arXiv:2603.02103v1 [math.OC] 2 Mar 2026

在结构化图上求解带指示变量的凸二次优化问题

Aaresh Bhathena* Salar Fattahi† Andr´es G´omez‡ Simge K¨u¸c¨ukyavuz §

摘要

本文研究凸二次最小化问题，其中每个连续变量都与一个二元指示变量相关联。我们关注这样一种结构化情形：二次项的海森矩阵为正定矩阵，且具有稀疏性。我们开发了一种精确的参数化动态规划算法，其计算复杂度明确取决于海森矩阵支撑图的树宽、体积增长以及适当的间隔参数。在适当的结构条件下，总体复杂度随问题维数线性增长。为展示我们方法的实际效果，我们提出了一种新颖的联合预测与异常值检测框架，将指数平滑扩展到含离群值的时间序列。针对合成数据集和真实数据集的计算实验表明，我们的方法显著优于先进求解器。

关键词： 混合整数二次规划、动态规划、异常值检测、指数平滑。

1 引言

我们考虑以下混合整数二次规划问题（MIQP），该问题由一个对称正定矩阵 $\boldsymbol{Q} \in \mathbb{R}^{n\times n}$ 和向量 $\boldsymbol{\lambda}$, $\boldsymbol{c} \in \mathbb{R}^n$定义：

$$\min_{\boldsymbol{x}\in\mathbb{R}^n, \boldsymbol{z}\in\{0,1\}^n} \quad \frac{1}{2}\boldsymbol{x}^\top \boldsymbol{Q}\boldsymbol{x} + \boldsymbol{c}^\top \boldsymbol{x} + \boldsymbol{\lambda}^\top \boldsymbol{z} \tag{1a}$$

$$\text{s.t.} \quad \boldsymbol{x}_i(1-\boldsymbol{z}_i) = 0 \qquad i = 1, 2, \ldots, n. \tag{1b}$$

在此问题中，二进制向量 $\boldsymbol{z} \in \{0,1\}^n$ 编码连续向量 $\boldsymbol{x} \in \mathbb{R}^n$ 的支撑集。具体而言，约束 $\boldsymbol{x}_i(1-\boldsymbol{z}_i) = 0$ 确保当 $\boldsymbol{z}_i = 0$ 时， $\boldsymbol{x}_i = 0$ ；而 $\boldsymbol{z}_i = 1$ 允许$\boldsymbol{x}_i \in \mathbb{R}$不受约束。向量 $\boldsymbol{\lambda} \in \mathbb{R}^n$ 作为分量正则化参数，促进 $\boldsymbol{x}$ 的稀疏性。我们始终假设，对于每个 $i = 1, \ldots, n$，都有 $\boldsymbol{\lambda}_i > 0$，其中 $\boldsymbol{\lambda}_i \leq 0$

*美国密歇根州安娜堡，密歇根大学工业与运筹工程系。aareshfb@umich.edu

†美国密歇根州安娜堡，密歇根大学工业与运筹工程系。fattahi@umich.edu

‡美国加利福尼亚州洛杉矶，南加州大学丹尼尔·J·爱泼斯坦工业与系统工程系。gomezand@usc.edu§美国伊利诺伊州埃文斯顿，西北大学工业工程与管理科学系。simge@northwestern.edu

这意味着在最优性条件下 $z_i = 1$。不失一般性，我们还通过将每个变量 $\boldsymbol{x}_i$ 重缩放为 $\boldsymbol{x}_i \to \boldsymbol{x}_i / \sqrt{\boldsymbol{Q}_{i,\ i}}$，将 $\boldsymbol{Q}$ 的对角元素归一化为 1。

本文聚焦于问题（1）中的一类实例，其中海森矩阵的稀疏模式$\boldsymbol{Q} \in \mathbb{R}^{n \times n}$ 与某个图的邻接矩阵重合——下文称之为支撑图——且该图具有特定的稀疏结构。具体而言，我们考虑具有有界树宽和多项式体积增长性质的支撑图。前者刻画图在多大程度上类似于树，后者则对任意给定节点周围固定图距离内所包含的节点数设置了上界。树宽是图论中的经典概念，而多项式体积增长则是一个较少见但同样有意义的结构性质；这两个概念将在第2节中正式介绍。

问题（1）出现在多个领域，包括稀疏回归 [12, 13, 27]，概率图模型 [43, 54, 33, 56, 50, 75, 74]，时间序列中的异常值检测 [8]，以及网络推断 [33, 60]。本文聚焦于时间序列中带异常值校正的预测问题；所提出的公式化模型自然地源于这一问题，并且特别有效。

1.1 我们贡献概述

我们提出的方法的核心是剪枝技术，能够高效地排除二进制向量 $\boldsymbol{z} \in \{0,1\}^n$的次优选择，从而将搜索空间的规模从指数级降至多项式级。考虑问题（1）的以下等价形式：

$$\min_{\substack{\boldsymbol{x}\in\mathbb{R}^n \\ \boldsymbol{z}\in\{0,1\}^n \\ \boldsymbol{x}\circ(1-\boldsymbol{z})=0}} \left\{ \frac{1}{2}\boldsymbol{x}^\top \boldsymbol{Q}\boldsymbol{x} + \boldsymbol{c}^\top \boldsymbol{x} + \boldsymbol{\lambda}^\top \boldsymbol{z} \right\} = \min_{\boldsymbol{x}\in\mathbb{R}^n} \left\{ \min_{\boldsymbol{z}\in\{0,1\}^n} \left\{ \frac{1}{2}\boldsymbol{x}^\top \left(\boldsymbol{Q} \circ \boldsymbol{z}\boldsymbol{z}^\top \right) \boldsymbol{x} + (\boldsymbol{c}\circ\boldsymbol{z})^\top \boldsymbol{x} + \boldsymbol{\lambda}^\top \boldsymbol{z} \right\} \right\}, \tag{2}$$

其中，$\circ$ 表示逐元素乘积。对每个固定的 $\boldsymbol{z} \in \{0,1\}^n$，定义凸二次函数 $p_{\boldsymbol{z}}(\boldsymbol{x}) := \frac{1}{2}\boldsymbol{x}^\top \left(\boldsymbol{Q} \circ \boldsymbol{z}\boldsymbol{z}^\top \right) \boldsymbol{x} + (\boldsymbol{c}\circ\boldsymbol{z})^\top \boldsymbol{x} + \boldsymbol{\lambda}^\top \boldsymbol{z}$后，上述问题可化简为以下两阶段优化问题：

$$\min_{\boldsymbol{x}\in\mathbb{R}^n} f(\boldsymbol{x}), \quad \text{where} \quad f(\boldsymbol{x}) = \min_{\boldsymbol{z}\in\{0,1\}^n} p_{\boldsymbol{z}}(\boldsymbol{x}). \tag{3}$$

上述两阶段形式对应于原问题（1）的两阶段求解策略：首先，通过消去二元变量 $\boldsymbol{z}$，刻画 $f : \mathbb{R}^n \to \mathbb{R}$，然后针对 $\boldsymbol{x}$直接优化 f。显然，第一阶段构成了计算瓶颈：有效求解该问题取决于能否有效刻画f。如上述重构所示，它是一个分段函数，最多由 2^n个凸二次函数片段组成。单独来看，这种方法相较于穷举所有 $\boldsymbol{z}$ 种配置并无明显优势。然而，我们证明，当海森矩阵 $\boldsymbol{Q}$ 具有特定的图结构时，可以通过系统地剪枝在所有可能的 $\boldsymbol{z}$ 取值下均可证明为次优的选择，从而大幅加速这一枚举过程 $\boldsymbol{x}$。

我们的剪枝策略基于两个关键思想。首先，由于 $\boldsymbol{Q}$ 是正定的，任何最优解 $\boldsymbol{x}^\star$ 的范数都必须有界；也就是说，对于某个常数 $U > 0$，有 $\|\boldsymbol{x}^\star\|_\infty \le U$。因此，只需在有界区域 $\mathcal{D} = \{\boldsymbol{x}$ 内刻画参数化成本 f：该区域包含最优解 $\|\boldsymbol{x}\|_\infty \le U\}$。其次，我们证明，在某些结构条件下，只需用多项式数量的二次片段 $p_{\boldsymbol{z}}$（满足 $\boldsymbol{z} \in \{0,1\}^n$）即可表示该有界区域内的 f

。这一结果背后的关键洞见是，对于任意两个稀疏模式 $\boldsymbol{z}_1$和 $\boldsymbol{z}_2 \in \{0,1\}^n$，如果它们存在显著重叠（通过 m-相似性的概念加以形式化；参见定义4），多项式函数 $p_{\boldsymbol{z}_1} - p_{\boldsymbol{z}_2}$ 的根会随重叠长度呈指数级增长。因此，当重叠足够长时，这些根便落在 $\mathcal{D}$之外，这意味着在该区域内，$p_{\boldsymbol{z}_1}(\boldsymbol{x}) > p_{\boldsymbol{z}_2}(\boldsymbol{x})$ 或$p_{\boldsymbol{z}_1}(\boldsymbol{x}) < p_{\boldsymbol{z}_2}(\boldsymbol{x})$。在这种情况下，可以安全地舍弃（剪枝）其中一个二次片段及其对应的稀疏模式，而不会影响对参数化成本 f的刻画。

尽管参数化成本 f 存在高效表示令人鼓舞，但这本身并不能保证能够高效地构造这种表示。为此，我们提出了一种高效算法，称为参数化算法，用于在矩阵$\boldsymbol{Q}$具有小宽度树分解的情形下刻画 f。我们提出的参数化算法通过在 $\boldsymbol{Q}$的树分解上运行动态规划（DP），高效地构造 f。所提方法的整体计算复杂度取决于树分解的宽度、支撑图的体积增长，以及问题的一种恰当间隔概念；我们将在后续章节中逐一详细讨论这些因素。

在实践中，我们的参数化算法比现成求解器快几个数量级，只需数秒至数分钟即可求解多达 20,000 个变量的问题，远超现有现成求解器的能力。我们还通过将该框架应用于时间序列预测，进一步展示了其实际价值。借助一种新的混合整数二次规划表述，我们开发了一种指数平滑扩展方法，可同时进行预测和异常值检测。计算结果表明，在真实世界数据上，这种整合提升了预测性能，并增强了对离群值的鲁棒性。

1.2 相关工作

对于一般的正定矩阵 $\boldsymbol{Q}$，问题（1）已知是 NP困难的 [24]。因此，现有方法通常依赖序贯凸松弛（通常通过割平面技术加以强化），或利用特定的结构性质，使特殊情形变得易于求解。

基于序贯凸松弛的方法。 基于凸松弛的经典方法采用$Big\text{-}M$ 表述来编码指示变量 [37]。这些表述已被广泛用于混合整数二次与回归类问题，并常常被纳入分支定界技术 [12, 11, 26]。尽管对中小规模实例有效，Big-M 表述通常只能给出较弱的松弛，并受数值不稳定性影响，从而导致在大规模问题上可扩展性较差 [44]。一项重大突破出现在透视重构 技术的引入之时，该技术为可分混合整数二次规划提供了显著更紧的凸松弛。该方法最初由Stubbs [64] 提出，并在一系列工作中得到进一步发展 [4, 35, 40], ，此后透视重构已成为面向稀疏与结构化优化的现代算法的基石 [11, 73, 38, 71, 72]。在实践中，这些技术构成了诸如 Gurobi等最先进商业求解器的骨干，这些求解器通过专门的割平面、预求解约简和基于透视的凸化模块，将其集成到混合整数二次规划求解器中。然而，尽管这些方法有效，此类基于松弛的方法仍必须嵌入分支定界或分支割平面框架之中，其指数级最坏情况复杂度仍然使它们对大规模问题在计算上难以承受。

针对结构化 Q的方法。 鉴于该问题具有指数级最坏情况复杂度，另一条研究路线考察了如下实例，其中 $\boldsymbol{Q}$ 具有能够实现更高效算法的结构性质。例子包括如下情形： $\boldsymbol{Q}$ 为对角 [22], 斯蒂尔切斯 [7, 43, 54], 秩一 [63, 42],允许稀疏分解 $\boldsymbol{Q} = \boldsymbol{Q}_0^\top \boldsymbol{Q}_0$ [27]，或呈现其他特殊结构 [52, 25]。与这一研究方向密切相关的，是对$\boldsymbol{Q}$中带状或树结构稀疏模式加以利用的研究。在特殊情形下，当 $\boldsymbol{Q}$ 为三对角时，Liu 等人 [53] 提出了一种基于最短路径表述的动态规划算法，可在 $\mathcal{O}(n^2)$ 时间和内存内恢复精确解。在此基础上，G´omez 等人 [39]将该方法推广到一般带状矩阵，并开发了 全多项式时间近似方案 （全多项式时间近似方案），用于问题（1）。

作为对 Liu 等人提出的动态规划方法的扩展， [53], Bhathena 等人 [14] 证明，问题（1）在 $\mathcal{O}(n^2)$ 时间内可以精确求解，前提是 $\boldsymbol{Q}$ 的稀疏图是一条路径或一棵树。这一结果已表明，树分解为超越树状图、拓展问题的可处理性提供了一个有原则的框架。然而，Bhathena 等人 [14] 的方法依赖于对单变量凸二次函数共轭函数的精细刻画，而这一性质难以推广到树图以外的情形。

基于小树宽的方法。Halin 最早引入了小树宽图[41] ，随后 Robertson 和 Seymour 对其进行了推广， [61] [参见30，了解相关历史]。有界树宽的概念是算法图论的基础，因为它刻画了能够高效应用动态规划方法的图类 [19]。自 20 世纪 80 年代末以来，人们已充分认识到，若将若干经典的难解组合问题限制在小树宽图上，例如独立集、图着色和哈密顿回路，便可通过动态规划方法在多项式时间内求解 [9, 18, 5]。

相比之下，问题（1）具有混合整数结构，同时涉及二元和连续决策变量。在这一更广泛的背景下，树分解已在多项式优化和混合整数优化领域得到广泛研究 [69, 16, 55, 48, 70], ，其中包括与我们的问题密切相关的问题子类。具体而言，Bienstock 和 Chen [15] 研究了二次优化问题，其中不相交变量块的稀疏性由受耦合约束限制的二元指示变量控制。当海森矩阵 $\boldsymbol{Q}$ 的树宽有界时，他们的框架适用于问题（1）。不过，他们的方法只能得到全多项式时间近似方案算法；据我们所知，文献中尚未报告所提出算法的实际实现。

1.3 概述

本文其余部分安排如下。第2 节介绍必要的预备知识和背景。第3 节介绍一项应用：基于指数平滑和异常值检测的时间序列预测。第4节介绍一种通过依次刻画参数化成本、以指数时间求解问题（1）的（低效）动态规划方法。尽管这一动态规划形式并不实用，但它为我们在第5节提出的高效剪枝策略奠定了基础。第6 节为所提出算法的正确性和运行时间提供理论保证。第7 节报告了合成实例以及来自 NAB 数据集的真实世界时间序列数据上的数值结果。最后，第8 节总结了主要贡献。

2 预备知识与背景

矩阵和向量分别用粗体大写和小写符号表示（例如， $\boldsymbol{Q}$ 和 $\boldsymbol{c}$），而标量则用非粗体符号表示（例如， n）。给定矩阵 $\boldsymbol{Q} \in \mathbb{R}^{n\times n}$和索引集 $\mathcal{I}$、$\mathcal{J} \subseteq \{1,\dots,n\}$，我们用$\boldsymbol{Q}_{\mathcal{I},\mathcal{J}}$表示 $\boldsymbol{Q}$的子矩阵，该子矩阵由索引为 $\mathcal{I}$ 的行和索引为 $\mathcal{J}$的列组成。类似地，对于向量 $\boldsymbol{c} \in \mathbb{R}^n$，我们用 $\boldsymbol{c}_{\mathcal{J}}$ 表示$\boldsymbol{c}$ 中索引属于 $\mathcal{J}$的子向量。我们用 $\mathbb{I}(x)$表示 $\mathbb{R}$上的指示函数；当 $x=0$ 时，其值为0，否则为1。对于对称矩阵 $\boldsymbol{Q} \in \mathbb{R}^{n\times n}$，令 $\mu_{\min}(\boldsymbol{Q})$和 $\mu_{\max}(\boldsymbol{Q})$分别表示其最小和最大特征值。 $\boldsymbol{Q}$ 的谱条件数定义为 $\kappa_2(\boldsymbol{Q}) := \mu_{\max}(\boldsymbol{Q})/\mu_{\min}(\boldsymbol{Q})$。类似地， $\boldsymbol{Q}$ 在诱导 ∞-范数下的条件数定义为 $\kappa_\infty(\boldsymbol{Q}) = \|\boldsymbol{Q}^{-1}\|_\infty \|\boldsymbol{Q}\|_\infty$。省略自变量时，即写作 $\mu_{\min}$、 $\mu_{\max}$、κ_2或 κ_∞时，这些量表示海森矩阵 $\boldsymbol{Q}$的特征值和条件数。 $\boldsymbol{S}$ 的逐元素ℓ_1,1-范数定义为$\|\boldsymbol{S}\|_{1,1} := \sum_{i=1}^n \sum_{j=1}^n |\boldsymbol{S}_{ij}|$。若对所有 $|i-j|>w$都有 $\boldsymbol{S}_{ij}=0$，则称 $\boldsymbol{S}$ 为带宽为 $w \in \mathbb{Z}_+$ 的带状矩阵。我们用 $f^\star$ 表示问题（1）的最优目标值，并用（$\boldsymbol{x}^\star$， $\boldsymbol{z}^\star$）表示其对应的最优解。

给定一个对称矩阵 $\boldsymbol{Q} \in R^{n\times n}$，其支撑图记为 $\mathrm{supp}(\boldsymbol{Q})$，它是一个简单无权图 $\mathsf{G} = (\mathcal{V}_\mathsf{G}, \mathcal{E}_\mathsf{G})$，顶点集为 $\mathcal{V}_\mathsf{G} = \{1,2,\dots,n\}$。当且仅当 $\boldsymbol{Q}_{ij} \neq 0$ for $i \neq j$ 时，边 $(i, j) \in \mathcal{E}_\mathsf{G}$ 存在。对于任意两个节点 $u, v \in \mathcal{V}_\mathsf{G}$，令 $\mathrm{dist}(u, v)$ 表示 $\mathrm{supp}(\boldsymbol{Q})$ 中 u 与 v 之间最短路径的长度。更一般地，对于节点子集 $\mathcal{I}, \mathcal{J} \subseteq \mathcal{V}_\mathsf{G}$，我们定义 $\mathrm{dist}(\mathcal{I}, \mathcal{J}) = \min_{i\in\mathcal{I},\, j\in\mathcal{J}} \{\mathrm{dist}(i, j)\}$。

2.1 树分解与树宽

定义 1 （树分解）。 图 $\mathsf{G} = (\mathcal{V}_\mathsf{G}, \mathcal{E}_\mathsf{G})$ 的树分解是一个二元组 $(\mathsf{T}\ \mathcal{B})$，其中 $\mathsf{T} = (\mathcal{V}_\mathsf{T}, \mathcal{E}_\mathsf{T})$ 是一棵树，并且 $\mathcal{B} = \{\mathcal{B}_u : u \in \mathcal{V}_\mathsf{T}\}$ 是 $\mathcal{V}_\mathsf{G}$ 的一个子集族（袋），满足以下条件：

1. $\mathcal{V}_\mathsf{G} = \bigcup_{u\in\mathcal{V}_\mathsf{T}} \mathcal{B}_u$。也就是说，G 的每个节点至少出现在一个袋中。

2. 对于每条边 $(i, j) \in \mathcal{E}_\mathsf{G}$，都存在一个袋 $\mathcal{B}_u$，其中同时包含 i 和 j。

3. 对于每个节点 $i \in \mathcal{V}_\mathsf{G}$，集合 $\{u \in \mathcal{V}_\mathsf{T} : i \in \mathcal{B}_u\}$ 在 T 中诱导出一棵连通子树。也就是说，包含 i 的袋在 T 中构成一棵连通子树。

树分解的宽度定义为 $\max_{u\in\mathcal{V}_\mathsf{T}} \{|\mathcal{B}_u| - 1\}$。一个图可能有许多不同的树分解。平凡分解将 $\mathcal{V}_\mathsf{G}$ 的所有顶点放入同一个袋中，得到宽度 $|\mathcal{V}_\mathsf{G}| - 1$。一个更有信息量的结构度量是树宽，记为 ω，定义为 G 的所有树分解中的最小宽度。直观而言，树宽衡量一个图与树的接近程度。例如，树的树宽为 1。串并联图的树宽为 2。另一方面，完全稠密图与树状结构相去甚远，其树宽达到最大值 $|\mathcal{V}_\mathsf{G}| - 1$。

接下来，我们引入平衡树分解的概念。

定义2 （平衡树分解）。 设 $(\mathsf{T}\ \mathcal{B})$ 是一个宽度为 ω 的树分解。若满足以下条件，则称 $(\mathsf{T}\ \mathcal{B})$ 为平衡树分解：

1. T 的每个袋都恰好包含 $\omega + 1$ 个节点。

2. 对于每条边 $(u,\ v)\ \in \mathcal{E}_{\mathsf{T}}$，对应的袋满足 $|\mathcal{B}_u \cap \mathcal{B}_v| = \omega$。

寻找宽度尽可能小的最优树分解——也就是宽度等于树宽的树分解 ω——是 NP困难问题 [6]。不过，对于树宽有界的图，最优树分解的识别和构造都可以在线性时间内完成 [20]。除了这些理论结果之外，诸如减少填充和嵌套剖分等实用启发式算法在实践中通常能够得到宽度接近最优的分解 [45]。此外，任何宽度为 ω 的树分解，都可以通过向已有袋中添加节点，并在必要时插入或移除袋，将其转换为宽度相同的平衡树分解。图1展示了一个示例。平衡树分解始终存在；给定一个宽度为 ω的树分解后，可以通过所谓的“优美”树分解在 $\mathcal{O}(n)$内得到；定义和构造方法见 [21, 第 2] 节。

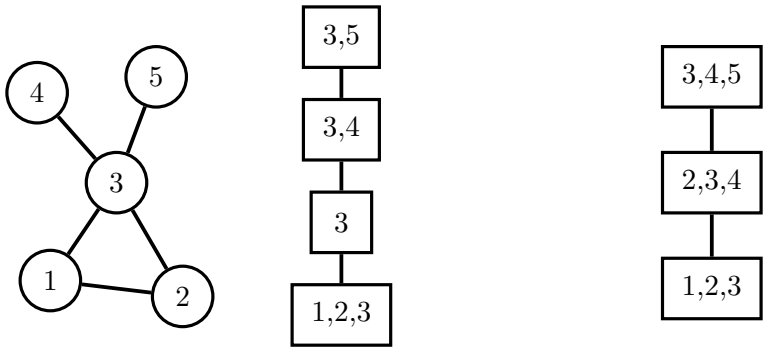

图1：一个违反定义2的图及其树分解（左），以及对应的平衡树分解（右）。

在本文中，我们假设已获得 supp(Q) 的一个宽度为 ω 的平衡树分解。例如，对于我们提出的 ESOC 表述，supp(Q) 的树宽为 2，且可以轻松获得对应的平衡树分解（见图 4 ）。需要指出的是，我们提出的方法不要求使用宽度最小的树分解；任何宽度足够小的分解均可满足要求。

接下来，我们描述一种过程，利用 supp(Q) 的平衡树分解 T 对其节点进行标记。后文将会说明，这种标记对于确立所提出的算法的效率至关重要。我们首先对 T 中的袋进行标记。根据平衡树分解的第二个性质，(T $\mathcal{B}$) 包含 $n-\omega$ 个袋。为简单起见，我们假设 T 中的边均自然地指向指定的根。在此约定下，每个袋可以有多个父节点，但至多有一个子节点。我们用 $\mathsf{child}_{\mathsf{T}}(u)$ 表示 T 中袋 u 的子节点，并用 $\mathsf{par}_{\mathsf{T}}(u)$ 表示其父节点集合。根据拓扑排序为袋赋予标签：对于每个袋 u，要求 $u < \mathsf{child}_{\mathsf{T}}(u)$。由于 T 包含 $n-\omega$ 个袋，根包获得标签 $n-\omega$。由于 T 是无环的，这种拓扑标号总是存在，并且可以在 $\mathcal{O}(n)$ 时间内计算，使用 [3, 内存。算法 3.8]。

现在，我们根据 supp(Q) 的已标记平衡树分解，为其中的节点标记。对于任意非根包 u，根据平衡树分解的定义，集合 $\mathcal{B}_u \setminus \mathcal{B}_{\mathsf{child}_{\mathsf{T}}(u)}$ 中包含 supp(Q) 的唯一一个节点，我们将其标记为 u。T 的根包中的节点则任意标记为剩余标签 $\{n-\omega+1,\ldots,n\}$。图 2 展示了这一标号方案。由于 $\mathcal{B}_u$ 和 $\mathcal{B}_{\mathsf{child}_{\mathsf{T}}(u)}$ 的大小均为 $\omega+1$，计算每个集合差

需要 $\mathcal{O}(\omega^2)$ 时间。对所有包重复此操作，总标记成本为 $\mathcal{O}(n\omega^2)$ 时间。所需内存为 $\mathcal{O}(n\omega)$，相当于存储每个包的标签。

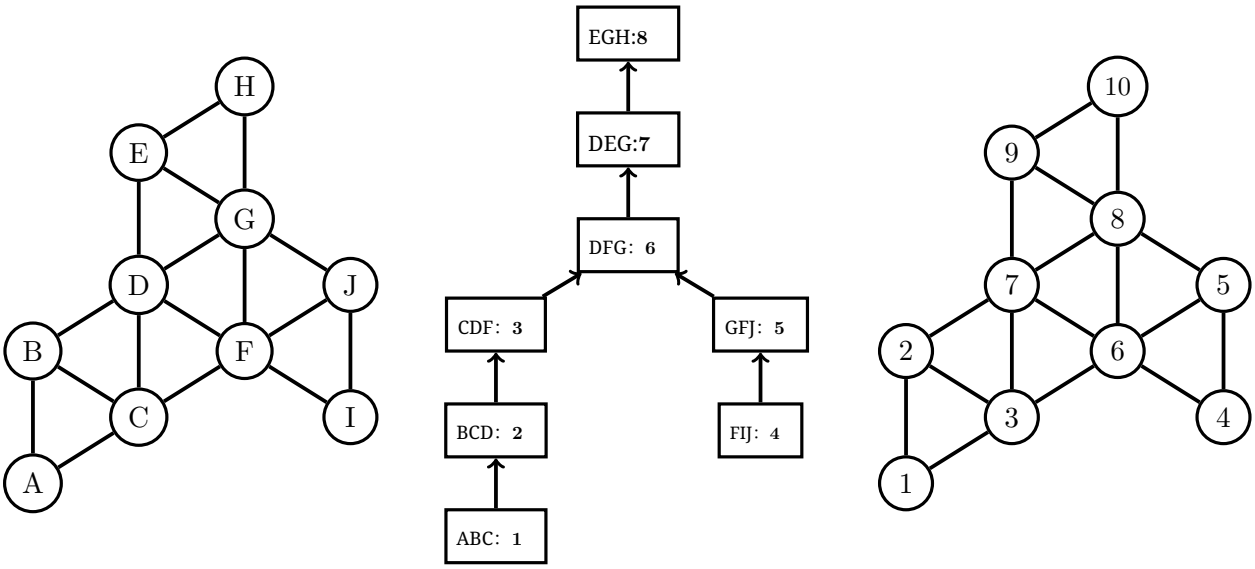

图 2：左图展示了一个节点任意标记的图 G。中间的面板展示了 G 的一种平衡树分解，其中各袋按照拓扑排序进行标记（标签以粗体显示）。右侧面板展示了通过所述过程得到的 G 中节点的标记。

对于任意满足 $u \le n-\omega$ 的节点 $u \in \mathsf{supp}(\boldsymbol{Q})$，我们将 $\mathsf{supp}_u(\boldsymbol{Q})$ 定义为 $\mathsf{supp}(\boldsymbol{Q})$ 的子图，该子图由包含在 T 中最大的子树内的节点诱导而成；该子树包含袋 $\mathcal{B}_u$ 及其所有祖先。对于 $u > n-\omega$，令 $\mathsf{supp}_u(\boldsymbol{Q}) := \mathsf{supp}(\boldsymbol{Q})$。对于任意 u，$\mathsf{supp}_u(\boldsymbol{Q})$ 的树宽不超过 ω [30, 引理 12.4.1]。例如，对于图 2 所示的图，$\mathsf{supp}_5(\boldsymbol{Q})$ 是由 $\mathcal{B}_5 \cup \mathcal{B}_4 = \{5,6,8\} \cup \{4,5,6\}$ 中的节点诱导的子图，而 $\mathsf{supp}_6(\boldsymbol{Q})$ 则由 $\bigcup_{i=1}^{6} \mathcal{B}_i$ 中的节点诱导。对于节点 $u \in \mathsf{supp}(\boldsymbol{Q})$，我们将 $\boldsymbol{Q}_{[u]}$ 记作由 $\mathsf{supp}_u(\boldsymbol{Q})$ 中节点索引的 $\boldsymbol{Q}$ 的主子矩阵；显然，$\mathsf{supp}(\boldsymbol{Q}_{[u]}) = \mathsf{supp}_u(\boldsymbol{Q})$。同理，$\boldsymbol{c}_{[u]}$ 和 $\boldsymbol{\lambda}_{[u]}$ 分别表示 $\boldsymbol{c}$ 和 $\boldsymbol{\lambda}$ 在这些索引上的子向量。我们用 $\mathcal{J}_u$ 表示 $\mathsf{supp}_u(\boldsymbol{Q})$ 中除去 $\mathcal{B}_u$ 中节点后的节点集合，并令 $n_u := |\mathcal{J}_u|$。回忆一下，树分解 T 包含$n-\omega$ 个袋，其中 $\mathcal{B}_{n-\omega}$ 指定为根。稍微滥用一下记号，我们定义辅助集合 $\mathcal{B}_u := \{u, u+1, \ldots, n\}$，其中 $u \in \{n-\omega+1, \ldots, n\}$。尽管这些集合不属于原始树分解，为方便起见，我们仍称它们为袋。对于 $u > n-\omega$，我们还定义 $\mathsf{par}_T(u) := \{u-1\}$。最后，对于任意 u，我们定义 $\tau_u := \min\{\omega, n-u\}$，从而确保 $|\mathcal{B}_u| = \tau_u + 1$。当 u 可由上下文明确时，我们简写为 τ。

2.2 局部参数化代价

回顾一下，$\boldsymbol{Q}_{[u]} \in \mathbb{R}^{n_u+\tau+1 \times n_u+\tau+1}$ 是 $\boldsymbol{Q}$ 的一个主子矩阵，由 $\mathcal{J}_u \cup \mathcal{B}_u$索引。令$\pi_u$: $\mathcal{J}_u \cup \mathcal{B}_u \to \{1, \ldots n_u+\tau+1\}$ 为规范索引映射，它将每个 $i \in \mathcal{J}_u \cup \mathcal{B}_u$映射到其在子矩阵 $\boldsymbol{Q}_{[u]}$中的对应行/列位置。对于任意 $u \in \{1, \ldots, n\}$ 和

$\tau = \min\{\omega, n-u\}$，局部参数化代价 f_u： $\mathbb{R}^{\tau+1} \to \mathbb{R}$定义为

$$f_u(\boldsymbol{\alpha}_{\mathcal{B}_u}) := \min_{\boldsymbol{x}\in\mathbb{R}^{n_u+\tau+1}, \boldsymbol{z}\in\{0,1\}^{n_u}} \quad \frac{1}{2}\boldsymbol{x}^\top \boldsymbol{Q}_{[u]}\boldsymbol{x} + \boldsymbol{c}_{[u]}^\top \boldsymbol{x} + \boldsymbol{\lambda}_{\mathcal{J}_u}^\top \boldsymbol{z} \tag{4a}$$

$$\text{s.t.} \quad \boldsymbol{x}_i(1-\boldsymbol{z}_i) = 0 \qquad i \in \pi_u(\mathcal{J}_u) \tag{4b}$$

$$\boldsymbol{x}_{\pi_u(\mathcal{B}_u)} = \boldsymbol{\alpha}_{\mathcal{B}_u}. \tag{4c}$$

我们注意到， $\boldsymbol{\alpha}_{\mathcal{B}_u}$ 中的下标在数学上并非必需；不过，正如后文将会看到的，它有助于使后续论证更加简洁清晰。直观来说， f_u 表示在将与 $\mathcal{B}_u$相关联的局部连续变量固定为 $\boldsymbol{\alpha}_{\mathcal{B}_u} \in \mathbb{R}^{\tau+1}$后，在 $\text{supp}_u(\boldsymbol{Q})$ 上定义的子问题的最优值。正如后文将要解释的，这种局部参数化代价在子问题之间建立了结构化联系，并使动态规划算法能够通过树分解高效地传递信息。

下面的引理证明，局部参数化代价可以表示为至多 2^{n_u} 个强凸二次函数的逐点最小值。该引理的证明见附录A.1。

引理1。任取 $u \in \{1, 2, \ldots, n\}$。局部参数化代价 f_u： $\mathbb{R}^{\tau+1} \to \mathbb{R}$ 可写为

$$f_u(\boldsymbol{\alpha}_{\mathcal{B}_u}) = \min_{\boldsymbol{s}\in\{0,1\}^{n_u}} \{p_{u,\boldsymbol{s}}(\boldsymbol{\alpha}_{\mathcal{B}_u})\} \tag{5}$$

其中，对于每个 $\boldsymbol{s} \in \{0,1\}^{n_u}$、 $p_{u,\ \boldsymbol{s}}(\boldsymbol{\alpha})$ 都是一个强凸二次函数。特别地，令$\mathcal{J}_{u,\ \boldsymbol{s}} = \{i \in \mathcal{J}_u \mid \boldsymbol{s}_i = 1\}$。则$p_{u,\ \boldsymbol{s}}(\boldsymbol{\alpha}_{\mathcal{B}_u})$ 由下式给出

$$p_{u,\boldsymbol{s}}(\boldsymbol{\alpha}_{\mathcal{B}_u}) = \frac{1}{2}\boldsymbol{\alpha}_{\mathcal{B}_u}\boldsymbol{A}_{u,\boldsymbol{s}}\boldsymbol{\alpha}_{\mathcal{B}_u} + \boldsymbol{b}_{u,\boldsymbol{s}}^\top \boldsymbol{\alpha}_{\mathcal{B}_u} + d_{u,\boldsymbol{s}},$$

$$where \begin{cases} \boldsymbol{A}_{u,\boldsymbol{s}} & = \boldsymbol{Q}_{\mathcal{B}_u,\mathcal{B}_u} - \boldsymbol{Q}_{\mathcal{B}_u,\mathcal{J}_{u,\boldsymbol{s}}} \left(\boldsymbol{Q}_{\mathcal{J}_{u,\boldsymbol{s}},\mathcal{J}_{u,\boldsymbol{s}}}\right)^{-1} \boldsymbol{Q}_{\mathcal{B}_u,\mathcal{J}_{u,\boldsymbol{s}}}^\top \\ \boldsymbol{b}_{u,\boldsymbol{s}} & = \boldsymbol{c}_{\mathcal{B}_u} - \boldsymbol{c}_{\mathcal{J}_{u,\boldsymbol{s}}}^T \left(\boldsymbol{Q}_{\mathcal{J}_{u,\boldsymbol{s}},\mathcal{J}_{u,\boldsymbol{s}}}\right)^{-1} \boldsymbol{Q}_{\mathcal{B}_u,\mathcal{J}_{u,\boldsymbol{s}}}^\top \\ d_{u,\boldsymbol{s}} & = -\frac{1}{2}\boldsymbol{c}_{\mathcal{J}_{u,\boldsymbol{s}}}^\top (\boldsymbol{Q}_{\mathcal{J}_{u,\boldsymbol{s}},\mathcal{J}_{u,\boldsymbol{s}}})^{-1}\boldsymbol{c}_{\mathcal{J}_{u,\boldsymbol{s}}} + \sum_{i\in\mathcal{J}_{u,\boldsymbol{s}}} \boldsymbol{\lambda}_i. \end{cases}$$

由于 f_u 可以表示为一组二次函数的最小值，原则上，可以通过保存与每个二次片段 $p_{u,\boldsymbol{s}}$ 对应的系数（$\boldsymbol{A}_{u,\boldsymbol{s}}$,$\boldsymbol{b}_{u,\boldsymbol{s}}$,$d_{u,\boldsymbol{s}}$）将其存储在内存中。存储每个 $p_{u,\boldsymbol{s}}$ 的系数需要 $\mathcal{O}(\omega^2)$ 内存。然而，由于稀疏模式 $\boldsymbol{s}$ 的数量会随着 $\text{supp}_u(\boldsymbol{Q})$ 中节点 n_u 的数量呈指数级增长，存储 f_u 所需的总内存也会呈指数级增长。显然，这种直接存储方式并不实用。

当 $\text{supp}(\boldsymbol{Q})$ 是一棵树（即 $\omega = 1$）时，Bhathena 等人 [14] 证明，参数化成本 f_u 可以表示为仅有 $\mathcal{O}(n_u)$ 个二次片段的最小值，因此只需线性内存。 然而，对于一般情形 $\omega > 1$，是否存在 f_u 的紧凑表示仍是一个开放问题。在本研究中，我们利用问题的其他结构性质来应对这一挑战，例如其体积增长；这些性质将在下文介绍。

2.3 图的体积增长

对于 $\text{supp}(\boldsymbol{Q})$ 中的任意节点 u ，定义 $\mathcal{V}_{u,\ m} := \{\, i \in \mathcal{J}_u \mid \text{dist}(i,\ \mathcal{B}_u) \le m \,\}$ 为 $\mathcal{J}_u$ 中距袋 $\mathcal{B}_u$ 不超过 m 的节点集合。该集合称为 $\mathcal{B}_u$ 在诱导子图 $\text{supp}_u(\boldsymbol{Q})$ 中的 m-邻域 。令 $\Delta_m := \max_{u\in\{1,\ldots,n\}} |\mathcal{V}_{u,\ m}|$。 $\text{supp}(\boldsymbol{Q})$ 的体积增长函数 刻画了 Δ_m 随 m 增长的速度。

假设 **tion 1 (多项式体积增长)。** 存在常数 γ, $\delta > 0$ 使得

$$\Delta_m \leq \delta\, m^{\gamma} \qquad \text{for all } m \geq 1.$$

假设 1 与一致多项式体积增长的概念密切相关 [31, 49],尽管精确定义可能略有不同。该假设对许多经过充分研究的图族成立。例子包括具有 $\mathcal{O}(1)$ 个叶子的树、环图、网格图，以及带状矩阵的支撑图。相比之下，具有 $\mathcal{O}(n)$ 个叶子的树可能不满足该假设。值得注意的例子是完全二叉树，对其有 $\Delta_m = \mathcal{O}(2^m)$，以及星图，其中 $\Delta_1 = n - 2$。

3 应用：带离群值校正的指数平滑

为引出我们针对结构化图上凸二次优化的方法框架，现介绍时间序列分析中的一项实际应用。具体而言，我们采用将预测与异常值检测融为一体的统一视角。在经典指数平滑这一简单而灵活的时间序列预测方法的基础上，我们将其扩展为基于混合整数二次规划的表述，以有效处理离群值。下一小节回顾指数平滑并讨论其局限性，从而引出我们提出的扩展。

3.1 指数平滑

指数平滑涵盖了一系列方法，广泛应用于运筹学中具有不同特征（如水平、趋势和季节性）的时间序列。通过为过去的观测值分配指数级递减的权重，这些方法能够平滑短期波动，同时保留长期结构，因此对于动态环境中的预测和决策具有重要价值。其应用范围涵盖零售和库存管理 [59, 17, 36], 市场分析与政策设计 [47, 66], 以及质量控制 [65]；更多示例请参见 [46] 。

在各种指数平滑方法中，简单指数平滑 （SES）——也称为单指数平滑——提供了一种基础形式。形式化地，设 $\{\boldsymbol{y}_t\}_{t=1}^T$表示一个单变量时间序列信号。平滑序列 $\{\boldsymbol{x}_t\}_{t=1}^T$ 递归定义如下：

$$\boldsymbol{x}_t = \beta \boldsymbol{y}_t + (1-\beta)\boldsymbol{x}_{t-1} \quad \text{for } t = 2, \ldots, T; \tag{SES}$$

初始值为 $\boldsymbol{x}_1 = \boldsymbol{y}_1$。平滑参数 $\beta \in$ （0，1）控制分配给最新观测值的权重。 β 取值越大，该方法对变化的响应性越强，但也越容易受到噪声影响；取值越小，输出越平滑，但响应性会有所降低。

为说明 SES 的局限性，我们考察云服务器网络流量的时间序列，该流量以字节为单位每五分钟测量一次。此数据集由 Amazon CloudWatch 收集，并可通过 Numenta异常基准（NAB）公开获取 [1]。图3（上排）展示了将 SES 应用于该时间序列的结果，其中平滑因子为 $\beta = 0.5$、0。2、0。05。圈出的点表示 NAB 标记的离群值。当 β取值较大时，平滑后的序列更贴近原始数据，但也因此无法排除离群值。反之，当 β 取值较小时，离群值的影响会减弱，但平滑结果相对于底层序列会出现滞后。这一现象揭示了 SES 的一个根本局限：它无法同时抑制离群值并快速适应底层信号的变化。鉴于 SES 固有的这种权衡，我们提出一种鲁棒扩展，称为带离群值校正的指数平滑（ESOC）。我们主要关注 SES，但也指出，我们的形式化方法可以自然地扩展到其他变体，包括双重指数平滑和三重指数平滑。

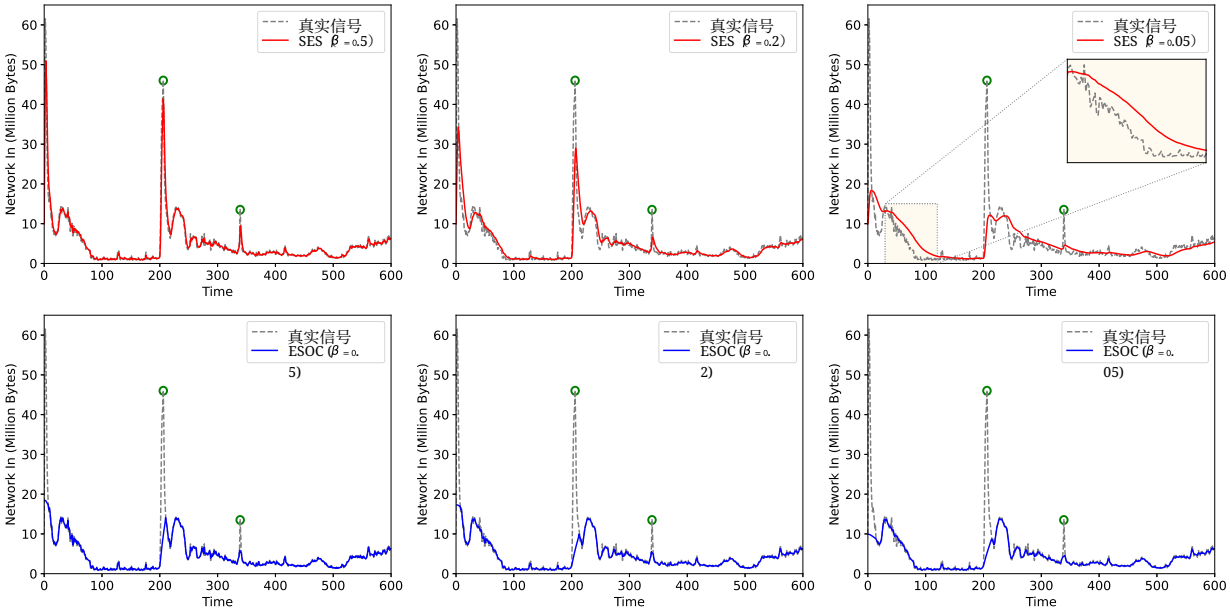

图3：**上排：**将 SES 应用于包含来自 Numenta异常基准（NAB）的异常的信号，平滑参数分别为 $\beta = 0.5$、0.2 和 0.05（从左到右）。由 NAB真实标签识别出的离群值以圆圈标出。较低的 β 值会降低对离群值的敏感度，但会导致相对于底层信号出现明显滞后。**下排：**ESOC 能够在所有 β 值下有效检测并去除离群值，同时保持与真实信号对齐并避免滞后。

3.2 带离群值校正的指数平滑

现在，我们介绍一种 SES 的混合整数二次规划扩展，它明确考虑了离群值。在此框架中，向量 $\boldsymbol{x} \in \mathbb{R}^T$ 表示平滑信号，而 $\boldsymbol{o} \in \mathbb{R}^T$ 是用于捕捉离群值的稀疏向量。相应的优化问题为

$$\begin{aligned} \min_{\boldsymbol{x},\boldsymbol{o}\in\mathbb{R}^T,\boldsymbol{z}\in\{0,1\}^T} \quad & \sum_{t=1}^{T} (\boldsymbol{y}_t - \boldsymbol{x}_t - \boldsymbol{o}_t)^2 + \sum_{t=1}^{T} \boldsymbol{\lambda}_t \boldsymbol{z}_t \\ \text{s.t.} \quad & \boldsymbol{x}_t = \beta(\boldsymbol{y}_t - \boldsymbol{o}_t) + (1-\beta)\boldsymbol{x}_{t-1} && \text{for } t = 2, \ldots, T \\ & \boldsymbol{o}_t(1 - \boldsymbol{z}_t) = 0 && \text{for } t = 1, \ldots, T, \end{aligned}$$

其中，$\boldsymbol{z} \in \{0,1\}^T$ 捕捉 $\boldsymbol{o} \in \mathbb{R}^T$ 的稀疏模式，并由非负正则化向量 $\boldsymbol{\lambda} \in \mathbb{R}^T$ 控制。当 $\boldsymbol{z}_t = 0$ 时，样本 $\boldsymbol{y}_t$ 被视为无噪声，从而得到标准的简单指数平滑递推。当 $\boldsymbol{z}_t = 1$ 时，变量 $\boldsymbol{o}_t$ 可以取非零值，因此能够校正含噪观测 $\boldsymbol{y}_t$。

我们不将指数平滑动态作为硬约束，而是采用松弛形式，在目标函数中对偏离该约束的情况进行惩罚，如下所示：

:

$$\min_{\boldsymbol{x},\boldsymbol{o}\in\mathbb{R}^T} \sum_{t=1}^{T}(\boldsymbol{y}_t-\boldsymbol{x}_t-\boldsymbol{o}_t)^2+\sum_{t=1}^{T}\boldsymbol{\lambda}_t\boldsymbol{z}_t+\mu_1\sum_{t=2}^{T}(\beta(\boldsymbol{y}_t-\boldsymbol{o}_t)+(1-\beta)\boldsymbol{x}_{t-1}-\boldsymbol{x}_t)^2+\mu_2\|\boldsymbol{o}\|_2^2 \quad \text{(ESOC)}$$
$$\text{s.t.} \quad \boldsymbol{o}_t(1-\boldsymbol{z}_t)=0 \qquad \text{for } t=1,\dots,T,$$

其中 $\mu_1 \geq 0$ 用于惩罚偏离指数平滑动态的情况。附加项$\mu_2\|\boldsymbol{o}\|_2^2$，其中 $\mu_2 \geq 0$，可确保目标函数的海森矩阵为正定。需要指出的是，尽管这种正则化可能会使 $\boldsymbol{o}$略微收缩， μ_2 但在实践中，其取值非常小，因此影响可以忽略不计，同时还能显著提升数值稳定性。由此得到的松弛形式符合问题（1）的结构，其海森矩阵的稀疏模式的树宽为 2，且随 $\Delta_m \leq 3m$ 呈线性体积增长（见图4）。因此，（ESOC）属于可使用本文提出的参数化算法高效求解的问题类别。

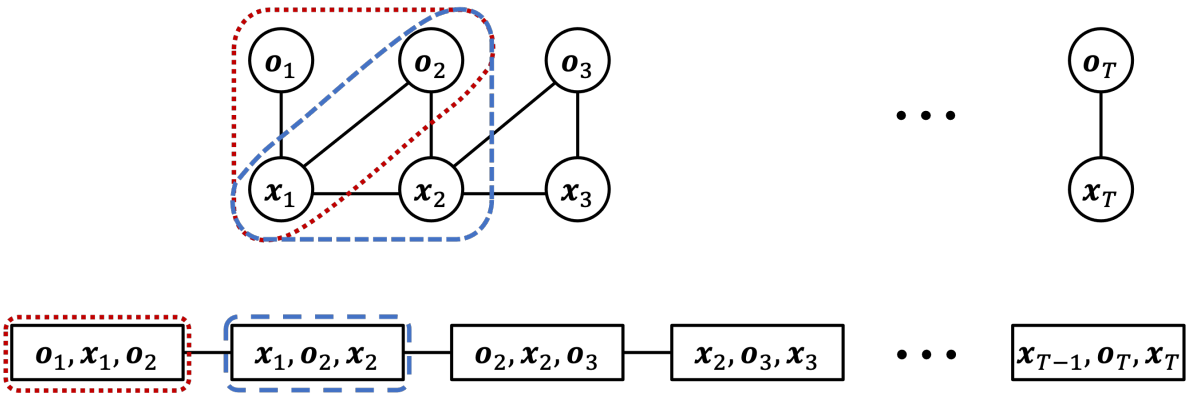

图4：ESOC 模型的海森矩阵支撑图（上）及对应的树分解（下）。

本节最后给出初步的实证结果；详细的实验研究见第 7.3 节。图3（第二行）展示了通过我们的参数化算法，在不同平滑参数 β 下得到的 ESOC 解。如前所述，SES 在异常值敏感性和响应性之间存在权衡。相比之下，ESOC 在所有 β 取值下都能有效抑制离群值，同时不牺牲响应性。为量化这一比较，我们报告两种方法对应的预测均方误差（均方误差）值。对于$\beta \in \{0.5, 0.2, 0.05\}$，SES 的均方误差值分别为 16.89、9.87 和 25.70；ESOC 对应的值则为 0.32、0.46 和 0.22。在每种情况下，ESOC 的均方误差都更低，意味着预测准确性有所提升。

接下来，我们评估ESOC的计算可扩展性。具体而言，我们使用 $\beta \in \{0.05, 0.2, 0.5\}$，将模型应用于五个来自 NAB 数据集的真实时间序列，包括 AWS CloudWatch 指标和交通速度数据。对于每个由此产生的优化问题，我们分别使用 GUROBI（一种先进的商用混合整数二次规划求解器）和我们所提出的算法进行求解。为了评估可扩展性，我们通过截取各时间序列来改变问题规模 T，并记录相应的求解时间。对于每种 T，我们求解 15 个实例（五个信号，每个信号对应三个 β 取值）。图5 报告了这些运行时间比较结果：我们的算法能够将所有实例求至最优，并且平均运行时间为

在 $T = 1000$ 时为 54 秒，而 GUROBI 的运行时间会随着 T 大幅增加，并且当超过 $T = 500$ 后会超出一小时的限制。

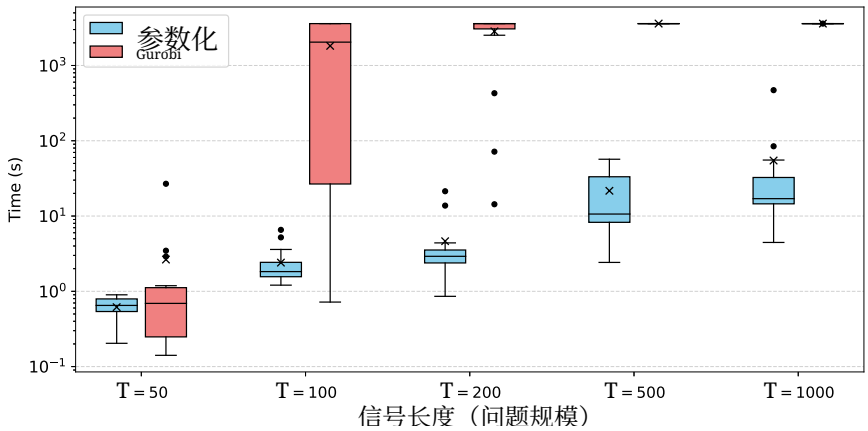

图5：所提出的算法与 GUROBI 在 NAB 数据集中的五个真实时间序列上的运行时间比较，使用平滑参数。通过截断改变信号长度。对于每个信号，平滑参数均设为 $\beta \in \{0.05, 0.2, 0.5\}$。“×”表示平均运行时间。GUROBI在运行一小时后终止，因此在较大的信号长度下取值保持平坦。

4 通过局部参数化代价进行动态规划

回顾式 (4) 中局部参数化代价 f_u 的定义。一个关键观察是，一旦局部变量 $\boldsymbol{x}_{\pi_u(\mathcal{B}_u)}$ 固定为 $\boldsymbol{\alpha}_{\mathcal{B}_u}$，定义在 $\text{supp}_u(\boldsymbol{Q})$ 上的子问题便分解为相互独立的分量，每个分量都与 T 中以 u 的某个父节点为根的子树内的变量相关联。这种分解直接源于树分解的运行交性质：如果节点 v 出现在两个袋 $\mathcal{B}_i$ 和 $\mathcal{B}_j$ 中，那么它也必须出现在 T 中连接 $\mathcal{B}_i$ 与 $\mathcal{B}_j$ 的路径上的每个袋中。因此，子图 $\text{supp}_u(\boldsymbol{Q})$ 中的节点仅通过袋 $\mathcal{B}_u$ 中的节点与图的其余部分相连，而固定 $\boldsymbol{x}_{\pi_u(\mathcal{B}_u)}$ 会消除所有剩余的耦合。以下引理将这一观察形式化，并给出局部参数化代价 f_u 的显式递归表示，该表示由树分解中 u 的父节点所对应的量构成。该引理的证明见附录 A.2。

引理2。 对于任意节点 u，局部参数化代价 $f_u: \mathbb{R}^{\tau+1} \to \mathbb{R}$ 满足

$$f_u(\boldsymbol{\alpha}_{\mathcal{B}_u}) = h_u(\boldsymbol{\alpha}_{\mathcal{B}_u}) + \sum_{v \in \text{par}_T(u)} \left(g_v(\boldsymbol{\alpha}_{\mathcal{B}_v \setminus v}) - \phi_v(\boldsymbol{\alpha}_{\mathcal{B}_v \setminus v}) \right), \tag{6}$$

其中，函数 $h_u: \mathbb{R}^{\tau+1} \to \mathbb{R}$ 和 $g_v, \phi_v: \mathbb{R}^{\tau} \to \mathbb{R}$ 定义如下

$$h_u(\boldsymbol{\alpha}_{\mathcal{B}_u}) := \frac{1}{2} \boldsymbol{\alpha}_{\mathcal{B}_u}^\top \boldsymbol{Q}_{\mathcal{B}_u, \mathcal{B}_u} \boldsymbol{\alpha}_{\mathcal{B}_u} + \boldsymbol{c}_{\mathcal{B}_u}^\top \boldsymbol{\alpha}_{\mathcal{B}_u}, \tag{7a}$$

$$g_v(\boldsymbol{\alpha}_{\mathcal{B}_v \setminus v}) := \min_{\boldsymbol{x}_v \in \mathbb{R}} \left\{ f_v(\boldsymbol{x}_v, \boldsymbol{\alpha}_{\mathcal{B}_v \setminus v}) + \boldsymbol{\lambda}_v \mathbb{I}(\boldsymbol{x}_v) \right\}, \tag{7b}$$

$$\phi_v(\boldsymbol{\alpha}_{\mathcal{B}_v \setminus v}) := \frac{1}{2} \boldsymbol{\alpha}_{\mathcal{B}_v \setminus v}^\top \boldsymbol{Q}_{\mathcal{B}_v \setminus v, \mathcal{B}_v \setminus v} \boldsymbol{\alpha}_{\mathcal{B}_v \setminus v} + \boldsymbol{c}_{\boldsymbol{\alpha}_{\mathcal{B}_v \setminus v}}^\top \boldsymbol{\alpha}_{\boldsymbol{\alpha}_{\mathcal{B}_v \setminus v}}. \tag{7c}$$

上述方程可解释为 u 阶段的 DP 更新 [10, 第 2], 章，它将袋 $\mathcal{B}_u$ 的局部参数化代价（或所谓的后续代价）表示为其父袋 $\{f_v$ 的局部参数化代价的函数： $v \in \text{par}_\mathcal{T}(u)\}$。(6) 背后的直觉很自然：要刻画局部参数化代价 f_u，需要根据其父袋中的局部代价进行三项调整。首先，必须消去 $\mathcal{B}_v$ 中未出现在 $\mathcal{B}_u$ 中的变量，这通过函数 g_v 定义中的 $\boldsymbol{\alpha}_v$ 最小化来实现。其次，必须加上与 $\mathcal{B}_u$ 中变量相关的代价，这部分由函数 h_u 表示。最后，通过 ϕ_v 去除与 $\mathcal{B}_v \backslash v$ 中剩余变量相关的代价，以避免在各父袋之间重复计数。

由此自然得到一种动态规划算法：该算法从叶节点向根节点遍历树分解，并在每个袋 $\mathcal{B}_u$处递归计算参数化成本 f_u 和函数 g_u。由于f_u 定义为凸二次函数的最小值，因此 g_u 也可以表示为二次函数的最小值，但这些二次函数不一定是凸函数。此外，如果 f_u 由 N 个二次片段组成，那么 g_u 至多包含 $2N$ 个二次片段。片段数量翻倍，是因为 g_u的定义中包含指示函数 $\mathbb{I}(\boldsymbol{x}_v)$；该指示函数会根据 $\mathbb{I}(\boldsymbol{x}_v)$的取值，将 f_u 的每个二次片段拆分为两个不同的二次函数。此外，标号方案保证，在计算任何依赖于某个 g_u 的参数化成本之前，都会先计算该 g_u。计算出f_n后，即可得到最优代价：

$$f^\star = \min_{\boldsymbol{\alpha}_n \in \mathbb{R}} \{f_n(\boldsymbol{\alpha}_n) + \boldsymbol{\lambda}_n \mathbb{I}(\boldsymbol{\alpha}_n)\}.$$

得到最优代价后，可以从根包向叶袋遍历，恢复其余局部参数。该动态规划算法见算法1。下述命题证明了该动态规划算法的正确性和计算复杂度。请注意，该算法的复杂度关于二元决策变量的数量呈指数增长 n，因此按当前形式可能不实用。

算法 1 问题 (1) 的动态规划

输入： 问题 (1) 以及 $\text{supp}(\boldsymbol{Q})$ 的平衡树分解 $\omega\mathsf{T}$，其宽度为 ω；**输出：** 问题 ($f^\star$1) 的最优代价及其最优解 $(\boldsymbol{x}^\star, \boldsymbol{z}^\star)$。

1：按照第2.1节中描述的方案为 $\text{supp}(\boldsymbol{Q})$ 的节点编号；2：设置 $f_1(\boldsymbol{\alpha}_{\mathcal{B}_1}) = \frac{1}{2}\boldsymbol{\alpha}_{\mathcal{B}_1}^\top \boldsymbol{Q}_{\mathcal{B}_1}$, $_{\mathcal{B}_1}\boldsymbol{\alpha}_{\mathcal{B}_1} + \boldsymbol{\alpha}_{\mathcal{B}_1}^\top \boldsymbol{c}_{\mathcal{B}_1}$.3:– 对于 $u = 1, \ldots, n-1$，执行 4：通过公式 (7b) 从 f_u 计算 g_u；5：通过公式 (6) 计算 f_{u+1}；6：循环结束 7：得到 $f^\star = \min_{\boldsymbol{\alpha}_n \in \mathbb{R}} \{f_n(\boldsymbol{\alpha}_n) + \boldsymbol{\lambda}_n \mathbb{I}(\boldsymbol{\alpha}_n)\}$, $\boldsymbol{x}_n^\star = \text{argmin}_{\boldsymbol{\alpha}_n \in \mathbb{R}} \{f_n(\boldsymbol{\alpha}_n) + \boldsymbol{\lambda}_n \mathbb{I}(\boldsymbol{\alpha}_n)\}$，以及 $\boldsymbol{z}_n^\star = \mathbb{I}(\boldsymbol{x}_n^\star)$; 8：对于 $u = n-1, \ldots, 1$，执行 9：设置 $\boldsymbol{x}_u^\star = \text{argmin}_{\boldsymbol{\alpha}_u \in \mathbb{R}} \left\{ f_u(\boldsymbol{\alpha}_u, \boldsymbol{x}_{\mathcal{B}_u \backslash u}^\star) + \boldsymbol{\lambda}_u \mathbb{I}(\boldsymbol{\alpha}_u) \right\}$ 和 $\boldsymbol{z}_u^\star = \mathbb{I}(\boldsymbol{x}_u^\star)$; 10：循环结束 11：返回 $f^\star$ 和 $(\boldsymbol{x}^\star, \boldsymbol{z}^\star)$;

命题 1. 算法 *1* 恢复问题 (1) 的最优解，所需 $O(\delta\ \omega^2\ n\ 2^n)$ 时间以及 $O(\omega^2\ n\ 2^n)$ 内存。

证明。 我们从正确性证明开始。

正确性证明。 该算法计算局部参数化代价 f_u，对应于每个节点 $u \in \{1, \dots, n\}$。对于基本情况 $u = 1$，不存在先前的子问题；因此，该算法计算 f_1，依据第 2 行，这与式（4）中定义的局部参数化代价一致。对于 $u > 1$，该算法归纳应用引理 2，从而正确构造局部参数化代价函数 f_u。在构造 f_n 之后，该算法计算最优代价$f^\star$，并通过回溯局部参数化代价得到一个最优解（$\boldsymbol{x}^\star$，$\boldsymbol{z}^\star$）。

复杂度证明。 我们通过逐步检查算法来分析其运行时间。第 1 行的标记步骤需要 $\mathcal{O}(n\omega^2)$ 时间和 $\mathcal{O}(n\omega)$ 内存，如第 2.1节所示。初始化第一个局部参数化代价 f_1 需要 $\mathcal{O}(\omega^2)$ 时间和内存（第 2 行）。

接下来，我们分析第一个 `for` 循环（第 3–6 行）的复杂度。对于每个 $u = 1, \dots, n-1$，函数 g_u 可通过分别对 f_u 的每个片段在有和无指示变量的情况下求最小值得到。由于 f_u 至多包含 2^{n_u} 个片段（引理 1），计算 g_u 需要$\mathcal{O}((2\omega^2)2^{n_u}) = \mathcal{O}(\omega^2\ 2^{n_u})$ 时间和内存。在第 5 行中，函数 f_{u+1} 根据（6）计算，为方便起见，我们在此重写如下：

$$f_{u+1}(\boldsymbol{\alpha}_{\mathcal{B}_{u+1}}) = h_{u+1}(\boldsymbol{\alpha}_{\mathcal{B}_{u+1}}) + \sum_{v \in \mathrm{par}_{\mathsf{T}}(u+1)} \Big(g_v(\boldsymbol{\alpha}_{\mathcal{B}_v \setminus v}) - \phi_v(\boldsymbol{\alpha}_{\mathcal{B}_v \setminus v}) \Big).$$

函数 h_{u+1} 和集合 $\{\phi_v \colon v \in \mathrm{par}_{\mathsf{T}}(u+1)\}$ 是单片段二次函数。它们分别利用方程（7a）和（7c）计算，每个都需要 $\mathcal{O}(\omega^2)$ 时间和内存。此外，函数 $\{g_v \colon v \in \mathrm{par}_{\mathsf{T}}(u+1)\}$ 已在第 4 行计算完成。构造 f_{u+1} 的一个片段时，需要从每个父项 $g_v - \phi_v$ 中各选取一个片段，其中 $v \in \mathrm{par}_{\mathsf{T}}(u+1)$均满足条件。由于 $|\mathrm{par}_{\mathsf{T}}(u+1)| \le \Delta_1 \le \delta$，计算每个片段需要 $\mathcal{O}(\delta\ \omega^2)$ 时间和$\mathcal{O}(\omega^2)$ 内存。根据引理1，f_{u+1} 至多有 $2^{n_{u+1}}$ 个片段；因此，计算 f_{u+1}的总成本为 $\mathcal{O}(\delta\ \omega^2\ 2^{n_{u+1}})$ 时间和 $\mathcal{O}(\omega^2\ 2^{n_{u+1}})$ 内存。由于第一个 for 循环迭代 $n-1$ 次，其成本为 $\mathcal{O}\big(\sum_{u=1}^{n-1} \delta\ \omega^2\ 2^{n_{u+1}}\big) = \mathcal{O}(n\ \delta\ \omega^2\ 2^n)$ 时间和 $\mathcal{O}(n\ \omega^2\ 2^n)$ 内存。

获取 $f^\star$、$\boldsymbol{x}_n^\star$和 $\boldsymbol{z}_n^\star$ 需要第 7 行的 $\mathcal{O}(2^n)$ 时间，因为 f_n 最多由 2^n 个片段组成。第二个 `for` 循环（第 8–10 行）的每次迭代需要 $\mathcal{O}(\omega^2\ 2^{n_u})$ 时间；为简洁起见，我们省略细节。综合以上步骤，可知该算法的运行时间为 $\mathcal{O}(\delta\ \omega^2\ n\ 2^n)$，内存占用为$\mathcal{O}(\omega^2\ n\ 2^n)$。

□

5 剪枝

用于求解问题（1）的动态规划方法很快可能变得难以处理，因为刻画局部参数化代价所需的二次函数数量可能呈指数级增长。本节旨在应对这一挑战。我们首先给出以下引理，证明问题（1）的最优解位于一个有界区域内。证明依赖于稍后介绍的一个结果（引理 5），因此推迟到附录 A.3。

引理 3. 设 （$\boldsymbol{x}^\star$、$\boldsymbol{z}^\star$）是问题 （1）的一个最优解。定义 $C_1 \colon = \frac{1}{\mu_{\min}}$最大值$\left\{1, \frac{(1+\sqrt{\kappa_2})^2}{2\kappa_2}\right\}$ 以及 $\rho \colon = \frac{\sqrt{\kappa_2}-1}{\sqrt{\kappa_2}+1}$。我们有 $\|\boldsymbol{x}^\star\|_\infty \le U$，其中：

- $U = \frac{2wC_1}{1-\rho} \|\boldsymbol{c}\|_\infty$ 若 $\boldsymbol{Q}$ 是带宽为 w的带状矩阵；
- $U = \frac{\delta\gamma! C_1}{(1-\rho)^{\gamma+1}} \|\boldsymbol{c}\|_\infty$ 如果满足多项式体积增长（假设1）。

上述引理的一个重要特点是，它为最优解提供逐元素的 ℓ_∞范数界，而不是更常见的 ℓ_2范数界。正如后文将解释的，这种更精细的逐坐标控制对我们的算法设计至关重要。Bertsimas等人 [12,定理 2。1] 也推导出了 ℓ_∞范数界，但他们的结果没有利用 $\boldsymbol{Q}$的稀疏结构，因此其规模会随问题维数增长。相比之下，引理3利用了这一结构，并在给定假设下避免了显式的维数依赖性。

尽管如此，引理3 中的界在实践中仍可能过于保守。在许多应用中，很容易获得更紧的实例特定界。例如，在带异常值校正的指数平滑模型中，可以安全地将 $U = \text{最大值}_t |\boldsymbol{y}_t|$设为观测信号绝对值的最大值。这类界通常要紧得多。我们的计算结果表明，即使是一般界在实践中也依然有效，但采用更紧的界可带来显著的计算收益。

虽然要刻画整个 $\mathbb{R}^{\tau+1}$ 上的局部参数化代价 f_u 可能需要指数级数量的二次函数，但上述引理表明，只需在有界区域 $\mathcal{D} = \{\boldsymbol{x}$ 内刻画该函数即可：$\|\boldsymbol{x}\|_\infty \le U\}$。回顾（5），这意味着，任何满足 $p_{u,\boldsymbol{s}}(\boldsymbol{\alpha}_{\mathcal{B}_u}) > f_u(\boldsymbol{\alpha}_{\mathcal{B}_u})$ 在区域 $\mathcal{D}_u = \{\boldsymbol{\alpha}_{\mathcal{B}_u}: \|\boldsymbol{\alpha}_{\mathcal{B}_u}\|_\infty \le U\}$ 内成立的二次函数 $p_{u,\boldsymbol{s}}$，都可以安全地舍弃。

定义3（相关与不相关函数）。设 $f_u(\boldsymbol{\alpha}_{\mathcal{B}_u}) = \min_{s\in\{0,1\}^{n_u}} \{p_{u,\boldsymbol{s}}(\boldsymbol{\alpha}_{\mathcal{B}_u})\}$。若函数 $p_{u,\boldsymbol{s}}$ 满足 $p_{u,\boldsymbol{s}}(\boldsymbol{\alpha}_{\mathcal{B}_u}) > f_u(\boldsymbol{\alpha}_{\mathcal{B}_u})$ 在区域 $\mathcal{D}_u = \{\boldsymbol{\alpha}_{\mathcal{B}_u}: \|\boldsymbol{\alpha}_{\mathcal{B}_u}\|_\infty \le U\}$ 内成立，则称其为不相关函数。反之，若 $p_{u,\boldsymbol{s}}(\boldsymbol{\alpha}_{\mathcal{B}_u}) = f_u(\boldsymbol{\alpha}_{\mathcal{B}_u})$ 对某个 $\boldsymbol{\alpha}_{\mathcal{B}_u} \in \mathcal{D}_u$ 成立，则称 $p_{u,\boldsymbol{s}}$ 为相关函数。

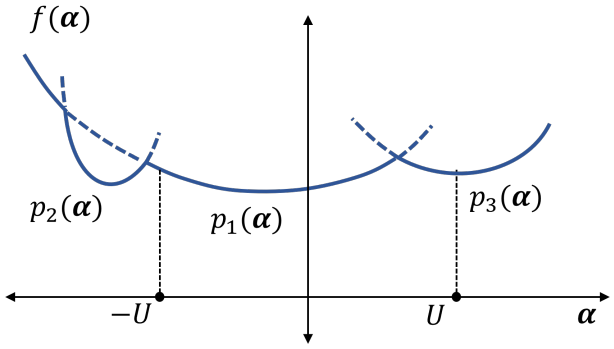

图 6：分段二次函数 $f(\boldsymbol{\alpha}) = \min\{p_1(\boldsymbol{\alpha}) p_2(\boldsymbol{\alpha}) p_3(\boldsymbol{\alpha})\}$ 由蓝色实线表示。这里，p_2是不相关的，因为对于每个 $-U \le \boldsymbol{\alpha} \le U$，$p_2(\boldsymbol{\alpha}) > f(\boldsymbol{\alpha})$。

图6 展示了一个用于区分相关函数和不相关函数的示例。为了区分二者，我们给出以下引理，它为两个二次函数的交点提供了一个下界。该引理的证明见附录 A.4。

引理 4 p_1 p_2 $\mathbb{R}^{\tau+1} \to \mathbb{R}$ $p_1(\boldsymbol{\alpha}) = \frac{1}{2}\boldsymbol{\alpha}^\top \boldsymbol{A}_1 \boldsymbol{\alpha} + \boldsymbol{b}_1^\top \boldsymbol{\alpha} +$。设，：是两个形式为 ( ) : d_1 $p_2(\boldsymbol{\alpha}) = \frac{1}{2}\boldsymbol{\alpha}^\top \boldsymbol{A}_2 \boldsymbol{\alpha} + \boldsymbol{b}_2^\top \boldsymbol{\alpha} + d_2$ $\bar{a} \ge \frac{1}{2}\|\boldsymbol{A}_1 - \boldsymbol{A}_2\|_{1\,1}$ $\bar{b} \ge \|\boldsymbol{b}_1 - \boldsymbol{b}_2\|_1$ $|d_1 - d_2| \ge \bar{d}$ 和( ) : 的二次函数。设¯„ 和。

设 $\hat{\boldsymbol{\alpha}}$ 是方程 $p_1(\boldsymbol{\alpha}) - p_2(\boldsymbol{\alpha}) = 0$的一个实根。若不存在实根，则令 $\hat{\boldsymbol{\alpha}} = +\infty$。于是，我们有

$$\|\hat{\boldsymbol{\alpha}}\|_\infty \geq L(p_1, p_2), \qquad where\ L(p_1, p_2) := \begin{cases} \frac{-\bar{b}+\sqrt{\bar{b}^2+4\bar{a}\bar{d}}}{2\bar{a}} & if\ \bar{a} \neq 0, \\ \frac{\bar{d}}{\bar{b}} & otherwise. \end{cases} \tag{8}$$

上述引理提供了识别不相关函数的判据：给定一对稀疏模式 $\boldsymbol{s}^{(1)}, \boldsymbol{s}^{(2)} \in \{0,1\}^{n_u}$ 及其对应的二次函数 $p_{u,\boldsymbol{s}^{(1)}}$ 和 $p_{u,\boldsymbol{s}^{(2)}}$，可以利用引理 4 计算 $L(p_{u,\boldsymbol{s}^{(1)}}, p_{u,\boldsymbol{s}^{(2)}})$。如果 $L(p_{u,\boldsymbol{s}^{(1)}}, p_{u,\boldsymbol{s}^{(2)}}) > U$ 且 $p_{u,\boldsymbol{s}^{(1)}}(\mathbf{0}) > p_{u,\boldsymbol{s}^{(2)}}(\mathbf{0})$（或者 $p_{u,\boldsymbol{s}^{(1)}}(\mathbf{0} < p_{u,\boldsymbol{s}^{(2)}}(\mathbf{0})$，分别如此），则可以将 $p_{u,\boldsymbol{s}^{(1)}}$（或者 $p_{u,\boldsymbol{s}^{(2)}}$，分别如此）判定为不相关。

在确立了识别不相关函数的判据后，我们现在说明如何将该判据纳入算法 1。对于某个 $1 \leq u \leq n-1$，假设局部参数化函数 f_{u+1} 已在第5行计算得到，且由 N 个二次片段组成；也就是说，$f_{u+1}(\boldsymbol{\alpha}_{\mathcal{B}_{u+1}}) = \min_{\boldsymbol{s} \in \mathcal{P}_{u+1}} \{p_{u+1,\boldsymbol{s}}(\boldsymbol{\alpha}_{\mathcal{B}_{u+1}})\}$，其中 $|\mathcal{P}_{u+1}| = N$。为识别并剔除该集合中的不相关片段，可以在第5行之后立即插入一个剪枝子程序。该子程序针对每一对 $\boldsymbol{s}^{(1)}, \boldsymbol{s}^{(2)} \in \mathcal{P}_{u+1}$ 计算 $L(p_{u,\boldsymbol{s}^{(1)}}, p_{u,\boldsymbol{s}^{(2)}})$，并移除其对应函数 $p_{u,\boldsymbol{s}^{(1)}}$ 满足 $L(p_{u,\boldsymbol{s}^{(1)}}, p_{u,\boldsymbol{s}^{(2)}}) > U$ 且 $p_{u,\boldsymbol{s}^{(1)}}(\mathbf{0}) > p_{u,\boldsymbol{s}^{(2)}}(\mathbf{0})$的所有索引 $\boldsymbol{s}^{(1)} \in \mathcal{P}_{u+1}$。改进后的算法见算法2。此版本与算法 1 的关键区别在于加入了 PRUNE 子程序（算法 3），该子程序执行上述剪枝操作。对于含有 N个二次片段的函数，PRUNE 子程序需要对这些二次片段进行 $\mathcal{O}(N^2)$ 次两两比较。每次比较都需要计算相应的 L，其计算所需的时间和内存均为 $\mathcal{O}(\omega^2)$。因此，PRUNE 子程序的总体时间复杂度和内存复杂度分别为$\mathcal{O}(\omega^2 N^2)$和 $\mathcal{O}(\omega^2 N)$。顺带一提，当 $\boldsymbol{Q}$ 的树分解是一条路径时，该子程序的时间复杂度可以改进为 $\mathcal{O}(\omega^2 N)$。关于这一特殊情形的进一步讨论留待第 7 节和附录 B。

算法2 有界树宽图上的矩阵参数化算法 输入：问题（1）以及 supp($\boldsymbol{Q}$) 的树分解 T，其宽度为 ω，上界为 U；输出：问题（1）的最优代价 $f^\star$ 和最优解 $(\boldsymbol{x}^\star, \boldsymbol{z}^\star)$；1：按照第2.1节所述方案标记 supp($\boldsymbol{Q}$) 的节点；2：设置 $f_1(\boldsymbol{\alpha}_{\mathcal{B}_1}) = \frac{1}{2}\boldsymbol{\alpha}_{\mathcal{B}_1}^\top \boldsymbol{Q}_{\mathcal{B}_1,\mathcal{B}_1}\boldsymbol{\alpha}_{\mathcal{B}_1} + \boldsymbol{\alpha}_{\mathcal{B}_1}^\top \boldsymbol{c}_{\mathcal{B}_1}$.3：对于 $u = 1, \ldots, n-1$，执行 4：通过方程（7b）根据 f_u 计算 g_u；5：通过方程（6）计算 f_{u+1}；6：设置 $f_{u+1} =$ PRUNE(f_{u+1}, U)；7：结束循环 8：得到 $f^\star = \min_{\boldsymbol{\alpha}_n \in \mathbb{R}} \{f_n(\boldsymbol{\alpha}_n) + \boldsymbol{\lambda}_n \mathbb{I}(\boldsymbol{\alpha}_n)\}$, $\boldsymbol{x}_n^\star = \operatorname{argmin}_{\boldsymbol{\alpha}_n \in \mathbb{R}} \{f_n(\boldsymbol{\alpha}_n) + \boldsymbol{\lambda}_n \mathbb{I}(\boldsymbol{\alpha}_n)\}$，以及 $\boldsymbol{z}_n^\star = \mathbb{I}(\boldsymbol{x}_n^\star)$；9：对于 $u = n-1, \ldots, 1$，执行 10：设置 $\boldsymbol{x}_u^\star = \operatorname{argmin}_{\boldsymbol{\alpha}_u \in \mathbb{R}} \left\{ f_u(\boldsymbol{\alpha}_u, \boldsymbol{x}_{\mathcal{B}_u \setminus u}^\star) + \boldsymbol{\lambda}_u \mathbb{I}(\boldsymbol{\alpha}_u) \right\}$，以及 $\boldsymbol{z}_u^\star = \mathbb{I}(\boldsymbol{x}_u^\star)$；11：结束循环 12：返回 $f^\star$ 和 $(\boldsymbol{x}^\star, \boldsymbol{z}^\star)$；

尽管下一节将详细讨论这种剪枝策略的理论保证，但这里我们先简要强调该方法在实践中的有效性。图7展示了

算法 3 PRUNE(f, U)

输入： 由二次片段 $\{p_s\}_{s\in\mathcal{I}}$ 和常数 U 表征的函数 f。

输出：剪枝后的函数 f^{prune}

将 $\mathcal{S}^{\text{prune}}$ 初始化为空列表；
将 $\mathcal{I}$ 转换为列表并存储在 $\mathcal{S}$ 中；
当 $\mathcal{S} \neq \emptyset$ **时执行**
    将 i 设为 $\mathcal{S}$ 的第一个元素，并从 $\mathcal{S}$中删除 i；
    将 i 添加到 $\mathcal{S}^{\text{prune}}$；　▷ 假设 p_i 是相关的
    对于 $j \in \mathcal{S}$，**执行**
        计算 $L(p_i, p_j)$;
        如果 $L > U$，**则**　▷要么 p_i 要么p_j 是不相关的
            如果 $p_i(0) < p_j(0)$ **那么**　▷ p_j 是不相关的
                从 $\mathcal{S}$中删除 j；
            else　▷ p_i 是不相关的
                从 $\mathcal{S}^{\text{prune}}$中删除 i；
                break;　▷ 退出 for 循环
            结束 if
        结束 if
    结束 for
结束 while
返回 f^{prune} 由二次片段 $\{p_i\}_{i\in\mathcal{S}^{\text{prune}}}$ 表征；

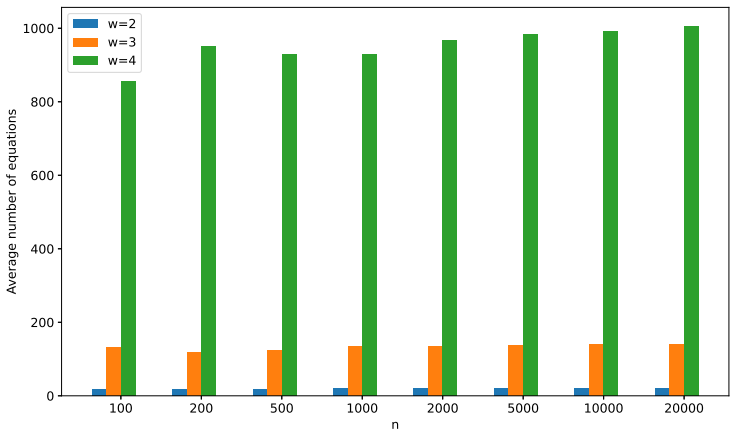

图7：针对不同的 n 和 w 值，剪枝参数算法计算的方程数量。所报告的结果为 5 次试验的平均值。

剪枝子程序在不同规模 n（100 到 20,000）以及不同带宽 w （2 到 4，对应的树宽为 2 到 4）的合成带状矩阵上的性能。图中显示了经过剪枝后保留的二次方程数量，该数量是在局部参数化代价上取平均，并对五次独立试验取平均所得。值得注意的是，即便对于规模最大的实例 $n = 20,000$，剪枝后的二次方程平均数量也不超过 1,100，而不进行剪枝时，二次方程数量则按 $2^{20,000}$ 的规模增长。这表明，剪枝子程序能够有效消除不相关的方程。

6 理论分析

带剪枝的所提参数化算法在实证中的表现表明，在指数级数量的二次方程中，只有一小部分与刻画局部参数化成本 f_u 相关。这一观察意味着，对于大多数二次函数对 $(p_{u,\boldsymbol{s}^{(1)}}, p_{u,\boldsymbol{s}^{(2)}})$，它们的交点位于相关区域 $\mathcal{D}_u = \{\boldsymbol{\alpha}_{\mathcal{B}_u}$ 之外：$\|\boldsymbol{\alpha}_{\mathcal{B}_u}\|_\infty \le U\}$。本节将从理论上论证这一关键观察。特别是，我们将证明，稀疏模式 $\boldsymbol{s}^{(1)}, \boldsymbol{s}^{(2)} \in \{0,1\}^{n_u}$ 越相似，其差值 $p_{u,\boldsymbol{s}^{(1)}} - p_{u,\boldsymbol{s}^{(2)}}$ 的根就越大。我们使用 m-相似性这一概念来量化两个稀疏模式的相似程度。对于任意袋 $\mathcal{B}_u$ 和正整数 m，回顾一下，$\mathcal{V}_{u,m} = \{i \in \mathcal{J}_u \mid \mathrm{dist}(i, \mathcal{B}_u) \le m\}$ 是诱导子图 $\mathrm{supp}_u(\boldsymbol{Q})$ 中 $\mathcal{B}_u$ 的 m-邻域，其中 $\mathcal{J}_u$ 是 $\mathrm{supp}_u(\boldsymbol{Q})$ 中不属于 $\mathcal{B}_u$ 的节点集合。回顾一下，$\pi : \mathcal{I} \to \{1, \dots, |\mathcal{I}|\}$ 是规范索引映射，它为每个 $i \in \mathcal{I}$ 指定其在子矩阵 $\boldsymbol{Q}_{\mathcal{I},\mathcal{I}}$ 中对应的行/列位置。

定义4 （m-相似性）。 若对于所有 $i \in \mathcal{V}_{u,m}$，均有 $\boldsymbol{s}^{(1)}_{\pi_u(i)} = \boldsymbol{s}^{(2)}_{\pi_u(i)}$，则称两个稀疏模式，$\boldsymbol{s}^{(2)} \in \{0,1\}^{n_u}$ 关于袋 $\mathcal{B}_u$ 是 m-相似的。当上下文中参考袋 $\mathcal{B}_u$ 明确时，我们简称 $\boldsymbol{s}^{(1)}$ 和 $\boldsymbol{s}^{(2)}$ 是 m- 相似的。

我们将证明，对于两个 m-相似的稀疏模式 $\boldsymbol{s}^{(1)}, \boldsymbol{s}^{(2)} \in \{0,1\}^{n_u}$，差值 $p_{u,\boldsymbol{s}^{(1)}} - p_{u,\boldsymbol{s}^{(2)}}$ 的根的范数会随 m 指数级地增长。为此，我们首先确立 $\boldsymbol{Q}^{-1}$ 的一个关键衰减性质。具体而言，我们证明，$\boldsymbol{Q}$ 的任意主子矩阵的逆矩阵中的元素，都会随着其支撑图中相应节点之间距离的增加而指数级地衰减。

引理 5。 设 $\boldsymbol{Q} \in \mathbb{R}^{n \times n}$ 为一个对称正定矩阵，$\mathcal{I} \subseteq \{1, \dots, n\}$ 为其行/列的任意子集。对于任意 i、$j \in \mathcal{I}$，有

$$\left| \left[\boldsymbol{Q}^{-1}_{\mathcal{I},\mathcal{I}} \right]_{\pi(i),\pi(j)} \right| \le C_1 \rho^{\mathrm{dist}(i,j)}, \quad where \quad C_1 := \frac{1}{\mu_{\min}} \max\left\{1, \frac{(1+\sqrt{\kappa_2})^2}{2\kappa_2}\right\}, \ and \ \rho := \frac{\sqrt{\kappa_2}-1}{\sqrt{\kappa_2}+1}. \tag{9}$$

证明见附录 A.5。带状矩阵的逆矩阵的指数型非对角衰减已得到广泛研究 [28]。上述引理将这一结果扩展到了更一般的情形：逆矩阵的衰减结构由矩阵的稀疏模式决定。

根据引理 1 可知，局部参数化代价 f_u 可以表示为 $\min_{\boldsymbol{s} \in \{0,1\}^{n_u}} \{ p_{u,\boldsymbol{s}}(\boldsymbol{\alpha}_{\mathcal{B}_u}) \}$，

其中每个 $p_{u,\ \boldsymbol{s}}$ 的定义如下：

$$
\begin{aligned}
&p_{u,\boldsymbol{s}}(\boldsymbol{\alpha}_{\mathcal{B}_u}) = \frac{1}{2}\boldsymbol{\alpha}_{\mathcal{B}_u}\boldsymbol{A}_{u,\boldsymbol{s}}\boldsymbol{\alpha}_{\mathcal{B}_u} + \boldsymbol{b}_{u,\boldsymbol{s}}^{\top}\boldsymbol{\alpha}_{\mathcal{B}_u} + d_{u,\boldsymbol{s}}, \\
&\text{where}\begin{cases}
\boldsymbol{A}_{u,s} &= \boldsymbol{Q}_{\mathcal{B}_u,\mathcal{B}_u} - \boldsymbol{Q}_{\mathcal{B}_u,\mathcal{J}_{u,\boldsymbol{s}}}\left(\boldsymbol{Q}_{\mathcal{J}_{u,\boldsymbol{s}},\mathcal{J}_{u,\boldsymbol{s}}}\right)^{-1}\boldsymbol{Q}_{\mathcal{B}_u,\mathcal{J}_{u,\boldsymbol{s}}}^{\top} \\
\boldsymbol{b}_{u,s} &= \boldsymbol{c}_{\mathcal{B}_u} - \boldsymbol{c}_{\mathcal{J}_{u,\boldsymbol{s}}}^{T}\left(\boldsymbol{Q}_{\mathcal{J}_{u,\boldsymbol{s}},\mathcal{J}_{u,\boldsymbol{s}}}\right)^{-1}\boldsymbol{Q}_{\mathcal{B}_u,\mathcal{J}_{u,\boldsymbol{s}}}^{\top} \\
d_{u,s} &= -\frac{1}{2}\boldsymbol{c}_{\mathcal{J}_{u,\boldsymbol{s}}}^{\top}(\boldsymbol{Q}_{\mathcal{J}_{u,\boldsymbol{s}},\mathcal{J}_{u,\boldsymbol{s}}})^{-1}\boldsymbol{c}_{\mathcal{J}_{u,\boldsymbol{s}}} + \sum_{i\in\mathcal{J}_{u,\boldsymbol{s}}}\lambda_i.
\end{cases}
\end{aligned}
\tag{10}
$$

接下来，我们根据 $\mathcal{J}_{u,\boldsymbol{s}} = \{i \in \mathcal{J}_u \mid s_i = 1\}$ 中的元素与袋 $\mathcal{B}_u$ 的图距离对其进行划分。对于整数 $m > 0$，令

$$
\mathcal{V}_{u,s,m} = \mathcal{J}_{u,\boldsymbol{s}} \cap \mathcal{V}_{u,m}, \qquad \mathcal{W}_{u,s,m} = \mathcal{J}_{u,\boldsymbol{s}} \setminus \mathcal{V}_{u,m}. \tag{11}
$$

由于我们的子序列分析对 $u \in \{1,\ldots,n\}$、$\boldsymbol{s} \in \{0,1\}^{n_u}$ 和 $m \in \{1,\ldots,n_u\}$ 的任意选择都成立，为简便起见，我们去掉下标，记为 $\mathcal{V} = \mathcal{V}_u$、$\boldsymbol{s}$、$m$ 和 $\mathcal{W} = \mathcal{W}_u$、$\boldsymbol{s}$、$m$。直观上，集合 $\mathcal{V}$ 包含 $\mathcal{J}_u$、$\boldsymbol{s}$ 中与袋 $\mathcal{B}_u$ 内节点的距离不超过 m 的所有节点，而 $\mathcal{W}$ 收集 $\mathcal{J}_u$、$\boldsymbol{s}$ 中与 $\mathcal{B}_u$ 的距离超过 m 的所有节点。作为说明性示例，考虑图 2 中的图。当 $u = 8$、$m = 2$ 和 $\boldsymbol{s} = [1,1,\ldots,1]^{\top}$ 时，我们得到 $\mathcal{V} = \{2,3,\ldots,7\}$ 和 $\mathcal{W} = \{1\}$。

给定 $\mathcal{V}$ 和 $\mathcal{W}$，矩阵$\boldsymbol{Q}_{\mathcal{J}_u、\boldsymbol{s}、\mathcal{J}_u、\boldsymbol{s}}$ 及其逆矩阵具有以下分块结构

$$
\boldsymbol{Q}_{\mathcal{J}_{u,\boldsymbol{s}},\mathcal{J}_{u,\boldsymbol{s}}} = \begin{bmatrix} \boldsymbol{Q}_{\mathcal{W},\mathcal{W}} & \boldsymbol{Q}_{\mathcal{W},\mathcal{V}} \\ \boldsymbol{Q}_{\mathcal{V},\mathcal{W}} & \boldsymbol{Q}_{\mathcal{V},\mathcal{V}} \end{bmatrix}, \tag{12}
$$

$$
\boldsymbol{Q}^{-1}_{\mathcal{J}_{u,\boldsymbol{s}},\mathcal{J}_{u,\boldsymbol{s}}} = \begin{bmatrix} (\boldsymbol{Q}_{\mathcal{J}_{u,\boldsymbol{s}},\mathcal{J}_{u,\boldsymbol{s}}}/\boldsymbol{Q}_{\mathcal{V},\mathcal{V}})^{-1} & -(\boldsymbol{Q}_{\mathcal{J}_{u,\boldsymbol{s}},\mathcal{J}_{u,\boldsymbol{s}}}/\boldsymbol{Q}_{\mathcal{V},\mathcal{V}})^{-1}\boldsymbol{Q}_{\mathcal{W},\mathcal{V}}\boldsymbol{Q}^{-1}_{\mathcal{V},\mathcal{V}} \\ -\boldsymbol{Q}^{-1}_{\mathcal{V},\mathcal{V}}\boldsymbol{Q}_{\mathcal{V},\mathcal{W}}(\boldsymbol{Q}_{\mathcal{J}_{u,\boldsymbol{s}},\mathcal{J}_{u,\boldsymbol{s}}})^{-1} & \boldsymbol{Q}^{-1}_{\mathcal{V},\mathcal{V}} + \boldsymbol{Q}^{-1}_{\mathcal{V},\mathcal{V}}\boldsymbol{Q}_{\mathcal{V},\mathcal{W}}(\boldsymbol{Q}_{\mathcal{J}_{u,\boldsymbol{s}},\mathcal{J}_{u,\boldsymbol{s}}}/\boldsymbol{Q}_{\mathcal{V},\mathcal{V}})^{-1}\boldsymbol{Q}_{\mathcal{W},\mathcal{V}}\boldsymbol{Q}^{-1}_{\mathcal{V},\mathcal{V}} \end{bmatrix}, \tag{13}
$$

其中，$\boldsymbol{Q}_{\mathcal{J}_u,\boldsymbol{s},\mathcal{J}_u,\boldsymbol{s}}/\boldsymbol{Q}_{\mathcal{V},\mathcal{V}} := \boldsymbol{Q}_{\mathcal{W},\mathcal{W}} - \boldsymbol{Q}_{\mathcal{W},\mathcal{V}}\boldsymbol{Q}^{-1}_{\mathcal{V}},_{\mathcal{V}}\boldsymbol{Q}_{\mathcal{V},\mathcal{W}}$是 $\boldsymbol{Q}_{\mathcal{J}_u,\boldsymbol{s},\mathcal{J}_u,\boldsymbol{s}}$ 中分块 $\boldsymbol{Q}_{\mathcal{W},\mathcal{W}}$ 的舒尔补。分析 $\boldsymbol{Q}^{-1}_{\mathcal{J}_u}$ $,\boldsymbol{s},\mathcal{J}_u,\boldsymbol{s}$ 的块结构的主要动机在于，根据式 (10)，系数 $\boldsymbol{A}_{u,\boldsymbol{s}}$ 和 $\boldsymbol{b}_{u,\boldsymbol{s}}$ 包含涉及分块乘积 $\boldsymbol{Q}_{\mathcal{B}_u}$, $_{\mathcal{V}}\boldsymbol{Q}^{-1}_{\mathcal{V}}$ $,_{\mathcal{V}}\boldsymbol{Q}_{\mathcal{V},\mathcal{W}}$ 的项；正如下一条引理中形式化表述的那样，该分块乘积的范数随 m 指数级地衰减。

引理 6。我们有

$$
\left\|\boldsymbol{Q}_{\mathcal{B}_u,\mathcal{V}}\boldsymbol{Q}^{-1}_{\mathcal{V},\mathcal{V}}\boldsymbol{Q}_{\mathcal{V},\mathcal{W}}\right\|_2 \le \frac{2\mu_{\max}^2}{\mu_{\min}}\sqrt{\Delta_1\Delta_m}\rho^{m-1}, \tag{14}
$$

其中 $0 < \rho < 1$ 是引理 5中的常数。

证明。令 $\pi\colon \mathcal{V} \to \{1,\ldots,|\mathcal{V}|\}$ 为索引映射，将每个 $i \in \mathcal{V}$映射到其在子矩阵 $\boldsymbol{Q}_{\mathcal{V},\mathcal{V}}$中的对应行/列位置。令 $\mathcal{I} = \{i \in \mathcal{V} \mid \mathrm{dist}(i,\mathcal{B}_u) = 1\}$，且 $\mathcal{J} = \{j \in \mathcal{V} \mid \mathrm{dist}(j,\mathcal{B}_u) = m\}$。这里，$\mathcal{I}$ 是 $\mathcal{V}$ 中与 $\mathcal{B}_u$相邻的节点集合，而 $\mathcal{J}$ 是与 $\mathcal{B}_u$的距离为 m 的节点集合。可以写成

$$
\boldsymbol{Q}_{\mathcal{B}_u,\mathcal{V}}\boldsymbol{Q}^{-1}_{\mathcal{V},\mathcal{V}}\boldsymbol{Q}_{\mathcal{V},\mathcal{W}} = \boldsymbol{Q}_{\mathcal{B}_u,\pi(\mathcal{I})}\left[\boldsymbol{Q}^{-1}_{\mathcal{V},\mathcal{V}}\right]_{\pi(\mathcal{I}),\pi(\mathcal{J})}\boldsymbol{Q}_{\pi(\mathcal{J}),\mathcal{W}}.
$$

事实上，若 $\mathcal{I}$或 $\mathcal{J}$为空，则 $\boldsymbol{Q}_{\mathcal{B}_u,\mathcal{V}}\boldsymbol{Q}_{\mathcal{V}}^{-1},_{\mathcal{V}}\boldsymbol{Q}_{\mathcal{V},\mathcal{W}}=0$。因此，不失一般性，我们假设两个集合均非空。根据引理 5，对于任意 $i\in\mathcal{I}$ 和 $j\in\mathcal{J}$，有

$$\left|\left[\boldsymbol{Q}_{\mathcal{V},\mathcal{V}}^{-1}\right]_{\pi(i),\pi(j)}\right|\leq C_1\rho^{\mathrm{dist}(i,j)}=C_1\rho^{m-1}.$$

这意味着

$$\left\|\left[\boldsymbol{Q}_{\mathcal{V},\mathcal{V}}^{-1}\right]_{\pi(\mathcal{I}),\pi(\mathcal{J})}\right\|_2\leq\sqrt{|\mathcal{I}||\mathcal{J}|}C_1\rho^{m-1}\leq\sqrt{\Delta_1\Delta_m}C_1\rho^{m-1}.$$

最后一个不等式源于 $\mathcal{J}\subseteq\mathcal{V}\subseteq\mathcal{V}_{u,m}$ 这一事实，由此可推出 $|\mathcal{J}|\leq|\mathcal{V}_{u,m}|\leq\Delta_m$。根据相同的推理，我们有 $|\mathcal{I}|\leq\Delta_1$。因此，

$$\begin{aligned}\left\|\boldsymbol{Q}_{\mathcal{B}_u,\mathcal{V}}\boldsymbol{Q}_{\mathcal{V},\mathcal{V}}^{-1}\boldsymbol{Q}_{\mathcal{V},\mathcal{W}}\right\|_2&=\left\|\boldsymbol{Q}_{\mathcal{B}_u,\pi(\mathcal{I})}\left[\boldsymbol{Q}_{\mathcal{V},\mathcal{V}}^{-1}\right]_{\pi(\mathcal{I}),\pi(\mathcal{J})}\boldsymbol{Q}_{\pi(\mathcal{J}),\mathcal{W}}\right\|_2\\&\leq\|\boldsymbol{Q}_{\mathcal{B}_u,\pi(\mathcal{I})}\|_2\left\|\left[\boldsymbol{Q}_{\mathcal{V},\mathcal{V}}^{-1}\right]_{\pi(\mathcal{I}),\pi(\mathcal{J})}\right\|_2\|\boldsymbol{Q}_{\pi(\mathcal{J}),\mathcal{W}}\|_2\\&\leq\mu_{\max}^2\left\|\left[\boldsymbol{Q}_{\mathcal{V},\mathcal{V}}^{-1}\right]_{\pi(\mathcal{I}),\pi(\mathcal{J})}\right\|_2\\&\leq\mu_{\max}^2\sqrt{\Delta_1\Delta_m}C_1\rho^{m-1}\\&=\frac{2\mu_{\max}^2}{\mu_{\min}}\sqrt{\Delta_1\Delta_m}\rho^{m-1}.\end{aligned}$$

第二个不等式利用了这样一个性质：$\mu_{\max}$ 是 $\boldsymbol{Q}$ 的每个子矩阵的谱范数上界。最后一个不等式源于 $\frac{(1+\sqrt{\kappa_2})^2}{2\kappa_2}\leq 2$ 这一事实，由此可推出 $C_1=\frac{1}{\mu_{\min}}\max\left\{1,\frac{(1+\sqrt{\kappa_2})^2}{2\kappa_2}\right\}\leq\frac{2}{\mu_{\min}}$。□

上述引理意味着 $\left\|\boldsymbol{Q}_{\mathcal{B}_u,\mathcal{V}}\boldsymbol{Q}_{\mathcal{V}}^{-1},_{\mathcal{V}}\boldsymbol{Q}_{\mathcal{V},\mathcal{W}}\right\|_2$ 随着 m 的增大而近乎指数级地衰减，前提是 Δ_1 和 Δ_m 不会指数级增长。事实证明，在多项式体积增长（假设1）条件下，这一条件得以保证。该性质对于建立最终保证至关重要。

基于上述结果，我们现在可以证明，对于任意两个 m-相似稀疏模式$\boldsymbol{s}^{(1)}$，$\boldsymbol{s}^{(2)}\in\{0,1\}^{n_u}$其对应二次片段$p_{u,\ \boldsymbol{s}^{(1)}}$ 的二次系数和线性系数$p_{u,\ \boldsymbol{s}^{(2)}}$ 彼此指数级接近。

引理 7。设 $\boldsymbol{s}^{(1)}$，$\boldsymbol{s}^{(2)}\in\{0,1\}^{n_u}$ 是两个 m-相似稀疏模式，并设 $p_{u,\ \boldsymbol{s}^{(1)}}$ 和 $p_{u,\ \boldsymbol{s}^{(2)}}$ 是它们在 (10) 中定义的对应二次片段。则以下界成立：

$$\begin{aligned}\|\boldsymbol{A}_{u,\boldsymbol{s}^{(1)}}-\boldsymbol{A}_{u,\boldsymbol{s}^{(2)}}\|_{1,1}&\leq 4\kappa_2^4\ (\omega+1)^{3/2}\ \Delta_1\Delta_m\ \rho^{2m-2},\\\|\boldsymbol{b}_{u,\boldsymbol{s}^1}-\boldsymbol{b}_{u,\boldsymbol{s}^2}\|_1&\leq 4\kappa_2^2(1+\kappa_\infty)\ U\ (\omega+1)^{3/2}\ \sqrt{\Delta_1\Delta_m}\ \rho^{m-1}.\end{aligned}$$

证明。对于每个满足 $i\in\{1,2\}$ 的 $\boldsymbol{s}^{(i)}$，考虑集合 $\mathcal{V}_{u,\boldsymbol{s}^{(i)},m}:=\mathcal{J}_{u,\boldsymbol{s}^{(i)}}\cap\mathcal{V}_{u,m}$ 和 $\mathcal{W}_{u,\boldsymbol{s}^{(i)},m}:=\mathcal{J}_{u,\boldsymbol{s}^{(i)}}\setminus\mathcal{V}_{u,m}$。由于 $\boldsymbol{s}^{(1)}$ 和 $\boldsymbol{s}^{(2)}$ 是 m-相似的，因此可得 $\mathcal{V}_{u,\boldsymbol{s}^{(1)},m}=\mathcal{V}_{u,\boldsymbol{s}^{(2)},m}$。为方便记号，我们省略下标，并记 $\mathcal{V}:=\mathcal{V}_{u,\boldsymbol{s}^{(1)},m}=\mathcal{V}_{u,\boldsymbol{s}^{(2)},m}$, $\mathcal{W}^{(1)}:=\mathcal{W}_{u,\boldsymbol{s}^{(1)},m}$，且

$\mathcal{W}^{(2)} := \mathcal{W}_{u,\boldsymbol{s}^{(2)},m}$。结合（$p_{u,\boldsymbol{s}^{(1)}}$ 和$p_{u,\boldsymbol{s}^{(2)}}$）中（10）的定义，以及（$\boldsymbol{Q}^{-1}_{\mathcal{J}_{u,\boldsymbol{s}^{(i)}},\mathcal{J}_{u,\boldsymbol{s}^{(i)}}}$）中（13）的 的块结构，可得

$$\begin{aligned}\boldsymbol{A}_{u,\boldsymbol{s}^{(1)}} - \boldsymbol{A}_{u,\boldsymbol{s}^{(2)}} =& -\boldsymbol{Q}_{\mathcal{B}_u,\mathcal{V}}\boldsymbol{Q}^{-1}_{\mathcal{V},\mathcal{V}}\boldsymbol{Q}_{\mathcal{V},\mathcal{W}^{(1)}}(\boldsymbol{Q}_{\mathcal{J}_{u,\boldsymbol{s}^{(1)}},\mathcal{J}_{u,\boldsymbol{s}^{(1)}}}/\boldsymbol{Q}_{\mathcal{V},\mathcal{V}})^{-1}\boldsymbol{Q}_{\mathcal{W}^{(1)},\mathcal{V}}\boldsymbol{Q}^{-1}_{\mathcal{V},\mathcal{V}}\boldsymbol{Q}_{\mathcal{V},\mathcal{B}_u}\\ &+\boldsymbol{Q}_{\mathcal{B}_u,\mathcal{V}}\boldsymbol{Q}^{-1}_{\mathcal{V},\mathcal{V}}\boldsymbol{Q}_{\mathcal{V},\mathcal{W}^{(2)}}(\boldsymbol{Q}_{\mathcal{J}_{u,\boldsymbol{s}^{(2)}},\mathcal{J}_{u,\boldsymbol{s}^{(2)}}}/\boldsymbol{Q}_{\mathcal{V},\mathcal{V}})^{-1}\boldsymbol{Q}_{\mathcal{W}^{(2)},\mathcal{V}}\boldsymbol{Q}^{-1}_{\mathcal{V},\mathcal{V}}\boldsymbol{Q}_{\mathcal{V},\mathcal{B}_u},\end{aligned}$$

其中我们利用了如下事实：$\boldsymbol{Q}_{\mathcal{B}_u,\mathcal{J}_{u,\boldsymbol{s}^{(i)}}} = \left[\boldsymbol{Q}_{\mathcal{B}_u,\mathcal{W}^{(i)}}\ \boldsymbol{Q}_{\mathcal{B}_u,\mathcal{V}}\right]$ 以及$\boldsymbol{Q}_{\mathcal{B}_u,\mathcal{W}^{(i)}} = 0$。上述等式给出

$$\begin{aligned}\|\boldsymbol{A}_{u,\boldsymbol{s}^{(1)}} - \boldsymbol{A}_{u,\boldsymbol{s}^{(2)}}\|_2 \le& \left\|\boldsymbol{Q}_{\mathcal{B}_u,\mathcal{V}}\boldsymbol{Q}^{-1}_{\mathcal{V},\mathcal{V}}\boldsymbol{Q}_{\mathcal{V},\mathcal{W}^{(1)}}(\boldsymbol{Q}_{\mathcal{J}_{u,\boldsymbol{s}^{(1)}},\mathcal{J}_{u,\boldsymbol{s}^{(1)}}}/\boldsymbol{Q}_{\mathcal{V},\mathcal{V}})^{-1}\boldsymbol{Q}_{\mathcal{W}^{(1)},\mathcal{V}}\boldsymbol{Q}^{-1}_{\mathcal{V},\mathcal{V}}\boldsymbol{Q}_{\mathcal{V},\mathcal{B}_u}\right\|_2\\ &+\left\|\boldsymbol{Q}_{\mathcal{B}_u,\mathcal{V}}\boldsymbol{Q}^{-1}_{\mathcal{V},\mathcal{V}}\boldsymbol{Q}_{\mathcal{V},\mathcal{W}^{(2)}}(\boldsymbol{Q}_{\mathcal{J}_{u,\boldsymbol{s}^{(2)}},\mathcal{J}_{u,\boldsymbol{s}^{(2)}}}/\boldsymbol{Q}_{\mathcal{V},\mathcal{V}})^{-1}\boldsymbol{Q}_{\mathcal{W}^{(2)},\mathcal{V}}\boldsymbol{Q}^{-1}_{\mathcal{V},\mathcal{V}}\boldsymbol{Q}_{\mathcal{V},\mathcal{B}_u}\right\|_2\\ \le& \left\|\boldsymbol{Q}_{\mathcal{B}_u,\mathcal{V}}\boldsymbol{Q}^{-1}_{\mathcal{V},\mathcal{V}}\boldsymbol{Q}_{\mathcal{V},\mathcal{W}^{(1)}}\right\|_2^2\ \left\|(\boldsymbol{Q}_{\mathcal{J}_{u,\boldsymbol{s}^{(1)}},\mathcal{J}_{u,\boldsymbol{s}^{(1)}}}/\boldsymbol{Q}_{\mathcal{V},\mathcal{V}})^{-1}\right\|_2\\ &+\left\|\boldsymbol{Q}_{\mathcal{B}_u,\mathcal{V}}\boldsymbol{Q}^{-1}_{\mathcal{V},\mathcal{V}}\boldsymbol{Q}_{\mathcal{V},\mathcal{W}^{(2)}}\right\|_2^2\ \left\|(\boldsymbol{Q}_{\mathcal{J}_{u,\boldsymbol{s}^{(2)}},\mathcal{J}_{u,\boldsymbol{s}^{(2)}}}/\boldsymbol{Q}_{\mathcal{V},\mathcal{V}})^{-1}\right\|_2.\end{aligned}$$

由引理6，我们有

$$\max\left\{\left\|\boldsymbol{Q}_{\mathcal{B}_u,\mathcal{V}}\boldsymbol{Q}^{-1}_{\mathcal{V},\mathcal{V}}\boldsymbol{Q}_{\mathcal{V},\mathcal{W}^{(1)}}\right\|_2, \left\|\boldsymbol{Q}_{\mathcal{B}_u,\mathcal{V}}\boldsymbol{Q}^{-1}_{\mathcal{V},\mathcal{V}}\boldsymbol{Q}_{\mathcal{V},\mathcal{W}^{(2)}}\right\|_2\right\} \le \frac{2\mu_{\max}^2}{\mu_{\min}}\sqrt{\Delta_1\Delta_m}\rho^{m-1}.$$

另一方面，

$$\begin{aligned}\frac{1}{\mu_{\min}(\boldsymbol{Q})} = \mu_{\max}\left(\boldsymbol{Q}^{-1}\right) = \max_{\boldsymbol{\zeta}\neq\boldsymbol{0}}\left\{\frac{\boldsymbol{\zeta}^\top\boldsymbol{Q}^{-1}\boldsymbol{\zeta}}{\boldsymbol{\zeta}^\top\boldsymbol{\zeta}}\right\} \ge& \max_{\bar{\boldsymbol{\zeta}}\neq\boldsymbol{0}}\left\{\frac{\bar{\boldsymbol{\zeta}}^\top\left[\boldsymbol{Q}^{-1}\right]_{\mathcal{W}^{(i)},\mathcal{W}^{(i)}}\bar{\boldsymbol{\zeta}}}{\bar{\boldsymbol{\zeta}}^\top\bar{\boldsymbol{\zeta}}}\right\}\\ =& \left\|(\boldsymbol{Q}_{\mathcal{J}_{u,\boldsymbol{s}^{(i)}},\mathcal{J}_{u,\boldsymbol{s}^{(i)}}}/\boldsymbol{Q}_{\mathcal{V},\mathcal{V}})^{-1}\right\|_2.\end{aligned} \tag{15}$$

综合上述不等式，我们得到

$$\|\boldsymbol{A}_{u,\boldsymbol{s}^{(1)}} - \boldsymbol{A}_{u,\boldsymbol{s}^{(2)}}\|_2 \le \frac{4\mu_{\max}^4}{\mu_{\min}^3}\ \Delta_1\Delta_m\ \rho^{2m-2} \le 4\kappa_2^4\ \Delta_1\Delta_m\ \rho^{2m-2},$$

其中最后一个不等式源于如下事实：$\mu_{\min} \le 1$，因此，$\mu_{\max} \le \kappa_2$。在注意到$\|\boldsymbol{A}_{\boldsymbol{s}^1} - \boldsymbol{A}_{\boldsymbol{s}^2}\|_{1,1} \le (\omega+1)^{3/2}\|\boldsymbol{A}_{u,s^{(1)}} - \boldsymbol{A}_{u,s^{(2)}}\|_2$·之后，第一项陈述的证明即告完成。

接下来，我们给出第二项陈述的证明。同样，由 （10）以及.....的块结构 $\boldsymbol{Q}^{-1}_{\mathcal{J}_{u,\boldsymbol{s}^{(i)}},\mathcal{J}_{u,\boldsymbol{s}^{(i)}}}$ 在（13）中，我们有

$$\begin{aligned}\boldsymbol{b}_{u,\boldsymbol{s}^{(1)}} - \boldsymbol{b}_{u,\boldsymbol{s}^{(2)}} =& -\left(\boldsymbol{c}_{\mathcal{W}^{(1)}} - \boldsymbol{Q}_{\mathcal{W}^{(1)},\mathcal{V}}\boldsymbol{Q}^{-1}_{\mathcal{V},\mathcal{V}}\boldsymbol{c}_{\mathcal{V}}\right)^\top\left(\boldsymbol{Q}_{\mathcal{J}_{u,\boldsymbol{s}^{(1)}},\mathcal{J}_{u,\boldsymbol{s}^{(1)}}}/\boldsymbol{Q}_{\mathcal{V},\mathcal{V}}\right)^{-1}\boldsymbol{Q}_{\mathcal{W}^{(1)},\mathcal{V}}\boldsymbol{Q}^{-1}_{\mathcal{V},\mathcal{V}}\boldsymbol{Q}_{\mathcal{V},\mathcal{B}_u}\\ &+\left(\boldsymbol{c}_{\mathcal{W}^{(2)}} - \boldsymbol{Q}_{\mathcal{W}^{(2)},\mathcal{V}}\boldsymbol{Q}^{-1}_{\mathcal{V},\mathcal{V}}\boldsymbol{c}_{\mathcal{V}}\right)^\top\left(\boldsymbol{Q}_{\mathcal{J}_{u,\boldsymbol{s}^{(2)}},\mathcal{J}_{u,\boldsymbol{s}^{(2)}}}/\boldsymbol{Q}_{\mathcal{V},\mathcal{V}}\right)^{-1}\boldsymbol{Q}_{\mathcal{W}^{(2)},\mathcal{V}}\boldsymbol{Q}^{-1}_{\mathcal{V},\mathcal{V}}\boldsymbol{Q}_{\mathcal{V},\mathcal{B}_u},\end{aligned}$$

这意味着

$$\begin{aligned}&\|\boldsymbol{b}_{u,\boldsymbol{s}^{(1)}} - \boldsymbol{b}_{u,\boldsymbol{s}^{(2)}}\|_\infty\\ &\le \left\|\boldsymbol{c}_{\mathcal{W}^{(1)}} - \boldsymbol{Q}_{\mathcal{W}^{(1)},\mathcal{V}}\boldsymbol{Q}^{-1}_{\mathcal{V},\mathcal{V}}\boldsymbol{c}_{\mathcal{V}}\right\|_\infty\left\|\left(\boldsymbol{Q}_{\mathcal{J}_{u,\boldsymbol{s}^{(1)}},\mathcal{J}_{u,\boldsymbol{s}^{(1)}}}/\boldsymbol{Q}_{\mathcal{V},\mathcal{V}}\right)^{-1}\right\|_\infty\left\|\boldsymbol{Q}_{\mathcal{W}^{(1)},\mathcal{V}}\boldsymbol{Q}^{-1}_{\mathcal{V},\mathcal{V}}\boldsymbol{Q}_{\mathcal{V},\mathcal{B}_u}\right\|_\infty\\ &+\left\|\boldsymbol{c}_{\mathcal{W}^{(2)}} - \boldsymbol{Q}_{\mathcal{W}^{(2)},\mathcal{V}}\boldsymbol{Q}^{-1}_{\mathcal{V},\mathcal{V}}\boldsymbol{c}_{\mathcal{V}}\right\|_\infty\left\|\left(\boldsymbol{Q}_{\mathcal{J}_{u,\boldsymbol{s}^{(2)}},\mathcal{J}_{u,\boldsymbol{s}^{(2)}}}/\boldsymbol{Q}_{\mathcal{V},\mathcal{V}}\right)^{-1}\right\|_\infty\left\|\boldsymbol{Q}_{\mathcal{W}^{(2)},\mathcal{V}}\boldsymbol{Q}^{-1}_{\mathcal{V},\mathcal{V}}\boldsymbol{Q}_{\mathcal{V},\mathcal{B}_u}\right\|_\infty.\end{aligned}$$

现在我们分别控制右侧的各项。对于任意 $i = 1,2$，有

$$\left\|\boldsymbol{c}_{\mathcal{W}^{(i)}} - \boldsymbol{Q}_{\mathcal{W}^{(i)},\mathcal{V}}\boldsymbol{Q}_{\mathcal{V},\mathcal{V}}^{-1}\boldsymbol{c}_{\mathcal{V}}\right\|_\infty \le \|\boldsymbol{c}\|_\infty + \|\boldsymbol{Q}\|_\infty \left\|\boldsymbol{Q}_{\mathcal{V},\mathcal{V}}^{-1}\right\|_\infty \|\boldsymbol{c}_{\mathcal{V}}\|_\infty \le \|\boldsymbol{c}\|_\infty + U\ \|\boldsymbol{Q}\|_\infty.$$

第一个不等式由诱导 ∞-范数的次乘性以及 $\left\|\boldsymbol{Q}_{\mathcal{W}^{(i)},\ \mathcal{V}}\right\|_\infty \le \|\boldsymbol{Q}\|_\infty$这一事实得出。第二个不等式由 $\left\|\boldsymbol{Q}_{\mathcal{V}}^{-1},\ _{\mathcal{V}}\right\|_\infty \|\boldsymbol{c}_{\mathcal{V}}\|_\infty \le U$得出（参见引理3的证明）。类似地，由（13）可得

$$\left\|\left(\boldsymbol{Q}_{\mathcal{J}_{u,\boldsymbol{s}^{(i)}},\mathcal{J}_{u,\boldsymbol{s}^{(i)}}}/\boldsymbol{Q}_{\mathcal{V},\mathcal{V}}\right)^{-1}\right\|_\infty = \left\|\left[\boldsymbol{Q}^{-1}\right]_{\mathcal{W}^{(i)},\mathcal{W}^{(i)}}\right\|_\infty \le \left\|\boldsymbol{Q}^{-1}\right\|_\infty.$$

最后，

$$\left\|\boldsymbol{Q}_{\mathcal{W}^{(i)},\mathcal{V}}\boldsymbol{Q}_{\mathcal{V},\mathcal{V}}^{-1}\boldsymbol{Q}_{\mathcal{V},\mathcal{B}_u}\right\|_\infty \le \sqrt{\omega+1}\left\|\boldsymbol{Q}_{\mathcal{W}^{(i)},\mathcal{V}}\boldsymbol{Q}_{\mathcal{V},\mathcal{V}}^{-1}\boldsymbol{Q}_{\mathcal{V},\mathcal{B}_u}\right\|_2 \le \sqrt{\omega+1}\ \frac{2\mu_{\max}^2}{\mu_{\min}}\sqrt{\Delta_1\Delta_m}\rho^{m-1},$$

最后一个不等式中，我们使用了引理6。结合这些界，我们得到

$$\begin{aligned}\|\boldsymbol{b}_{u,\boldsymbol{s}^{(1)}} - \boldsymbol{b}_{u,\boldsymbol{s}^{(2)}}\|_\infty &\le 2\left(\|\boldsymbol{c}\|_\infty + U\ \|\boldsymbol{Q}\|_\infty\right)\ \left\|\boldsymbol{Q}^{-1}\right\|_\infty \sqrt{\omega+1}\ \frac{2\mu_{\max}^2}{\mu_{\min}}\sqrt{\Delta_1\Delta_m}\rho^{m-1}\\ &\le 4\kappa_2^2(1+\kappa_\infty)U\ \sqrt{\omega+1}\ \sqrt{\Delta_1\Delta_m}\rho^{m-1},\end{aligned}$$

第二个不等式由 $\left\|\boldsymbol{Q}^{-1}\right\|_\infty\|\boldsymbol{c}\|_\infty \le U$，$\kappa_\infty = \|\boldsymbol{Q}\|_\infty\left\|\boldsymbol{Q}^{-1}\right\|_\infty$和 $\frac{\mu_{\max}^2}{\mu_{\min}} \le \kappa_2^2$得出。注意到$\|\boldsymbol{b}_{u,\boldsymbol{s}^{(1)}} - \boldsymbol{b}_{u,\boldsymbol{s}^{(2)}}\|_1 \le (\omega+1)\|\boldsymbol{b}_{u,\boldsymbol{s}^{(1)}} - \boldsymbol{b}_{u,\boldsymbol{s}^{(2)}}\|_\infty$后，证明即告完成 □

我们的下一个目标是分析一对 m-相似稀疏模式$\boldsymbol{s}^{(1)}$ 和 $\boldsymbol{s}^{(2)}$对应的项 $d_{u,\boldsymbol{s}^{(1)}} - d_{u,\boldsymbol{s}^{(2)}}$。从式 (10) 中 $d_{u,\boldsymbol{s}}$ 的表达式可以看出，与二次系数和线性系数 $\boldsymbol{A}_{u,\boldsymbol{s}}$ 及 $\boldsymbol{b}_{u,\boldsymbol{s}}$不同，常数项 $d_{u,\boldsymbol{s}}$ 不依赖于与袋 $\mathcal{B}_u$相关的 $\boldsymbol{Q}$ 的任何子矩阵。因此，与二次项和线性项不同，引理 6 无法用于证明$|d_{u,\boldsymbol{s}^{(1)}} - d_{u,\boldsymbol{s}^{(2)}}|$呈衰减趋势。正是这一观察使我们的剪枝策略得以奏效：如引理 7 所述，在假设1中的多项式增长条件下，$\|\boldsymbol{A}_{u,\boldsymbol{s}^{(1)}} - \boldsymbol{A}_{u,\boldsymbol{s}^{(2)}}\|_1$,1 和$\|\boldsymbol{b}_{u,\boldsymbol{s}^{(1)}} - \boldsymbol{b}_{u,\boldsymbol{s}^{(2)}}\|_1$ 都随 m 指数级快速衰减，而$|d_{u,\boldsymbol{s}^{(1)}} - d_{u,\boldsymbol{s}^{(2)}}|$ 则不会表现出这种衰减。因此，利用引理 4，我们可以得出结论：对于任意一对 m-相似稀疏模式 $\boldsymbol{s}^{(1)}$ 和 $\boldsymbol{s}^{(2)}$，量 $L(p_{u,\boldsymbol{s}^{(1)}},p_{u,\boldsymbol{s}^{(2)}})$ 会随 m 指数级增长。因此，当 m足够大时，函数 $p_{u,\boldsymbol{s}^{(1)}}$ 或 $p_{u,\boldsymbol{s}^{(2)}}$ 中至少有一个会变得不相关。下一个引理将这一想法形式化。

引理 8. 设 $\boldsymbol{s}^{(1)}$、$\boldsymbol{s}^{(2)} \in \{0,1\}^{n_u}$ 是两个 m-相似的稀疏模式，并令 $p_{u,\boldsymbol{s}^{(1)}}$ 和 $p_{u,\boldsymbol{s}^{(2)}}$ 分别表示它们在 (10) 中定义的对应二次片段。假设多项式体积增长条件（假设 *1*）成立，且 $|d_{u,\boldsymbol{s}^{(1)}} - d_{u,\boldsymbol{s}^{(2)}}| \ge \eta$, 其中某个 $0 < \eta \le 1$。假设

$$m \ge \max\left\{\frac{2\log\left(8U^2/\eta\right) + 2\log\left(\delta(\omega+1)^{3/2}\right) + 2\log\left(\kappa_2^2(1+\kappa_\infty)\right)}{\log(1/\rho)} + 2,\ \frac{2\gamma}{\log(1/\rho)},\ \left(\frac{\gamma}{\log(1/\rho)}\right)^2\right\}. \tag{16}$$

然后， $L(\boldsymbol{s}^{(1)},\boldsymbol{s}^{(2)}) \ge U$；特别是， $\boldsymbol{s}^{(1)}$ 或 $\boldsymbol{s}^{(2)}$ 中至少有一个是不相关的。

在证明上述引理之前，我们先给出以下辅助论断，它将在证明中发挥重要作用。

断言 1. 假设 $\Gamma > 0$。那么，对于任意 $m \geq \max\{1, 2/\Gamma, (1/\Gamma)^2\}$，有 $\Gamma m \geq \log(m)$。

证明。首先，假设 $\Gamma \geq 1$。那么，容易验证，对于 $m > 0$，有 $\Gamma m \geq m \geq \log(m)$。现在，考虑 $0 < \Gamma < 1$ 的情形。定义 $m = \xi/\Gamma$，其中 $\xi \geq \max\{2, 1/\Gamma\}$。于是，我们有

$$\Gamma m - \log(m) = \xi - \log(\xi) - \log(1/\Gamma) \geq \log(\xi) - \log(1/\Gamma) \geq 0,$$

其中，第一个不等式由以下事实得出：对于 $\xi \geq 2$，有 $\xi \geq 2\log(\xi)$；最后一个不等式则由以下事实得出：$\xi \geq \max\{2, 1/\Gamma\} \geq 1/\Gamma$。 □

证明。引理的证明 8。在多项式增长条件下，我们有 $\Delta_1 \leq \delta$ 且 $\Delta_m \leq \delta m^\gamma$。结合引理 7，可得

$$\begin{aligned}
\|\boldsymbol{A}_{u,\boldsymbol{s}^{(1)}} - \boldsymbol{A}_{u,\boldsymbol{s}^{(2)}}\|_{1,1} &\leq 4\kappa_2^4\ (\omega+1)^{3/2}\ \delta^2 m^\gamma\ \rho^{2m-2} \\
\|\boldsymbol{b}_{u,\boldsymbol{s}^1} - \boldsymbol{b}_{u,\boldsymbol{s}^2}\|_1 &\leq 4\kappa_2^2(1+\kappa_\infty)\ U\ (\omega+1)^{3/2}\ \delta m^{\gamma/2}\ \rho^{m-1}
\end{aligned}$$

应用引理 4 并取 $\bar{a} = 4\kappa_2^4\,(\omega+1)^{3/2}\,\delta^2 m^\gamma\,\rho^{2m-2} > 0, \bar{b} = 4\kappa_2^2(1+\kappa_\infty)\ U\,(\omega+1)^{3/2}\,\delta m^{\gamma/2}\,\rho^{m-1}$，¯且 $d = \eta$，得到

$$L(\boldsymbol{s}^{(1)}, \boldsymbol{s}^{(2)}) = \frac{-\bar{b} + \sqrt{\bar{b}^2 + 4\bar{a}\bar{d}}}{2\bar{a}} = \frac{4\bar{a}\bar{d}}{2\bar{a}\left(\bar{b} + \sqrt{\bar{b}^2 + 4\bar{a}\bar{d}}\right)} \geq \frac{\bar{d}}{\sqrt{\bar{b}^2 + 4\bar{a}\bar{d}}} \geq \frac{\bar{d}}{\bar{b} + 2\sqrt{\bar{a}\bar{d}}}.$$

¯ ¯代入 $\bar{a}$、$\bar{b}$ 和 d 的表达式后，我们得到

$$\begin{aligned}
\frac{\bar{d}}{\bar{b} + 2\sqrt{\bar{a}\bar{d}}} &\geq \frac{\eta}{4\kappa_2^2(1+\kappa_\infty)\ U\ (\omega+1)^{3/2}\ \delta m^{\gamma/2}\ \rho^{m-1} + 4\sqrt{\eta}\ \kappa_2^2\ (\omega+1)^{3/4}\ \delta m^{\gamma/2}\ \rho^{m-1}} \\
&= \frac{\eta}{4\delta(\omega+1)^{3/2}\left(\kappa_2^2(1+\kappa_\infty)U + \sqrt{\eta}\ \kappa_2^2\ (\omega+1)^{-3/4}\right)}\ m^{-\gamma/2}\ \rho^{-m+1} \\
&\geq K m^{-\gamma/2}\rho^{-m+1}, \quad \text{where} \quad K := \frac{\eta}{8U\ \delta(\omega+1)^{3/2}\ \kappa_2^2(1+\kappa_\infty)}.
\end{aligned}$$

在最后一个不等式中，我们利用了以下事实：$0 < \eta \leq 1$ 以及 $(\omega+1)^{-3/4} \leq 1$，由此可得$\sqrt{\eta}\ (\omega+1)^{-3/4} \leq 1$。此外，我们还利用了以下事实：$\kappa_2^2 \leq \kappa_2^2\ (1+\kappa_\infty)\ U$。要完成证明，只需证明 $K m^{-\gamma/2}\rho^{-m+1} \geq U$。为此，我们首先注意到

$$\begin{aligned}
K m^{-\gamma/2}\rho^{-m+1} \geq U &\iff \log\left(m^{-\gamma/2}\rho^{-m+1}\right) \geq \log\left(U/K\right) \\
&\iff -\log(m) + (m-1)\frac{2\log\left(1/\rho\right)}{\gamma} \geq \frac{2\log\left(U/K\right)}{\gamma} \\
&\iff -\log(m) + m\frac{2\log\left(1/\rho\right)}{\gamma} \geq \frac{2\log\left(U/K\right)}{\gamma} + \frac{2\log(1/\rho)}{\gamma}.
\end{aligned}$$

令 $\Gamma := \log(1/\rho)\gamma$。根据我们的假设，$m \geq \max\{1, 2/\Gamma, (1/\Gamma)^2\}$。因此，命题1 可推出$\Gamma m \geq \log(m)$，从而得到

$$-\log(m) + m\frac{2\log(1/\rho)}{\gamma} \geq 2\frac{\log(U/K)}{\gamma} + 2\frac{\log(1/\rho)}{\gamma} \Longleftarrow m\frac{\log(1/\rho)}{\gamma} \geq 2\frac{\log(U/K)}{\gamma} + 2\frac{\log(1/\rho)}{\gamma}$$
$$\Longleftarrow m \geq \frac{2\log(U/K)}{\log(1/\rho)} + 2.$$

根据 K的定义，可以验证，最后一个不等式可由（16）推出。

□

为利用上述引理，并获得剪枝后每个局部参数化代价所保留的二次片段最大数量的全局界，我们引入（k, η）-间隔的概念。在后续论证中，我们固定 m 为满足（16）的任意量。对于任意 $u \in \{1, \ldots, n\}$ 和任意$\boldsymbol{\xi} \in \{0,1\}^{|\mathcal{V}_{u,\,m}|}$，定义$\mathcal{C}_{u,\,\boldsymbol{\xi}} := \{\boldsymbol{s} \in \{0,1\}^{n_u} : \boldsymbol{s}_{\mathcal{V}_{u,\,m}} = \boldsymbol{\xi}\}$ 为所有 m-相似稀疏模式的集合，其中在 $\mathrm{supp}_u(\boldsymbol{Q})$ 中，袋 $\mathcal{B}_u$ 的 m-邻域内的条目均与 $\boldsymbol{\xi}$一致。等价地，$\mathcal{C}_{u,\,\boldsymbol{\xi}}$ 是由关系 $\boldsymbol{s} \sim \boldsymbol{s}'$ 当且仅当$\boldsymbol{s}_{\mathcal{V}_{u,\,m}} = \boldsymbol{s}'_{\mathcal{V}_{u,\,m}}$所诱导的等价类。回忆$|\mathcal{V}_{u,\,m}| \leq \Delta_m$；因此，这类等价类至多有 2^{Δ_m} 个。

定义5 （η-最优集与（k, η）-间隔）**.** 给定任意 $\eta \geq 0$、$u \in \{1, \ldots, n\}$和 $\boldsymbol{\xi} \in \{0,1\}^{|\mathcal{V}_{u,m}|}$，等价类 $\mathcal{C}_{u,\boldsymbol{\xi}}$ 的 η-最优集定义为

$$\mathcal{R}_{u,\boldsymbol{\xi},\eta} := \left\{ \boldsymbol{s} \in \mathcal{C}_{u,\boldsymbol{\xi}} : d_{u,\boldsymbol{s}} - \min_{\boldsymbol{z} \in \mathcal{C}_{u,\boldsymbol{\xi}}} \{d_{u,\boldsymbol{z}}\} \leq \eta \right\}.$$

我们称该问题具有（k, η）-间隔，如果对每一个 $u \in \{1, \ldots, n\}$ 和$\boldsymbol{\xi} \in \{0,1\}^{|\mathcal{V}_{u,m-1}|}$，$|\mathcal{R}_{u,\boldsymbol{\xi},\eta}| \leq k$ 成立。

为了直观理解 η-最优集与（k, η）-间隔的概念，可以证明，由其定义（10），$d_{u,\boldsymbol{s}}$ 与如下二次规划的最优值一致

$$d_{u,\boldsymbol{s}} = \min_{\boldsymbol{x} \in \mathbb{R}^{|\mathcal{J}_{u,\boldsymbol{s}}|}} \left\{ \tfrac{1}{2}\boldsymbol{x}^\top \boldsymbol{Q}_{\mathcal{J}_{u,\boldsymbol{s}},\mathcal{J}_{u,\boldsymbol{s}}}\boldsymbol{x} + \boldsymbol{c}_{\mathcal{J}_{u,\boldsymbol{s}}}^\top \boldsymbol{x} + \sum_{i \in \mathcal{J}_{u,\boldsymbol{s}}} \boldsymbol{\lambda}_i \right\}.$$

该值来自在设置 $\boldsymbol{x}_{\mathcal{B}_u} = \boldsymbol{0}$ 并根据稀疏模式 $\boldsymbol{s}$ 固定二进制向量 $\boldsymbol{z} \in \{0,1\}^{n_u}$ 后，定义在 $\mathrm{supp}_u(\boldsymbol{Q})$ 上的子问题 **(4)**，即对于 $i \in \mathcal{J}_u$、$\boldsymbol{s}$ 和 $i \notin \mathcal{J}_u$、$\boldsymbol{s}$ 分别取 $\boldsymbol{z}_i = 1$ 和 $\boldsymbol{z}_i = 0$。因此，$\min_{\boldsymbol{z} \in \mathcal{C}_{u,\boldsymbol{\xi}}} \{d_{u,\boldsymbol{z}}\}$ 表示所有这类稀疏模式属于等价类 $\mathcal{C}_{u,\boldsymbol{\xi}}$ 的受限子问题中的最优值。相应的 η-最优集合 $\mathcal{R}_{u,\boldsymbol{\xi},\eta}$ 由那些对应子问题值与该最小值之差不超过 η 的稀疏模式组成。换言之，$\mathcal{R}_{u,\boldsymbol{\xi},\eta}$ 包含等价类 $\mathcal{C}_{u,\boldsymbol{\xi}}$ 中所有对于节点 u 处的局部子问题而言近似最优的二进制赋值（最优性间隙为 η）。例如，如果$|\mathcal{R}_{u,\boldsymbol{\xi},\eta}| = 1$，那么相应的子问题将从 $\mathcal{C}_{u,\boldsymbol{\xi}}$ 中得到唯一的最优稀疏模式，并具有 η 间隔。这一观察促使我们定义 (k, η)-间隔：如果对于每个袋 $\mathcal{B}_u$ 和每个等价类 $\mathcal{C}_{u,\boldsymbol{\xi}}$，相应的 η-最优集合的基数至多为 k，则称该问题具有 (k, η)-间隔；也就是说，至多有 k 个解，其对应的最优值位于最小值的 η 间隔之内。

这一假设充当一种结构性的“间隔”条件，用于控制几乎无法区分的离散解的多重性。从概念上讲，这类似于统计学习中的间隔条件——例如经典的 Tsybakov低噪声假设 [67]及其各种改进形式 [29]——这些条件限制了决策边界附近点的质量。我们的框架通过对 η-最优集合的大小施加上界来适应该思想。

事实上，所提出剪枝策略的有效性关键取决于（k，η）-间隔的参数。根据引理8，一旦片段落在 η-间隔之外，剪枝策略便能保证识别出不相关片段。因此，在每个等价类$\mathcal{C}_{u,\ \boldsymbol{\xi}}$中，剪枝后最多有 k 个方程仍然相关。由于等价类最多有 $2^{|\mathcal{V}_{u,\ m}|}$ 个，剪枝步骤后保留的方程总数至多为 $k\ 2^{|\mathcal{V}_{u,\ m}|}$。以下引理将对此作出形式化表述。

引理9。 假设问题 (1) 具有 $(k,\ \eta)$-间隔，且假设*1*中的多项式增长条件成立。对于每个 $u \in \{1, \ldots, n-1\}$，令 $\mathcal{P}_{u+1}$ 表示剪枝步骤后（算法第 *6* 行）的剪枝后的索引集。则， $\mathcal{P}_{u+1}$ 满足 $|\mathcal{P}_{u+1}| \leq k\ 2^{|\mathcal{V}_{u+1,\ m}|}$，其中 m 是满足 (16) 的最小整数。

证明。 任取 $\boldsymbol{\xi} \in \{0,1\}^{|\mathcal{V}_{u+1,m}|}$，并考虑相应的等价类 $\mathcal{C}_{u+1,\boldsymbol{\xi}}$，根据定义，其中的元素彼此 m-相似。根据引理 8，对于任意两个稀疏模式 $\boldsymbol{s}^{(1)}, \boldsymbol{s}^{(2)} \in \mathcal{C}_{u+1,\boldsymbol{\xi}}$，若满足 $|d_{u+1,\boldsymbol{s}^{(1)}} - d_{u+1,\boldsymbol{s}^{(2)}}| \geq \eta$，则这两个模式中至少有一个是不相关的，因此会在剪枝步骤中被移除。因此，剪枝后，$\mathcal{C}_{u+1,\boldsymbol{\xi}}$ 中仅保留常数项 $d_{u+1,\boldsymbol{s}}$ 与该等价类中的最小值相差不超过 η 的稀疏模式，即

$$\mathcal{R}_{u+1,\boldsymbol{\xi},\eta} = \Big\{ \boldsymbol{s} \in \mathcal{C}_{u+1,\boldsymbol{\xi}} : d_{u+1,\boldsymbol{s}} - \min_{\boldsymbol{z} \in \mathcal{C}_{u+1,\boldsymbol{\xi}}} d_{u+1,\boldsymbol{z}} \leq \eta \Big\}.$$

根据 (k,η)-间隔假设，对于每个 $\boldsymbol{\xi}$，均有$|\mathcal{R}_{u+1,\boldsymbol{\xi},\eta}| \leq k$，因此剪枝后每个等价类中至多剩下 k 个稀疏模式。每个 $\boldsymbol{\xi} \in \{0,1\}^{|\mathcal{V}_{u+1,m}|}$ 都定义了唯一的等价类 $\mathcal{C}_{u+1,\boldsymbol{\xi}}$，因此总共有 $2^{|\mathcal{V}_{u+1,m}|}$ 个这样的等价类。对所有等价类求和可得 $|\mathcal{P}_{u+1}| \leq k\ 2^{|\mathcal{V}_{u+1,m}|}$, 由此得证。 □

有了上述引理，我们现在可以给出关于算法 2 的正确性和运行时间的主定理。

定理1。 假设问题 (1) 具有 (k,η)-间隔，且假设*1*中的多项式增长条件 *1* 成立。那么，所提出的参数化算法（算法*2*）可在 (1)时间内解决问题 $\mathcal{O}(n\omega^2\ \delta\ k^{2\delta}\ 4^{\Delta_{m+1}})$，并使用 $\mathcal{O}(n\omega^2\ k^{\delta}\ 2^{\Delta_{m+1}})$ 内存，其中

$$m = \left\lceil \max\left\{ \frac{2\log\left(8U^2/\eta\right) + 2\log\left(\delta(\omega+1)^{3/2}\right) + 2\log\left(\kappa_2^2(1+\kappa_\infty)\right)}{\log(1/\rho)} + 2,\ \frac{2\gamma}{\log(1/\rho)},\ \left(\frac{\gamma}{\log(1/\rho)}\right)^2 \right\} \right\rceil.$$

证明。 我们先证明该算法的正确性，然后分析其时间和内存复杂度。

正确性证明。令（$\boldsymbol{x}^\star$，$\boldsymbol{z}^\star$）为问题（1）的一个最优解，并令 $f^\star$ 为其目标值。为证明正确性，只需说明 $\boldsymbol{z}^\star_{\mathcal{J}_n}$ 属于局部参数化代价 f_n的剪枝集 $\mathcal{P}_n$。证明采用反证法。假设 $\boldsymbol{z}^\star_{\mathcal{J}_n} \notin \mathcal{P}_n$。那么存在一个索引 $u \leq n$，使得最优解在节点

在 $\mathcal{J}_u$ 中，即 $\boldsymbol{z}^\star_{\mathcal{J}_u}$，其稀疏模式在剪枝步骤中被丢弃。令 u 表示满足 $\boldsymbol{z}^\star_{\mathcal{J}_u} \notin \mathcal{P}_u$ 的最小索引。由于 $\boldsymbol{z}^\star_{\mathcal{J}_u}$ 在剪枝步骤中被丢弃，根据剪枝过程的定义，相应的片段 $p_{u,\ \boldsymbol{z}^\star_{\mathcal{J}_u}}$ 不相关。因此，存在 $\bar{\boldsymbol{s}} \in \mathcal{P}_u$，使得对于所有 $\|\boldsymbol{\alpha}\|_\infty < U$，均有 $p_{u,\ \boldsymbol{z}^\star_{\mathcal{J}_u}}(\boldsymbol{\alpha}) > p_{u,\ \bar{\boldsymbol{s}}}(\boldsymbol{\alpha})$。

为导出矛盾，我们在最优解（$\boldsymbol{x}^\star$，$\boldsymbol{z}^\star$）处计算问题（1）的目标函数。在此之前，我们将 supp($\boldsymbol{Q}$) 中的节点划分为三个互不相交的集合：$\mathcal{A} = \{1, \ldots, n\} \backslash (\mathcal{J}_u \cup \mathcal{B}_u)$、$\mathcal{B}_u$ 和 $\mathcal{J}_u$。根据树分解的性质，supp($\boldsymbol{Q}$) 中的 $\mathcal{A}$ 与 $\mathcal{J}_u$ 之间没有边；因此 $\boldsymbol{Q}_{\mathcal{A},\ \mathcal{J}_u} = \boldsymbol{Q}^\top_{\mathcal{J}_u,\ \mathcal{A}} = \boldsymbol{0}$。对于这一划分，$\boldsymbol{Q}$ 和 $\boldsymbol{c}$ 具有如下分块形式：

$$\boldsymbol{Q} = \begin{bmatrix} \boldsymbol{Q}_{\mathcal{A},\mathcal{A}} & \boldsymbol{Q}_{\mathcal{A},\mathcal{B}_u} & \boldsymbol{0} \\ \boldsymbol{Q}_{\mathcal{B}_u,\mathcal{A}} & \boldsymbol{Q}_{\mathcal{B}_u,\mathcal{B}_u} & \boldsymbol{Q}_{\mathcal{B}_u,\mathcal{J}_u} \\ \boldsymbol{0} & \boldsymbol{Q}_{\mathcal{J}_u,\mathcal{B}_u} & \boldsymbol{Q}_{\mathcal{J}_u,\mathcal{J}_u} \end{bmatrix}, \quad \boldsymbol{c} = \begin{bmatrix} \boldsymbol{c}_{\mathcal{A}} \\ \boldsymbol{c}_{\mathcal{B}_u} \\ \boldsymbol{c}_{\mathcal{J}_u} \end{bmatrix}.$$

在最优解处计算问题（1）的目标函数，得到：

$$\begin{aligned} f^\star &= \frac{1}{2}(\boldsymbol{x}^\star)^\top \boldsymbol{Q}\boldsymbol{x}^\star + \boldsymbol{c}^\top \boldsymbol{x}^\star + \boldsymbol{\lambda}^\top \boldsymbol{z}^\star \\ &= \frac{1}{2}(\boldsymbol{x}^\star_{\mathcal{A}})^\top \boldsymbol{Q}_{\mathcal{A},\mathcal{A}} \boldsymbol{x}^\star_{\mathcal{A}} + \boldsymbol{c}^\top_{\mathcal{A}} \boldsymbol{x}^\star_{\mathcal{A}} + \sum_{i \in \mathcal{A}} \lambda_i \boldsymbol{z}^\star_i + \boldsymbol{x}^\star_{\mathcal{A}} Q_{\mathcal{A},\mathcal{B}_u} \boldsymbol{x}^\star_{\mathcal{B}_u} + \sum_{i \in \mathcal{B}_u} \lambda_i \boldsymbol{z}^\star_i + p_{u,\boldsymbol{z}^\star_{\mathcal{J}_u}}(\boldsymbol{x}^\star_{\mathcal{B}_u}) \\ &> \frac{1}{2}(\boldsymbol{x}^\star_{\mathcal{A}})^\top \boldsymbol{Q}_{\mathcal{A},\mathcal{A}} \boldsymbol{x}^\star_{\mathcal{A}} + \boldsymbol{c}^\top_{\mathcal{A}} \boldsymbol{x}^\star_{\mathcal{A}} + \sum_{i \in \mathcal{A}} \lambda_i \boldsymbol{z}^\star_i + \boldsymbol{x}^\star_{\mathcal{A}} Q_{\mathcal{A},\mathcal{B}_u} \boldsymbol{x}^\star_{\mathcal{B}_u} + \sum_{i \in \mathcal{B}_u} \lambda_i \boldsymbol{z}^\star_i + p_{u,\bar{\boldsymbol{s}}}(\boldsymbol{x}^\star_{\mathcal{B}_u}). \end{aligned}$$

这表明，对于二元变量而言，向量 $\bar{\boldsymbol{z}} = \begin{bmatrix} \boldsymbol{z}^\star_{\mathcal{A}} & \boldsymbol{z}^\star_{\mathcal{B}_u} & \bar{\boldsymbol{s}} \end{bmatrix}^\top$ 是一个严格更优的选择，从而与（$\boldsymbol{x}^\star$，$\boldsymbol{z}^\star$）的最优性相矛盾。

复杂度证明。 为分析运行时间，我们考察算法的每个步骤。拓扑排序及随后的标记分别可在 $\mathcal{O}(n)$ 和 $\mathcal{O}(n\omega^2)$ 时间内完成，且所需内存为 $\mathcal{O}(n\omega)$（参见第 2.1 节）。初始化 f_1 需要 $\mathcal{O}(\omega^2)$ 时间和内存（第 2 行）。

接下来，我们分析第一个 for 循环（第3–7行）的复杂度。对于每个 $u = 1$，.....，$n-1$，函数 g_u 可通过分别在有无指示变量的情况下最小化 f_u 的每个片段得到，总代价为 $\mathcal{O}(|\mathcal{P}_u|\omega^2)$ 的时间和内存。另一方面，根据引理9，我们有 $|\mathcal{P}_u| \le k\ 2^{|\mathcal{V}_u}$，$_m|$。这意味着计算 g_u 的代价可由$\mathcal{O}($ $\omega^2\ k\ 2^{|\mathcal{V}_u,m|}) = \mathcal{O}(\omega^2\ k\ 2^{\Delta_m})$界定。此外，给定 $\{g_v\}_{v \in \mathrm{par}_T(u+1)}$，第5行中的函数 f_{u+1} 可根据（6）计算，我们在此将其改写为：

$$f_{u+1}(\boldsymbol{\alpha}_{\mathcal{B}_{u+1}}) = h_{u+1}(\boldsymbol{\alpha}_{\mathcal{B}_{u+1}}) + \sum_{v \in \mathrm{par}_T(u+1)} \Big(g_v(\boldsymbol{\alpha}_{\mathcal{B}_{u+1} \backslash v}) - \phi_v(\boldsymbol{\alpha}_{\mathcal{B}_{u+1} \backslash v}) \Big).$$

正如算法的运行时间分析所确立的，1，h_{u+1} 与 $\{\phi_v :\ v \in \mathrm{par}_T(u+1)\}$ 是单片段二次函数，并可在 $\mathcal{O}(\omega^2)$ 的时间和内存内计算。

接下来，我们转向计算 g_v。由于 g_v 的片段数量至多是对应 f_v 的两倍，因此每个 g_v 至多包含 $2k\ 2^{|\mathcal{V}_v}$，$_m|$ 个片段。因此，只需考虑 $g_v - \phi_v$ 中针对 $v \in \mathrm{par}_T(u+1)$ 的所有片段组合，即可构造 f_{u+1}，从而得到

至多

$$\prod_{v\in \mathrm{par}_T(u+1)} 2k\ 2^{|\mathcal{V}_{v,m}|} = k^{|\mathrm{par}_T(u+1)|}\ 2^{\sum_{v\in \mathrm{par}_T(u+1)}(|\mathcal{V}_{v,m}|+1)}$$

片段。事实上，我们有 $|\mathrm{par}_T(u+1)| \le \Delta_1 \le \delta$。此外，根据 $\mathcal{V}_{u+1}$ 的定义，$m+1$，我们有 $\mathcal{V}_{u+1,m+1} = \bigcup_{v\in \mathrm{par}_T(u+1)}(\mathcal{V}_{v,m}\cup\{v\})$。由于集合 $\{\mathcal{V}_{v,m}\cup\{v\}\}_{v\in \mathrm{par}_T(u+1)}$ 互不相交，因此可得 $|\mathcal{V}_{u+1,m+1}| = \sum_{v\in \mathrm{par}_T(u+1)}(|\mathcal{V}_{v,m}|+1)$。这意味着在剪枝之前，$f_{u+1}$ 至多包含 $k^\delta\ 2^{|\mathcal{V}_{u+1,m+1}|}$ 个片段，并且可在 $\mathcal{O}(\delta\ \omega^2\ k^\delta\ 2^{|\mathcal{V}_{u+1,m+1}|})$ 时间内构造完成，使用 $\mathcal{O}(\omega^2\ k^\delta\ 2^{|\mathcal{V}_{u+1,m+1}|})$ 内存。最后，第 6 行调用剪枝子程序 PRUNE，其运行时间为 $\mathcal{O}\left(\omega^2\ \left(k^\delta\ 2^{|\mathcal{V}_{u+1,m+1}|}\right)^2\right) = \mathcal{O}\left(\omega^2\ k^{2\delta}\ 4^{\Delta_{m+1}}\right)$，内存开销为 $\mathcal{O}\left(\omega^2\ k^\delta\ 2^{|\mathcal{V}_{u+1,m+1}|}\right) = \mathcal{O}\left(\omega^2\ k^\delta\ 2^{\Delta_{m+1}}\right)$，如第 5 节所述。由于第一个 for 循环执行 $n-1$ 次迭代，因此总开销为 $\mathcal{O}(n\omega^2\ \delta\ k^{2\delta}\ 4^{\Delta_{m+1}})$ 时间和 $\mathcal{O}(n\omega^2\ k^\delta\ 2^{\Delta_{m+1}})$ 内存。

最后，在第 $f^\star$、$\boldsymbol{x}_n^\star$ 和 $\boldsymbol{z}_n^\star$ 行中获取结果8；第二个 `for` 循环（第 9–11 行）的每次迭代也可在 $\mathcal{O}(\omega^2\ 2^{\Delta_m})$ 时间和内存内完成，为简洁起见，我们省略细节。综合所有这些步骤，可得该算法的运行时间为 $\mathcal{O}(n\omega^2\ \delta\ k^{2\delta}\ 4^{\Delta_{m+1}})$，内存占用为 $\mathcal{O}(n\omega^2\ k^\delta\ 2^{\Delta_{m+1}})$。 □

为进一步理解定理 1，我们聚焦于稀疏图 $\mathrm{supp}(\boldsymbol{Q})$ 具有线性体积增长的特殊问题类别，即假设 1 在 $\gamma = 1$ 下成立，因为在这种情况下，我们的复杂度界形式更为简洁。

推论1。 假设问题 (1) 具有 $(k,\ \eta)$-间隔，并满足线性体积增长（假设 *1* 在 $\gamma = 1$ 下成立）。那么，所提出的参数化算法（算法 *2*）可在问题 (1) 中求解

$$\mathcal{O}\left(n\omega^2\ \delta\ k^{2\delta}\ \max\left\{\left(\frac{U^2}{\eta}\ \delta(\omega+1)^{3/2}\ \kappa_2^2(1+\kappa_\infty)\right)^{\frac{4\delta}{\log(1/\rho)}},\ 4^{\delta\theta}\right\}\right),$$

时间，其中 $\theta := \max\left\{\frac{2}{\log(1/\rho)},\ \frac{1}{2^{\log(1/\rho)}}\right\}$。

证明。 该结果可由定理 1 在设定 $\gamma = 1$ 后直接得到。 □

上述推论的重要意义在于，对于固定的树宽 ω、间隔参数 $(k,\ \eta)$、条件数 κ_2、κ_∞ 以及体积增长参数 $(\delta,\ \gamma)$，运行时间随 n 线性增长。在下一节中，我们将考察这种理论依赖在实践中体现到何种程度。这里，我们结合两个已有保证的重要特例来阐述这一结果：叶子数有界的树，以及带宽为 w 的带状矩阵。前者的树宽等于 1，后者的树宽至多为 w。

- **树结构问题**。 当 $\mathrm{supp}(\boldsymbol{Q})$ 是一棵树时，[14] 表明，参数化算法的一个特化版本可在 $\mathcal{O}(n^2$ 时间和内存内求解问题（1）。尽管理论界关于 n 呈二次增长，但其中报告的实测性能近乎线性。上述推论对此现象提供了部分解释：对于叶子数有界的树，在较弱的（n、$k\eta$）间隔条件下，理论复杂度实际上关于 n 呈线性增长。这一观察使理论保证与实测性能相一致，并阐明了树结构为何特别有利于我们的方法。

- **带状矩阵**。 假设 $\boldsymbol{Q}$ 是带宽为 w 的带状矩阵。对于特殊情况$w = 2$（对应三对角结构），[53] 证明了问题（1）可在 $\mathcal{O}(n^2)$ 时间内达到最优解。最近，G´omez 等人 [39] 为带状 $\boldsymbol{Q}$（带宽为 $w \geq 2$）的问题（1）开发了全多项式时间近似方案，在 n 和 $\|\boldsymbol{c}\|_\infty/\epsilon$均固定的条件下，可在关于 ϵ和 n 的多项式时间内计算出 ϵ-精度解。在上述间隔假设下，我们的结果给出了精确算法，并降低了这两类问题对 n 的依赖。

7 数值实验

接下来，我们将在一系列合成数据和真实场景实例上评估所提算法的性能。在小节 7.1 中，我们介绍实验设置和实现细节。在小节 7.2 中，我们使用合成数据考察算法对各种问题参数的依赖性。最后，在小节 7.3 中，我们使用真实世界数据，展示所提方法在带异常值校正的指数平滑问题上的有效性。

7.1 方法与设置

所有实验均在一台配备 8 核 3.0 GHz Xeon Gold 6154 处理器和 16 GB 内存的计算机上进行。我们将所提出的参数化算法与Gurobi v10.0.2 进行比较。对于所有 Gurobi 运行，我们均设置一小时的时间限制，并在最优性间隙低于 0.01% 时终止求解器。如果 Gurobi在此时间限制内未能达到 0.01% 或更小的最优性间隙，我们将报告求解器达到的最佳最优性间隙。表格和图中报告的所有结果均为五次独立试验的平均值。本文案例研究所用算法的 Python 实现及相关代码可从以下地址获取：

https://github.com/aareshfb/Treewidth-Parametric-Algorithm.

- **Gurobi** 我们将问题 (1) 重构为

$$\min_{\boldsymbol{x}\in\mathbb{R}^n,\, \boldsymbol{z}\in\{0,1\}^n} \quad \frac{1}{2}\boldsymbol{x}^\top \boldsymbol{Q}\boldsymbol{x} + \boldsymbol{c}^\top \boldsymbol{x} + \boldsymbol{\lambda}^\top \boldsymbol{z}, \tag{17a}$$

$$\text{s.t.} \quad -U\boldsymbol{z}_i \leq \boldsymbol{x}_i \leq U\boldsymbol{z}_i, \qquad i = 1, 2, \ldots, n. \tag{17b}$$

其中， U 是我们在参数化算法中使用的同一个上界。我们注意到，该问题也可以进行透视重构 [34]。我们尝试在 Gurobi 中显式纳入这一重构；然而，我们发现求解器的默认形式始终优于我们手动实现的透视重构。我们推测，这是因为 Gurobi 在内部更有效地利用了基于透视的强化。因此，本文报告的所有数值结果均基于Gurobi的内置配置。

- **参数化方法** 我们将 Gurobi 与我们提出的参数化算法（算法 2）的性能进行比较。我们注意到，剪枝子程序（算法 3）的运行时间为 $\mathcal{O}$（$\omega^2 N^2$），因为它穷举比较构成局部参数化代价的所有 N 二次片段

。在 $\boldsymbol{Q}$ 的支撑图具有路径形式的树分解这一特殊情况下（例如带状矩阵），我们可以利用一种改进的剪枝过程，只需单次遍历这些片段，因此仅需进行 $\mathcal{O}(N)$ 次比较。当二次函数以每一对相邻函数均具有 m-相似性的方式存储时，这一改进即可实现；当树分解具有路径结构时，无需额外成本即可保证这一点。该实现的详细信息将在附录 B中给出。

7.2 合成生成实例的结果

回顾定理 1，参数化算法的运行时间随问题规模线性增长，随间隔参数（k，η）、体积增长参数 γ 以及条件数参数 κ_2 和 κ_∞呈多项式增长，并随体积增长参数 δ呈指数级增长。本小节通过大量合成实例，对这些依赖关系进行实证研究。

海森矩阵 $\boldsymbol{Q}$ 构造为 $\boldsymbol{Q} = \boldsymbol{Y}^\top \boldsymbol{Y} + \nu I$，其中 $\boldsymbol{Y}$是一个上三角矩阵，其带内元素从 $[-1, 1]$ 均匀采样，带外元素为零（即，对于 $i \le j \le i + w$，$\boldsymbol{Y}_{ij} \sim \mathcal{U}(-1,\ 1)$，其他情况下为零）。简单计算可知，supp($\boldsymbol{Q}$)是一个带状矩阵，带宽为 w。参数 $\nu > 0$ 用于调节 $\boldsymbol{Q}$的条件数。向量 $\boldsymbol{c}$ 的各个元素独立生成，且服从 $[-10, 10]$上的均匀分布。除非另有说明，正则化参数 $\boldsymbol{\lambda}$ 的各个元素独立生成，且服从区间 (3.5，4.5)上的均匀分布，这通常会得到非零项比例约为 50% 的解。在一些实验中，我们改变 ν（以改变条件数）和 $\boldsymbol{\lambda}$（以改变稀疏度），同时保持 $\boldsymbol{Q}$ 和 $\boldsymbol{c}$的构造不变，仍按上述方式进行。

问题规模的影响。 在第一个实验中，我们考察参数化算法在不同 n的问题上的性能。表 1 报告了带宽 $w = 2$ 和$w = 4$下的运行时间。对于超过 $n = 200$的实例， Gurobi 无法在一小时的时间限制内求得该实例的最优解（以 TL 表示）。相比之下，参数化算法可以在几秒内求解规模大得多的实例。为验证正确性，我们报告了两种方法的最优目标值。可以看出，参数化算法得到的目标值至少与 Gurobi 得到的目标值一样好（在某些情况下甚至严格更好）。此外，对于参数化算法，我们报告了剪枝后保留的二次片段平均数量。可以看出，该数量仅随 n小幅增加，展现了所提出剪枝方法在消除不相关片段方面的有效性。

接下来，我们在更广泛的问题规模和带宽范围内评估参数化算法的运行时间。图 8 （左）报告了 $w = 2$、3、4、5 且 n 最大达到 20、000 的性能。在这些实验中，我们调节 ν 使得 $\kappa_2 \approx 6.50$。我们观察到，对于 $w = 4$，参数化算法能够在一小时内求解规模最大的实例；而对于 $w = 5$，它能够在一小时内求解 n 最大达到 2 、 000 的实例。对于 w 的所有取值，运行时间曲线的斜率几乎相同，这表明带宽会影响总体运行时间的尺度，但不会影响其对 n的线性依赖关系，这与推论 1一致。

带宽的影响。接下来，我们单独考察带宽 w 对参数化算法运行时间的影响。图8 （右）展示了运行时间随 w 变化的情况，问题规模为$n = 50$，200，条件数为 $\kappa_2 \approx 4.63$，其中 w 取值范围为 2 到 6。该对数线性图表明，运行时间随 w 呈指数级增长，这与推论 1 一致。

表 1：不同规模的性能比较

w	指标	方法	$n = 100$	$n = 200$	$n = 500$	$n = 1000$	$n = 2000$
2	条件数（κ_2）	—	$\kappa_2 \approx 7.10$	$\kappa_2 \approx 7.43$	$\kappa_2 \approx 7.43$	$\kappa_2 \approx 8.11$	$\kappa_2 \approx 8.02$
	时间（秒）	参数化 Gurobi	0.17 11.01	0.35 565.43	0.91 TL	1.82 TL	3.70 TL
	目标值	参数化 Gurobi	-588.89 -588.89	-1143.57 -1143.57	-3033.48 -3033.28	-6363.05 -6362.32	-12367.60 -12365.20
	平均方程数	参数化	23	23	25	25	25
	分支定界节点 最优性间隙	Gurobi	31326 0.00%	2963376 0.00%	4882224 1.11%	2938436 1.30%	1169022 1.50%
4	条件数 (κ_2)	—	$\kappa_2 \approx 6.18$	$\kappa_2 \approx 6.20$	$\kappa_2 \approx 6.18$	$\kappa_2 \approx 6.43$	$\kappa_2 \approx 6.79$
	时间（秒）	参数化 Gurobi	7.82 9.20	17.11 1634.54	42.45 TL	83.88 TL	176.98 TL
	目标值	参数化 Gurobi	-277.17 -277.17	-547.57 -547.57	-1383.69 -1383.44	-2886.91 -2885.99	-5837.25 -5833.97
	平均方程数	参数化	995	1103	1082	1092	1139
	分支定界节点 最优性间隙	Gurobi	13265 0.00%	3162950 0.41%	3669416 1.82%	2583014 2.32%	1017048 2.51%

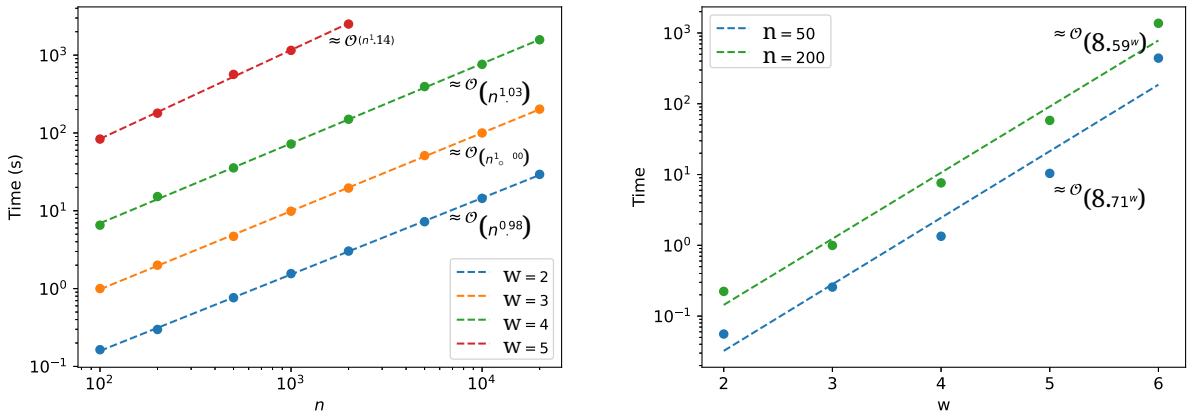

图 8：（左）固定 w 时，参数化算法的运行时间随 n 的变化情况，以双对数坐标显示。实测运行时间随 n 线性增长。（右）固定 n 时，运行时间随 w 的变化情况，以对数-线性坐标显示。近似线性的趋势表明，运行时间主要指数依赖于 w。

条件数的影响。对于 $n = 1,000$，我们改变条件数 κ_2 ，同时保持带宽 w 不变。正则化参数 $\boldsymbol{\lambda}$ 由独立且均匀分布于 (2，3) 的元素生成，以确保不同实例的最优解具有相近的稀疏度。图9（左）展示了对应的运行时间 $w = 2$，3、4、5。根据推论1，算法的运行时间随 κ_2呈多项式增长，且增长指数随带宽 w 增大而增大。这一经验观察与我们的理论结果完全一致。对于较小的带宽 $w = 2$，运行时间按 $\mathcal{O}(\kappa_2^{1.31})$ 的尺度增长；而较大的带宽则会带来显著更陡的增长。例如， $w = 5$ 呈现出约为 $\mathcal{O}(\kappa_2^{4.20})$ 的尺度增长。

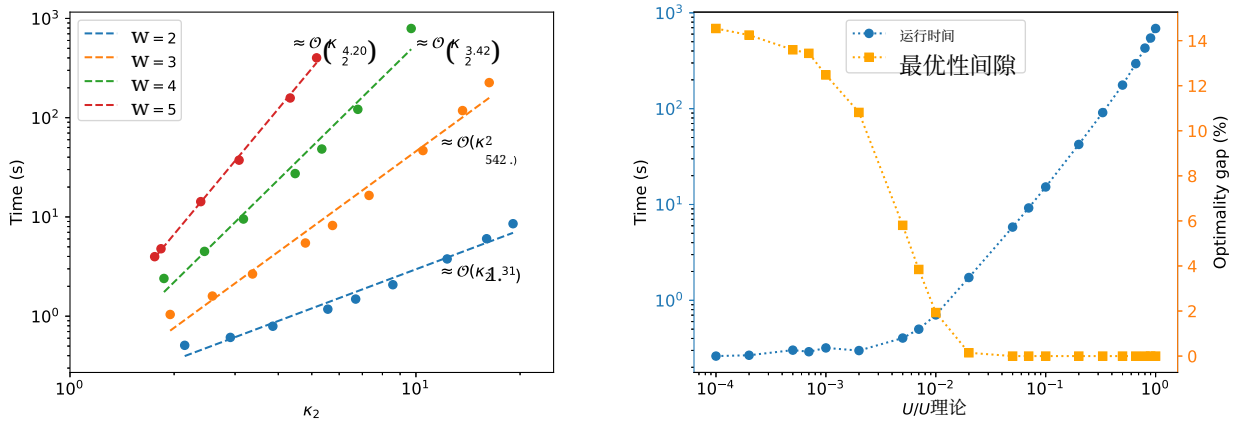

图9：（左）在固定 w时，参数化算法的运行时间作为 κ_2 的函数，以双对数坐标显示。经验运行时间随 κ_2呈多项式增长。（右）运行时间（左轴）与最优性间隙（右轴）作为归一化参数 U/U^{theory}的函数，以半对数坐标显示。减小 U 会降低运行时间，但代价是更大的最优性间隙。

解的上界的影响。 在接下来的实验中，我们研究 U的选择如何影响参数化算法的性能。令 U^{theory} 表示 U 的取值，该取值由引理3给出。对于这一选择，最优解满足 $\|\boldsymbol{x}^\star\|_\infty \le U^{\text{theory}}$，因此参数化算法保证能够恢复精确最优解。当 $U < U^{\text{theory}}$时，剪枝子程序会更加激进，从而降低算法的运行时间。然而，这种运行时间上的改善可能以剔除潜在最优稀疏模式为代价，从而导致非零的最优性间隙。

图 9 （右）展示了运行时间（蓝色圆点，左侧 y 轴）和最优性间隙（橙色方块，右侧 y 轴）随归一化参数 U/U^{theory}变化的情况。在此实验中，我们固定 $n = 1,000$， $w = 4$ 以及 $\kappa_2 \approx 9.69$。正如预期，当 $U/U^{\text{theory}} = 1$时，算法能够恢复精确的最优解。当 U/U^{theory} 降至 1 以下时，运行时间随之缩短。有趣的是，当$U/U^{\text{theory}} \ge 0.05$时，算法始终能够恢复最优解，而运行时间则从 600 秒大幅降至仅 6 秒。这表明， U的理论界在实际应用中可能较为保守。

稀疏度参数的影响。 下一项实验旨在评估稀疏度参数 $\boldsymbol{\lambda}$ 的选择如何影响参数化算法的性能。我们将问题规模固定为 $n = 1,\ 000$，带宽固定为 $w = 4$，条件数约为 $\kappa_2 \approx 6.56$。¯每次试验中，每个坐标 λ_i 都设为相同的常数值，记作 λ。结果见表 2。随着最优解变得更加稠密，参数化算法的运行时间略有增加；当解中非零元素比例（NZ）为 93.7% 时，运行时间达到 135.41 秒。相比之下， Gurobi 仅在这一个稠密情形下返回了最优解。在¯所有其他设置中， Gurobi 均超时，且报告的最优性间隙随着 λ急剧增大。对于极度稀疏的解（即非零元素比例仅为 0.2%）， Gurobi 报告的最优性间隙为$+\infty$。

表 2：不同正则化参数下的性能比较

指标	**方法**	$\bar{\boldsymbol{\lambda}} = \mathbf{0.07}$ 非零元素比例 ≈ 93.7%	$\bar{\boldsymbol{\lambda}} = \mathbf{1}$ 非零元素比例 ≈ 75.2%	$\bar{\boldsymbol{\lambda}} = \mathbf{5}$ 非零元素比例 ≈ 50.8%	$\bar{\boldsymbol{\lambda}} = \mathbf{7}$ 非零元素比例 ≈ 35.0%	$\bar{\boldsymbol{\lambda}} = \mathbf{14}$ 非零元素比例 ≈ 7.38%	$\bar{\boldsymbol{\lambda}} = \mathbf{20}$ 非零元素比例 ≈ 0.2%
时间（秒）	参数化 Gurobi	135.41 10.04	119.85 TL	86.71 TL	67.44 TL	36.23 TL	21.97 TL
分支定界节点 最优性间隙	Gurobi	1 0.00%	1780599 0.15%	2549880 2.26%	2840523 9.63%	4025755 210.25%	4416035 ∞

非零元素比例指最优解 $\boldsymbol{x}^\star$ 中非零元素所占的百分比。

这一行为凸显了参数化算法的一项关键优势。在 ESOC 等应用中，通常只有一小部分观测值受到离群噪声的污染。正是在这些情形下，当 Gurobi 返回最优性间隙较大的解时，参数化算法能以小得多的运行时间恢复最优解。

超越带状结构。接下来，我们表明，小树宽所蕴含的结构性质可以在简单带状结构之外加以利用。为此，我们比较本算法两种变体的性能：一种充分利用问题的低树宽结构，另一种仅利用带状结构。我们针对 $\boldsymbol{Q}$ 构造实例，其中树宽固定为 $\omega = 2$，而带宽在 $w = \{3, 4, 5\}$范围内变化。如图 10所示。为生成矩阵 $\boldsymbol{Q}$ 并使其具有所需性质，我们首先按照上一节所述步骤，构造一个具有指定带宽的对称正定矩阵。随后，我们有选择地将某些非对角元素置零，以将树宽降至 $\omega = 2$，同时保留原有的带状结构。

表 3 报告了在固定树宽 $\omega = 2$ 和带宽$w = 3$、4、5 时我们的方法的性能。所有实验的问题规模均为 $n = 1,\ 000$，对应的条件数 κ_2 也在表中报告。标记为“参数化”的行对应于将参数化算法应用于宽度为 $\omega = 2$的树分解。标记为“带状”的行对应于仅在带状结构假设下应用相同的参数化算法；在这种情况下，诱导树分解的宽度等于 $w > 2$。如表所示，利用图较小的树宽（而不仅仅是其带状结构），可显著降低运行时间和二次片段的平均数量。相比之下， Gurobi 无法在一小时时间限制内求解这些实例中的任何一个。

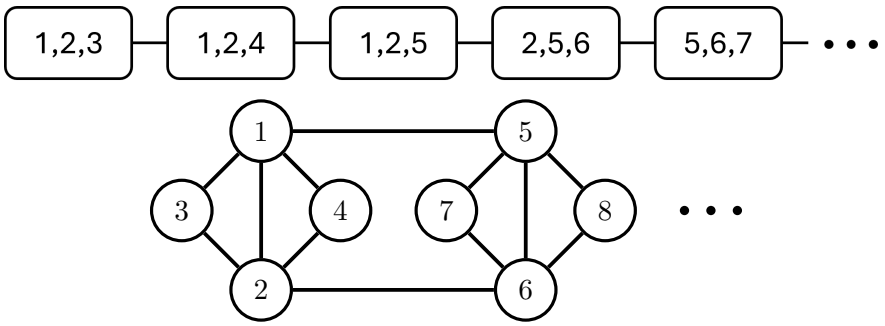

图 10：由带宽为 $w = 4$、树宽为 $\omega = 2$ 的带状矩阵导出的图的树分解。通过消除边并保留原始带宽，可降低树宽。

表 3：不同带宽下的性能比较

指标	方法	$w = 3$	$w = 4$	$w = 5$
条件数（κ_2）	—	8.71	9.25	8.12
时间（秒）	参数化	8.60	48.58	167.83
	带状	8.52	64.36	658.24
	Gurobi	TL	TL	TL
平均方程数	参数化	81	442	1477
	带状	123	848	3385
分支定界节点	Gurobi	3193465	3594783	3552258
最优性间隙		10.02%	14.02%	12.67%

7.3 真实世界实例的结果

接下来，我们应用所提出的参数化算法，求解第 3.2 节介绍的带离群值校正的指数平滑问题（ESOC）。除了运行时间外，我们还关注该模型相对于简单指数平滑（SES）的预测准确度。

数据集 我们的实验使用了来自 Numenta 异常基准（NAB） [1] 的四个真实世界时间序列。NAB 是一个旨在评估流数据异常检测算法的数据集，其中包含来自不同领域的真实世界时间序列数据，例如云基础设施指标和交通数据。有关该数据集、评估协议和相关方法的详细说明，请参阅 [51, 2]。我们的实验使用的具体数据集如下：

1. ec2 cpu utilization 53ea38.csv；简称 CPU-1；2. ec2 cpu utilization ac20cd.csv；简称 CPU-2；3. rds cpu utilization e47b3b.csv；简称 CPU-3；4. speed 7578.csv；简称 Traffic。

下面简要介绍这些数据集。

CPU 利用率数据 前三个数据集由从亚马逊云科技（AWS）服务器收集的 CPU 利用率百分比组成。在这类数据集中，异常可能源于工作负载突然激增或分布式拒绝服务（DDoS）攻击 [32]。前两个数据集对应通用计算实例（亚马逊 EC2），第三个数据集则来自数据库服务器（亚马逊 RDS）。每个数据集包含 2，000 个 CPU 利用率测量值，记录间隔为 5 分钟。如图 11的前三行所示，这些时间序列呈现出不同的时间模式，为评估 ESOC模型的准确性提供了丰富多样的行为样本。

交通数据 最终数据集由明尼苏达州双子城都会区的平均交通速度测量值组成，包含 1,127 条以不规则时间间隔记录的观测值。准确的短期交通速度预测对于预判不断变化的道路状况并实现有效的交通管理至关重要 [58]。在此背景下，异常可能预示着造成交通阻塞的重大交通事故 [76, 68]。该数据集显示在图 11 的最后一行，并被纳入以评估所提出的方法在采样不规则性和噪声特征与 CPU 利用率信号显著不同的数据上的表现。

实验设置 令 $\boldsymbol{y} \in \mathbb{R}^T$ 表示长度为 T 的时间序列，其中 $\boldsymbol{y}_t$ 表示时间 t处的观测值。我们的目标是在每个时间 t 生成单步点预测。我们将所得预测分别记为 SES 的 $\hat{}\boldsymbol{y}_t^{\mathrm{SES}}$ 和 ESOC 的 $\hat{}\boldsymbol{y}_t^{\mathrm{ESOC}}$。对于这两种模型，时间 $t+1$ 处的单步预测均由前一时间t [46]的平滑值得到。具体而言，对于 SES，预测由 $\hat{}\boldsymbol{y}_{t+1}^{\mathrm{SES}} := \boldsymbol{x}_t^{\mathrm{SES}} = \beta \boldsymbol{y}_t + (1-\beta)\boldsymbol{x}_{t-1}^{\mathrm{SES}}$ 给出，其中 $\boldsymbol{x}_t^{\mathrm{SES}}$表示指数平滑信号。对于 ESOC，预测定义为 $\hat{}\boldsymbol{y}_{t+1}^{\mathrm{ESOC}} := \boldsymbol{x}_t^{\mathrm{ESOC}}$，其中 $\boldsymbol{x}^{\mathrm{ESOC}} \in \mathbb{R}^T$ 是通过求解 ESOC 得到的平滑信号。为评估 SES 的预测准确度，我们使用均方误差（均方误差）：

$$\mathrm{MSE}^{\mathrm{SES}} = \frac{1}{T-1}\sum_{t=2}^{T}\left(\hat{\boldsymbol{y}}_t^{\mathrm{SES}} - \boldsymbol{y}_t\right)^2. \tag{18}$$

与（SES）中的模型不同，（ESOC）中的模型能够识别离群值并降低其影响。因此，我们仅在无离群值区域上评估其预测准确度。具体而言，我们定义

$$\mathrm{MSE}^{\mathrm{ESOC}} = \frac{1}{T-1-\sum_{t=2}^{T}\mathbb{I}(\boldsymbol{o}_t)}\sum_{t=2}^{T}\left(1-\mathbb{I}(\boldsymbol{o}_t)\right)\left(\hat{\boldsymbol{y}}_t^{\mathrm{ESOC}} - \boldsymbol{y}_t\right)^2, \tag{19}$$

其中 $\mathbb{I}(\boldsymbol{o}_t)$是一个指示函数，当 $\boldsymbol{y}_t$ 被识别为离群值时等于1，否则等于0。

两种模型的参数均通过对预定义的候选值集合进行网格搜索来选取。为进行参数调优，我们将每条时间序列划分为由前一半观测值组成的训练集，$\{\boldsymbol{y}_t\}_{t=1}^{\lfloor T/2 \rfloor}$，以及由其余观测值组成的测试集，$\{\boldsymbol{y}_t\}_{t=\lfloor T/2 \rfloor+1}^{T}$。对于每种方法及参数配置，我们计算训练集上的均方误差，并选择使该量最小的参数。使用所选参数，我们在整个信号上运行每种算法，并通过计算测试集上的均方误差来比较性能，依据（18）对（SES）以及（19）对（ESOC）。

对于（ESOC）中的模型，我们对以下参数进行网格搜索：$\beta \in \{0.01, 0.1, 0.2, \ldots, 0.9, 0.99\}$ 和 $\boldsymbol{\lambda}_t = \lambda$，$t = 1$，……，$T$，其中 $\lambda \in \{10^{-5}, 5\times10^{-5}, 10^{-4}, 5\times10^{-4}, 10^{-3}, 5\times10^{-3}, 10^{-2}, 5\times10^{-2}\}$，

同时固定 $\mu_1 = 1.2$ 和 $\mu_2 = 0.001$。为避免退化解，我们将网格搜索限制在这样的参数设置范围内：被分类为异常的观测值少于 10%。由于（SES）中的模型只有一个可调参数 β，我们使用与上文相同的候选集来选择 β。

结果表4 报告了使用在训练集上获得的参数得到的平滑信号的训练与测试均方误差。对于（ESOC），我们还报告了检测到的离群值百分比。在所有情况下，（ESOC）的均方误差均低于（SES），表明其能够抑制异常并产生更准确的预测。值得注意的是，测试均方误差的最大改进出现在交通和CPU-3 信号上。我们注意到，检测到的离群值比例仅在训练阶段被约束在10%以下。本表中的结果是通过使用训练中学习到的参数，对整个信号（训练和测试）求解优化问题而得到的。因此，对于交通信号，观测到的离群值比例超过了该阈值。

表4：SES与ESOC的均方误差比较。

信号	分段	MSE		检测到的离群值
		SES	ESOC	
CPU-1	训练	0.0101	0.0063	3.2%
	Test	0.0106	0.0068	2.8%
CPU-2	训练	9.1930	3.0941	2.9%
	Test	5.3983	3.1840	3.0%
CPU-3	训练	6.4085	0.1404	6.0%
	Test	0.8021	0.1649	16.6%
交通	训练	20.2650	6.7490	11.5%
	Test	65.5931	6.0920	20.7%

表 5 比较了 Gurobi 与所提出的参数化算法在求解（ESOC）时的运行时间。对于所有实例，Gurobi 均达到一小时时间限制；因此我们仅报告终止时的最优性间隙。相比之下，参数化算法对所有实例均返回可证明的最优解。大多数数据集在约30秒内得到求解，其中一个明显更大的实例（CPU-3）约需958秒。总体而言，这些结果表明参数化算法始终能够获得最优解，并且效率远高于 Gurobi。

表5：参数化算法与 Gurobi 在求解ESOC时的性能比较。

指标	方法	CPU-1	CPU-2	CPU-3	交通
信号长度（T）	—	2000	2000	2000	1127
时间（秒）	参数化	21.24	24.94	958.93	27.10
	Gurobi	TL	TL	TL	TL
分支定界节点	Gurobi	2077648	3909773	1609698	7360291
最优性间隙		78.88%	2.04%	1.06%	1.94%

最后，图 11 展示了使用模型（SES）和（ESOC）在全部四个数据集上获得的恢复信号。

每一行对应一个数据集，左列显示 SES 的输出，右列显示 ESOC 的输出。在所有面板中，观测信号均与

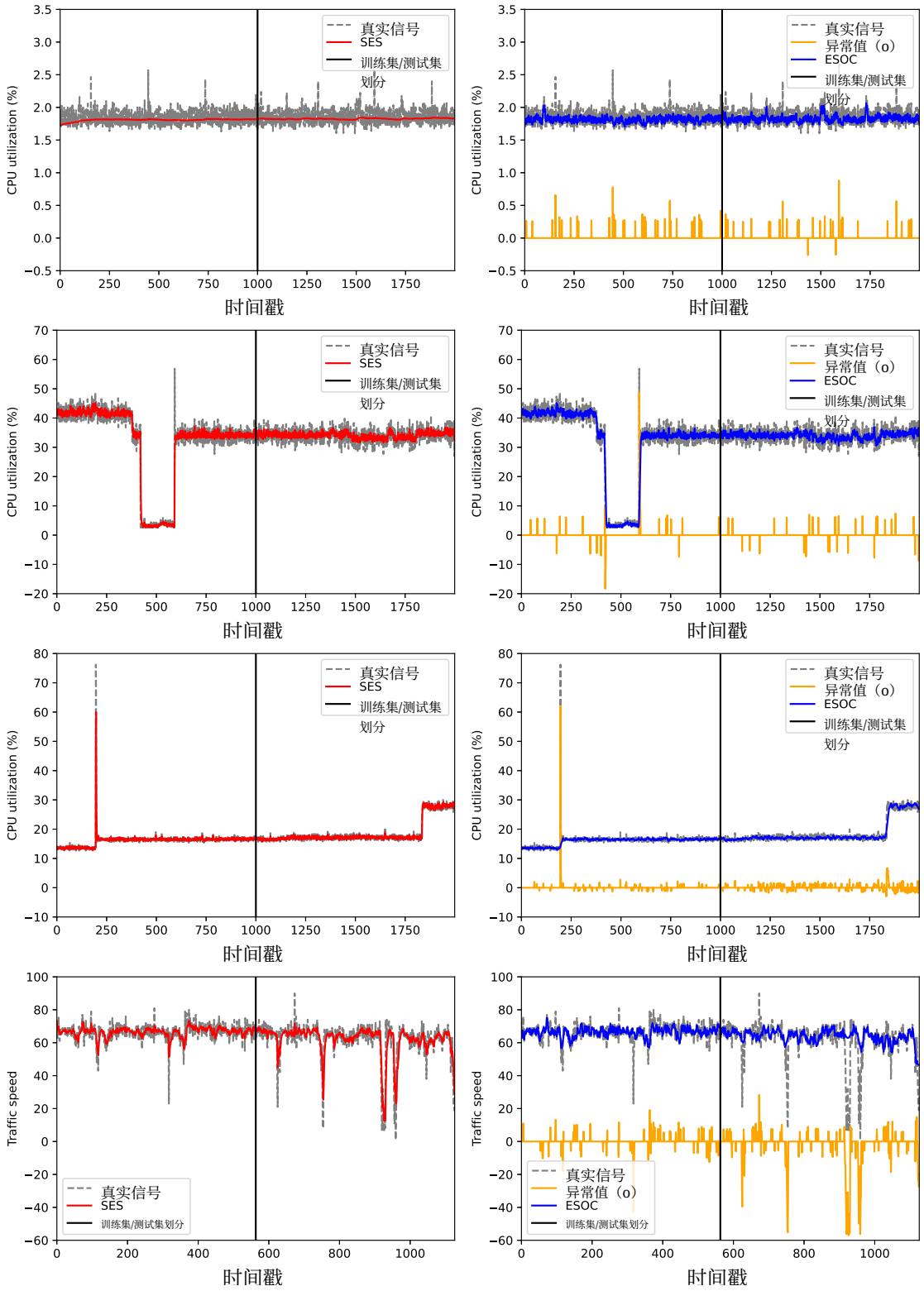

图11：四个具有代表性的真实世界信号上 SES（左列）和 ESOC（右列）的比较。每一行对应一个不同的信号实例，同时展示了观测信号和平滑估计。

相应的平滑估计，而对于 ESOC，估计的离群值向量 $\boldsymbol{o}$ 以橙色突出显示。

对于第一个数据集（CPU-1），训练集中的趋势能够代表整个序列。在 SES 下，由于 β 的取值较小（在训练数据上选取），预测相对平滑，但在序列早期表现出明显的滞后，此时递推主要由初始观测 $\boldsymbol{y}_1$ 主导（见第 3.1 节）。相比之下，来自 ESOC 的预测适应得更快，更紧密地跟踪底层信号，并且通过显式建模离群值，对初始值 $\boldsymbol{y}_1$的敏感性更低。

对于第二个数据集（CPU-2），训练段与测试段表现出明显不同的行为。训练部分包含突变（介于 $t = 420$ 与 $t = 600$ 之间），随后出现一个较大的离群值，而这两者在测试集（信号的后半段）中均不存在。虽然 SES 和 ESOC 都削弱了该离群值，但 ESOC 对底层趋势给出了更平滑的估计。如表 4 所示，这使得 ESOC 的测试均方误差大幅降低。

对于第三个数据集（CPU-3），训练集包含一个较大的离群值，随后出现一次跳变（约在 $t \approx 240$），而测试集则多出现一次跳变。事实上，SES 能够适应这些跳变，但会将较大的离群值纳入预测中。相比之下，ESOC 成功识别并去除离群值，准确跟踪跳变，并对潜在趋势给出更平滑的估计。

最后，对于第四个数据集（交通），信号的平滑度较低，在训练集和测试集中都呈现频繁的短期波动。在时间范围末端，出现了两个较大的尖峰。虽然 SES 保留了大小波动，包括较大的尖峰，ESOC 则能捕捉潜在趋势和较小的变化，同时有效剔除较大的尖峰。

8 结论

总而言之，本文表明，通过利用图结构稀疏性和间隔条件，可以高效求解带有指示变量的凸二次优化问题。借助这些性质，所提出的参数化算法在适当假设下可实现多项式时间复杂度和空间复杂度；在某些情形下，甚至可达到线性时间复杂度和空间复杂度，在大规模实例上显著优于通用混合整数求解器。除理论保证外，该框架还为指数平滑提供了一种原则性扩展，可用于联合预测和异常值检测，并在具有挑战性的真实世界时间序列数据集上展现出强大的实证性能。

9 代码和数据披露

代　　码以及支撑本文数值实验的数据可在以下地址找到

https://github.com/aareshfb/Treewidth-Parametric-Algorithm.

请注意，所提供的实现仅适用于树分解为路径的实例。实验中使用的包含真实世界数据的Numenta异常基准（NAB） [1] 可通过以下网址访问

https://github.com/numenta/NAB/tree/master/data。

致谢

本研究部分获得了美国国家科学基金会资助（项目编号 2152776、2337776）、美国海军研究办公室资助（项目编号 N00014-26-1-2074、N00014-22-1-2602、N00014-26-12117），以及美国空军科学研究办公室资助（项目编号 FA9550-24-1-0086）。

参考文献

[1] Subutai Ahmad 和 Alexander Lavin。《Numenta 异常基准》。https://github.com/numenta/NAB，2015年。[2] Subutai Ahmad、Alexander Lavin、Scott Purdy 和 Zuha Agha。面向流数据的无监督实时异常检测。《神经计算》，262：134–147，2017年。[3] Ravindra K Ahuja、Thomas L Magnanti 和 James B Orlin。《网络流》。普伦蒂斯·霍尔，1988年。[4] M Selim Akt¨urk、Alper Atamt¨urk 和 Sinan G¨urel。具有可控加工时间的机器-作业分配的强锥二次重构。《运筹学快报》，37(3)：187–191，2009年。[5] Stefan Arnborg 和 Andrzej Proskurowski。针对部分k树上NP困难问题的线性时间算法。《离散应用数学》，23(1)：11–24，1989年。[6] Stefan Arnborg、Derek G Corneil 和 Andrzej Proskurowski。寻找k树嵌入的复杂性。《SIAM代数离散方法杂志》，8(2)：277–284，1987年。[7] Alper Atamt¨urk 和 Andr´es G´omez。基于M矩阵和指示变量的二次优化强公式。《Mathematical Programming》，170(1)：141–176，2018年。[8] Alper Atamt¨urk、Andr´es G´omez 和 Shaoning Han。稀疏光滑信号估计：ℓ_0-形式的凸化。《Journal of Machine Learning Research》，22(52)：1–43，2021年。[9] Marshall W. Bern、Eugene L. Lawler 和 Alice L Wong。可分解图最优子图的线性时间计算。《算法杂志》，8(2)：216–235，1987年。[10] Dimitri Bertsekas。《动态规划与最优控制：第一卷》，第4卷。Athena Scientific，2012年。[11] Dimitris Bertsimas 和 Bart Van Parys。稀疏高维回归。《The Annals of Statistics》，48(1)：300–323，2020年。[12] Dimitris Bertsimas、Angela King 和 Rahul Mazumder。透过现代优化视角进行最佳子集选择。《The Annals of Statistics》，44(2)：813–852，2016年。[13] Dimitris Bertsimas、Vassilis Digalakis Jr、Michael Lingzhi Li 和 Omar Skali Lami。稀疏约束下的缓变回归。《运筹学》，73(3)：1581–1597，2025年。[14] Aaresh Bhathena、Salar Fattahi、Andr´es G´omez 和 Simge K¨uc¸¨ukyavuz。一种用于求解树上带指示变量的凸二次优化问题的参数化方法。《Mathematical Programming》，第1–46页，2025年。

[15] Daniel Bienstock 和 Tongtong Chen。指示条件下具有结构化稀疏性的凸二次规划求解。arXiv预印本 arXiv:2411.11722，2024年。[16] Daniel Bienstock 和 Gonzalo Munoz。多项式优化问题的线性规划公式。《SIAM Journal on Optimization》，28(2)：1121–1150，2018年。[17] Baki Billah、Maxwell L King、Ralph D Snyder 和 Anne B Koehler。用于预测的指数平滑模型选择。《国际预测学杂志》，22(2)：239–247，2006年。[18] Hans L Bodlaender。树宽有界图上的动态规划。载于《自动机、语言与编程：第15届国际研讨会，芬兰坦佩雷，1988年7月11日至15日，论文集15》，第105–118页。施普林格，1988年。[19] Hans L Bodlaender。《树宽导览》，第92卷。出版商不详，1992年。[20] Hans L. Bodlaender。寻找小树宽树分解的线性时间算法。《SIAM计算期刊》，25(6)：1305–1317，1996年。[21] Hans L Bodlaender 和 Ton Kloks。求解图的路径宽和树宽的高效构造性算法。《算法杂志》，21(2)：358–402，1996年。[22] Sebasti´an Ceria 和 Jo˜ao Soares。用于析取凸优化的凸规划。《Mathematical Programming》，86(3)：595–614，1999年。[23] Charalambos A Charalambides。《枚举组合学》。Chapman and Hall/CRC，2018年。[24] Xiaojun Chen、Dongdong Ge、Zizhuo Wang 和 Yinyu Ye。无约束最小化的复杂性。《Mathematical Programming》，143(1-2)：371–383，2014年。[25] Abhimanyu Das 和 David Kempe。线性回归中子集选择的算法。载于第四十届ACM计算理论年会论文集，第45–54页，2008年。[26] Antoine Dedieu、Hussein Hazimeh 和 Rahul Mazumder。稀疏分类器学习：连续优化与混合整数优化视角。《Journal of Machine Learning Research》，22 (135)：1–47，2021年。[27] Alberto Del Pia、Santanu S Dey 和 Robert Weismantel。稀疏矩阵中的子集选择。《SIAM Journal on Optimization》，30(2)：1173–1190，2020年。[28] Stephen Demko、William F Moss 和 Philip W Smith。带状矩阵逆矩阵的衰减率。《计算数学》，43(168)：491–499，1984年。[29] Ilias Diakonikolas 和 Nikos Zarifis。学习带Massart噪声的间隔半空间的近最优算法。《神经信息处理系统进展》，37：88605–88626，2024年。[30] Reinhard Diestel。《图论》，第173卷。Springer Nature，2025年。[31] Farzam Ebrahimnejad 和 James R Lee。关于一致多项式增长的平面图。《概率论及相关领域》，180(3)：955–984，2021年。

[32] Mahmoud Said El Sayed、Nhien-An Le-Khac、Marianne A Azer 和 Anca D Jurcut。针对软件定义网络中 DDoS 攻击的基于流的异常检测方法与特征选择方法。IEEE《认知通信与网络汇刊》，8(4)：1862–1880，2022年。[33] Salar Fattahi 和 Andres G´omez。稀疏变化高斯马尔可夫随机场的可扩展推断。《神经信息处理系统进展》，34：6529–6541，2021年。[34] Antonio Frangioni 和 Claudio Gentile。面向一类凸 0–1 混合整数规划的透视割。Mathematical Programming，106(2)：225–236，2006年。[35] Antonio Frangioni 和 Claudio Gentile。透视松弛重构的计算比较：二阶锥规划与割平面。《运筹学快报》，37(3)：206–210，2009年。[36] Everette S Gardner Jr.。指数平滑：现状——第二部分。《国际预测学杂志》，22(4)：637–666，2006年。[37] Fred Glover。非线性整数问题的改进线性整数规划形式。《管理科学》，22(4)：455–460，1975年。[38] Andr´es G´omez 和 Weijun Xie。关于带指示变量的二次约束的注记：凸包描述与透视松弛。《运筹学快报》，52：107059，2024年。[39] Andres G´omez、Shaoning Han 和 Leonardo Lozano。带状矩阵和指示变量下二次优化问题的实时求解。arXiv 预印本 arXiv:2405.03051，2024年。[40] Oktay G¨unl¨uk 和 Jeff Linderoth。带指示变量的混合整数非线性规划的透视重构。Mathematical Programming，124(1)：183–205，2010年。[41] Rudolf Halin。图的 S 函数。《几何学杂志》，8(1-2)：171–186，1976年。[42] Shaoning Han 和 Andr´es G´omez。带指示变量的低秩函数的紧凑扩展形式。《运筹学数学》，50(3)：1992–2016，2025年。[43] Shaoning Han、Andr´es G´omez 和 Jong-Shi Pang。关于组合马尔可夫随机场的多项式时间可解性。arXiv 预印本 arXiv:2209.13161，2022年。[44] Hussein Hazimeh、Rahul Mazumder 和 Ali Saab。大规模稀疏回归：根植于一阶优化的分支定界。Mathematical Programming，196(1)：347–388，2022年。[45] Pinar Heggernes。图的极小三角剖分：综述。《离散数学》，306(3)：297–317，2006年。[46]Rob Hyndman、Anne B Koehler、J Keith Ord 和 Ralph D Snyder。指数平滑预测：状态空间方法。Springer Science & Business Media，2008年。[47] Esma Kahraman 和 Ozlem Akay。主要金属全球价格预测中的指数平滑方法比较。《矿产经济学》，36(3)：427–435，2023。

[48] Masakazu Kojima、Sunyoung Kim 和 Hayato Waki。多项式平方和的稀疏性。Mathematical Programming, 103(1)：45–62，2005年。[49] George Kontogeorgiou 和 Martin Winter。（随机）中间均匀增长的树。arXiv 预印本 arXiv:2212.01883，2022年。[50] Simge K¨u˙c¨ukyavuz、Ali Shojaie、Hasan Manzour、Linchuan Wei 和 Hao-Hsiang Wu。用于贝叶斯网络学习的一致二阶锥整数规划。《Journal of Machine Learning Research》，24(322)：1–38，2023。[51] Alexander Lavin 和 Subutai Ahmad。评估实时异常检测算法——Numenta 异常基准。在 2015 IEEE 第 14 届国际机器学习与应用会议（ICMLA）论文集，第 38–44 页。IEEE，2015年。[52] Jisun Lee、Andr´es G´omez 和 Alper Atamt¨urk。带指示变量的多周期二次规划凸化。arXiv 预印本 arXiv:2412.17178，2024年。[53] Peijing Liu、Salar Fattahi、Andr´es G´omez 和 Simge K¨u˙c¨ukyavuz。带指示变量的凸二次优化的基于图的分解方法。Mathematical Programming，200(2)：669–701，2023年。[54] Peijing Liu、Alper Atamt¨urk、Andr´es G´omez 和 Simge K¨u˙c¨ukyavuz。带 Stieltjes 矩阵和指示变量的二次优化问题的多面体分析。即将发表于 Mathematical Programming，第 1–27 页，2025年。[55] Ramtin Madani、Somayeh Sojoudi、Ghazal Fazelnia 和 Javad Lavaei。使用凸优化寻找稀疏线性矩阵不等式的低秩解。SIAM Journal on Optimization，27(2)：725–758，2017年。[56] Hasan Manzour、Simge K¨u˙c¨ukyavuz、Hao-Hsiang Wu 和 Ali Shojaie。从连续数据中学习有向无环图的整数规划。INFORMS 优化期刊，3(1)：46–73，2021年。[57] G¨unter Meinardus。函数逼近：理论与数值方法，第 13 卷。Springer Science & Business Media，2012年。[58] Yi Ouyang、Richard Y Zhang、Javad Lavaei 和 Pravin Varaiya。大规模交通信号偏移优化。IEEE 网络系统控制汇刊，7(3)：1176–1187，2020年。[59] S Ratnasari、R Zakaria 等。采用五参数指数平滑进行需求预测。在 IOP 会议系列：材料科学与工程，第 495 卷，第 012014 页。IOP Publishing，2019年。[60] Visweswaran Ravikumar、Aaresh Bhathena、Wajd N Al-Holou、Salar Fattahi 和 Arvind Rao。使用离散惩罚高效推断动态基因调控网络。arXiv 预印本 arXiv:2507.23106，2025年。[61] Neil Robertson 和 Paul D. Seymour。图小 II：树宽的算法方面。《算法杂志》，7(3)：309–322，1986年。[62] Walter Rudin。泛函分析。McGraw–Hill，1973年。

[63] Soroosh Shafiee 和 Fatma Kılınc-Karzan。带指示变量的秩一函数约束优化：S. shafiee, f. kılınc-karzan。Mathematical Programming，208(1):533–579，2024年。[64] Robert Alan Stubbs。混合0-1凸规划的分支切割法。博士论文，西北大学，1996年。[65] Bilal Taghezouit、Fouzi Harrou、Ying Sun、Amar Hadj Arab 和 Cherif Larbes。用于光伏系统监测的双指数方案：一种简单有效的检测策略。Solar Energy，214:337–354，2021年。[66] Quang Thanh Tran、Li Hao 和 Quang Khai Trinh。蜂窝网络流量建模与预测中指数平滑方法的综合研究。Concurrency and Computation: Practice and Experience，32(23):e5602，2020年。[67] Alexander B Tsybakov。统计学习中分类器的最优聚合。The Annals of Statistics，32(1):135–166，2004年。[68] Franco Van Wyk、Yiyang Wang、Anahita Khojandi 和 Neda Masoud。自动驾驶车辆中的实时传感器异常检测与识别。IEEE Transactions on Intelligent Transportation Systems，21(3):1264–1276，2019年。[69] Martin J Wainwright 和 Michael I Jordan。Sherali-Adams 和 Lasserre 松弛精确性的基于树宽的条件。技术报告，技术报告 671，加州大学伯克利分校，2004年。[70] Hayato Waki、Sunyoung Kim、Masakazu Kojima 和 Masakazu Muramatsu。具有结构化稀疏性的多项式优化问题的平方和与半定规划松弛。SIAM Journal on Optimization，17(1):218–242，2006年。[71] Linchuan Wei、Andr´es G´omez 和 Simge K¨u¸c¨ukyavuz。带指示变量的约束凸优化问题的理想表述。Mathematical Programming，192(1): 57–88，2022年。[72] Linchuan Wei、Alper Atamt¨urk、Andr´es G´omez 和 Simge K¨u¸c¨ukyavuz。带指示变量的凸二次优化问题的凸包。Mathematical Programming，204(1): 703–737，2024年。[73] Weijun Xie 和 Xinwei Deng。稀疏岭回归的可扩展算法。SIAM Journal on Optimization，30(4):3359–3386，2020年。[74] Tong Xu、Simge K¨u¸c¨ukyavuz、Ali Shojaie 和 Armeen Taeb。用于从高斯模型中学习贝叶斯网络的渐近最优坐标下降算法。即将在 Journal of Machine Learning Research 上发表，2025年。URL https://arxiv.org/abs/2408.11977。[75] Tong Xu、Armeen Taeb、Simge K¨u¸c¨ukyavuz 和 Ali Shojaie。基于不可识别高斯模型学习有向无环图的整数规划。Biometrika，112(3):asaf032，04 2025年。

[76] Chao Zhao、Xiaokun Chang、Tian Xie、Hamido Fujita 和 Jian Wu。基于无监督异常检测的道路交通事故风险评估方法。Applied Intelligence，53(1): 369–384，2023。

A 省略的证明

A.1 引理1的证明

回忆一下， $\mathcal{J}_u$ 定义为 $\text{supp}_u(\boldsymbol{Q})$ 中的节点集合，但不包括 $\mathcal{B}_u$ 中的节点。定义$p_{u,\ \boldsymbol{s}}(\boldsymbol{\alpha}_{\mathcal{B}_u})$为

$$p_{u,\boldsymbol{s}}(\boldsymbol{\alpha}_{\mathcal{B}_u}) = \min_{\boldsymbol{x}\in\mathbb{R}^{n_u}} \frac{1}{2}\boldsymbol{\alpha}_{\mathcal{B}_u}^\top \boldsymbol{Q}_{\mathcal{B}_u,\mathcal{B}_u}\boldsymbol{\alpha}_{\mathcal{B}_u} + \boldsymbol{c}_{\mathcal{B}_u}^\top \boldsymbol{\alpha}_{\mathcal{B}_u} + \left(\frac{1}{2}\boldsymbol{x}^\top \boldsymbol{Q}_{\mathcal{J}_u,\mathcal{J}_u}\boldsymbol{x} + \boldsymbol{\alpha}_{\mathcal{B}_u}^\top \boldsymbol{Q}_{\mathcal{B}_u,\mathcal{J}_u}\boldsymbol{x} + \boldsymbol{c}_{\mathcal{J}_u}^\top \boldsymbol{x} + \boldsymbol{\lambda}_{\mathcal{J}_u}^\top \boldsymbol{s}\right)$$

$$\text{s.t.}\quad x_i(1-s_i) = 0 \quad i \in \mathcal{J}_u.$$

容易验证， $f_u(\boldsymbol{\alpha}_{\mathcal{B}_u}) = \min_{\boldsymbol{s}\in\{0,1\}^{n_u}}\{p_{u,\boldsymbol{s}}(\boldsymbol{\alpha}_{\mathcal{B}_u})\}$。因此，接下来需要刻画每个 $\boldsymbol{s} \in \{0,1\}^{n_u}$ 对应的 $p_{u,\ \boldsymbol{s}}(\boldsymbol{\alpha}_{\mathcal{B}_u})$ 的显式形式，并证明它是强凸二次函数。

首先，令 $\mathcal{J}_{u,\ s} = \{i \in \mathcal{J}_u \mid s_i = 1\}$。因此，我们有

$$\begin{aligned} p_{u,\boldsymbol{s}}(\boldsymbol{\alpha}_{\mathcal{B}_u}) =& \frac{1}{2}\boldsymbol{\alpha}_{\mathcal{B}_u}^\top \boldsymbol{Q}_{\mathcal{B}_u,\mathcal{B}_u}\boldsymbol{\alpha}_{\mathcal{B}_u} + \boldsymbol{c}_{\mathcal{B}_u}^\top \boldsymbol{\alpha}_{\mathcal{B}_u} \\ &+ \min_{\boldsymbol{x}\in\mathbb{R}^{|\mathcal{J}_{u,\boldsymbol{s}}|}} \left\{\frac{1}{2}\boldsymbol{x}^\top \boldsymbol{Q}_{\mathcal{J}_{u,\boldsymbol{s}},\mathcal{J}_{u,\boldsymbol{s}}}\boldsymbol{x} + \boldsymbol{\alpha}_{\mathcal{B}_u}^\top \boldsymbol{Q}_{\mathcal{B}_u,\mathcal{J}_{u,\boldsymbol{s}}}\boldsymbol{x} + \boldsymbol{c}_{\mathcal{J}_{u,\boldsymbol{s}}}^\top \boldsymbol{x} + \sum_{i\in\mathcal{J}_{u,\boldsymbol{s}}} \boldsymbol{\lambda}_i\right\}. \end{aligned}$$

根据卡鲁什-库恩-塔克条件，可得

$$\begin{aligned} p_{u,\boldsymbol{s}}(\boldsymbol{\alpha}_{\mathcal{B}_u}) =& \frac{1}{2}\boldsymbol{\alpha}_{\mathcal{B}_u}^\top \left(\boldsymbol{Q}_{\mathcal{B}_u,\mathcal{B}_u} - \boldsymbol{Q}_{\mathcal{B}_u,\mathcal{J}_{u,\boldsymbol{s}}}\left(\boldsymbol{Q}_{\mathcal{J}_{u,\boldsymbol{s}},\mathcal{J}_{u,\boldsymbol{s}}}\right)^{-1}\boldsymbol{Q}_{\mathcal{B}_u,\mathcal{J}_{u,\boldsymbol{s}}}^\top\right)\boldsymbol{\alpha}_{\mathcal{B}_u} \\ &+ \left(\boldsymbol{c}_{\mathcal{B}_u} - \boldsymbol{c}_{\mathcal{J}_{u,\boldsymbol{s}}}^T\left(\boldsymbol{Q}_{\mathcal{J}_{u,\boldsymbol{s}},\mathcal{J}_{u,\boldsymbol{s}}}\right)^{-1}\boldsymbol{Q}_{\mathcal{B}_u,\mathcal{J}_{u,\boldsymbol{s}}}^\top\right)^\top \boldsymbol{\alpha}_{\mathcal{B}_u} \\ &+ \left(-\frac{1}{2}\boldsymbol{c}_{\mathcal{J}_{u,\boldsymbol{s}}}^\top(\boldsymbol{Q}_{\mathcal{J}_{u,\boldsymbol{s}},\mathcal{J}_{u,\boldsymbol{s}}})^{-1}\boldsymbol{c}_{\mathcal{J}_{u,\boldsymbol{s}}} + \sum_{i\in\mathcal{J}_{u,\boldsymbol{s}}}\lambda_i\right). \end{aligned}$$

此外，注意到 $\left(\boldsymbol{Q}_{\mathcal{B}_u,\mathcal{B}_u} - \boldsymbol{Q}_{\mathcal{B}_u,\mathcal{J}_{u,\boldsymbol{s}}}\left(\boldsymbol{Q}_{\mathcal{J}_{u,\boldsymbol{s}},\mathcal{J}_{u,\boldsymbol{s}}}\right)^{-1}\boldsymbol{Q}_{\mathcal{B}_u,\mathcal{J}_{u,\boldsymbol{s}}}^\top\right)$ 是$\boldsymbol{Q}_{\mathcal{J}_{u,\ \boldsymbol{s}}\cup\{\mathcal{B}_u\},\ \mathcal{J}_{u,\ \boldsymbol{s}}\cup\{\mathcal{B}_u\}}$的舒尔补，由于 $\boldsymbol{Q}$正定，因此它也是正定的。证毕。

□

A.2 引理2的证明

首先，我们推导出每个 $v \in \text{part}(u)$ 对应的 g_v 的显式表达式。接下来，我们证明定义 f_u 的优化问题可以分解为针对每个 v 的独立子问题。

最后，我们将每个子问题中的 g_v 表达式代入，得到所述的 f_u 方程。

我们从第一步开始，推导 g_v 的显式表达式。固定 $v \in \mathrm{par}(u)$，并考虑局部参数化代价 f_v。按照 $\boldsymbol{x}_{\pi_v(\mathcal{B}_v)} = \boldsymbol{\alpha}_{\mathcal{B}_v}$ 设置变量，使其满足 4c（约束 f_v 定义中的约束条件），可得：

$$f_v(\boldsymbol{\alpha}_{\mathcal{B}_v}) = \frac{1}{2}\boldsymbol{\alpha}_{\mathcal{B}_v}^\top \boldsymbol{Q}_{\mathcal{B}_v,\mathcal{B}_v}\boldsymbol{\alpha}_{\mathcal{B}_v} + \boldsymbol{c}_{\mathcal{B}_v}^\top \boldsymbol{\alpha}_{\mathcal{B}_v} + \min_{\boldsymbol{x}\in\mathbb{R}^{n_v},\boldsymbol{z}\in\{0,1\}^{n_v}} \left\{ \frac{1}{2}\boldsymbol{x}_{\pi_v(\mathcal{J}_v)}^\top \boldsymbol{Q}_{\mathcal{J}_v,\mathcal{J}_v}\boldsymbol{x}_{\pi_v(\mathcal{J}_v)} + \boldsymbol{x}_{\pi_v(\mathcal{J}_v)}^\top \boldsymbol{Q}_{\mathcal{J}_v,\mathcal{B}_v}\boldsymbol{\alpha}_{\mathcal{B}_v} + \boldsymbol{c}_{\mathcal{J}_v}^\top \boldsymbol{x}_{\pi_v(\mathcal{J}_v)} + \boldsymbol{\lambda}_{\mathcal{J}_v}^\top \boldsymbol{z}_{\pi_v(\mathcal{J}_v)} \right\} \tag{23a}$$

$$\text{s.t.} \quad \boldsymbol{x}_i(1-\boldsymbol{z}_i) = 0 \qquad \forall i \in \pi_v(\mathcal{J}_v). \tag{23b}$$

分离出变量 $\boldsymbol{\alpha}_v$，$\frac{1}{2}\boldsymbol{\alpha}_{\mathcal{B}_v}^\top \boldsymbol{Q}_{\mathcal{B}_v}$ 的$\mathcal{B}_v$ $\boldsymbol{\alpha}_{\mathcal{B}_v} + \boldsymbol{c}_{\mathcal{B}_v}^\top \boldsymbol{\alpha}_{\mathcal{B}_v}$ 贡献后，可以将其改写为

$$\begin{aligned}\frac{1}{2}\boldsymbol{\alpha}_{\mathcal{B}_v}^\top \boldsymbol{Q}_{\mathcal{B}_v,\mathcal{B}_v}\boldsymbol{\alpha}_{\mathcal{B}_v} + \boldsymbol{c}_{\mathcal{B}_v}^\top \boldsymbol{\alpha}_{\mathcal{B}_v} &= \frac{1}{2}\boldsymbol{\alpha}_{\mathcal{B}_v\setminus v}^\top \boldsymbol{Q}_{\mathcal{B}_v\setminus v,\mathcal{B}_v\setminus v}\boldsymbol{\alpha}_{\mathcal{B}_v\setminus v} + \boldsymbol{c}_{\mathcal{B}_v\setminus v}^\top \boldsymbol{\alpha}_{\mathcal{B}_v\setminus v} \\ &\quad + \frac{1}{2}\boldsymbol{Q}_{v,v}\boldsymbol{\alpha}_v^2 + \boldsymbol{\alpha}_v \boldsymbol{Q}_{v,\mathcal{B}_v\setminus v}\boldsymbol{\alpha}_{\mathcal{B}_v\setminus v} + \boldsymbol{c}_v\boldsymbol{\alpha}_v \\ &= \phi_v(\boldsymbol{\alpha}_{\mathcal{B}_v\setminus v}) + \frac{1}{2}\boldsymbol{Q}_{v,v}\boldsymbol{\alpha}_v^2 + \boldsymbol{\alpha}_v \boldsymbol{Q}_{v,\mathcal{B}_v\setminus v}\boldsymbol{\alpha}_{\mathcal{B}_v\setminus v} + \boldsymbol{c}_v\boldsymbol{\alpha}_v. \end{aligned} \tag{24}$$

Si类似地，项 $\boldsymbol{x}_{\pi_v(\mathcal{J}_v)}^\top \boldsymbol{Q}_{\mathcal{J}_v,\ \mathcal{B}_v}\boldsymbol{\alpha}_{\mathcal{B}_v}$ 出现在 (23a) 中，可以分解为 s

$$\boldsymbol{x}_{\pi_v(\mathcal{J}_v)}^\top \boldsymbol{Q}_{\mathcal{J}_v,\mathcal{B}_v}\boldsymbol{\alpha}_{\mathcal{B}_v} = \boldsymbol{x}_{\pi_v(\mathcal{J}_v)}^\top \boldsymbol{Q}_{\mathcal{J}_v,\mathcal{B}_v\setminus v}\boldsymbol{\alpha}_{\mathcal{B}_v\setminus v} + \boldsymbol{x}_{\pi_v(\mathcal{J}_v)}^\top \boldsymbol{Q}_{\mathcal{J}_v,v}\boldsymbol{\alpha}_v. \tag{25}$$

现在回顾 g_v 的定义：

$$g_v(\boldsymbol{\alpha}_{\mathcal{B}_v\setminus v}) = \min_{\boldsymbol{x}_v\in\mathbb{R}} \left\{ f_v(\boldsymbol{x}_v, \boldsymbol{\alpha}_{\mathcal{B}_v\setminus v}) + \boldsymbol{\lambda}_v \mathbb{I}(\boldsymbol{x}_v) \right\}.$$

将 (24) 和 (25) 连同 (23) 代入 g_v 的表达式，可得

$$\begin{aligned} g_v(\boldsymbol{\alpha}_{\mathcal{B}_v\setminus v}) = \phi_v(\boldsymbol{\alpha}_{\mathcal{B}_v\setminus v}) + \min_{\boldsymbol{x}\in\mathbb{R}^{n_v+1},\boldsymbol{z}\in\{0,1\}^{n_v+1}} \Big\{ & \frac{1}{2}\boldsymbol{x}_{\pi_v(\mathcal{J}_v)}^\top \boldsymbol{Q}_{\mathcal{J}_v,\mathcal{J}_v}\boldsymbol{x}_{\pi_v(\mathcal{J}_v)} + \frac{1}{2}\boldsymbol{Q}_{v,v}\boldsymbol{x}_{\pi_v(v)}^2 + \\ & \boldsymbol{x}_{\pi_v(\mathcal{J}_v)}^\top \boldsymbol{Q}_{\mathcal{J}_v,\mathcal{B}_v\setminus v}\boldsymbol{\alpha}_{\mathcal{B}_v\setminus v} + \boldsymbol{x}_{\pi_v(v)}\boldsymbol{Q}_{v,\mathcal{B}_v\setminus v}\boldsymbol{\alpha}_{\mathcal{B}_v\setminus v} + \\ & \boldsymbol{x}_{\pi_v(\mathcal{J}_v)}^\top \boldsymbol{Q}_{\mathcal{J}_v,v}\boldsymbol{x}_{\pi_v(v)} + \boldsymbol{c}_v\boldsymbol{x}_{\pi_v(v)} + \boldsymbol{c}_{\mathcal{J}_v}^\top \boldsymbol{x}_{\pi_v(\mathcal{J}_v)} + \\ & \boldsymbol{\lambda}_v \boldsymbol{z}_{\pi_v(v)} + \boldsymbol{\lambda}_{\mathcal{J}_v}^\top \boldsymbol{z}_{\pi_v(\mathcal{J}_v)} \Big\} \end{aligned} \tag{26a}$$

$$\text{s.t.} \quad \boldsymbol{x}_i(1-\boldsymbol{z}_i) = 0 \qquad \forall i \in \pi_v(\mathcal{J}_v \cup \{v\}). \tag{26b}$$

推导出 g_v的表达式后，我们现在证明引理中给出的方程 f_u。

诱导子图 $\mathrm{supp}_u(\boldsymbol{Q})$ 中的节点可以表示为不交并 $\mathcal{B}_u \cup \mathcal{J}_u$，而集合 $\mathcal{J}_u$ 还可以进一步分解为

$$\mathcal{J}_u = \bigcup_{v\in\mathrm{par}_\mathsf{T}(u)} (\mathcal{J}_v \cup \{v\}).$$

为了形式化描述这一分解的性质，我们给出以下命题。

命题 2。 对于任意两个不同的 $v_i, v_j \in \mathrm{par}_{\mathsf{T}}(u)$，以下结论成立

1. $\left(\mathcal{J}_{v_i} \cup \{v_i\}\right) \cap \left(\mathcal{J}_{v_j} \cup \{v_j\}\right) = \emptyset$；*2.* 对于任意 $i \in \mathcal{J}_{v_i} \cup \{v_i\}$ 和 $j \in \mathcal{J}_{v_j} \cup \{v_j\}$，我们有 $\boldsymbol{Q}_{\bar{i},\bar{j}} = 0$。

命题 *2* 的证明。 反设存在 $\ell \in \left(\mathcal{J}_{v_i} \cup \{v_i\}\right) \cap \left(\mathcal{J}_{v_j} \cup \{v_j\}\right)$。根据树分解的运行交性质（定义 1 的第三个条件），包含 ℓ 的袋所构成的集合在 T 中诱导出一棵连通子树。由于袋 $\mathcal{B}_u$ 位于连接以 v_i 和 v_j 为根的子树的唯一路径上，因此可得 $\ell \in \mathcal{B}_u$。另一方面，$J_{v_i} \cup \{v_i\} \subset \mathcal{J}_u$ 和 $\mathcal{J}_{v_j} \cup \{v_j\} \subset \mathcal{J}_u$；此外，根据构造，$\mathcal{J}_u \cap \mathcal{B}_u = \emptyset$。因此，$(\mathcal{J}_{v_i} \cup v_i) \cap \mathcal{B}_u = \emptyset$ 和 $(\mathcal{J}_{v_j} \cup v_j) \cap \mathcal{B}_u = \emptyset$。这与 $\ell \in \mathcal{B}_u$ 矛盾。因此，$\ell \in \left(\mathcal{J}_{v_i} \cup \{v_i\}\right) \cap \left(\mathcal{J}_{v_j} \cup \{v_j\}\right) = \emptyset$，从而完成第一项陈述的证明。

接下来证明第二项陈述。固定 $i \in \mathcal{J}_{v_i} \cup \{v_i\}$ 和 $j \in \mathcal{J}_{v_j} \cup \{v_j\}$。反设 $\boldsymbol{Q}_{\bar{i},\bar{j}} \neq 0$。那么，根据定义 4 的第二个条件，必定存在一个袋，其中同时包含 i 和 j。下面证明这个袋不可能存在。

根据该命题的第一项陈述，集合 $\{\mathcal{J}_v \cup \{v\}\}_{v \in \mathrm{par}_{\mathsf{T}}(u)}$ 两两不相交。因此，由于 $i \in \mathcal{J}_{v_i} \cup \{v_i\}$，我们有 $i \notin \mathcal{J}_v \cup \{v\}, \forall v \in \mathrm{par}_{\mathsf{T}}(u) \backslash \{v_i\}$，类似地，$j \notin \mathcal{J}_v \cup \{v\}, \forall v \in \mathrm{par}_{\mathsf{T}}(u) \backslash \{v_j\}$。因此，树分解中可能同时包含 i 和 j 的袋只有 $\mathcal{B}_u$ 以及袋 $\{\mathcal{B}_v\}_{v \in \mathrm{par}_{\mathsf{T}}(u)}$。然而，根据构造，$\mathcal{B}_u$ 既不包含 i 也不包含 j。因此，必定存在某个 $v \in \mathrm{par}_{\mathsf{T}}(u)$，使得 $\mathcal{B}_v$ 同时包含 $\bar{i}$ 和 $\bar{j}$。现在，根据树分解的运行交性质，包含 i 的袋所构成的集合形成 T 的一个连通子树，而包含 j的袋也具有同样性质。由于 $\mathcal{B}_v$ 同时包含两个节点，且 u 与 v相邻，可知 $\mathcal{B}_u$必定属于这两个连通子树中的至少一个。特别地，这意味着 $i \in \mathcal{B}_u$ 或 $j \in \mathcal{B}_u$。这与 $\mathcal{B}_u$ 既不包含 i 也不包含 j 这一事实矛盾，由此完成第二项陈述的证明。 □

考虑节点 f_u 处的局部参数化代价 u。与 (23) 类似，我们有

$$f_u(\boldsymbol{\alpha}_{\mathcal{B}_u}) = \frac{1}{2}\boldsymbol{\alpha}_{\mathcal{B}_u}^\top \boldsymbol{Q}_{\mathcal{B}_u,\mathcal{B}_u}\boldsymbol{\alpha}_{\mathcal{B}_u} + \boldsymbol{c}_{\mathcal{B}_u}^\top \boldsymbol{\alpha}_{\mathcal{B}_u} + \min_{\boldsymbol{x}\in\mathbb{R}^{n_u}, \boldsymbol{z}\in\{0,1\}^{n_u}} \left\{ \frac{1}{2}\boldsymbol{x}_{\pi_u(\mathcal{J}_u)}^\top \boldsymbol{Q}_{\mathcal{J}_u,\mathcal{J}_u}\boldsymbol{x}_{\pi_u(\mathcal{J}_u)} + \boldsymbol{x}_{\pi_u(\mathcal{J}_u)}^\top \boldsymbol{Q}_{\mathcal{J}_u,\mathcal{B}_u}\boldsymbol{\alpha}_{\mathcal{B}_u} + \boldsymbol{c}_{\mathcal{J}_u}^\top \boldsymbol{x}_{\pi_u(\mathcal{J}_u)} + \boldsymbol{\lambda}_{\mathcal{J}_u}^\top \boldsymbol{z}_{\pi_u(\mathcal{J}_u)} \right\}$$
$$\text{s.t.} \quad \boldsymbol{x}_i(1 - \boldsymbol{z}_i) = 0 \qquad \forall i \in \pi_u(\mathcal{J}_u).$$

由 (7a) 可知，目标函数的前两项与 $h_u(\boldsymbol{\alpha}_{\mathcal{B}_u})$ 一致。现在我们转向目标函数中的其余项。首先，注意到

$$\begin{aligned}\frac{1}{2}\boldsymbol{x}_{\pi_u(\mathcal{J}_u)}^\top \boldsymbol{Q}_{\mathcal{J}_u,\mathcal{J}_u}\boldsymbol{x}_{\pi_u(\mathcal{J}_u)} &= \sum_{v\in\mathrm{par}_{\mathsf{T}}(u)} \left(\frac{1}{2}\boldsymbol{x}_{\pi_u(\mathcal{J}_v\cup v)}^\top \boldsymbol{Q}_{\mathcal{J}_v\cup v,\mathcal{J}_v\cup v}\boldsymbol{x}_{\pi_u(\mathcal{J}_v\cup v)}\right) \\ &= \sum_{v\in\mathrm{par}_{\mathsf{T}}(u)} \left(\frac{1}{2}\boldsymbol{x}_{\pi_u(\mathcal{J}_v)}^\top \boldsymbol{Q}_{\mathcal{J}_v,\mathcal{J}_v}\boldsymbol{x}_{\pi_u(\mathcal{J}_v)} + \frac{1}{2}\boldsymbol{Q}_{v,v}\boldsymbol{x}_{\pi_u(v)}^2 + \boldsymbol{x}_{\pi_u(v)}\boldsymbol{Q}_{v,\mathcal{J}_v}x_{\pi_u(\mathcal{J}_v)}\right).\end{aligned}$$

该分解使用了断言2，其中指出并集 $\mathcal{J}_u = \bigcup_{v\in\mathrm{par}_{\mathsf{T}}(u)}(\mathcal{J}_v \cup \{v\})$ 是不相交的，并且交叉块满足 $\boldsymbol{Q}_{\mathcal{J}_{v_i}\cup\{v_i\},\mathcal{J}_{v_j}\cup\{v_j\}} = \boldsymbol{0}$，对于所有互异的 $v_i, v_j \in \mathrm{par}_{\mathsf{T}}(u)$。

接下来，我们考虑耦合变量$\boldsymbol{x}_{\pi_u(\mathcal{J}_u)}$与$\boldsymbol{\alpha}_{\mathcal{B}_u}$的双线性项。

$$\begin{aligned}
\boldsymbol{x}_{\pi_u(\mathcal{J}_u)}^\top \boldsymbol{Q}_{\mathcal{J}_u,\mathcal{B}_u}\boldsymbol{\alpha}_{\mathcal{B}_u} &= \sum_{v\in\mathrm{par}_\mathsf{T}(u)} \left(\boldsymbol{x}_{\pi_u(\mathcal{J}_v\cup v)}^\top \boldsymbol{Q}_{\mathcal{J}_v\cup v,\mathcal{B}_u}\boldsymbol{\alpha}_{\mathcal{B}_u}\right) \\
&= \sum_{v\in\mathrm{par}_\mathsf{T}(u)} \left(\boldsymbol{x}_{\pi_u(\mathcal{J}_v)}^\top \boldsymbol{Q}_{\mathcal{J}_v,\mathcal{B}_u}\boldsymbol{\alpha}_{\mathcal{B}_u} + \boldsymbol{x}_{\pi_u(v)}^\top \boldsymbol{Q}_{v,\mathcal{B}_u}\boldsymbol{\alpha}_{\mathcal{B}_u}\right) \\
&= \sum_{v\in\mathrm{par}_\mathsf{T}(u)} \left(\boldsymbol{x}_{\pi_u(\mathcal{J}_v)}^\top \boldsymbol{Q}_{\mathcal{J}_v,\mathcal{B}_v\setminus v}\boldsymbol{\alpha}_{\mathcal{B}_v\setminus v} + \boldsymbol{x}_{\pi_u(v)}^\top \boldsymbol{Q}_{v,\mathcal{B}_v\setminus v}\boldsymbol{\alpha}_{\mathcal{B}_v\setminus v}\right).
\end{aligned}$$

要理解第三个等式，首先注意到，对于任意 $v\in\mathrm{par}_\mathsf{T}(u)$，袋 $\mathcal{B}_u$ 都可以作不相交分解 $\mathcal{B}_u = (\mathcal{B}_v\setminus\{v\})\cup(\mathcal{B}_u\setminus\mathcal{B}_v)$，此外，$\boldsymbol{Q}_{\mathcal{J}_v,\ \mathcal{B}_u\setminus\mathcal{B}_v}=\boldsymbol{0}$，因为 $\mathcal{J}_v$ 中的节点与 $\mathcal{B}_u\setminus\mathcal{B}_v$ 中的节点之间没有边。

现在来看其余各项：

$$\boldsymbol{c}_{\mathcal{J}_u}^\top \boldsymbol{x}_{\pi_u(\mathcal{J}_u)} + \boldsymbol{\lambda}_{\mathcal{J}_u}^\top \boldsymbol{z}_{\pi_u(\mathcal{J}_u)} = \sum_{v\in\mathrm{par}_\mathsf{T}(u)} \left(\boldsymbol{c}_{\mathcal{J}_v}^\top \boldsymbol{x}_{\pi_u(\mathcal{J}_v)} + \boldsymbol{c}_v \boldsymbol{x}_{\pi_u(v)} + \boldsymbol{\lambda}_{\mathcal{J}_v}^\top \boldsymbol{z}_{\pi_u(\mathcal{J}_v)} + \boldsymbol{\lambda}_v \boldsymbol{z}_{\pi_v(v)}\right).$$

将上面推导出的各项代入 f_u 的表达式，我们得到

$$\begin{aligned}
f_u(\boldsymbol{\alpha}_{\mathcal{B}_u}) = h_u(\boldsymbol{\alpha}_{\mathcal{B}_u}) + \\
\min_{\boldsymbol{x}\in\mathbb{R}^{n_u},\boldsymbol{z}\in\{0,1\}^{n_u}} & \sum_{v\in\mathrm{par}_\mathsf{T}(u)} \Big(\frac{1}{2}\boldsymbol{x}_{\pi_u(\mathcal{J}_v)}^\top \boldsymbol{Q}_{\mathcal{J}_v,\mathcal{J}_v}\boldsymbol{x}_{\pi_u(\mathcal{J}_v)} + \frac{1}{2}\boldsymbol{Q}_{v,v}\boldsymbol{x}_{\pi_u(v)}^2 + \boldsymbol{x}_{\pi_u(v)}\boldsymbol{Q}_{v,\mathcal{J}_v}x_{\pi_u(\mathcal{J}_v)} \\
& + \boldsymbol{x}_{\pi_u(\mathcal{J}_v)}^\top \boldsymbol{Q}_{\mathcal{J}_v,\mathcal{B}_v\setminus v}\boldsymbol{\alpha}_{\mathcal{B}_v\setminus v} + \boldsymbol{x}_{\pi_u(v)}^\top \boldsymbol{Q}_{v,\mathcal{B}_v\setminus v}\boldsymbol{\alpha}_{\mathcal{B}_v\setminus v} \\
& + \boldsymbol{c}_{\mathcal{J}_v}^\top \boldsymbol{x}_{\pi_u(\mathcal{J}_v)} + \boldsymbol{c}_v \boldsymbol{x}_{\pi_u(v)} + \boldsymbol{\lambda}_{\mathcal{J}_v}^\top \boldsymbol{z}_{\pi_u(\mathcal{J}_v)} + \boldsymbol{\lambda}_v \boldsymbol{z}_{\pi_u(v)}\Big) \\
\text{s.t.} \quad & \boldsymbol{x}_i(1-\boldsymbol{z}_i) = 0 \qquad \forall i\in\pi_u(\mathcal{J}_u).
\end{aligned}$$

由于求和中的每一项仅取决于 $\mathcal{J}_v\cup\{v\}$ 中的变量，因此该优化问题可分解为针对每个 $v\in\mathrm{par}_\mathsf{T}(u)$ 的独立子问题。此外，每个子问题都等于 $g_v(\boldsymbol{\alpha}_{\mathcal{B}_v\setminus v}) - \phi_v(\boldsymbol{\alpha}_{\mathcal{B}_v\setminus v})$，由（26）可知。因此，我们得出

$$f_u(\boldsymbol{\alpha}_{\mathcal{B}_u}) = h_u(\boldsymbol{\alpha}_{\mathcal{B}_u}) + \sum_{v\in\mathrm{par}_\mathsf{T}(u)} g_v(\mathcal{B}_v\setminus v) - \phi_v(\boldsymbol{\alpha}_{\mathcal{B}_v\setminus v}).$$

□

A.3 引理 3 的证明

令 $\mathcal{J}=\{i:\ \boldsymbol{z}_i^\star=1\}$。那么，$\boldsymbol{x}^\star = (\boldsymbol{Q}_{\mathcal{J},\ \mathcal{J}})^{-1}\boldsymbol{c}_\mathcal{J}$，这意味着

$$\|\boldsymbol{x}^\star\|_\infty = \left\|(\boldsymbol{Q}_{\mathcal{J},\mathcal{J}})^{-1}\boldsymbol{c}_\mathcal{J}\right\|_\infty \le \left\|(\boldsymbol{Q}_{\mathcal{J},\mathcal{J}})^{-1}\right\|_\infty \|\boldsymbol{c}_\mathcal{J}\|_\infty \le \left\|(\boldsymbol{Q}_{\mathcal{J},\mathcal{J}})^{-1}\right\|_\infty \|\boldsymbol{c}\|_\infty.$$

现在我们推导 $\|(\boldsymbol{Q}_{\mathcal{J},\ \mathcal{J}})^{-1}\|_\infty$ 的上界。根据引理 5，

$$\|(\boldsymbol{Q}_{\mathcal{J},\mathcal{J}})^{-1}\|_\infty \le \max_{i\in\mathcal{J}} \sum_{j\in\mathcal{J}} \left|[\boldsymbol{Q}_{\mathcal{I},\mathcal{I}}^{-1}]_{\pi(i),\pi(j)}\right| \le \max_{i\in\mathcal{J}} \sum_{j\in\mathcal{J}} C_1\rho^{\mathrm{dist}(i,j)}.$$

令 $N_{r,\ i}$ 表示 supp($\boldsymbol{Q}$) 中与节点 i距离恰好为 r 的节点数量，并设N_r :=max$_i\{N_{r,i}\}$。于是

$$\max_{i\in\mathcal{J}} \sum_{j\in\mathcal{J}} C_1\rho^{\mathrm{dist}(i,j)} \ \leq\ C_1 \sum_{r=0}^{n} N_r\rho^r.$$

我们区分两种情形：

- 若 $\boldsymbol{Q}$ 是带宽为 w的带状，则对所有 $0 \leq r \leq n$有 $N_r \leq 2w$ 。因此，

$$\|(\boldsymbol{Q}_{\mathcal{J},\mathcal{J}})^{-1}\|_\infty \leq C_1 \sum_{r=0}^{n} 2w\,\rho^r \leq 2wC_1 \sum_{r=0}^{\infty} \rho^r = \frac{2wC_1}{1-\rho}.$$

- 若supp($\boldsymbol{Q}$)满足多项式体积增长条件（假设1），则 $N_r \leq \delta r^\gamma$。因此，

$$\|(\boldsymbol{Q}_{\mathcal{J},\mathcal{J}})^{-1}\|_\infty \leq C_1 \sum_{r=0}^{n} \delta r^\gamma\rho^r \leq \delta C_1 \sum_{r=0}^{\infty} r^\gamma\rho^r \leq \frac{\delta\gamma!\,C_1}{(1-\rho)^{\gamma+1}},$$

其中最后一个不等式使用了经典界$\sum_{r=0}^{\infty} r^\gamma\rho^r \leq \quad ^\gamma_{!\,(1-\rho)^{\gamma+1}}$；例如参见 [23, 命题1.4.4]。

□

A.4 引理4的证明

Consi 考虑二次方程 $p_1(\boldsymbol{\alpha}) - p_2(\boldsymbol{\alpha}) = 0$。令 $\hat{}\boldsymbol{\alpha}$ 表示其解。那么

$$\begin{aligned}
& p_1(\hat{\boldsymbol{\alpha}}) - p_2(\hat{\boldsymbol{\alpha}}) = 0 \\
\Longrightarrow\ & \frac{1}{2}\hat{\boldsymbol{\alpha}}^\top(\boldsymbol{A}_1 - \boldsymbol{A}_2)\hat{\boldsymbol{\alpha}} + (\boldsymbol{b}_1 - \boldsymbol{b}_2)^\top\hat{\boldsymbol{\alpha}} + (d_1 - d_2) = 0 \\
\Longrightarrow\ & \frac{1}{2}\|\boldsymbol{A}_1 - \boldsymbol{A}_2\|_{1,1}\|\hat{\boldsymbol{\alpha}}\|_\infty^2 + \|\boldsymbol{b}_1 - \boldsymbol{b}_2\|_1\|\hat{\boldsymbol{\alpha}}\|_\infty + (d_1 - d_2) \geq 0 \\
\Longrightarrow\ & \bar{a}\|\hat{\boldsymbol{\alpha}}\|_\infty^2 + \bar{b}\|\hat{\boldsymbol{\alpha}}\|_\infty \geq \bar{d},
\end{aligned}$$

其中，不失一般性，我们假设 $d_2 \geq d_1$。我们考虑两种情况。事实上，若 $\bar{a} = 0$，则

$$\|\boldsymbol{\alpha}\|_\infty \geq \frac{\bar{d}}{\bar{b}}.$$

另一方面，如果 $\bar{a} > 0$，则有

$$\begin{aligned}
& \|\hat{\boldsymbol{\alpha}}\|_\infty^2 + \frac{\bar{b}}{\bar{a}}\|\hat{\boldsymbol{\alpha}}\|_\infty \geq \frac{\bar{d}}{\bar{a}} \\
\Longrightarrow\ & \|\hat{\boldsymbol{\alpha}}\|_\infty^2 + \frac{\bar{b}}{\bar{a}}\|\hat{\boldsymbol{\alpha}}\|_\infty + \frac{\bar{b}^2}{4\bar{a}^2} \geq \frac{\bar{d}}{\bar{a}} + \frac{\bar{b}^2}{4\bar{a}^2} \\
\Longrightarrow\ & \|\hat{\boldsymbol{\alpha}}\|_\infty \geq -\frac{\bar{b}}{2\bar{a}} + \sqrt{\frac{\bar{d}}{\bar{a}} + \frac{\bar{b}^2}{4\bar{a}^2}}.
\end{aligned}$$

重新排列各项后，得到

$$\|\hat{\boldsymbol{\alpha}}\|_\infty \geq \frac{-\bar{b}+\sqrt{\bar{b}^2+4\bar{a}\bar{d}}}{2\bar{a}}.$$

综合这两种情况，我们得出

$$\|\hat{\boldsymbol{\alpha}}\|_\infty \geq \begin{cases} \frac{-\bar{b}+\sqrt{\bar{b}^2+4\bar{a}\bar{d}}}{2\bar{a}} & \text{if } \bar{a}\neq 0, \\ \frac{\bar{d}}{\bar{b}} & \text{if } \bar{a}=0. \end{cases}$$

□

A.5 引理 5的证明

证明通过在紧区间上对函数 $\psi(x) = x^{-1}$ 进行多项式逼近来完成，为此我们使用以下结果。

命题 2 （的多项式逼近 x^{-1}，[57]）。设 $0 < a < b$。对于 $\ell \geq 0$，令 Π_ℓ 表示次数至多为 ℓ 的多项式集合，并定义

$$e_\ell([a,b]) := \inf_{\bar{p}\in\Pi_\ell} \max_{x\in[a,b]} \left|x^{-1} - \bar{p}(x)\right|.$$

令 $r = b/a$，并且

$$\rho := \frac{\sqrt{r}-1}{\sqrt{r}+1}.$$

Then

$$e_\ell([a,b]) = \frac{(1+\sqrt{r})^2}{2ar}\,\rho^{\ell+1}.$$

令 $a = \mu_{\min}(\boldsymbol{Q})$ $b = \mu_{\max}(\boldsymbol{Q})$，$C_0 \equiv \frac{(1+\sqrt{\kappa_2})^2}{2\kappa_2\mu_{\min}(\boldsymbol{Q})}$ 以及 $\sigma(\boldsymbol{Q})$ 表示 $\boldsymbol{Q}$ 的特征值集合。由于 $\boldsymbol{Q}$ 是正定且可逆的，因此有 $0 < a < b$。根据命题 2，存在多项式序列 $\bar{p}_\ell \in \Pi_\ell$ 满足

$$\max_{x\in[a,b]} \left\{\left|x^{-1} - \bar{p}_\ell(x)\right|\right\} = C_0\rho^{\ell+1}.$$

来自谱理论 [62]

$$\begin{aligned} \|\boldsymbol{Q}^{-1} - \bar{p}_\ell(\boldsymbol{Q})\|_2 &= \max_{x\in\sigma(A)} \left\{\left|x^{-1} - \bar{p}_\ell(x)\right|\right\} \\ &\leq \max_{x\in[a,b]} \left\{\left|x^{-1} - \bar{p}_\ell(x)\right|\right\} \\ &= C_0\rho^{\ell+1}. \end{aligned}$$

此外，对于每个整数 $k \geq 0$，我们有

$$(\boldsymbol{Q}^k)_{ij} = 0 \qquad \text{whenever } \operatorname{dist}(i,j) > k.$$

因此，如果 $\bar{p}_k \in \Pi_k$，那么 $\bar{p}_k(\boldsymbol{Q})$ 满足

$$\bar{p}_k(\boldsymbol{Q})_{ij} = 0 \qquad \text{whenever } \operatorname{dist}(i,j) > k,$$

因为 $\bar{p}_k(\boldsymbol{Q})$ 是 $\mathbf{I}$ $\boldsymbol{Q}$、……、 $\boldsymbol{Q}^k$ 的线性组合。

设 $i \neq j$ 并选取 $\ell = \mathrm{dist}(i,j) - 1$。则 $\bar{p}_\ell(\boldsymbol{Q})_{ij} = 0$，因此

$$\left|\boldsymbol{Q}^{-1}_{ij}\right| = \left|\boldsymbol{Q}^{-1}_{ij} - \bar{p}_\ell(\boldsymbol{Q})_{ij}\right| \leq \left\|\boldsymbol{Q}^{-1} - \bar{p}_\ell(\boldsymbol{Q})\right\|_2 \leq C_0 \rho^{\,\mathrm{dist}(i,j)}.$$

若 $i = j$，注意

$$\left|\boldsymbol{Q}^{-1}_{ii}\right| \leq \left\|\boldsymbol{Q}^{-1}\right\|_2 = \frac{1}{a},$$

这与所述界一致。至此证明完毕。 □

B PRUNE 的启发式方法

为确定相关函数集，PRUNE 的精确版本需要比较每一对二次函数，从而导致运行时间关于 N 是二次的，其中 N 是输入中函数的总数。

为缓解这一二次代价，我们给出一种启发式方法版本，见算法 4。该过程以单次遍历处理这些函数，仅需 $N-1$ 次比较。这种线性时间方法可能会保留一些形式上不相关的函数，但所保留函数的总数仍与引理 9 中所建立的界一致。

算法4启发式方法-PRUNE

输入： $\tilde{f}(\boldsymbol{\alpha}) \equiv [p_i(\boldsymbol{\alpha})]\ \forall i = 1, \ldots, N$以及常数 U。
输出： f^{prune}

$f^{\mathrm{prune}} \leftarrow [p_1]$ ▷ 将 p_1 添加到 f^{prune}
$j = 1$
</style}针对 $i = 2$ $, \ldots, N$ **do**
    计算 $L(p_i,\ p_j)$ 来自方程（8） ▷ $p_i - p_j = 0$ 的根的下界
    如果 $L(p_i,\ p_j) > U$ **则**
        如果 $d_i < d_j$ **那么**
            $f^{\mathrm{prune}} \leftarrow$删除 ($f_{\text{剪枝}}$，结束 ($f_{\text{剪枝}}$)) ▷ 从 f^{prune}中移除最后一个函数 p_j
            $f^{\mathrm{prune}} \leftarrow$追加 ($f_{\text{剪枝}}\ p_i$) ▷ 将 p_i 添加到 f^{prune}
            $j = i$
        结束如果
    else
        $f_{\text{剪枝}} \leftarrow$追加(f^{prune}, p_i)
        $j = i$
    结束如果
结束循环
返回 f^{prune}

当二次函数的存储顺序满足每一对相邻函数都具有 m-相似性时，这种改进便成为可能。在这种情况下，只需检查相邻函数对之间是否存在无关项即可。对于 $\boldsymbol{Q}$ 的树分解为路径的情形（包括带状情形），我们注意到，无需额外代价即可维持这种排序。然而，在一般情形下，按所需顺序存储函数需要额外步骤，这可能会增加计算

代价，使启发式方法效率低下。因此，针对一般情形，我们采用PRUNE算法3 中所述的版本。