbol{X}\in\mathcal{S}}\langle\boldsymbol{X}-\boldsymbol{S}^*,\boldsymbol{N}\rangle\le0\}$ 是 $\mathcal{S}$ 在 $\boldsymbol{S}^*$ 处的法锥，并且 $\nabla\|\boldsymbol{S}^*\|_{1,1}$ 是良定义的，因为 $\|\cdot\|_{1,1}=\langle\cdot,\mathbf{1}\mathbf{1}^\top\rangle$ 在 $\boldsymbol{S}^*\ge0$ 处可微。进一步计算可得

$$\nabla\|\boldsymbol{S}^*\|_{1,1}=\mathbf{1}\mathbf{1}^\top,\quad(\nabla_{\boldsymbol{S}}\mathbf{1}^\top\log(\boldsymbol{S}^*\mathbf{1}))_{ij}=\frac{1}{(\boldsymbol{S}^*\mathbf{1})_i}.$$

将其与（15）结合，并对两边与 $\boldsymbol{S}^*$ 取内积，可得

$$\sum_{i,j}(\boldsymbol{S}^*)_{ij}-\alpha\sum_{i,j}\frac{(\boldsymbol{S}^*)_{ij}}{(\boldsymbol{S}^*\mathbf{1})_i}+\langle\boldsymbol{\Lambda}_1,\boldsymbol{C}_\infty\boldsymbol{S}^*-\boldsymbol{S}^*\boldsymbol{C}_\infty\rangle+\langle\boldsymbol{\Lambda}_2,\boldsymbol{S}^*\rangle=0.\tag{16}$$

根据 $\mathcal{S}$ 的结构以及 $\Lambda_2\in\mathcal{N}_{\mathcal{S}}(\boldsymbol{S}^*)$ 这一事实，可知 $\langle\boldsymbol{\Lambda}_2,\ \boldsymbol{S}^*\rangle=0$。此外，注意到 $\boldsymbol{C}_\infty\boldsymbol{S}^*=\boldsymbol{S}^*\boldsymbol{C}_\infty$。因此，方程（16）可化简为所需结果：

$$\|\boldsymbol{S}^*\|_{1,1}=\sum_{i,j}(\boldsymbol{S}^*)_{ij}=\alpha\sum_{i,j}\frac{(\boldsymbol{S}^*)_{ij}}{(\boldsymbol{S}^*\mathbf{1})_i}=\alpha\sum_{i=1}^m\sum_{j=1}^m\frac{(\boldsymbol{S}^*)_{ij}}{(\boldsymbol{S}^*\mathbf{1})_i}=\alpha m.$$

rLogSpecT 的 KKT 条件表明，存在 $\lambda_1\ge0$、$\boldsymbol{\Lambda}_2\in\mathcal{N}_{\mathcal{S}}(\boldsymbol{S}_n^*)$ 和 $\boldsymbol{Q}\in\partial\|\boldsymbol{C}_n\boldsymbol{S}_n^*-\boldsymbol{S}_n^*\boldsymbol{C}_n\|_F$（即函数 $\boldsymbol{S}\mapsto\|\boldsymbol{C}_n\boldsymbol{S}-\boldsymbol{S}\boldsymbol{C}_n\|_F$ 在 $\boldsymbol{S}_n^*$ 处的次梯度），使得

$$\begin{cases}\nabla_{\boldsymbol{S}}(\|\boldsymbol{S}_n^*\|_{1,1}-\alpha\mathbf{1}^\top\log(\boldsymbol{S}_n^*\mathbf{1}))+\lambda_1\boldsymbol{Q}+\boldsymbol{\Lambda}_2=\mathbf{0},\\ \lambda_1(\|\boldsymbol{C}_n\boldsymbol{S}_n^*-\boldsymbol{S}_n^*\boldsymbol{C}_n\|_F-\delta_n)=0,\\ \boldsymbol{S}_n^*\in\mathcal{S}.\end{cases}\tag{17}$$

此外，根据凸次微分的定义，我们知道 $0\ge\|\boldsymbol{C}_n\boldsymbol{S}_n^*-\boldsymbol{S}_n^*\boldsymbol{C}_n\|_F-\langle\boldsymbol{Q},\ \boldsymbol{S}_n^*\rangle$。因此，对等式（17）两边与 $\boldsymbol{S}_n^*$ 作内积，

可得:

$$\begin{aligned}0 &= \sum_{i,j}(\boldsymbol{S}_n^*)_{ij} - \alpha m + \lambda_1\langle \boldsymbol{Q}, \boldsymbol{S}_n^*\rangle + \langle \boldsymbol{\Lambda}_2, \boldsymbol{S}_n^*\rangle \\ &\geq \sum_{i,j}(\boldsymbol{S}_n^*)_{ij} - \alpha m + \lambda_1\|\boldsymbol{C}_n\boldsymbol{S}_n^* - \boldsymbol{S}_n^*\boldsymbol{C}_n\|_F + \langle \boldsymbol{\Lambda}_2, \boldsymbol{S}_n^*\rangle \\ &= \sum_{i,j}(\boldsymbol{S}_n^*)_{ij} - \alpha m + \lambda_1\delta_n,\end{aligned}$$

这意味着 $\sum_{i,\ j}(\boldsymbol{S}_n^*)_{ij} \leq \alpha m - \lambda_1\delta_n \leq \alpha m$。第一条结论得证。

对于第二条结论，我们首先证明 v_n^*和 v^*大于 $\alpha m(1-\log\ \alpha)$。定义辅助函数 $g:\ \mathbb{R}\to\mathbb{R}$，使得对于任意 $x\in\mathbb{R}_+$，$g(x) := x - \alpha\log\ x$，其最小值在 α处取得。由于对于任意 $\boldsymbol{S}\in\mathcal{S}$,

$$f(\boldsymbol{S}) = \sum_{i=1}^{m} g\left(\sum_{j=1}^{m} S_{ij}\right),$$

其中 f 是 LogSpecT 中的目标函数，因此

$$f(\boldsymbol{S}) \geq \sum_{i=1}^{m} g(\alpha) = \alpha m(1-\log\alpha).$$

这意味着 v_n^*和 v^*大于 $\alpha m(1-\log\ \alpha)$。接下来，我们将证明 $v_n^* \leq v^*$。任取 LogSpecT 的一个最优解 $\boldsymbol{S}^*$。我们证明它对 rLogSpecT 是可行的。

$$\begin{aligned}\|\boldsymbol{C}_n\boldsymbol{S}^* - \boldsymbol{S}^*\boldsymbol{C}_n\| &= \|\boldsymbol{C}_n\boldsymbol{S}^* - \boldsymbol{C}_\infty\boldsymbol{S}^* + \boldsymbol{S}^*\boldsymbol{C}_\infty - \boldsymbol{S}^*\boldsymbol{C}_n\| \\ &\leq 2\|\boldsymbol{S}^*\|_{1,1}\|\boldsymbol{C}_n - \boldsymbol{C}_\infty\| \leq 2\alpha m\|\boldsymbol{C}_n - \boldsymbol{C}_\infty\| \leq \delta_n,\end{aligned}$$

其中，等号是由 $\boldsymbol{C}_\infty\boldsymbol{S}^* = \boldsymbol{S}^*\boldsymbol{C}_\infty$ 得出的，第一个不等式是因为$\|\boldsymbol{X}\boldsymbol{Y}\| \leq \|\boldsymbol{X}\|_F\|\boldsymbol{Y}\| \leq \|\boldsymbol{X}\|_{1,1}\|\boldsymbol{Y}\|$，第二个不等式由第一条结论得出，最后一个不等式则是由于 $\delta_n \geq 2\alpha m\|\boldsymbol{C}_n - \boldsymbol{C}_\infty\|$。因此，$\boldsymbol{S}^*$ 对 rLogSpecT 是可行的，这表明$v_n^* \leq v^*$。证明完毕。

定理证明4.1 &推论 4.4

D.1 截断豪斯多夫距离

本节介绍一种先进的优化技术，可高效分析鲁棒形式的恢复保证。在此之前，我们先介绍两个集合之间的截断豪斯多夫距离这一概念。

定义 D.1 (截断豪斯多夫距离 [RW21, 6.J])。对于任意 $\rho \geq 0$，两个集合 $\mathcal{C}$ 和 $\mathcal{D}$ 的截断豪斯多夫距离定义为:

$$\hat{\mathrm{d}}_\rho(\mathcal{C},\mathcal{D}) = \max\{\mathrm{dist}(\mathcal{C}\cap\mathbb{B}(\mathbf{0},\rho),\mathcal{D}), \mathrm{dist}(\mathcal{D}\cap\mathbb{B}(\mathbf{0},\rho),\mathcal{C})\}.$$

事实证明，在某些条件下，两个最小化问题的最优值之间的距离可以用其上图的截断豪斯多夫距离来界定。以下引理给出了这一结果。

引理 D.2 ([RW21, 定理 6。56])。 设 $\rho \in [0,\infty)$。假设扩展实值函数 f， $g:\mathbb{R}^n \to \overline{\mathbb{R}}$ 满足

- 下确界 f，下确界 $g \in [-\rho,\rho]$，
- 使其达到最小值 f，使其达到最小值 $g \subseteq \mathbb{B}$ $(0,\rho)$。

由此可得

$$|\inf f - \inf g| \le \hat{\mathrm{d}}_\rho(\operatorname{epi} f, \operatorname{epi} g).^{3} \tag{18}$$

ˆ 进一步假设 $\varepsilon > 2\mathrm{d}\rho(\text{epi } f, \text{epi } g)$，则有

$$\operatorname{dist}(\boldsymbol{x}_g^*, \varepsilon\text{-}\operatorname{argmin} f) \le \hat{\mathrm{d}}_\rho(\operatorname{epi} f, \operatorname{epi} g), \tag{19}$$

其中， ε-argmin f 是 f 的 ε-次优解集，定义为 ε-argmin $f := \{\boldsymbol{x} \in \mathbb{R}^n : f(\boldsymbol{x}) \le$ 下确界 $f + \varepsilon\}$，且 $\boldsymbol{x}_g^*$ 是 g的一个极小化点。

由上述引理可知，如果两个优化问题足够接近（就截断豪斯多夫距离而言），那么它们的最优值也应当彼此接近。因此，为了应用这一结果，我们需要明确地界定截断豪斯多夫距离，以下 Kenmochi 条件便解决了这一问题。

引理 D.3 (Kenmochi 条件 [RW21，命题 6。58])。 设 $\rho \in [0,\infty)$。那么，对于 f， g： $\mathbb{R}^n \to \overline{\mathbb{R}}$ 若上图非空，则有

$$\hat{\mathrm{d}}_\rho(\operatorname{epi} f, \operatorname{epi} g) = \inf\left\{\eta > 0 : \begin{array}{l} \inf\limits_{\mathbb{B}(\boldsymbol{x},\eta)} g \le \max\{f(\boldsymbol{x}), -\rho\} + \eta,\ \forall \boldsymbol{x} \in [f \le \rho] \cap \mathbb{B}(\boldsymbol{0},\rho) \\ \inf\limits_{\mathbb{B}(\boldsymbol{x},\eta)} f \le \max\{g(\boldsymbol{x}), -\rho\} + \eta,\ \forall \boldsymbol{x} \in [g \le \rho] \cap \mathbb{B}(\boldsymbol{0},\rho) \end{array}\right\},$$

其中 $[f \le \rho] := \{\boldsymbol{x} \in \mathbb{R}^n : f(\boldsymbol{x}) \le \rho\}$。

D.2 定理 4.1 的证明

在给出证明之前，我们先介绍以下引理。

引理 D.4 (Hoffman 误差界 [Hof52])。 考虑集合 $\mathcal{S} := \{\boldsymbol{x} \in \mathbb{R}^n : \boldsymbol{A}\boldsymbol{x} \le \boldsymbol{b}\}$。存在 $C > 0$ 使得对于任意 $\boldsymbol{x} \in \mathbb{R}^n$，都有

$$\operatorname{dist}(\boldsymbol{x}, \mathcal{S}) \le C \cdot \|(\boldsymbol{A}\boldsymbol{x} - \boldsymbol{b})_+\|_2.$$

为简洁起见，我们记

$$\begin{aligned} \bar{f}_n(\boldsymbol{S}) &:= \|\boldsymbol{S}\|_{1,1} - \alpha \boldsymbol{1}^\top \log(\boldsymbol{S}\boldsymbol{1}) + \iota_{\mathbb{R}_-}(\|\boldsymbol{C}_n\boldsymbol{S} - \boldsymbol{S}\boldsymbol{C}_n\|_F - \delta_n) + \iota_{\mathcal{S}}(\boldsymbol{S}), \\ \bar{f}(\boldsymbol{S}) &:= \|\boldsymbol{S}\|_{1,1} - \alpha \boldsymbol{1}^\top \log(\boldsymbol{S}\boldsymbol{1}) + \iota_{\{0\}}(\|\boldsymbol{C}_\infty\boldsymbol{S} - \boldsymbol{S}\boldsymbol{C}_\infty\|_F) + \iota_{\mathcal{S}}(\boldsymbol{S}). \end{aligned}$$

[3] 对于函数 f： $\mathbb{R}^n \to \overline{\mathbb{R}}$，其上图定义为 $\operatorname{epi} f := \{(\boldsymbol{x}, y) \mid y \ge f(\boldsymbol{x})\}$。

因此，优化问题 LogSpecT（分别地，rLogSpecT）等价于 $\inf f$ （分别地，$\inf f_n$）。

现在，我们要利用引理D.3 来界定 $\hat{d}_\rho(\operatorname{epi} f, \operatorname{epi} f_n)$。令 $\boldsymbol{S} \in \mathcal{S} \cap \mathbb{B}(\mathbf{0}, \rho)$ 满足

$$\bar{f}(\boldsymbol{S}) \le \rho \quad \text{and} \quad \boldsymbol{S}\boldsymbol{C}_\infty = \boldsymbol{C}_\infty \boldsymbol{S}.$$

于是，我们知道

$$\|\boldsymbol{S}\boldsymbol{C}_n - \boldsymbol{C}_n\boldsymbol{S}\|_F \le 2\|\boldsymbol{S}\|_F \|\boldsymbol{C}_n - \boldsymbol{C}_\infty\| \le 2\rho\|\boldsymbol{C}_n - \boldsymbol{C}_\infty\| \le \delta_n,$$

因此，$\boldsymbol{S}$ 属于 f_n 的定义域。于是，对于任意 $\eta > 0$，都有

$$\inf_{\mathbb{B}(\boldsymbol{S},\eta)} \bar{f}_n \le \bar{f}_n(\boldsymbol{S}) = \bar{f}(\boldsymbol{S}) \le \max\{\bar{f}(\boldsymbol{S}), -\rho\}, \quad \forall \boldsymbol{S} \in [\bar{f} \le \rho] \cap \mathbb{B}(\mathbf{0}, \rho). \tag{20}$$

在验证 Kenmochi 条件的反向之前，我们先考虑 $[f_n \le \rho] \cap \mathbb{B}(\mathbf{0}, \rho)$ 的非空性。由于

$$\delta_n \ge 2\rho\|\boldsymbol{C}_n - \boldsymbol{C}_\infty\| \ge 2\alpha m\|\boldsymbol{C}_n - \boldsymbol{C}_\infty\|,$$

根据命题 3.7，可得 $\|\boldsymbol{S}_n^*\|_{1,1} \le \alpha m \le \rho$ 和 $f_n^* \le f^* \le \rho$，这意味着 $[f_n \le \rho] \cap \mathbb{B}(\mathbf{0}, \rho)$ 非空。令 $\boldsymbol{S}_n \in [f_n \le \rho] \cap \mathbb{B}(\mathbf{0}, \rho)$。于是有

$$\boldsymbol{S}_n \in \mathcal{S} \quad \text{and} \quad \|\boldsymbol{C}_n\boldsymbol{S}_n - \boldsymbol{S}_n\boldsymbol{C}_n\|_F \le \delta_n.$$

因此可得

$$\|\boldsymbol{C}_\infty\boldsymbol{S}_n - \boldsymbol{S}_n\boldsymbol{C}_\infty\| \le 2\|\boldsymbol{S}_n\|_F\|\boldsymbol{C}_\infty - \boldsymbol{C}_n\| + \|\boldsymbol{C}_n\boldsymbol{S}_n - \boldsymbol{S}_n\boldsymbol{C}_n\|_F \le 2\rho\|\boldsymbol{C}_\infty - \boldsymbol{C}_n\| + \delta_n.$$

此外，注意到存在 $\beta > 0$ 使得 $(\boldsymbol{S}_n\mathbf{1})_i \ge \beta$ 对所有 $i \in [m]$ 因为 $f_n \le \rho$ 且 $\|\boldsymbol{S}_n\|_{1,1} - \alpha\mathbf{1}^\top\log(\boldsymbol{S}_n\mathbf{1}) \to \infty$ 当 $\boldsymbol{S}_n \to \mathbf{0}$时。因此，将引理 D.4 应用于线性系统

$$\tilde{\mathcal{S}} := \{\boldsymbol{S} \in \mathbb{R}^{m\times m} : \boldsymbol{S}\boldsymbol{C}_\infty = \boldsymbol{C}_\infty\boldsymbol{S},\ \boldsymbol{S} \in \mathcal{S},\ (\boldsymbol{S}\mathbf{1})_i \ge \beta,\ \forall i \in [m]\}$$

由此可得，存在 $\tilde{c} > 0$ 使得

$$\operatorname{dist}(\boldsymbol{S}_n, \tilde{\mathcal{S}}) \le \tilde{c} \cdot (2\rho\|\boldsymbol{C}_\infty - \boldsymbol{C}_n\| + \delta_n).$$

因此，存在 $\boldsymbol{S}$ 属于 f 的定义域，使得

$$\|\boldsymbol{S}_n - \tilde{\boldsymbol{S}}\|_F \le \tilde{c} \cdot (2\rho\|\boldsymbol{C}_\infty - \boldsymbol{C}_n\| + \delta_n) \quad \text{and} \quad (\tilde{\boldsymbol{S}}\mathbf{1})_i \ge \beta, \quad \forall i \in [m].$$

由于函数 $\boldsymbol{S} \mapsto \|\boldsymbol{S}\|_{1,1} - \alpha\mathbf{1}^\top\log(\boldsymbol{S}\mathbf{1})$ 在 $(\boldsymbol{S}\mathbf{1})_i \ge \beta$ 时局部 Lipschitz 连续，因此存在 $L > 0$，使得

$$\begin{aligned}
\bar{f}(\tilde{\boldsymbol{S}}) = \|\tilde{\boldsymbol{S}}\|_{1,1} - \alpha\mathbf{1}^\top\log(\tilde{\boldsymbol{S}}\mathbf{1}) &\le \|\boldsymbol{S}_n\|_{1,1} - \alpha\mathbf{1}^\top\log(\boldsymbol{S}_n\mathbf{1}) + L\|\boldsymbol{S}_n - \tilde{\boldsymbol{S}}\|_F \\
&= \bar{f}_n(\boldsymbol{S}_n) + L\|\boldsymbol{S}_n - \tilde{\boldsymbol{S}}\|_F \\
&\le \bar{f}_n(\boldsymbol{S}_n) + L\tilde{c} \cdot (2\rho\|\boldsymbol{C}_\infty - \boldsymbol{C}_n\| + \delta_n).
\end{aligned}$$

令 $c_1 \geq \max\{1, L\} \cdot \tilde{c}$，可得对于任意 $\boldsymbol{S}_n \in [f_n \leq \rho] \cap \mathbb{B}(\mathbf{0}, \rho)$

$$\inf_{\mathbb{B}(\boldsymbol{S}_n, \eta)} \bar{f} \leq \bar{f}(\tilde{\boldsymbol{S}}) \leq \bar{f}_n(\boldsymbol{S}_n) + c_1 \cdot (2\rho \|\boldsymbol{C}_\infty - \boldsymbol{C}_n\| + \delta_n) \leq \max\{\bar{f}_n(\boldsymbol{S}_n), -\rho\} + \eta, \tag{21}$$

其中 $\eta := c_1 \cdot (2\rho \|\boldsymbol{C}_\infty - \boldsymbol{C}_n\| + \delta_n)$。结合不等式 (20) 和 (21)，我们可以得出

$$\hat{\mathrm{d}}_\rho(\operatorname{epi} \bar{f}, \operatorname{epi} \bar{f}_n) \leq c_1 \cdot (2\rho \|\boldsymbol{C}_\infty - \boldsymbol{C}_n\| + \delta_n). \tag{22}$$

为了推出结论 (i) 和 (ii)，还需验证引理D.2中的条件。由于 $\rho \geq \alpha m$，命题 3.7 的第一个陈述表明，$\inf f$和 $\inf f_n$的最优解均位于 $\mathbb{B}(\mathbf{0}, \rho)$ 内。由于 $\rho \geq f^*$ 且 $-\rho \leq \alpha m(1 - \log \alpha)$，该命题的第二个陈述表明 $\inf f$，$\inf f_n \in [-\rho, \rho]$。因此，应用引理 D.2 即可完成前两个陈述的证明。

为证明结论 (iii)，我们先提出以下两个断言：

(a) $\mathcal{S}_0^* \mathbf{1}$ 是一个单元素集合，其元素记为 $\boldsymbol{S}^* \mathbf{1}$，

(b) 对于任意 $\bar{\varepsilon} \in [0, \infty)$，存在一个 $\delta(\bar{\varepsilon}) > 0$，使得对于所有 $0 \leq \varepsilon \leq \bar{\varepsilon}$ 和 $\boldsymbol{S}_\varepsilon \in \mathcal{S}_\varepsilon^*$，有

$$\|\boldsymbol{S}_\varepsilon \mathbf{1} - \boldsymbol{S}^* \mathbf{1}\|_2 \leq \delta(\bar{\varepsilon}) \cdot \sqrt{\varepsilon}. \tag{23}$$

结合上述结果并借助定理 4.1，我们可以推导出对所有 $\boldsymbol{S}_n^* \in \mathcal{S}^{n,*}$

$$\begin{aligned} \operatorname{dist}(\boldsymbol{S}_n^* \mathbf{1}, \mathcal{S}_0^* \mathbf{1}) = \|\boldsymbol{S}_n^* \mathbf{1} - \boldsymbol{S}^* \mathbf{1}\|_2 &\leq \|\boldsymbol{S}_n^* \mathbf{1} - \boldsymbol{S}_{2\varepsilon_n} \mathbf{1}\|_2 + \|\boldsymbol{S}_{2\varepsilon_n} \mathbf{1} - \boldsymbol{S}^* \mathbf{1}\|_2 \\ &\leq \sqrt{m} \operatorname{dist}(\boldsymbol{S}_n^*, \mathcal{S}_{2\varepsilon_n}^*) + \|\boldsymbol{S}_{2\varepsilon_n} \mathbf{1} - \boldsymbol{S}^* \mathbf{1}\|_2 \\ &\leq \tilde{c}_1 \varepsilon_n + \tilde{c}_2 \sqrt{\varepsilon_n}, \end{aligned}$$

其中 $\tilde{c}_1$、$\tilde{c}_2$ 为正数，且 $\boldsymbol{S}_{2\varepsilon_n} \in \mathcal{S}_{2\varepsilon_n}^*$ 满足 $\|\boldsymbol{S}_n^* - \boldsymbol{S}_{2\varepsilon_n}\|_F = \operatorname{dist}(\boldsymbol{S}_n^*, \mathcal{S}_{2\varepsilon_n}^*)$（其存在性由 $\mathcal{S}_\varepsilon^*$ 是凸且紧的保证）。因此，

$$\operatorname{dist}(\mathcal{S}^{n,*} \mathbf{1}, \mathcal{S}_0^* \mathbf{1}) \leq \tilde{c}_1 \varepsilon_n + \tilde{c}_2 \sqrt{\varepsilon_n}.$$

接下来，只需证明这些结论。定义辅助函数 $h : \mathbb{R}^m \to \mathbb{R}$，使得$h(\boldsymbol{x}) = \sum_{i=1}^m x_i - \alpha \sum_{i=1}^m \log x_i$，对每个 $\boldsymbol{x} \in \mathbb{R}_+^m$均成立。考虑以下优化问题：

$$\begin{aligned} &\min_{\boldsymbol{x}} \ h(\boldsymbol{x}) \\ &\text{s.t.} \ \ \boldsymbol{x} \in \{\boldsymbol{S}\mathbf{1} \in \mathbb{R}^m \mid \boldsymbol{S} \text{ that is feasible for LogSpecT}\}. \end{aligned} \tag{24}$$

为简洁起见，将（ ε-次优解集24）表示为 $\mathcal{H}_\varepsilon^*$。接下来，我们将首先证明 $\mathcal{S}_\varepsilon^* \mathbf{1} = \mathcal{H}_\varepsilon^*$，然后根据 h的严格凸性，得到所需的两个结论。

第一步是证明，问题（24）的最优函数值满足 $h^* = f^*$。由于显然 $\tilde{\boldsymbol{x}} = \boldsymbol{S}^* \mathbf{1}$对于（24）是可行的，因此 $h^* \leq h(\tilde{\boldsymbol{x}}) = f(\boldsymbol{S}^*) = f^*$。反设 $h^* < f^*$。由目标函数是强制的且连续这一事实，以及可行集是闭集可知，存在 $\tilde{\boldsymbol{S}}$，使其对于 LogSpecT 是可行的，且满足 $\boldsymbol{x}^* = \tilde{\boldsymbol{S}}\mathbf{1}$，其中 $\boldsymbol{x}^*$ 是（24）的一个最优解。由于 $h^* = h(\boldsymbol{x}^*) = h(\tilde{\boldsymbol{S}}\mathbf{1}) = f(\tilde{\boldsymbol{S}})$，这与以下事实矛盾

˜即 $f(\boldsymbol{S}) \geq f^*$。因此， $h^* = f^*$。接下来，我们将证明 $\mathcal{S}_\varepsilon^*\mathbf{1} = \mathcal{H}_\varepsilon^*$。考虑任意一个 ε-次优解 $\boldsymbol{S} \in \mathcal{S}_\varepsilon^*$，即，

$$h(\boldsymbol{S}\mathbf{1}) = f(\boldsymbol{S}) \leq f^* + \varepsilon = h^* + \varepsilon.$$

因此， $\boldsymbol{S}\mathbf{1} \in \mathcal{H}_\varepsilon^*$，这意味着 $\mathcal{S}_\varepsilon^*\mathbf{1} \subseteq \mathcal{H}_\varepsilon^*$。另一方面，对于任意 ε-次优解$\boldsymbol{x} \in \mathcal{H}_\varepsilon^*$，存在 $\boldsymbol{S}$，它对于 LogSpecT 是可行的，且满足 $\boldsymbol{x} = \boldsymbol{S}\mathbf{1}$。因此，

$$f(\boldsymbol{S}) = h(\boldsymbol{x}) \leq h^* + \varepsilon = f^* + \varepsilon.$$

这意味着 $\boldsymbol{S} \in \mathcal{S}_\varepsilon^*$，进而 $\mathcal{H}_\varepsilon^* \subseteq \mathcal{S}_\varepsilon^*\mathbf{1}$。因此， $\mathcal{H}_\varepsilon^* = \mathcal{S}_\varepsilon^*\mathbf{1}$。

由于 h 是严格凸的，其最优解集 $\mathcal{H}_0^*$ 是单元素集合。因此， $\mathcal{S}_0^*\mathbf{1} = \mathcal{H}_0^*$ 是单元素集合，这证明了第一个结论。对于第二个结论，我们知道，对于任意 $\boldsymbol{S}_\varepsilon \in \mathcal{S}_\varepsilon^*$，存在 $\boldsymbol{x}_\varepsilon \in \mathcal{H}_\varepsilon^*$ 使得

$$\|\boldsymbol{S}_\varepsilon\mathbf{1} - \boldsymbol{S}^*\mathbf{1}\|_2 = \|\boldsymbol{x}_\varepsilon - \boldsymbol{x}^*\|_2, \tag{25}$$

其中 $\boldsymbol{x}^* \in \mathcal{H}_0^*$。 h 的强制性表明 $\boldsymbol{x}_\varepsilon$ 和 $\boldsymbol{x}^*$ 有界。结合 h 在任意有界集上都是强凸的这一事实，可以得出存在 $\mu > 0$ 使得

$$h(\boldsymbol{x}_\varepsilon) \geq h(\boldsymbol{x}^*) + \langle \nabla h(\boldsymbol{x}^*), \boldsymbol{x}_\varepsilon - \boldsymbol{x}^* \rangle + \frac{1}{\mu}\|\boldsymbol{x}_\varepsilon - \boldsymbol{x}^*\|_2^2 \geq h(\boldsymbol{x}^*) + \frac{1}{\mu}\|\boldsymbol{x}_\varepsilon - \boldsymbol{x}^*\|_2^2, \tag{26}$$

其中，第二个不等式来自 $\boldsymbol{x}^*$ 的全局最优性。结合（25），可得 (26)

$$\|\boldsymbol{S}_\varepsilon\mathbf{1} - \boldsymbol{S}^*\mathbf{1}\|_2 = \|\boldsymbol{x}_\varepsilon - \boldsymbol{x}^*\|_2 \leq \sqrt{\mu(h(\boldsymbol{x}_\varepsilon) - h(\boldsymbol{x}^*))} \leq \sqrt{\mu\varepsilon}.$$

至此，所述结论的证明完毕。

D.3 推论 4.4 的证明

反设存在一个序列 $\{\boldsymbol{S}_n^*\}_n$，其中第 n 个元素是样本量为 n 时 rLogSpecT 的一个最优解，使得

$$\operatorname{dist}(\boldsymbol{S}_n^*, \mathcal{S}_0^*) \not\to 0.$$

根据命题 3.7，我们知道 $\{\boldsymbol{S}_n^*\}_n$ 有界，因此存在一个收敛子列。不失一般性，我们可以假设该序列本身收敛，且其极限点为 $\boldsymbol{S}^*$。注意到

$$\|\boldsymbol{C}_n\boldsymbol{S}_n^* - \boldsymbol{S}_n^*\boldsymbol{C}_n\|_F \leq \delta_n, \quad \boldsymbol{C}_n \to \boldsymbol{C}_\infty \quad \text{and} \quad \delta_n \to 0.$$

因此， $\boldsymbol{C}_\infty\boldsymbol{S}^* = \boldsymbol{S}^*\boldsymbol{C}_\infty$。这表明 $\boldsymbol{S}^*$ 对 LogSpecT 是可行的。接着，根据定理4.1，我们知道 $f(\boldsymbol{S}_n^*) = f_n^* \to f^*$，由此得到 $f(\boldsymbol{S}^*) = f^*$，因为 $f(\cdot) = \|\cdot\|_{1,\,1} - \alpha\mathbf{1}^\top\log(\cdot)\mathbf{1}$ 是连续的。结合 $\boldsymbol{S}^*$ 是可行的这一事实，我们得出 $\boldsymbol{S}^*$ 是 LogSpecT 的一个最优解。这进一步意味着 $\operatorname{dist}(\boldsymbol{S}_n^*, \mathcal{S}_0^*) \to 0$，这与假设矛盾。

D.4 引理 4.7 的证明

回顾生成模型（1）。由于 $\boldsymbol{w}$ 服从次高斯分布，可以证明，对每个 $t>0$,

$$\mathbb{P}(\|\boldsymbol{x}\|_2>t)\leq\mathbb{P}\left(\|\boldsymbol{w}\|_2>\frac{t}{\|\mathcal{H}(\boldsymbol{S})\|}\right)\leq Ce^{-v't^2},$$

对于某个正常数 v'，这意味着 $\boldsymbol{x}$ 也服从次高斯分布。因此，根据次高斯性质，$\|\boldsymbol{C}_n-\boldsymbol{C}_\infty\|$ 可以通过以下引理明确界定。

引理 D.5 ([Ver12，命题 2。1])。设 $\boldsymbol{x}_1$、$\boldsymbol{x}_2$、……、$\boldsymbol{x}_n\in\mathbb{R}^m$ 为次高斯、同分布且相互独立的随机向量，满足 $n>m$。则对所有 $\varepsilon>0$，有

$$\mathbb{P}\left(\left\|\frac{1}{n}\sum_{i=1}^n\boldsymbol{x}_i\boldsymbol{x}_i^\top-\mathbb{E}[\boldsymbol{x}\boldsymbol{x}^\top]\right\|_2\leq\varepsilon\right)\geq1-2e^{2m-l\varepsilon^2n},$$

对于某个常数 $l>0$。

令 $\varepsilon^2=(4/l)\log(2n)m/n$，引理 D.5 表明，以高概率（其下界为 $1-n^{-1}$），

$$\|\boldsymbol{C}_n-\boldsymbol{C}_\infty\|\leq\mathcal{O}\left(\sqrt{\frac{\log n}{n}}\right).$$

E　L-ADMM 的推导与收敛性分析　　s

本节介绍 rLogSpecT 的 L-ADMM 详细内容。

E.1 命题证明5.1

注意，最小化问题（7）关于 $\boldsymbol{Z}$ 和 $\boldsymbol{q}$是可分的，因此可拆分为两个子问题：

$$\min_{\boldsymbol{Z}\in\mathbb{B}(\boldsymbol{0},\delta_n)}\|\boldsymbol{C}_n\boldsymbol{S}^{(k)}-\boldsymbol{S}^{(k)}\boldsymbol{C}_n+\boldsymbol{\Lambda}^{(k)}/\rho-\boldsymbol{Z}\|_F^2, \tag{27}$$

$$\min_{\boldsymbol{q}}\ -\alpha\boldsymbol{1}^\top\log\boldsymbol{q}+\boldsymbol{\lambda}_2^{(k)\top}(\boldsymbol{q}-\boldsymbol{S}^{(k)}\boldsymbol{1})+\frac{\rho}{2}\|\boldsymbol{q}-\boldsymbol{S}^{(k)}\boldsymbol{1}\|_2^2. \tag{28}$$

对于问题（27），最优解是将 $\boldsymbol{C}_n\boldsymbol{S}^{(k)}-\boldsymbol{S}^{(k)}\boldsymbol{C}_n+\boldsymbol{\Lambda}^{(k)}/\rho$ 投影到$\mathbb{B}$（$\boldsymbol{0}$，δ_n）上，具体为

$$\boldsymbol{Z}^{(k+1)}=\min\left\{1,\frac{\delta_n}{\|\tilde{\boldsymbol{Z}}\|_F}\right\}\tilde{\boldsymbol{Z}}\ \text{ with }\ \tilde{\boldsymbol{Z}}=\boldsymbol{C}_n\boldsymbol{S}^{(k)}-\boldsymbol{S}^{(k)}\boldsymbol{C}_n+\boldsymbol{\Lambda}^{(k)}/\rho.$$

对于问题（28），一阶最优性条件为

$$-\alpha1/\boldsymbol{q}+\boldsymbol{\lambda}_2^{(k)}+\rho(\boldsymbol{q}-\boldsymbol{S}^{(k)}\boldsymbol{1})=0.$$

结合目标函数是凸函数这一事实，可以推知

$$\boldsymbol{q}^{(k+1)}=\frac{\tilde{\boldsymbol{q}}+\sqrt{\tilde{\boldsymbol{q}}^2+4\alpha/\rho\boldsymbol{1}}}{2}\ \text{ with }\ \tilde{\boldsymbol{q}}=\frac{1}{\rho}(\rho\boldsymbol{S}^{(k)}\boldsymbol{1}-\boldsymbol{\lambda}_2^{(k)}).$$

E.2 $\Pi_{\mathcal{S}}(\cdot)$ 的计算

$\boldsymbol{X}$ 到 $\mathcal{S}$ 的投影可以通过求解一个优化问题来计算：

$$\begin{aligned}\min_{\boldsymbol{S}} \ & \|\boldsymbol{X}-\boldsymbol{S}\|_F^2 \\ \text{s.t.} \ & \boldsymbol{S}^\top = \boldsymbol{S}, \\ & S_{ii}=0,\ i=1,2,\dots,m, \\ & S_{ij}\geq 0,\ \forall i,j,\end{aligned}$$

这等价于

$$\begin{aligned}\min \ & \sum_{i<j}\left((X_{ij}-S_{ij})^2+(X_{ji}-S_{ij})^2\right) \\ \text{s.t.} \ & S_{ij}\geq 0,\ \forall i<j, \\ & S_{ii}=0,\ \forall i.\end{aligned}$$

因此

$$(\Pi_{\mathcal{S}}(\boldsymbol{X}))_{ij}=\begin{cases}\dfrac{1}{2}\max\{0,X_{ij}+X_{ji}\}, & i\neq j,\\ 0, & i=j.\end{cases}$$

E.3 ρ 的停止准则和更新规则

我们遵循 [BPC+11] 中的步骤来更新 ρ，每次迭代均如此。类似地，我们将原始残差和对偶残差定义如下：

$$\begin{aligned}p_{\text{res}}^{(k+1)} &= \sqrt{\|\boldsymbol{Z}^{(k+1)}-\boldsymbol{C}_n\boldsymbol{S}^{(k+1)}+\boldsymbol{S}^{(k+1)}\boldsymbol{C}_n\|_F^2+\|\boldsymbol{q}^{(k+1)}-\boldsymbol{S}^{(k+1)}\mathbf{1}\|_2^2},\\ d_{\text{res}}^{(k+1)} &= \rho^{(k)}\left(\boldsymbol{C}_n(\boldsymbol{S}^{(k+1)}-\boldsymbol{S}^{(k)})-(\boldsymbol{S}^{(k+1)}-\boldsymbol{S}^{(k)})\boldsymbol{C}_n+\mathbf{1}^\top(\boldsymbol{S}^{(k+1)}-\boldsymbol{S}^{(k)})\mathbf{1}\right).\end{aligned}$$

更新 ρ 的目的是控制 p_{res} 和 d_{res} 的衰减速度，使它们之间的差异不会过大。为此，我们按照以下方案自适应地更新 ρ：

$$\rho^{(k+1)}:=\begin{cases}2\rho^{(k)}, & \text{if } p_{\text{res}}^{(k+1)}>5d_{\text{res}}^{(k+1)},\\ \rho^{(k)}/2, & \text{if } d_{\text{res}}^{(k+1)}>5p_{\text{res}}^{(k+1)},\\ \rho^{(k)}, & \text{otherwise.}\end{cases}$$

当 p_{res} 和 d_{res} 都小于阈值 $\varepsilon=10^{-5}$ 时，我们停止算法。

E.4 收敛性分析

定义 $\boldsymbol{D}:=$ 对角矩阵$(\mathbf{1}_m^\top,\dots,\mathbf{1}_m^\top)\in\mathbb{R}^{m\times m^2}$。那么，$\boldsymbol{D}$ 满足 $\boldsymbol{D}\text{vec}(\boldsymbol{S})=\boldsymbol{S}\mathbf{1}$ 和 $\|\boldsymbol{D}^\top\boldsymbol{D}\|=m$。记

$$\boldsymbol{Q}:=\tau\boldsymbol{I}-\boldsymbol{D}^\top\boldsymbol{D}-\boldsymbol{A}_n^\top\boldsymbol{A}_n.$$

于是，线性化交替方向乘子法更新式（8）$\boldsymbol{S}$ 可写为：

$$\min_{\boldsymbol{S}}\ L(\boldsymbol{S})+\frac{\rho}{2}\|\text{vec}(\boldsymbol{S})-\text{vec}(\boldsymbol{S}^{(k)})\|_{\boldsymbol{Q}},$$

其中，$\|\boldsymbol{x}\|_{\boldsymbol{Q}} := \boldsymbol{x}^\top \boldsymbol{Q}\boldsymbol{x}$。由于 $\tau > m + \|\boldsymbol{A}_n\|^2$，我们知道 $\boldsymbol{Q}$ 是正定的。因此，将（$\boldsymbol{Z}$，$\boldsymbol{q}$）视为一个变量后，我们可以应用 [ZBO11] 中的定理 4.2，直接得到该结果。

F 合成数据上的更多实验与讨论

为了公平比较 rSpecT 和 rLogSpecT，我们在具有相同图滤波器的 BA 图上测试 rSpecT，结果如图6所示。显然，rSpecT 在这些情况下会失效，且无法从样本量增加中获益。这是合理的，因为如图1所示，SpecT 在 BA 图上会失效，更不用说近似形式的 rSpecT 了。

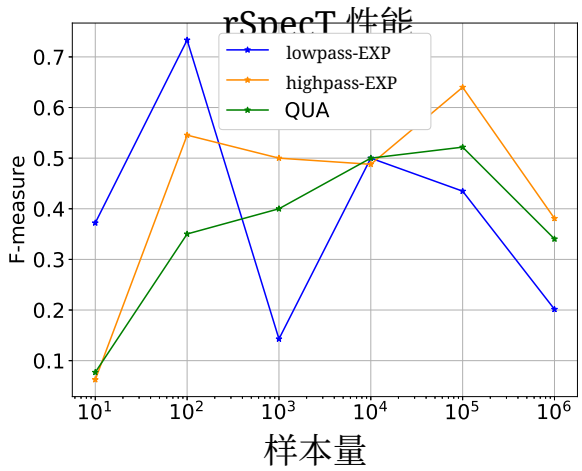

图 6：rSpecT 在 BA 图上的性能。

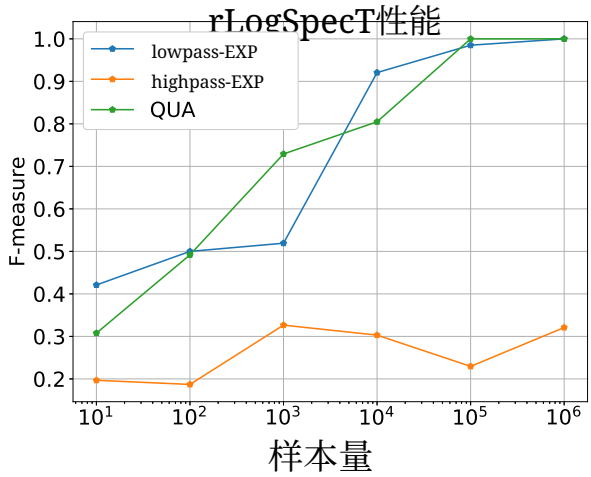

图 7：rLogSpecT 在 ER图 上的表现，$\delta_n = 20\sqrt{\log\, n/n}$。

我们进一步在观测到不同数量信号的 ER图 上测试 rLogSpecT。参数 δ_n 设为 $20\sqrt{\log n/n}$，结果如图 7所示。图中显示，对于非高通图滤波器，当样本量足够大时，rLogSpecT 能够实现近乎完美的恢复。此外，与在 BA 图上的表现相比，rLogSpecT 在 ER图 上效果更好。这一观察与图 1的结论一致：LogSpecT 在 ER图 上的表现优于在 BA 图上的表现。我们还注意到，低通图滤波器与高通图滤波器之间的差异非常显著。为检验 rLogSpecT 通常在低通图滤波器上表现更好的猜想，我们选取了不同的图滤波器 $\exp(t\boldsymbol{S})$，其中 t 取值范围为 -2至 2，并在 ER图 上开展实验。当图移位算子为邻接矩阵时，正的低通参数 t 对应低通图滤波器，而负值 t 对应高通图滤波器 [RWS20、HW22]。我们省略了 $t = 0$ 的情况，因为该滤波器不包含任何图信息（注意，$\exp(0\boldsymbol{S}) = \boldsymbol{I}$）。

然后，我们重复实验 50 次，并在图 8 中报告平均结果。低通图滤波器与高通图滤波器的性能比较表明，低通图滤波器的整体表现优于高通图滤波器。仔细观察结果可发现，当 t 的绝对值较小时，性能提升更快。

最终，绝对值较小的图滤波器 t 占据主导地位。这一观察很有意思，因为图 1 表明，图滤波器的选择对模型性能影响不大。一种解释是，低通图滤波器和高通图滤波器都会衰减图的某些频率，而 t 的绝对值越大，有限信号所携带的信息损失就越多。

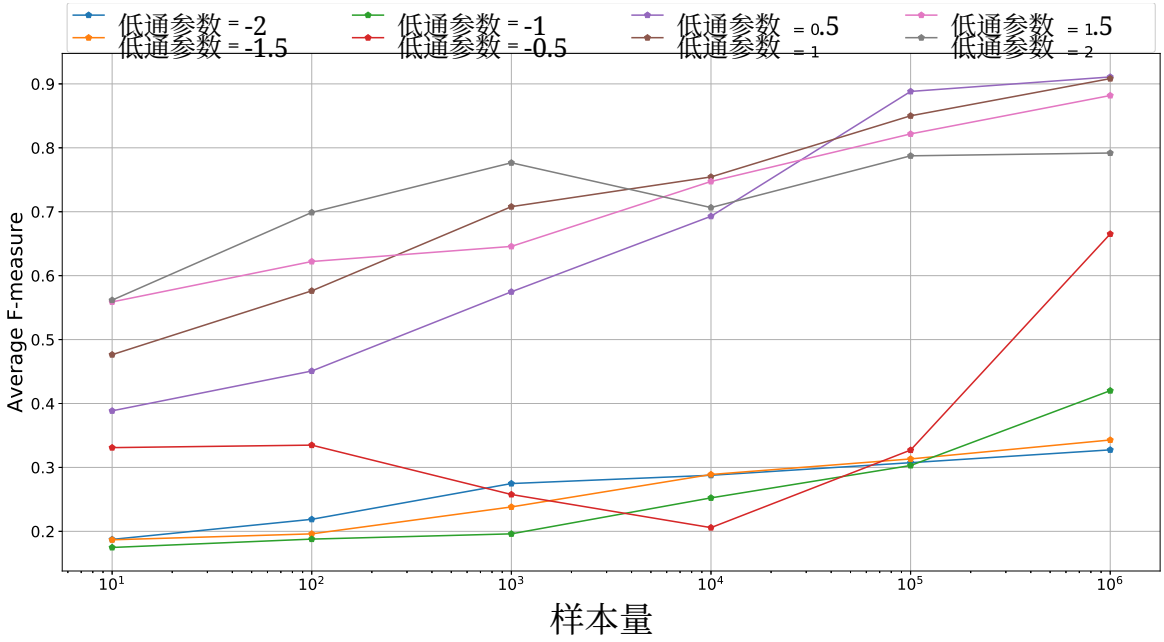

图 8：低通参数的影响：当 t 从 -2 变化到 2 时，图滤波器 $\exp(tS)$ 的性能差异。