



华南理工大学

South China University of Technology

本科毕业设计（论文）

基于稀疏性优化的卡尔曼滤波状态转移模型参数估计

学 院	数学学院
专 业	数学与应用数学（统计学）
学生姓名	邓肯
学生学号	202130320311
指导教师	曾才斌
提交日期	2025年5月26日

华南理工大学

学位论文原创性声明

本人郑重声明：所呈交的论文是本人在导师的指导下独立进行研究所取得的研究成果。除了文中特别加以标注引用的内容外，本论文不包含任何其他个人或集体已经发表或撰写的成果作品。对本文的研究做出重要贡献的个人和集体，均已在文中以明确方式标明。本人完全意识到本声明的法律后果由本人承担。

作者签名：邓肯

日期：2025 年 5 月 26 日

学位论文版权使用授权书

本学位论文作者完全了解学校有关保留、使用学位论文的规定，即：学校有权保存并向国家有关部门或机构送交论文的复印件和电子版，允许学位论文被查阅；学校可以公布学位论文的全部或部分内容，可以允许采用影印、缩印或其它复制手段保存、汇编学位论文。本人电子文档的内容和纸质论文的内容相一致。

作者签名：邓肯

日期：2025 年 5 月 26 日

指导教师签名：

日期：2025 年 5 月 26 日

作者联系电话：13632181278

电子邮箱：202130320311@mail.scut.edu

摘要

本文探讨了线性高斯状态空间模型 (LGSSM) 中的参数估计问题，重点研究了不同图结构下的转移矩阵估计。我们提出了 **GraphEM** 算法，这是一个将图推理和正则化技术集成到期望最大化 (EM) 算法中的新颖框架。通过引入拉普拉斯、高斯和混合先验等正则化项的先验知识，**GraphEM** 扩展了传统的 EM 方法，使其能够处理更广泛的图属性，包括稀疏性、平滑度和块结构。

我们的贡献主要有三方面：(1) 我们对转移矩阵背景下的图推理进行了全面的分析，重点突出了不同图结构（例如小世界网络、无标度图）对参数估计的影响。(2) 我们引入了 **GraphEM** 的两个新变体，分别结合了高斯先验和混合拉普拉斯-高斯先验，从而增强了算法的灵活性和性能。(3) 我们证明了组合不同的正则化项可以在不增加计算成本的情况下逼近最优解，这为正则化策略之间的权衡提供了新的见解。

在合成数据集和真实数据集上进行的大量实验验证了 **GraphEM** 的有效性。结果表明，该方法在跨多种图类型的估计精度和鲁棒性方面均优于传统方法。本研究提供的理论框架和实践见解为未来状态空间模型的图推理和正则化技术研究铺平了道路。

为了提高可重复性并促进进一步研究，我们在 <https://github.com/kdeng-gzcn/KalmanParamEstimation> 上提供了所提方法的开源实现以及实验代码和数据集。

关键词： 卡尔曼滤波，EM 算法，近端优化，图模型

Abstract

This paper addresses the problem of parameter estimation in Linear Gaussian State Space Models (LGSSMs) with a focus on estimating the transition matrix under various graph structures. We propose the GraphEM algorithm, a novel framework that integrates graphical inference and regularization techniques into the Expectation-Maximization (EM) algorithm. By incorporating prior knowledge through regularization terms such as Laplace, Gaussian, and mixed priors, GraphEM extends the traditional EM approach to handle a wider range of graph properties, including sparsity, smoothness, and block structures.

Our contributions are threefold: (1) We provide a comprehensive analysis of graphical inference in the context of transition matrices, highlighting the impact of different graph structures (e.g., small-world, scale-free) on parameter estimation. (2) We introduce two new variants of GraphEM, incorporating Gaussian and mixed Laplace-Gaussian priors, which enhance the algorithm’s flexibility and performance. (3) We demonstrate that combining different regularization terms can approximate the optimal solution without additional computational cost, offering new insights into the trade-offs between regularization strategies.

Extensive experiments on synthetic and real-world datasets validate the effectiveness of GraphEM. Our results show that the proposed method outperforms traditional approaches in terms of estimation accuracy and robustness across diverse graph types. The theoretical framework and practical insights provided in this work pave the way for future research in graphical inference and regularization techniques for state space models.

To promote reproducibility and facilitate further research, we provide an open-source implementation of the proposed methods, along with experimental code and datasets, at <https://github.com/kdeng-gzcn/KalmanParamEstimation>.

Keywords: Kalman Filter, EM Algorithm, Graphical Inference, Proximal Optimization

目 录

摘 要	I
Abstract	II
目 录	III
插图目录	V
表格目录	VI
第一章 绪论	1
1.1 问题背景	1
1.1.1 现有方法	1
1.1.2 创新方法	1
1.2 相关文献	2
第二章 状态空间模型	4
2.1 状态空间模型	4
2.1.1 用于最优过滤的贝叶斯推断	4
2.1.2 关键边际分布	4
2.2 线性高斯状态空间模型	5
2.2.1 卡尔曼滤波器	6
2.2.2 卡尔曼平滑器	7
2.3 状态空间模型中的参数估计	7
2.3.1 具有缺失参数的线性高斯状态空间模型	9
2.3.2 梯度下降算法	9
2.3.3 EM 算法	10
第三章 GraphEM 算法	14
3.1 引言	14
3.2 E 步骤中的计算	15
3.3 M 步骤中的优化	15
3.3.1 近端算子	16

3.3.2	Douglas-Rachford 迭代法	16
3.4	先验正则化	17
3.4.1	稳定性约束	17
3.4.2	正则项	18
第四章	数值实验	20
4.1	汽车的动态模型	20
4.2	不同图结构下的正则化分析	24
结论		29
参考文献		30
致谢		33
附录 A	其他图类型	34

插图目录

2-1	线性高斯状态空间模型的流程。蓝色块表示转换过程，红色块说明马尔可夫链属性的细节，绿色块表示观察过程。	6
4-1	观测轨迹与使用卡尔曼滤波器和平滑器估计得到的轨迹。	21
4-2	EM 算法在估计 \mathbf{A} 、 \mathbf{Q} 和 \mathbf{H} 时的性能可视化结果。	22
4-3	真实参数与估计参数 \mathbf{A} 、 \mathbf{Q} 、 \mathbf{H} 的图形对比。	23
4-4	不同正则化方法下，分块对角图的负对数似然收敛曲线。	27
4-5	分块对角图的真实转移矩阵与估计矩阵对比图。	28
A-1	小世界图的真实转移矩阵与估计矩阵对比图。	35
A-2	无标度图的真实转移矩阵与估计矩阵对比图。	36
A-3	二部图的真实转移矩阵与估计矩阵对比图。	37
A-4	环图的真实转移矩阵与估计矩阵对比图。	38
A-5	星型图的真实转移矩阵与估计矩阵对比图。	39

表格目录

3-1	常见凸约束集合及其投影算子示例。	18
3-2	先验类型、正则项及其在尺度参数 $\vartheta > 0$ 下的近端算子示例。	18
4-1	汽车模型的 EM 算法结果。基线负对数似然为 $\mathcal{L}_T(\boldsymbol{\theta}) = -261.989$ 。	21
4-2	不同图结构在各类正则方法下的实验结果。每种图类型中加粗的数值表示最佳结果。	26

第一章 绪论

1.1 问题背景

状态空间模型 (SSM) 广泛应用于工程 [1-2]、环境科学 [3]、信号处理 [4] 等各个领域的随机过程建模。SSM 利用隐藏变量来表示底层的马尔可夫过程，使其成为分析时间序列数据的强大工具。

当模型参数已知时，贝叶斯最优滤波器和平滑器为求解状态空间模型 [5] 提供了理论框架。这些工具能够进行准确的预测，为工业应用中的决策和战略制定提供参考。具体而言，贝叶斯滤波器能够以高可靠性预测状态空间模型中的长期行为。在线性高斯状态空间模型 (LGSSM) 中，该过程允许一个解析解，详见 [5]。对于非线性状态空间模型，粒子滤波器提供了一种实用的解决方案。

本文重点介绍线性高斯状态空间模型 (LGSSM)。在 LGSSM 中，诸如转移矩阵之类的参数对于理解卡尔曼滤波器的行为至关重要。转移矩阵封装了不同的马尔可夫链，其准确估计至关重要。然而，在实际场景中，模型参数通常未知，但具有有用的属性 [6]，因此参数估计是一项至关重要的任务。

1.1.1 现有方法

LGSSM 中的参数估计主要有三种方法。第一种方法是直接优化 [7]，它使用基于梯度的方法迭代逼近最优模型参数。第二种方法是期望最大化 (EM) 方法 [8]，它优化目标函数的上限以控制其行为 [5, 9]。第三种方法是马尔可夫链蒙特卡洛 (MCMC)，它通过采样来近似模型参数的分布。然而，这些方法都不是直接针对模型参数的最大后验 (MAP) 估计。结合先验分布对于理解系统隐藏的动态通常至关重要。

一种名为 GraphEM 的新算法通过将先验知识集成为正则化来解决这一限制 [10]。GraphEM 以 EM 方法为基础，利用其快速收敛和理论保证，同时结合了对模型参数（尤其是转移矩阵）的图推理。通过结合不同的马尔可夫过程，GraphEM 在估计现实世界中的转移矩阵（通常具有稀疏性等特性）方面表现出卓越的性能。

1.1.2 创新方法

GraphEM 引入了转移矩阵的先验知识，通过稀疏性正则化增强了算法。受转移矩阵图推理视角的启发，我们探索了除稀疏性之外的其他图属性 [11-12]。具体而言，我们通过引入新的先验知识类型来扩展 GraphEM，包括高斯先验以及结合拉普拉斯分布和

高斯分布的混合先验 [13]。通过对各种图类型（例如小世界图、无标度图、循环图和星形图）进行大量实验，我们证明了所提变体的多功能性和鲁棒性。

本文的贡献如下：

- **深入探索基于图结构的统计推断：**我们对过渡矩阵背景下的图推理进行了全面的分析，重点突出了不同图结构（例如小世界图、无标度图）对参数估计的影响。该分析为理解 GraphEM 在不同拓扑条件下的行为提供了新的见解。
- **提出 GraphEM 的新变体：**我们提出并实现了两种新的 GraphEM 变体：一种融合了高斯先验，另一种结合了拉普拉斯先验和高斯先验。这些变体扩展了原始方法的能力，使其能够处理除稀疏性之外的更广泛的图属性。
- **提出新的正则化策略：**通过比较拉普拉斯、高斯和混合先验的性能，我们确定了每种正则化策略的优势场景。此外，我们的分析表明，通过在我们的优化框架内组合不同类型的正则化项，有可能逼近最优解，而不会产生额外的计算成本。

1.2 相关文献

卡尔曼滤波器及其扩展的理论基础已非常完善，[5] 提供了全面的推导和分析。

卡尔曼滤波器的参数估计已得到广泛研究，其中期望最大化 (EM) 算法是最广泛使用的方法之一。EM 算法（如 [5] 所述）通过迭代优化似然函数，为估计模型参数提供了一个稳健的框架。然而，传统的 EM 方法通常缺乏融入先验知识的能力，而这对于提高实际应用中的估计精度至关重要。

参数估计领域的最新进展主要集中在通过集成图推理技术来增强 EM 算法。[10] 的开创性工作提出了 GraphEM，这是一种将转移矩阵的先验知识作为正则化项的新方法。该方法利用转移矩阵的稀疏结构，从而提升了各种应用中的估计性能。GraphEM 的有效性已在多个领域得到证实，包括股票市场预测 [14]、社交媒体趋势分析 [15] 和气候建模 [16]。

GraphEM 的进一步扩展已在时间序列预测和非凸优化领域得到探索。例如，[17] 将 GraphEM 应用于旅行商问题 (TSP) 场景，而 [18] 则研究了其在非凸设置下的性能。此外，[19] 提出了一种基于拉格朗日的 GraphEM 变体，用于非马尔可夫系统，进一步拓宽了其适用性。

近期的研究也致力于提升 GraphEM 的计算效率和可扩展性。例如，[20] 引入了 SPARJ 算法，提高了大规模系统中图推理的效率。该算法在 [21] 和 [22] 中得到了进一步扩展，其中集成了基于梯度的方法来处理高维转移矩阵。此外，[23] 提出了一种联合估计转移矩阵和噪声协方差的新方法，显著提高了估计精度。

与此同时，其他研究人员也在探索状态空间模型中参数估计的替代方法。例如，[24] 提出了一种用于变点检测的自适应调优方法，而 [25] 将 GraphEM 与循环神经网络 (RNN) 相结合，用于股票市场预测。此外，[26] 使用哈密顿蒙特卡罗方法研究了非线性系统的推理方法，[27] 则在图模型的背景下探索了格兰杰因果关系。

第二章 状态空间模型

2.1 状态空间模型

用数学术语来说，状态空间模型 (SSM) 中的最优滤波和平滑可以表述为统计反演问题。我们的目标是从一组噪声观测值 $\{\mathbf{y}_1, \mathbf{y}_2, \dots\}$ 中估计一个未知的向量值时间序列 $\{\mathbf{x}_0, \mathbf{x}_1, \mathbf{x}_2, \dots\}$ 。为了方便记述，我们定义 $\mathbf{x}_{0:T} = \{\mathbf{x}_0, \mathbf{x}_1, \dots, \mathbf{x}_T\}$ 和 $\mathbf{y}_{1:T} = \{\mathbf{y}_1, \mathbf{y}_2, \dots, \mathbf{y}_T\}$ 。

贝叶斯滤波解决了一般概率状态空间模型中的状态估计问题，这些模型的特征如下：

$$\mathbf{x}_k \sim p(\mathbf{x}_k | \mathbf{x}_{k-1}), \quad (2-1)$$

$$\mathbf{y}_k \sim p(\mathbf{y}_k | \mathbf{x}_k), \quad (2-2)$$

其中 $k = 1, 2, \dots$ ，其中：

- $\mathbf{x}_k \in \mathbb{R}^n$ 表示系统在时间步 k 的状态，
- $\mathbf{y}_k \in \mathbb{R}^m$ 表示时间步 k 的测量值 k ，
- $p(\mathbf{x}_k | \mathbf{x}_{k-1})$ 是动态模型，描述系统的随机演化。
- $p(\mathbf{y}_k | \mathbf{x}_k)$ 是测量模型，表示给定状态下观测值的似然值。

2.1.1 用于最优过滤的贝叶斯推断

在 SSM 的背景下，主要关注的是估计给定相应观测值的所有状态的联合后验分布。该分布由以下公式给出：

$$p(\mathbf{x}_{0:T} | \mathbf{y}_{1:T}) = \frac{p(\mathbf{y}_{1:T} | \mathbf{x}_{0:T})p(\mathbf{x}_{0:T})}{p(\mathbf{y}_{1:T})}, \quad (2-3)$$

其中：

- $p(\mathbf{x}_{0:T})$ 是状态的先验分布，
- $p(\mathbf{y}_{1:T} | \mathbf{x}_{0:T})$ 是给定状态下观测值的似然值，
- $p(\mathbf{y}_{1:T})$ 是归一化常数，确保后验分布积分为 1。

2.1.2 关键边际分布

在实践中，我们通常对由联合后验推导的特定边际分布感兴趣。这些分布包括：

- 滤波分布，它根据所有过去和当前的观测值估计当前状态：

$$\{p(\mathbf{x}_k | \mathbf{y}_{1:k}) | k = 1, \dots, T\} \quad (2-4)$$

- 预测分布，根据当前观测值预测未来状态：

$$\{p(\mathbf{x}_{k+n} | \mathbf{y}_{1:k}) | k = 1, \dots, T, n = 1, 2, \dots\} \quad (2-5)$$

- 平滑分布，利用所有可用观测值改进过去的状态估计：

$$\{p(\mathbf{x}_k | \mathbf{y}_{1:T}) | k = 1, \dots, T\} \quad (2-6)$$

这些分布构成了贝叶斯滤波和平滑的基础，从而能够在 SSM 中实现准确的状态估计和预测。

2.2 线性高斯状态空间模型

卡尔曼滤波器是贝叶斯滤波器的一个特例，其中动态模型和测量模型均为线性高斯模型。该模型由以下方程定义：

$$\mathbf{x}_k = \mathbf{A}_{k-1}\mathbf{x}_{k-1} + \mathbf{q}_{k-1}, \quad (2-7)$$

$$\mathbf{y}_k = \mathbf{H}_k\mathbf{x}_k + \mathbf{r}_k, \quad (2-8)$$

其中：

- $\mathbf{q}_{k-1} \sim \mathcal{N}(\mathbf{0}, \mathbf{Q}_{k-1})$ 是过程噪声，假设为具有零均值和协方差的高斯分布 \mathbf{Q}_{k-1} ，
- $\mathbf{r}_k \sim \mathcal{N}(\mathbf{0}, \mathbf{R}_k)$ 为测量噪声，假设为均值为零、协方差为 \mathbf{R}_k 的高斯分布。
- $\mathbf{x}_0 \sim \mathcal{N}(\mathbf{m}_0, \mathbf{P}_0)$ 为初始状态，假设为均值为 \mathbf{m}_0 、协方差为 \mathbf{P}_0 的高斯分布。
- \mathbf{A}_{k-1} 为状态转移矩阵。
- \mathbf{H}_k 是观测矩阵。

或者，该模型可以用概率术语表示为：

$$p(\mathbf{x}_k | \mathbf{x}_{k-1}) \sim \mathcal{N}(\mathbf{x}_k | \mathbf{A}_{k-1}\mathbf{x}_{k-1}, \mathbf{Q}_{k-1}), \quad (2-9)$$

$$p(\mathbf{y}_k | \mathbf{x}_k) \sim \mathcal{N}(\mathbf{y}_k | \mathbf{H}_k\mathbf{x}_k, \mathbf{R}_k) \quad (2-10)$$

图 2-1 展示了线性高斯状态空间模型 (LGSSM) 的流程。蓝色块表示状态转换过程，

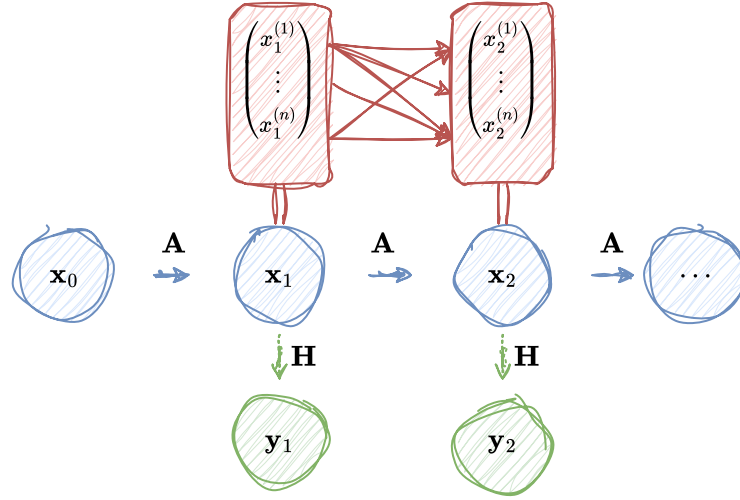


图 2-1 线性高斯状态空间模型的流程。蓝色块表示转换过程，红色块说明马尔可夫链属性的细节，绿色块表示观察过程。

描述系统随时间的变化。红色块突出了模型的马尔可夫属性，强调当前状态仅依赖于先前状态。绿色块表示观察过程，将状态映射到观察到的测量值。

2.2.1 卡尔曼滤波器

卡尔曼滤波器为线性高斯状态空间模型 (LGSSM) 中的状态估计问题提供了递归解决方案。卡尔曼滤波器涉及的关键分布是：

$$p(\mathbf{x}_k | \mathbf{y}_{1:k-1}) \sim \mathcal{N}(\mathbf{x}_k | \mathbf{m}_k^-, \mathbf{P}_k^-), \quad (2-11)$$

$$p(\mathbf{x}_k | \mathbf{y}_{1:k}) \sim \mathcal{N}(\mathbf{x}_k | \mathbf{m}_k, \mathbf{P}_k), \quad (2-12)$$

$$p(\mathbf{y}_k | \mathbf{y}_{1:k-1}) \sim \mathcal{N}(\mathbf{y}_k | \mathbf{H}_k \mathbf{m}_k^-, \mathbf{S}_k) \quad (2-13)$$

预测步骤

预测步骤使用动态模型将状态估计及其不确定性从上一个时间步传播到当前时间步。预测步骤的公式如下：

$$\mathbf{m}_k^- = \mathbf{A}_{k-1} \mathbf{m}_{k-1}, \quad (2-14)$$

$$\mathbf{P}_k^- = \mathbf{A}_{k-1} \mathbf{P}_{k-1} \mathbf{A}_{k-1}^\top + \mathbf{Q}_{k-1}. \quad (2-15)$$

其中， \mathbf{m}_k^- 和 \mathbf{P}_k^- 分别表示预测状态均值和协方差。

更新步骤

更新步骤结合新的观测值 \mathbf{y}_k 来优化预测状态估计。更新步骤的方程为：

$$\mathbf{v}_k = \mathbf{y}_k - \mathbf{H}_k \mathbf{m}_k^-, \quad (2-16)$$

$$\mathbf{S}_k = \mathbf{H}_k \mathbf{P}_k^- \mathbf{H}_k^\top + \mathbf{R}_k, \quad (2-17)$$

$$\mathbf{K}_k = \mathbf{P}_k^- \mathbf{H}_k^\top \mathbf{S}_k^{-1}, \quad (2-18)$$

$$\mathbf{m}_k = \mathbf{m}_k^- + \mathbf{K}_k \mathbf{v}_k, \quad (2-19)$$

$$\mathbf{P}_k = \mathbf{P}_k^- - \mathbf{K}_k \mathbf{S}_k \mathbf{K}_k^\top. \quad (2-20)$$

其中， \mathbf{v}_k 为残差， \mathbf{S}_k 为残差协方差， \mathbf{K}_k 为卡尔曼增益， \mathbf{m}_k 和 \mathbf{P}_k 分别为更新后的状态均值和协方差。

2.2.2 卡尔曼平滑器

卡尔曼平滑器扩展了卡尔曼滤波器，它使用所有可用的观测值 $\mathbf{y}_{1:T}$ 来估计每个时间步 k 的状态。具体来说，它计算平滑分布：

$$p(\mathbf{x}_k | \mathbf{y}_{1:T}) \sim \mathcal{N}(\mathbf{x}_k | \mathbf{m}_k^s, \mathbf{P}_k^s), \quad (2-21)$$

其中 \mathbf{m}_k^s 和 \mathbf{P}_k^s 分别是平滑后的状态均值和协方差。

向后递归方程

卡尔曼平滑器通过向后递归进行运算，从最终时间步长 T 开始，向后递归到初始时间步长 0。

2.3 状态空间模型中的参数估计

在贝叶斯框架中，缺失参数 $\boldsymbol{\theta} \in \mathbb{R}^d$ 被视为具有先验分布 $p(\boldsymbol{\theta})$ 的随机变量。包含这些缺失参数的状态空间模型 (SSM) 定义为：

$$\boldsymbol{\theta} \sim p(\boldsymbol{\theta}), \quad (2-22)$$

$$\mathbf{x}_0 \sim p(\mathbf{x}_0 | \boldsymbol{\theta}), \quad (2-23)$$

$$\mathbf{x}_k \sim p(\mathbf{x}_k | \mathbf{x}_{k-1}, \boldsymbol{\theta}), \quad (2-24)$$

$$\mathbf{y}_k \sim p(\mathbf{y}_k | \mathbf{x}_k, \boldsymbol{\theta}) \quad (2-25)$$

参数估计的基本方法基于贝叶斯规则，该规则将给定观测值的状态和参数的联合后验分布表示为：

$$p(\mathbf{x}_{0:T}, \boldsymbol{\theta} \mid \mathbf{y}_{1:T}) = \frac{p(\mathbf{y}_{1:T} \mid \mathbf{x}_{0:T}, \boldsymbol{\theta}) p(\mathbf{x}_{0:T} \mid \boldsymbol{\theta}) p(\boldsymbol{\theta})}{p(\mathbf{y}_{1:T})}, \quad (2-26)$$

其中：

- $p(\mathbf{x}_{0:T} \mid \boldsymbol{\theta}) = p(\mathbf{x}_0 \mid \boldsymbol{\theta}) \prod_{k=1}^T p(\mathbf{x}_k \mid \mathbf{x}_{k-1}, \boldsymbol{\theta})$ 是状态的先验分布，
- $p(\mathbf{y}_{1:T} \mid \mathbf{x}_{0:T}, \boldsymbol{\theta}) = \prod_{k=1}^T p(\mathbf{y}_k \mid \mathbf{x}_k, \boldsymbol{\theta})$ 是观测值。

参数估计的核心目标是找到最大化边缘后验分布的后验模态 $\boldsymbol{\theta}$ ：

$$\boldsymbol{\theta} = \arg \max_{\boldsymbol{\theta}} p(\boldsymbol{\theta} \mid \mathbf{y}_{1:T}) \quad (2-27)$$

这本质上是一个优化问题。然而，在解决这个问题之前，我们必须先解决计算 $p(\boldsymbol{\theta} \mid \mathbf{y}_{1:T})$ 的难题。

为了集中计算 $p(\boldsymbol{\theta} \mid \mathbf{y}_{1:T})$ ，即 $\boldsymbol{\theta}$ 的边缘后验分布，直接求解方法涉及对状态进行边缘化：

$$p(\boldsymbol{\theta} \mid \mathbf{y}_{1:T}) = \int p(\mathbf{x}_{0:T}, \boldsymbol{\theta} \mid \mathbf{y}_{1:T}) d\mathbf{x}_{0:T}. \quad (2-28)$$

遗憾的是，计算这个高维积分极具挑战性，而且随着测量值的增多，计算难度也会越来越大。为了解决这个问题，我们提出了一些参数估计方法，该方法可以近似边缘后验分布 $p(\boldsymbol{\theta} \mid \mathbf{y}_{1:T})$ ，而无需明确构建联合后验分布 $p(\mathbf{x}_{0:T}, \boldsymbol{\theta} \mid \mathbf{y}_{1:T})$ 。具体来说，我们关注：

$$p(\boldsymbol{\theta} \mid \mathbf{y}_{1:T}) \propto p(\mathbf{y}_{1:T} \mid \boldsymbol{\theta}) p(\boldsymbol{\theta}) \quad (2-29)$$

因此，问题简化为计算 似然值 $p(\mathbf{y}_{1:T} \mid \boldsymbol{\theta})$ 。（先验项 $p(\boldsymbol{\theta})$ 是根据先验知识选择的，而归一化常数 $p(\mathbf{y}_{1:T})$ 则被贝叶斯理论所避免。）

估计这种似然值的关键在于状态空间模型中的递归计算，也称为因式分解或预测误差分解：

$$p(\mathbf{y}_{1:T} \mid \boldsymbol{\theta}) = \prod_{k=1}^T p(\mathbf{y}_k \mid \mathbf{y}_{1:k-1}, \boldsymbol{\theta}), \quad (2-30)$$

其中：

- $p(\mathbf{y}_1 | \mathbf{y}_{1:0}, \boldsymbol{\theta}) \triangleq p(\mathbf{y}_1 | \boldsymbol{\theta})$ 。

处理 $p(\mathbf{y}_k | \mathbf{y}_{1:k-1}, \boldsymbol{\theta})$ 的核心思想是：

$$p(\mathbf{y}_k | \mathbf{y}_{1:k-1}, \boldsymbol{\theta}) = \int p(\mathbf{y}_k | \mathbf{x}_k, \boldsymbol{\theta}) p(\mathbf{x}_k | \mathbf{y}_{1:k-1}, \boldsymbol{\theta}) d\mathbf{x}_k, \quad (2-31)$$

$$p(\mathbf{x}_k | \mathbf{y}_{1:k}, \boldsymbol{\theta}) = \frac{p(\mathbf{y}_k | \mathbf{x}_k, \boldsymbol{\theta}) p(\mathbf{x}_k | \mathbf{y}_{1:k-1}, \boldsymbol{\theta})}{p(\mathbf{y}_k | \mathbf{y}_{1:k-1}, \boldsymbol{\theta})}, \quad (2-32)$$

其中：

- $p(\mathbf{y}_k | \mathbf{x}_k, \boldsymbol{\theta})$ 是测量模型，
- $p(\mathbf{x}_k | \mathbf{y}_{1:k-1}, \boldsymbol{\theta})$ 是状态 \mathbf{x}_k 的预测分布，由以下公式给出：

$$p(\mathbf{x}_k | \mathbf{y}_{1:k-1}, \boldsymbol{\theta}) = \int p(\mathbf{x}_k | \mathbf{x}_{k-1}, \boldsymbol{\theta}) p(\mathbf{x}_{k-1} | \mathbf{y}_{1:k-1}, \boldsymbol{\theta}) d\mathbf{x}_{k-1} \quad (2-33)$$

- $p(\mathbf{x}_k | \mathbf{y}_{1:k}, \boldsymbol{\theta})$ 是贝叶斯滤波器。

注意，上述方程求解了状态空间模型中的似然函数。对于线性高斯状态空间模型的特殊情况，通过利用卡尔曼滤波器的结果，似然的计算得到了显著的简化。

2.3.1 具有缺失参数的线性高斯状态空间模型

具有缺失参数 $\boldsymbol{\theta}$ 的线性高斯状态空间模型 (LGSSM) 定义为：

$$\mathbf{x}_k = \mathbf{A}(\boldsymbol{\theta}) \mathbf{x}_{k-1} + \mathbf{q}_{k-1}, \quad (2-34)$$

$$\mathbf{y}_k = \mathbf{H}(\boldsymbol{\theta}) \mathbf{x}_k + \mathbf{r}_k, \quad (2-35)$$

其中：

- $\mathbf{q}_{k-1} \sim \mathcal{N}(\mathbf{0}, \mathbf{Q}(\boldsymbol{\theta}))$ 和 $\mathbf{r}_k \sim \mathcal{N}(\mathbf{0}, \mathbf{R}(\boldsymbol{\theta}))$ 分别是过程噪声和测量噪声，
- $\mathbf{x}_0 \sim \mathcal{N}(\mathbf{m}_0(\boldsymbol{\theta}), \mathbf{P}_0(\boldsymbol{\theta}))$ 是初始状态分布，
- 假设模型参数 $\boldsymbol{\theta}$ 是时间不变的。

在接下来的章节中，我们将介绍应对 LGSSM 中计算和优化挑战的实用方法，具体包括梯度下降算法和期望最大化 (EM) 算法。

2.3.2 梯度下降算法

我们从方程 (2-13) 中知道， $p(\mathbf{y}_k | \mathbf{y}_{1:k-1}, \boldsymbol{\theta})$ 的分布是高斯分布：

$$p(\mathbf{y}_k | \mathbf{y}_{1:k-1}, \boldsymbol{\theta}) \sim \mathcal{N}(\mathbf{H}_k(\boldsymbol{\theta}) \mathbf{m}_k^-(\boldsymbol{\theta}), \mathbf{S}_k(\boldsymbol{\theta})).$$

因此，对数似然 $\log p(\mathbf{y}_{1:T} | \boldsymbol{\theta})$ 可以表示为从正态分布推导出的递归求和项：

$$\log p(\mathbf{y}_{1:T} | \boldsymbol{\theta}) = \sum_{k=1}^T \left(\frac{1}{2} \log |2\pi \mathbf{S}_k(\boldsymbol{\theta})| + \frac{1}{2} \mathbf{v}_k^T(\boldsymbol{\theta}) \mathbf{S}_k^{-1}(\boldsymbol{\theta}) \mathbf{v}_k(\boldsymbol{\theta}) \right), \quad (2-36)$$

其中：

- $\mathbf{v}_k(\boldsymbol{\theta})$ 和 $\mathbf{S}_k(\boldsymbol{\theta})$ 分别是残差值和残差协方差，定义见卡尔曼滤波器中的方程 (2-16) 和方程 (2-17)，
- m 是观测空间的维度，满足 $\mathbf{y} \in \mathbb{R}^m$ 。

令 $\ell(\boldsymbol{\theta}) \triangleq \log p(\mathbf{y}_{1:T} | \boldsymbol{\theta})$ 。为了最大化对数似然，我们使用梯度下降算法，该算法通过以下方式迭代更新参数估计 $\boldsymbol{\theta}$ ：

$$\boldsymbol{\theta}^{(i+1)} = \boldsymbol{\theta}^{(i)} - \eta \nabla_{\boldsymbol{\theta}} \ell(\boldsymbol{\theta}^{(i)}), \quad (2-37)$$

其中：

- $\eta > 0$ 是学习率，控制更新的步长，
- $\nabla_{\boldsymbol{\theta}} \ell(\boldsymbol{\theta})$ 是对数似然关于 $\boldsymbol{\theta}$ 的梯度，
- $\boldsymbol{\theta}^{(i)}$ 是第 i 次迭代中的参数估计。

梯度 $\nabla_{\boldsymbol{\theta}} \ell(\boldsymbol{\theta})$ 可以使用链式法则计算，利用卡尔曼滤波器的递归性质。具体而言， $\mathbf{v}_k(\boldsymbol{\theta})$ 和 $\mathbf{S}_k(\boldsymbol{\theta})$ 的梯度项由卡尔曼滤波器方程推导得到，整体梯度通过求和所有时间步 $k = 1, \dots, T$ 的贡献得到。

2.3.3 EM 算法

状态空间模型的 EM 算法 令 $q(\mathbf{x}_{0:T})$ 为状态 $\mathbf{x}_{0:T}$ 上的任意概率密度函数。对于对数似然 $\log p(\mathbf{y}_{1:T} | \boldsymbol{\theta})$ ，总是存在一个下界，如下所示：

$$\log p(\mathbf{y}_{1:T} | \boldsymbol{\theta}) \geq F(q(\mathbf{x}_{0:T}), \boldsymbol{\theta}), \quad (2-38)$$

其中二元函数 $F(\cdot)$ 定义为：

$$F(q(\mathbf{x}_{0:T}), \boldsymbol{\theta}) = \int q(\mathbf{x}_{0:T}, \boldsymbol{\theta}) \log \frac{p(\mathbf{x}_{0:T}, \mathbf{y}_{1:T} | \boldsymbol{\theta})}{q(\mathbf{x}_{0:T}, \boldsymbol{\theta})} d\mathbf{x}_{0:T}. \quad (2-39)$$

这自然引出了算法 1，用于求解这个优化问题。

Algorithm 1 简略 EM 算法**Output:** $\theta^{(N)}$

- 1: $\theta^{(0)} \leftarrow$ 随机值
- 2: **for** $n = 0, 1, \dots, N - 1$ **do**
- 3: E 步骤: $q^{(n+1)}(\cdot) \leftarrow \arg \max_{q(\cdot)} F(q(\cdot), \theta^{(n)})$ 。
- 4: M 步骤: $\theta^{(n+1)} \leftarrow \arg \max_{\theta} F(q^{(n+1)}(\cdot), \theta)$ 。
- 5: **end for**

Algorithm 2 状态空间模型的 EM 算法**Output:** $\theta^{(N)}$

- 1: $\theta^{(0)} \leftarrow$ 随机值
- 2: **for** $n = 0, 1, \dots, N - 1$ **do**
- 3: E 步骤: 计算 $Q(\theta, \theta^{(n)})$ 。
- 4: M 步骤: $\theta^{(n+1)} \leftarrow \arg \max_{\theta} Q(\theta, \theta^{(n)})$ 。
- 5: **end for**

为了求解算法 1 中的 E 步骤，我们直接给出最优解：

$$q^{(n+1)}(\mathbf{x}_{0:T}) = p(\mathbf{x}_{0:T} \mid \mathbf{y}_{1:T}, \theta^{(n)}), \quad (2-40)$$

$$\begin{aligned} F[q^{(n+1)}(\mathbf{x}_{0:T}), \theta] &= \int p(\mathbf{x}_{0:T} \mid \mathbf{y}_{1:T}, \theta^{(n)}) \log p(\mathbf{x}_{0:T}, \mathbf{y}_{1:T} \mid \theta) d\mathbf{x}_{0:T} \\ &\quad - \int p(\mathbf{x}_{0:T} \mid \mathbf{y}_{1:T}, \theta^{(n)}) \log p(\mathbf{x}_{0:T} \mid \mathbf{y}_{1:T}, \theta^{(n)}) d\mathbf{x}_{0:T}. \end{aligned} \quad (2-41)$$

注意，方程 (2-41) 中的第二项不依赖于 θ 。因此，算法 1 中的 M 步骤优化可以简化，只考虑方程 (2-41) 中的第一项作为对数似然的新下界，记作 $Q(\theta, \theta^{(n)})$ ：

$$Q(\theta, \theta^{(n)}) = \int p(\mathbf{x}_{0:T} \mid \mathbf{y}_{1:T}, \theta^{(n)}) \log p(\mathbf{x}_{0:T}, \mathbf{y}_{1:T} \mid \theta) d\mathbf{x}_{0:T}. \quad (2-42)$$

因此，我们得到了最终的迭代不等式，改编自方程 (2-38)：

$$\log p(\mathbf{y}_{1:T} \mid \theta) \geq Q(\theta, \theta^{(n)}). \quad (2-43)$$

这引出了简化版本的 EM 算法，用于估计状态空间模型的参数，如算法 2 所示。

线性高斯状态空间模型的 EM 算法 在线性高斯状态空间模型 (LGSSM) 的假设下, $Q(\theta, \theta^{(n)})$ 的表达式为:

$$\begin{aligned} Q(\theta, \theta^{(n)}) = & -\frac{1}{2} \log |2\pi \mathbf{P}_0(\theta)| - \frac{T}{2} \log |2\pi \mathbf{Q}(\theta)| - \frac{T}{2} \log |2\pi \mathbf{R}(\theta)| \\ & - \frac{1}{2} \text{tr} \{ \mathbf{P}_0^{-1}(\theta) [\mathbf{P}_0^s + (\mathbf{m}_0^s - \mathbf{m}_0(\theta))(\mathbf{m}_0^s - \mathbf{m}_0(\theta))^T] \} \\ & - \frac{T}{2} \text{tr} \{ \mathbf{Q}^{-1}(\theta) [\Sigma - \mathbf{C}\mathbf{A}^T(\theta) - \mathbf{A}(\theta)\mathbf{C}^T + \mathbf{A}(\theta)\Phi\mathbf{A}^T(\theta)] \} \\ & - \frac{T}{2} \text{tr} \{ \mathbf{R}^{-1}(\theta) [\mathbf{D} - \mathbf{B}\mathbf{H}^T(\theta) - \mathbf{H}(\theta)\mathbf{B}^T + \mathbf{H}(\theta)\Sigma\mathbf{H}^T(\theta)] \}, \end{aligned} \quad (2-44)$$

其中, 中间量 $\Sigma, \Phi, \mathbf{B}, \mathbf{C}, \mathbf{D}$ 计算为:

$$\Sigma = \frac{1}{T} \sum_{k=1}^T \mathbf{P}_k^s + \mathbf{m}_k^s [\mathbf{m}_k^s]^T, \quad (2-45)$$

$$\Phi = \frac{1}{T} \sum_{k=1}^T \mathbf{P}_{k-1}^s + \mathbf{m}_{k-1}^s [\mathbf{m}_{k-1}^s]^T, \quad (2-46)$$

$$\mathbf{B} = \frac{1}{T} \sum_{k=1}^T \mathbf{y}_k [\mathbf{m}_k^s]^T, \quad (2-47)$$

$$\mathbf{C} = \frac{1}{T} \sum_{k=1}^T \mathbf{P}_k^s \mathbf{G}_{k-1}^T + \mathbf{m}_k^s [\mathbf{m}_{k-1}^s]^T, \quad (2-48)$$

$$\mathbf{D} = \frac{1}{T} \sum_{k=1}^T \mathbf{y}_k \mathbf{y}_k^T, \quad (2-49)$$

这些量是由卡尔曼滤波器和卡尔曼平滑器的结果推导而来。

通过将梯度 $\frac{\partial Q(\theta, \theta^{(n)})}{\partial \theta}$ 设置为零, 我们得到以下闭式更新公式来更新参数:

$$\mathbf{m}_0^* = \mathbf{m}_0^s, \quad (2-50)$$

$$\mathbf{P}_0^* = \mathbf{P}_0^s + (\mathbf{m}_0^s - \mathbf{m}_0)(\mathbf{m}_0^s - \mathbf{m}_0)^T, \quad (2-51)$$

$$\mathbf{A}^* = \mathbf{C}\Phi^{-1}, \quad (2-52)$$

$$\mathbf{Q}^* = \Sigma - \mathbf{C}\mathbf{A}^T - \mathbf{A}\mathbf{C}^T + \mathbf{A}\Phi\mathbf{A}^T, \quad (2-53)$$

$$\mathbf{H}^* = \mathbf{B}\Sigma^{-1}, \quad (2-54)$$

$$\mathbf{R}^* = \mathbf{D} - \mathbf{H}\mathbf{B}^T - \mathbf{B}\mathbf{H}^T + \mathbf{H}\Sigma\mathbf{H}^T, \quad (2-55)$$

其中:

- θ^* 表示 $\arg \max_{\theta} Q(\theta, \theta^{(n)})$, 对应算法 2 中的 M 步骤。参数 θ 可以是 $\{\mathbf{A}, \mathbf{H}, \mathbf{Q}, \mathbf{R}, \mathbf{m}_0, \mathbf{P}_0\}$

Algorithm 3 线性高斯状态空间模型的 EM 算法**Output:** $\theta^{(N)}$

- 1: $\theta^{(0)} \leftarrow$ 随机值
- 2: **for** $n = 0, 1, \dots, N - 1$ **do**
- 3: E 步骤: 使用卡尔曼滤波器和卡尔曼平滑器的结果计算 $\Sigma, \Phi, \mathbf{B}, \mathbf{C}, \mathbf{D}$, 并使用 $\theta^{(n)}$ 。
- 4: M 步骤: 使用方程 (2-52), (2-54), (2-53), (2-55), (2-50), 和 (2-51) 更新 $\theta^{(n+1)} \leftarrow \arg \max_{\theta} Q(\theta, \theta^{(n)})$ 。
- 5: **end for**

的任意子集。

这引出了线性高斯状态空间模型的 EM 算法的实际实现, 如算法 3 所示。

第三章 GraphEM 算法

3.1 引言

本节中，我们提出了一个通用框架，用于在适当的先验假设下估计线性高斯状态空间模型 (LGSSM) 的转移矩阵 \mathbf{A} 。我们的目标是寻找 $\hat{\mathbf{A}}$ ，使得：

$$\hat{\mathbf{A}} = \arg \max_{\mathbf{A}} p(\mathbf{A} | \mathbf{y}_{1:T}), \quad (3-1)$$

其中：

- 由贝叶斯法则得： $p(\mathbf{A} | \mathbf{y}_{1:T}) \propto p(\mathbf{y}_{1:T} | \mathbf{A})p(\mathbf{A})$ 。

回忆在未引入先验信息之前，对数似然函数的下界由式 (2-43) 给出：

$$\log p(\mathbf{y}_{1:T} | \mathbf{A}) \geq Q(\mathbf{A}, \mathbf{A}^{(n)}) \quad (3-2)$$

为了引入先验信息，我们为 EM 框架导出一个新的下界：

$$\log p(\mathbf{y}_{1:T} | \mathbf{A}) + \log p(\mathbf{A}) \geq Q(\mathbf{A}, \mathbf{A}^{(n)}) + \log p(\mathbf{A}) \quad (3-3)$$

为适应最小化问题，我们引入新的记号。令损失函数为负对数似然：

$$\mathcal{L}_T(\mathbf{A}) \triangleq \mathcal{L}_0(\mathbf{A}) + \mathcal{L}_{1:T}(\mathbf{A}) \quad (3-4)$$

其中：

- $\mathcal{L}_0(\mathbf{A}) = -\log p(\mathbf{A})$ 是先验损失项，
- $\mathcal{L}_{1:T}(\mathbf{A}) = -\log p(\mathbf{y}_{1:T} | \mathbf{A})$ 是似然损失项。

采用该记号，我们的优化目标可由式 (3-1) 重写为：

$$\hat{\mathbf{A}} = \arg \min_{\mathbf{A}} \mathcal{L}_T(\mathbf{A}) \quad (3-5)$$

我们引入一个新的上界作为 GraphEM 算法中的目标函数：

$$\mathfrak{Q}(\mathbf{A}, \mathbf{A}^{(n)}) \triangleq -Q(\mathbf{A}, \mathbf{A}^{(n)}) + \lambda \mathcal{L}_0(\mathbf{A}) \quad (3-6)$$

其中 λ 用于控制先验损失项 $\mathcal{L}_0(\mathbf{A})$ 的权重。

Algorithm 4 简略 GraphEM 算法

Output: $\mathbf{A}^{(N)}$

- 1: $\mathbf{A}^{(0)} \leftarrow$ 随机初始化
 - 2: **for** $n = 0, 1, \dots, N - 1$ **do**
 - 3: E 步骤: 计算 $\mathfrak{Q}(\mathbf{A}, \mathbf{A}^{(n)})$ 。
 - 4: M 步骤: $\mathbf{A}^{(n+1)} \leftarrow \arg \min_{\mathbf{A}} \mathfrak{Q}(\mathbf{A}, \mathbf{A}^{(n)})$ 。
 - 5: **end for**
-

利用新记号重写式 (3-3)，并改变不等号方向，可得上界形式的 EM 结构：

$$\mathcal{L}_T(\mathbf{A}) \leq \mathfrak{Q}(\mathbf{A}, \mathbf{A}^{(n)}) \quad (3-7)$$

采用目标函数 $\mathfrak{Q}(\mathbf{A}, \mathbf{A}^{(n)})$ ，我们提出 GraphEM 算法，如算法 4 所示。

3.2 E 步骤中的计算

本节中，我们专注于计算目标函数 $\mathfrak{Q}(\mathbf{A}, \mathbf{A}^{(n)})$ 。

回忆，对于线性高斯状态空间模型 (LGSSM)，下界 $Q(\mathbf{A}, \mathbf{A}^{(n)})$ 由式 (2-44) 给出。由于我们只关心 \mathbf{A} ，该表达式可简化为：

$$Q(\mathbf{A}, \mathbf{A}^{(n)}) = -\frac{T}{2} \text{tr} \{ \mathbf{Q}^{-1} [\boldsymbol{\Sigma} - \mathbf{C}\mathbf{A}^\top - \mathbf{A}\mathbf{C}^\top + \mathbf{A}\boldsymbol{\Phi}\mathbf{A}^\top] \} + \text{ct.}_{\mathbf{A}}, \quad (3-8)$$

其中：

- $\boldsymbol{\Sigma} = \boldsymbol{\Sigma}(\mathbf{A}^{(n)})$ ，由式 (2-45) 计算，
- $\boldsymbol{\Phi} = \boldsymbol{\Phi}(\mathbf{A}^{(n)})$ ，由式 (2-46) 计算，
- $\mathbf{C} = \mathbf{C}(\mathbf{A}^{(n)})$ ，由式 (2-48) 计算，
- $\text{ct.}_{\mathbf{A}}$ 表示与 \mathbf{A} 无关的常数项。

利用式 (3-6)，一旦评估出 $\mathcal{L}_0(\mathbf{A})$ ，我们便可以计算 $\mathfrak{Q}(\mathbf{A}, \mathbf{A}^{(n)})$ 。这完成了 E 步骤中目标函数的计算。

3.3 M 步骤中的优化

本节的目标是找到使目标函数 $\mathfrak{Q}(\mathbf{A}, \mathbf{A}^{(n)})$ 最小化的矩阵 \mathbf{A} 。

在标准的 EM 算法中，该优化问题通常具有闭式解。然而，在我们的目标函数中引入了先验项，使问题变得更加复杂，因此需要采用专门的优化方法。这里，我们提出一种适用于该修正 EM 结构的求解方法。

3.3.1 近端算子

设 $f: \mathbb{R}^{n \times n} \rightarrow \mathbb{R}$ 是一个适当的、凸的、下半连续的函数。函数 f 在点 $\tilde{\mathbf{A}} \in \mathbb{R}^{n \times n}$ 的近端算子定义为：

$$\text{prox}_f(\tilde{\mathbf{A}}) = \arg \min_{\mathbf{A}} \left(f(\mathbf{A}) + \frac{1}{2} \|\mathbf{A} - \tilde{\mathbf{A}}\|_F^2 \right) \quad (3-9)$$

我们的目标函数由两项组成。首先，我们推导第一项 $-Q(\mathbf{A}, \mathbf{A}^{(n)})$ 的近端算子，其形式如下：

$$\text{prox}_{-Q(\tilde{\mathbf{A}}, \mathbf{A}^{(n)})}(\tilde{\mathbf{A}}) = \text{lyapunov}(\vartheta \mathbf{Q}^{-1}, \mathbf{\Phi}^{-1}, \tilde{\mathbf{A}} \mathbf{\Phi}^{-1} + \vartheta \mathbf{Q}^{-1} \mathbf{C} \mathbf{\Phi}^{-1}), \quad (3-10)$$

其中：

- $\vartheta > 0$ 是缩放参数，
- $\text{lyapunov}(X, Y, Z)$ 表示 Lyapunov 方程 $XW + WY = Z$ 的解。

进一步地，若 $\mathbf{Q} = \mathbf{I}_n \sigma_{\mathbf{Q}}^2$ ，则近端算子可简化为：

$$\text{prox}_{-Q(\tilde{\mathbf{A}}, \mathbf{A}^{(n)})}(\tilde{\mathbf{A}}) = \left(\frac{\vartheta T}{\sigma_{\mathbf{Q}}^2} \mathbf{C} + \tilde{\mathbf{A}} \right) \left(\frac{\vartheta}{\sigma_{\mathbf{Q}}^2} \mathbf{\Phi} + \mathbf{I}_n \right)^{-1}, \quad (3-11)$$

其中：

- $\sigma_{\mathbf{Q}}^2 > 0$ 是过程噪声的方差，
- $\mathbf{I}_n \in \mathbb{R}^{n \times n}$ 是单位矩阵。

3.3.2 Douglas-Rachford 迭代法

我们考虑如下优化问题：

$$\min_{\mathbf{A}} \mathfrak{Q}(\mathbf{A}, \mathbf{A}^{(n)}) \equiv \min_{\mathbf{A}} -Q(\mathbf{A}, \mathbf{A}^{(n)}) + \mathcal{L}_0(\mathbf{A}) \quad (3-12)$$

通过定义：

$$f_1(\mathbf{A}) \triangleq -Q(\mathbf{A}, \mathbf{A}^{(n)}), \quad f_2(\mathbf{A}) \triangleq \mathcal{L}_0(\mathbf{A}), \quad (3-13)$$

该问题可以使用 Douglas-Rachford 迭代法求解，其形式如下：

$$\mathbf{A}^{(i+1)} = \mathbf{A}^{(i)} + \alpha \left(\text{prox}_{\vartheta f_1} (2 \text{prox}_{\vartheta f_2}(\mathbf{A}^{(i)}) - \mathbf{A}^{(i)}) - \text{prox}_{\vartheta f_2}(\mathbf{A}^{(i)}) \right), \quad (3-14)$$

Algorithm 5 M 步中的 Douglas-Rachford 迭代法

Output: $\mathbf{A}^{(n+1)}$.

- 1: $\mathbf{A}^{(0)} \leftarrow \mathbf{A}^{(n)}$ 。
 - 2: **for** $i = 0, 1, 2, \dots$ **do**
 - 3: 计算 $\mathbf{Y}^{(i)} = \text{prox}_{\theta f_2}(\mathbf{A}^{(i)})$ 。
 - 4: 计算 $\mathbf{Z}^{(i)} = \text{prox}_{\theta f_1}(2\mathbf{Y}^{(i)} - \mathbf{A}^{(i)})$ 。
 - 5: 更新 $\mathbf{A}^{(i+1)} \leftarrow \mathbf{A}^{(i)} + \lambda(\mathbf{Z}^{(i)} - \mathbf{Y}^{(i)})$ 。
 - 6: **end for**
 - 7: $\mathbf{A}^{(n+1)} \leftarrow \mathbf{A}^{(i)}$ 。
-

其中我们设置 $\alpha = 1$ 。

在我们的具体问题中，两个近端算子分别为：

$$\text{prox}_{\theta f_1}(\mathbf{A}^{(i)}) \triangleq \arg \min_{\mathbf{A}} \left(-Q(\mathbf{A}, \mathbf{A}^{(n)}) + \frac{1}{2\theta} \|\mathbf{A} - \mathbf{A}^{(i)}\|^2 \right), \quad (3-15)$$

$$\text{prox}_{\theta f_2}(\mathbf{A}^{(i)}) \triangleq \arg \min_{\mathbf{A}} \left(\mathcal{L}_0(\mathbf{A}) + \frac{1}{2\theta} \|\mathbf{A} - \mathbf{A}^{(i)}\|^2 \right), \quad (3-16)$$

其中 $\text{prox}_{\theta f_1}(\mathbf{A}^{(i)})$ 和 $\text{prox}_{\theta f_2}(\mathbf{A}^{(i)})$ 的计算方法可参考前一节以及表 3-2。

Douglas-Rachford 方法在解决结构化优化问题中表现出色，适用于凸和非凸场景。用于 M 步优化的显式迭代步骤列于算法 5 中。

3.4 先验正则化

在 GraphEM 算法中，先验正则化在优化过程中起着关键作用，有助于提升稳定性和性能。本节将介绍用于算法中的稳定性约束及多种可选的正则项。这些约束与正则项对于引导优化方向至关重要，能够确保转移矩阵 \mathbf{A} 满足期望的性质，如稀疏性、平滑性或结构限制等。

3.4.1 稳定性约束

为了确保算法的稳定性，我们对初始转移矩阵 \mathbf{A} 施加约束。具体而言，我们要求 \mathbf{A} 的谱范数被一个常数 $\delta < 1$ 所约束。该约束确保系统动力学保持稳定，并使优化过程具有良好的收敛性。约束集合 \mathcal{S} 定义如下：

$$\mathcal{S} = \{\mathbf{A} \in \mathbb{R}^{n \times n} \mid \|\mathbf{A}\|_2 \leq \delta < 1\}, \quad (3-17)$$

其中 $\delta \in (0, 1)$ 是用户设定的参数，用于控制系统的稳定性。

表 3-1 给出了常见凸约束集合的示例及其对应的投影算子。这些约束对于在优化过

表 3-1 常见凸约束集合及其投影算子示例。

约束类型	\mathcal{S}	$\text{Proj}_{\mathcal{S}}(\mathbf{A})$
谱范数约束	$\ \mathbf{A}\ _2 \leq \delta$	$\mathbf{U}^\top \text{diag}((\text{sign}(s_i) \min(s_i , \delta))_{i \leq n}) \mathbf{V}$
元素范围约束	$(a_{ij} \in [a_{\min}, a_{\max}])_{i,j \leq n}$	$(\min(\max(a_{ij}, a_{\min}), a_{\max}))_{i,j \leq n}$
Frobenius 范数约束	$\ \mathbf{A}\ _F \leq \delta$	$(a_{ij}(1 - \frac{\delta}{\max(\ \mathbf{A}\ _F, \delta)}))_{i,j \leq n}$

 表 3-2 先验类型、正则项及其在尺度参数 $\vartheta > 0$ 下的近端算子示例。

先验类型	$\mathcal{L}_0(\mathbf{A})$	$\text{prox}_{\vartheta \mathcal{L}_0}(\mathbf{A})$
Laplace	$\ \mathbf{A}\ _1$	$(\text{sign}(a_{ij}) \max(0, a_{ij} - \lambda \vartheta))_{i,j \leq n}$
Block-Laplace	$\ \mathbf{A}\ _{2,1} = \sum_{b=1}^B \ \mathbf{a}(b)\ _2$	$(\left(1 - \frac{\lambda \vartheta}{\max(\ \mathbf{a}(b)\ _2, \lambda \vartheta)}\right) \mathbf{a}(b))_{b \leq B}$
Gaussian	$\frac{1}{2} \ \mathbf{A}\ _F^2$	$(\frac{a_{ij}}{1 + \lambda \vartheta})_{i,j \leq n}$
Laplace + Gaussian	$\ \mathbf{A}\ _1 + \frac{1}{2} \ \mathbf{A}\ _F^2$	$(\text{sign}(\frac{a_{ij}}{1 + \lambda \vartheta}) \max(0, \frac{a_{ij}}{1 + \lambda \vartheta} - \frac{\lambda \vartheta}{1 + \lambda \vartheta}))_{i,j \leq n}$

程中保持转移矩阵 \mathbf{A} 位于可行区域内至关重要。

3.4.2 正则项

除了稳定性约束外，我们还引入惩罚项以对转移矩阵 \mathbf{A} 进行正则化。这些惩罚项来源于对 \mathbf{A} 结构的先验知识，并以 $\mathcal{L}_0(\mathbf{A})$ 的形式融入目标函数中。不同的正则项鼓励 \mathbf{A} 具备不同的性质，例如稀疏性、块稀疏性或平滑性等。

表 3-2 列举了常用的先验类型、对应的正则项 $\mathcal{L}_0(\mathbf{A})$ 及其相关的近端算子。这些正则项在引导优化过程中发挥着关键作用，确保所估计的转移矩阵 \mathbf{A} 与其潜在图结构保持一致。

稳定性约束与正则项对于确保 GraphEM 算法产生具有意义且可解释的结果至关重要。稳定性约束保证了系统动力学的稳定性，而正则项则融入了关于转移矩阵 \mathbf{A} 结构的先验知识。

这些先验在处理不同类型图结构时尤其有用，例如小世界网络、无标度网络或二部图等，有助于算法适应图结构的特定性质。这种适配性使得算法能够有效捕捉图中节点之间的复杂关系，从而提升推断的准确性。

通过结合稳定性约束与正则项，GraphEM 算法为估计转移矩阵提供了一个灵活的框架，可广泛应用于时间序列分析、网络推断等多个领域。正则项的选择依赖于具体问题和对转移矩阵性质的需求，从而使该算法能够高度适配不同应用场景，尤其是在面对

具有不同连接模式和拓扑结构的图时，能够更精确地捕捉潜在的动态过程。

第四章 数值实验

4.1 汽车的动态模型

在本节中，我们通过一个真实场景示例展示卡尔曼滤波器的强大能力及其捕捉马尔可夫过程行为的效果。我们考虑一个汽车的动态模型，该模型在二维坐标系中模拟汽车的运动，并跟踪其位置与速度。转移矩阵 \mathbf{A} 定义如下：

$$\begin{pmatrix} x_k^{(1)} \\ x_k^{(2)} \\ x_k^{(3)} \\ x_k^{(4)} \end{pmatrix} = \underbrace{\begin{pmatrix} 1 & 0 & \Delta t & 0 \\ 0 & 1 & 0 & \Delta t \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}}_{\mathbf{A}} \begin{pmatrix} x_{k-1}^{(1)} \\ x_{k-1}^{(2)} \\ x_{k-1}^{(3)} \\ x_{k-1}^{(4)} \end{pmatrix} + \mathbf{q}_{k-1}, \quad (4-1)$$

其中 \mathbf{q}_{k-1} 是一个离散时间高斯噪声过程，其均值为零，协方差矩阵 \mathbf{Q} 定义为：

$$\mathbf{Q} = \begin{pmatrix} \frac{q_1^c \Delta t^3}{3} & 0 & \frac{q_1^c \Delta t^2}{2} & 0 \\ 0 & \frac{q_2^c \Delta t^3}{3} & 0 & \frac{q_2^c \Delta t^2}{2} \\ \frac{q_1^c \Delta t^2}{2} & 0 & q_1^c \Delta t & 0 \\ 0 & \frac{q_2^c \Delta t^2}{2} & 0 & q_2^c \Delta t \end{pmatrix}. \quad (4-2)$$

该动态系统的线性高斯状态空间模型（LGSSM）完整形式为：

$$\mathbf{x}_k = \mathbf{A}_{k-1} \mathbf{x}_{k-1} + \mathbf{q}_{k-1}, \quad (4-3)$$

$$\mathbf{y}_k = \mathbf{H}_k \mathbf{x}_k + \mathbf{r}_k, \quad (4-4)$$

其中模型参数设置如下：

- \mathbf{A}_{k-1} 按照式 (4-1) 定义，
- $\mathbf{H}_k = \mathbf{I}_4$ ，
- $\mathbf{r}_k \sim \mathcal{N}(\mathbf{0}, \sigma_r^2 \mathbf{I}_4)$ ，其中 $\sigma_r = 0.1$ ，
- $\mathbf{x}_0 \sim \mathcal{N}(\mathbf{m}_0, \mathbf{P}_0)$ ，其中 $\mathbf{m}_0 = \mathbf{0}$ ， $\mathbf{P}_0 = \sigma_p^2 \mathbf{I}_4$ ， $\sigma_p = 0.1$ 。

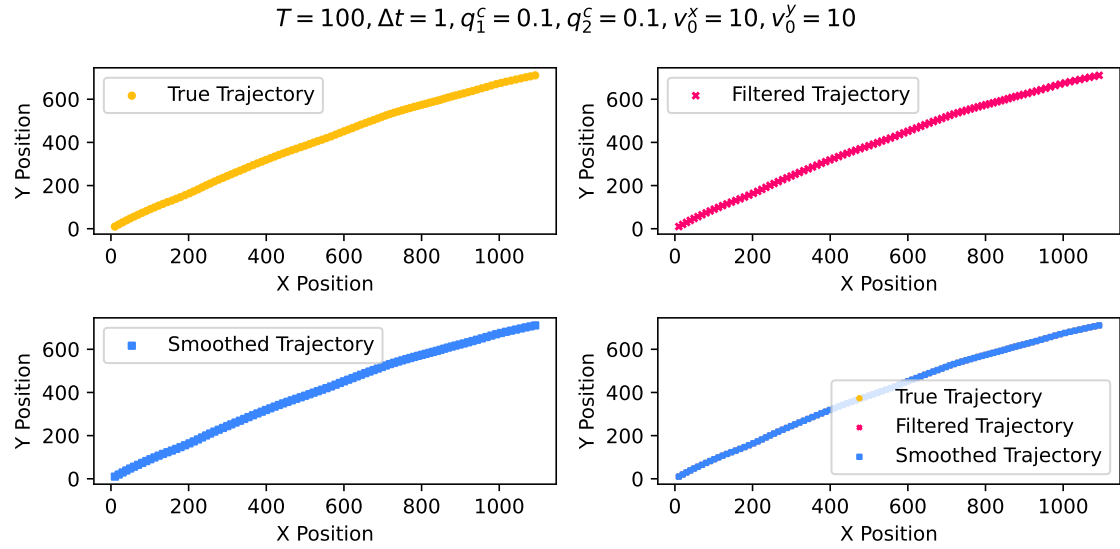


图 4-1 观测轨迹与使用卡尔曼滤波器和平滑器估计得到的轨迹。

表 4-1 汽车模型的 EM 算法结果。基线负对数似然为 $\mathcal{L}_T(\theta) = -261.989$ 。

待估参数	$\mathcal{L}_T(\hat{\theta})$	$\ \hat{\theta} - \theta\ _F$
A	-273.250	0.248
Q	-247.424	0.308
H	-169.839	0.642

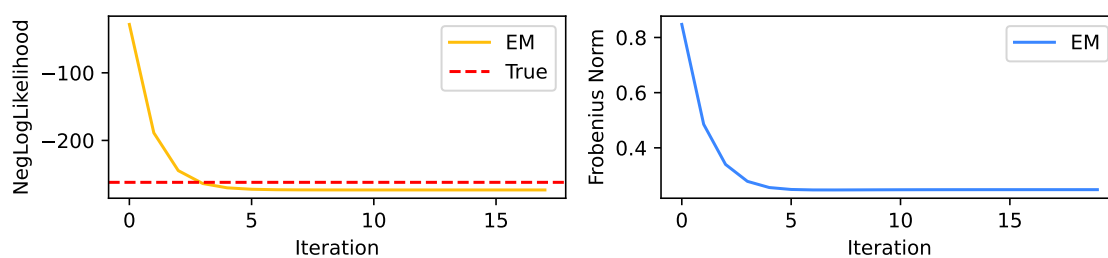
该模型表示的是匀速运动，其中状态向量的前两个分量对应位置，后两个分量对应速度。卡尔曼滤波器与平滑器为估计观测数据背后的潜在状态变量提供了闭式解法。

图 4-1 展示了实验结果，绘制了使用卡尔曼滤波器和平滑器所估计的轨迹。该图直观展示了这些方法在准确估计潜在状态变量方面的有效性。

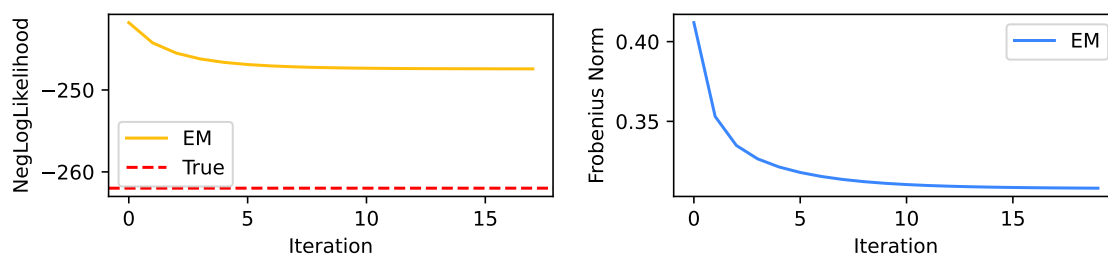
我们还使用经典的期望最大化 (EM) 算法对该汽车模型的参数进行估计。EM 算法是状态空间模型中广泛应用的参数估计方法，我们将在该示例中评估其性能。

表 4-1 总结了使用 EM 算法对参数 **A**、**Q** 和 **H** 的估计结果。表中列出了每个估计参数对应的负对数似然值 $\mathcal{L}_T(\hat{\theta})$ 以及与真实参数之间的 Frobenius 范数 $\|\hat{\theta} - \theta\|_F$ 。

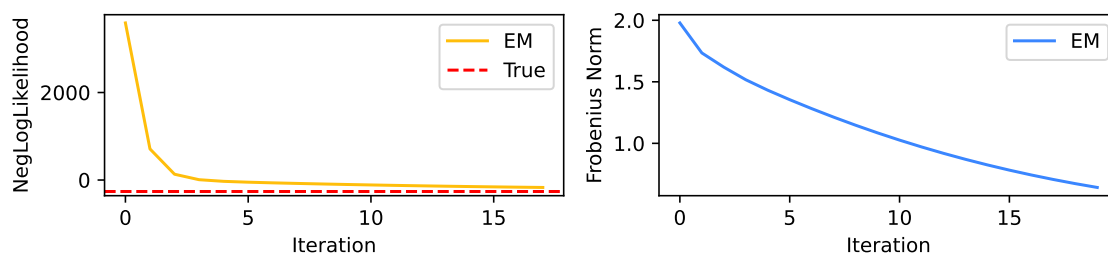
为进一步直观展示参数估计过程，图 4-2 展示了 EM 迭代中每个参数对应的负对数似然与 Frobenius 范数的变化趋势。此外，图 4-3 给出了各参数的图结构对比图，分别显示了真实值与估计结果之间的差异。



a) \mathbf{A} 的负对数似然与 Frobenius 范数变化。

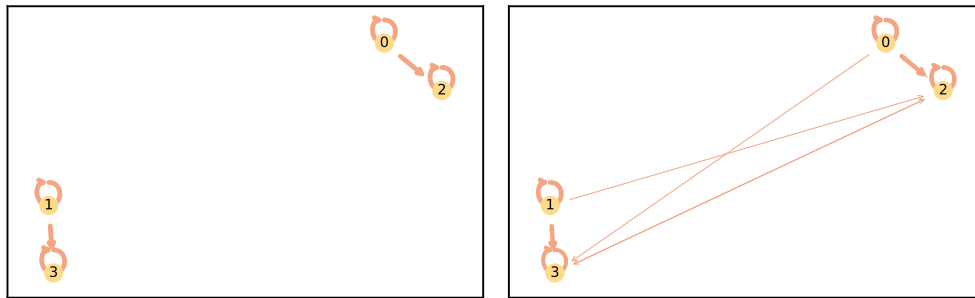


b) \mathbf{Q} 的负对数似然与 Frobenius 范数变化。

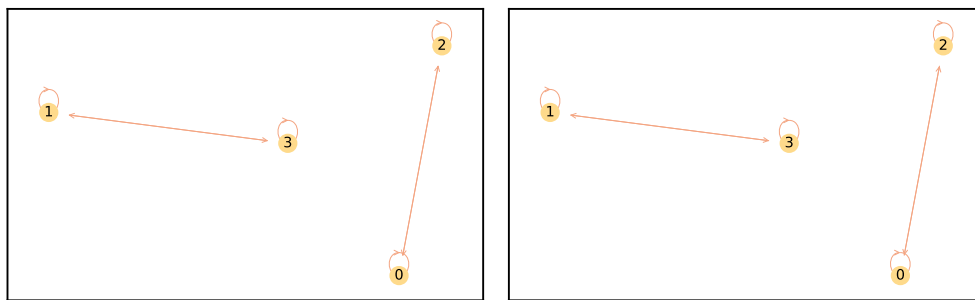


c) \mathbf{H} 的负对数似然与 Frobenius 范数变化。

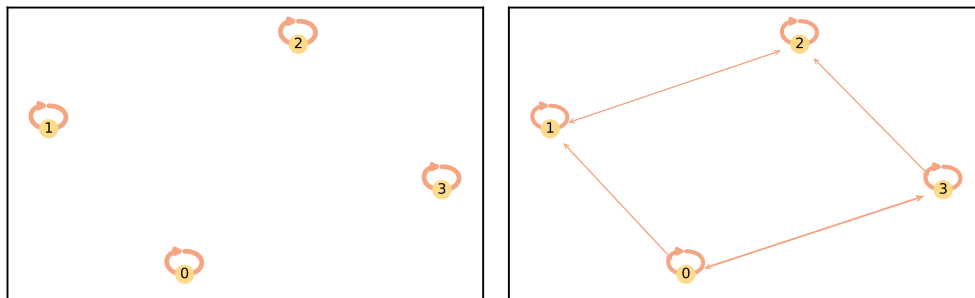
图 4-2 EM 算法在估计 \mathbf{A} 、 \mathbf{Q} 和 \mathbf{H} 时的性能可视化结果。



a) \mathbf{A} 的图结构表示：左为真实值，右为估计值。



b) \mathbf{Q} 的图结构表示：左为真实值，右为估计值。



c) \mathbf{H} 的图结构表示：左为真实值，右为估计值。

图 4-3 真实参数与估计参数 \mathbf{A} 、 \mathbf{Q} 、 \mathbf{H} 的图形对比。

4.2 不同图结构下的正则化分析

在本节中，我们介绍实验中使用的不同图结构的数学特性。这些结构在 GraphEM 算法中起到了定义转移矩阵正则化先验的关键作用。下文将对每种图类型进行详细分析，并说明其生成所采用的具体参数。

分块对角结构 (Blockwise-Diagonal) 分块对角矩阵是一类特殊的稀疏矩阵，其非零元素局限于对角线上的若干块中。数学形式如下：

$$\mathbf{A} = \begin{pmatrix} \mathbf{A}_1 & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{A}_2 & \cdots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & \mathbf{A}_k \end{pmatrix},$$

其中 \mathbf{A}_i 是子矩阵， $\mathbf{0}$ 表示相应维度的零矩阵。该结构常用于建模解耦或弱耦合的子系统。

小世界图 (Small-World) 小世界图具有较高的聚类系数与较短的平均路径长度。在实验中，我们使用 Watts-Strogatz 模型生成小世界图，设节点数 $n = 16$ ，每个节点连接 $k = 4$ 个最近邻节点，重连概率为 $p = 0.3$ 。小世界图的邻接矩阵 \mathbf{A} 通常具有局部连接与少量远程连接的混合特性，实现稀疏性与连通性的平衡。

无标度图 (Scale-Free) 无标度图的节点度分布满足幂律形式： $P(k) \sim k^{-\gamma}$ ，其中 k 是节点度， γ 是常数。我们使用 Barabási-Albert 模型生成无标度图，设 $n = 16$ ，每步增加 $m = 2$ 条边。该图的邻接矩阵 \mathbf{A} 高度稀疏，仅有少数枢纽节点具有较高的度数。

二部图 (Bipartite) 二部图是一类可将顶点划分为两个不相交集 U 与 V 的图，其中每条边仅连接 U 与 V 中的节点。实验中，我们生成一个完全二部图，节点数 $n = 16$ ，两个子集分别为 $n_1 = \lfloor n/2 \rfloor$ 、 $n_2 = \lfloor n/2 \rfloor$ 。其邻接矩阵形式如下：

$$\mathbf{A} = \begin{pmatrix} \mathbf{0} & \mathbf{B} \\ \mathbf{B}^\top & \mathbf{0} \end{pmatrix},$$

其中 \mathbf{B} 是表示两个集合间连接关系的矩阵。该结构适用于建模两个不同实体间的关联关系。

环图 (Cycle) 环图由一个封闭的循环组成, 每个节点恰好与两个其他节点相连。我们生成节点数为 $n = 16$ 的环图。其邻接矩阵为循环矩阵 (circulant matrix):

$$\mathbf{A} = \begin{pmatrix} 0 & 1 & 0 & \cdots & 1 \\ 1 & 0 & 1 & \cdots & 0 \\ 0 & 1 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & 0 & 0 & \cdots & 0 \end{pmatrix}.$$

环图常用于建模周期性或循环性结构。

星型图 (Star) 星型图由一个中心节点与所有其他节点连接而成, 其他节点之间没有连接。实验中, 我们生成 $n = 16$ 节点的星型图。其邻接矩阵为:

$$\mathbf{A} = \begin{pmatrix} 0 & 1 & 1 & \cdots & 1 \\ 1 & 0 & 0 & \cdots & 0 \\ 1 & 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & 0 & 0 & \cdots & 0 \end{pmatrix}.$$

星型图适用于建模中心化系统或星状网络结构 (hub-and-spoke)。

表 4-2 总结了 EM 算法及其 GraphEM 变体 (包括 Laplace、Gaussian 和 Laplace+Gaussian 正则化) 在不同图结构上的性能表现, 图类型包括分块对角、小世界、无标度、二部图、环图和星型图。表中报告了每种方法在每种图类型下的负对数似然 $\mathcal{L}_T(\hat{\mathbf{A}})$ 以及 Frobenius 范数 $\|\hat{\mathbf{A}} - \mathbf{A}\|_F$ 。

从表中可以看出, GraphEM 的各类变体在 Frobenius 范数方面均优于标准 EM 算法。具体而言:

- *GraphEM Laplace+Gaussian* 方法在所有图类型中均表现最佳, 取得了最小的 Frobenius 范数。这表明将 Laplace 与 Gaussian 先验结合能够提供更稳健的正则化框架。
- *GraphEM Laplace* 方法相较标准 EM 同样具有显著提升, 尤其在降低 Frobenius 范数方面效果明显, 说明稀疏性正则化在刻画图结构方面是有效的。
- *GraphEM Gaussian* 方法相比标准 EM 稍有提升, 但整体效果不如 Laplace 及 Laplace + Gaussian 变体, 进一步强调了诱导稀疏性的先验在图结构正则化中的重要性。

图 4-4 展示了分块对角图在不同正则方法下的负对数似然收敛情况。从图中可以看

表 4-2 不同图结构在各类正则方法下的实验结果。每种图类型中加粗的数值表示最佳结果。

图类型	方法	$\mathcal{L}_T(\hat{\mathbf{A}})$	$\ \hat{\mathbf{A}} - \mathbf{A}\ _F$
分块对角	EM	-21288.702	0.611
	GraphEM Laplace	-21262.253	0.575
	GraphEM Gaussian	-21287.225	0.608
	GraphEM Laplace+Gaussian	-21244.464	0.566
小世界	EM	-2237.305	2.489
	GraphEM Laplace	-2195.235	2.181
	GraphEM Gaussian	-2235.439	2.424
	GraphEM Laplace+Gaussian	-2193.343	2.140
无标度	EM	-2234.821	2.176
	GraphEM Laplace	-2197.950	2.005
	GraphEM Gaussian	-2233.167	2.138
	GraphEM Laplace+Gaussian	-2195.138	1.974
二部图	EM	-2228.962	2.366
	GraphEM Laplace	-2187.252	2.170
	GraphEM Gaussian	-2227.116	2.310
	GraphEM Laplace+Gaussian	-2185.598	2.127
环图	EM	-2279.460	2.257
	GraphEM Laplace	-2244.706	2.022
	GraphEM Gaussian	-2277.679	2.206
	GraphEM Laplace+Gaussian	-2243.051	1.984
星型图	EM	-2287.181	2.517
	GraphEM Laplace	-2253.640	2.240
	GraphEM Gaussian	-2285.723	2.456
	GraphEM Laplace+Gaussian	-2252.407	2.196

出，所有 EM 系列方法的收敛值均低于真实参数对应的负对数似然值，进一步验证了其估计效果的有效性。

为了进行更细致的对比，图 4-5 可视化展示了分块对角图的真实转移矩阵与估计转移矩阵。结果显示，GraphEM 的 Laplace 与 Laplace+Gaussian 方法所得估计最接近真实值，伪影更少，且更好地保留了块对角结构。

实验结果表明，将先验知识以正则化方式融入 EM 算法是有效的。具体而言：

- *Laplace* 先验在促进稀疏性方面尤其有效，这对捕捉小世界图与无标度图等图结构至关重要。
- *Gaussian* 先验提供了更平滑的估计，但在保持稀疏性方面较弱，因此更适用于如二部图和环图这类相对稠密的结构。
- *Laplace+Gaussian* 先验综合了两种先验的优势，在所有图类型中表现最佳。这表明混合正则策略在图结构参数估计中具有高度通用性和效果。

这些发现突显了 GraphEM 在多种图结构上的优越性能。与标准 EM 算法相比，

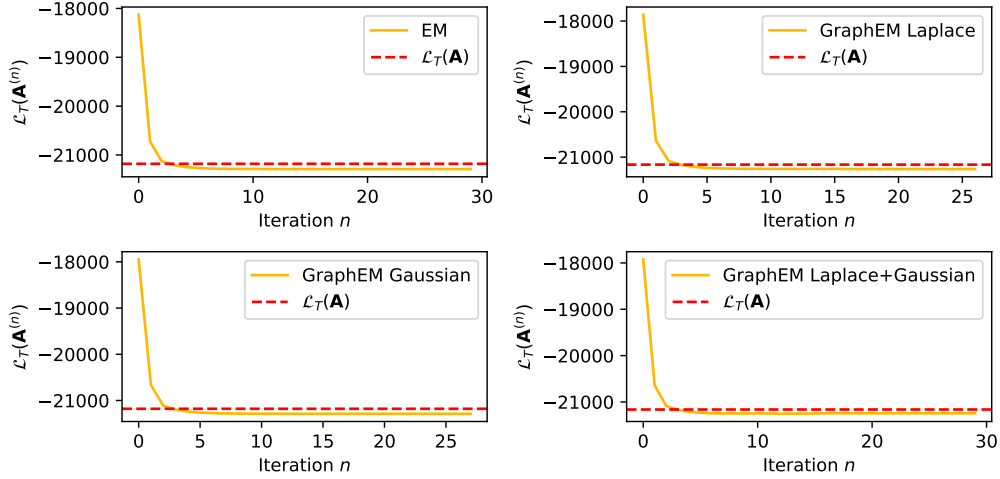


图 4-4 不同正则化方法下，分块对角图的负对数似然收敛曲线。

GraphEM 不仅保持了相当的收敛速度，而且在估计精度上有显著提升，进一步验证了其有效性。值得注意的是，在所有图类型中，在 **Laplace** 先验基础上引入 **Gaussian** 正则项的做法始终能进一步提升表现。这一现象可有两种解释：(1) 稀疏性仍是这些图结构中的主导特征；或 (2) 算法对具体图类型具有一定鲁棒性。在这两种情形下，**Gaussian** 正则所引入的稳定性都能带来性能提升。

此外，我们还观察到，在 **GraphEM** 中将 **Laplace** 与 **Gaussian** 先验结合后，能够“无折中”地叠加两者的优点，而不会引入额外的优化计算开销。这一结果尤其具有前景意义，表明 **GraphEM** 的迭代优化框架可以无缝整合多种正则项。这一发现为未来研究开辟了新方向，即通过累积不同类型的正则项，在不增加计算复杂度的前提下迭代逼近最优解。

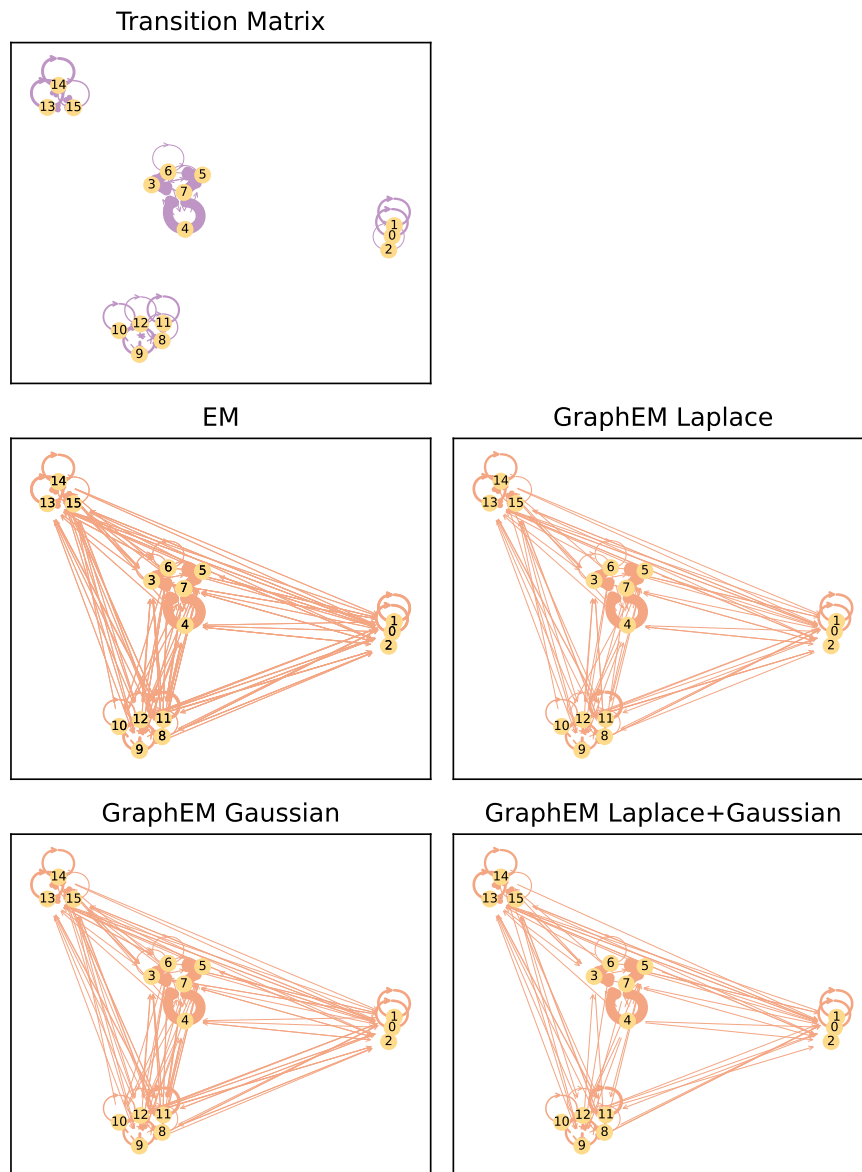


图 4-5 分块对角图的真实转移矩阵与估计矩阵对比图。

结论

本文研究了 GraphEM 算法在线性高斯状态空间模型 (LGSSM) 参数估计中的应用。GraphEM 是一种新颖的算法，它将先验知识整合进期望最大化 (EM) 框架中。通过引入图推理与正则化技术，我们对原始 GraphEM 方法进行了扩展，使其能够处理更广泛的图结构与正则化策略，包括 Laplace、Gaussian 及混合 Laplace+Gaussian 先验。

一个具有前景的结果是，GraphEM 框架能够无缝集成多种正则项。Laplace 与 Gaussian 先验的组合实现了两者优势的“无折中”叠加，而不引入额外的优化开销。这一特性为未来的研究开辟了新方向：通过累积多样化的正则项，可以在不增加计算复杂度的前提下，迭代地逼近最优解。

未来的研究方向包括发展自适应正则化策略，根据观测数据与图结构动态调整 Laplace 与 Gaussian 先验的权重。此外，进一步研究 GraphEM 在其他图类型上的表现，如随机图、分层图与动态图，有助于验证其通用性。对混合 Laplace+Gaussian 正则方法的收敛性与鲁棒性进行理论分析，也将有助于深入理解其有效性。最后，将 GraphEM 应用于金融、气候建模与社交网络分析等实际问题中，有望展示其在参数估计中的实用价值与潜在影响。

参考文献

- [1] HAMILTON J D. State-space models[J]. Handbook of econometrics, 1994, 4: 3039-3080.
- [2] KIM C J, NELSON C R. State-space models with regime switching: classical and gibbs-sampling approaches with applications[M]. MIT press, 2017.
- [3] CHOI B. Urban temperature and electricity demand: probabilistic modeling and adaptive decision-making[D]. Carnegie Mellon University, 2023.
- [4] DABUSH L, ROUTTENBERG T. Kalman filter for tracking network dynamic[C]// ICASSP 2024-2024 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). IEEE, 2024: 13216-13220.
- [5] SÄRKKÄ S, SVENSSON L. Institute of mathematical statistics textbooks: Bayesian filtering and smoothing[M]. 2nd ed. Cambridge University Press, 2023.
- [6] WATTS D J, STROGATZ S H. Collective dynamics of ‘small-world’ networks[J]. nature, 1998, 393(6684): 440-442.
- [7] GUPTA N, MEHRA R. Computational aspects of maximum likelihood estimation and reduction in sensitivity function calculations[J]. IEEE transactions on automatic control, 1974, 19(6): 774-783.
- [8] NEAL R M, HINTON G E. A view of the em algorithm that justifies incremental, sparse, and other variants[M]//Learning in graphical models. Springer, 1998: 355-368.
- [9] SHUMWAY R H, STOFFER D S. An approach to time series smoothing and forecasting using the em algorithm[J]. Journal of time series analysis, 1982, 3(4): 253-264.
- [10] CHOUZENOUX E, ELVIRA V. Graphem: Em algorithm for blind kalman filtering under graphical sparsity constraints[C]//ICASSP 2020-2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). IEEE, 2020: 5840-5844.
- [11] GRANGER C W. Investigating causal relations by econometric models and cross-spectral methods[J]. Econometrica: journal of the Econometric Society, 1969: 424-438.
- [12] RAVAZZI C, TEMPO R, DABBENE F. Learning influence structure in sparse social networks[J]. IEEE Transactions on Control of Network Systems, 2017, 5(4): 1976-1986.

-
- [13] ZOU H, HASTIE T. Regularization and variable selection via the elastic net[J]. Journal of the Royal Statistical Society Series B: Statistical Methodology, 2005, 67(2): 301-320.
- [14] SHARMA S, ELVIRA V, CHOUZENOUX E, et al. Recurrent dictionary learning for state-space models with an application in stock forecasting[J]. Neurocomputing, 2021, 450: 1-13.
- [15] DAHIYA S, SHARMA S, SAHNAN D, et al. Would your tweet invoke hate on the fly? forecasting hate intensity of reply threads on twitter[C]//Proceedings of the 27th ACM SIGKDD Conference on Knowledge Discovery & Data Mining. 2021: 2732-2742.
- [16] ELVIRA V, CHOUZENOUX E, CERDA J, et al. Graphs in state-space models for granger causality in climate science[A]. 2023.
- [17] ELVIRA V, CHOUZENOUX É. Graphical inference in linear-gaussian state-space models [J]. IEEE Transactions on Signal Processing, 2022, 70: 4757-4771.
- [18] CHOUZENOUX E, ELVIRA V. Graphit: Iterative reweighted l1 algorithm for sparse graph inference in state-space models[C]//ICASSP 2023-2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). IEEE, 2023: 1-5.
- [19] CHOUZENOUX E, ELVIRA V. Graphical inference in non-markovian linear-gaussian state-space models[C]//ICASSP 2024-2024 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). IEEE, 2024: 13141-13145.
- [20] COX B, ELVIRA V. Parameter estimation in sparse linear-gaussian state-space models via reversible jump markov chain monte carlo[C]//2022 30th European Signal Processing Conference (EUSIPCO). IEEE, 2022: 797-801.
- [21] COX B, ELVIRA V. Sparse bayesian estimation of parameters in linear-gaussian state-space models[J]. IEEE Transactions on Signal Processing, 2023, 71: 1922-1937.
- [22] COX B, CHOUZENOUX E, ELVIRA V. Graphgrad: Efficient estimation of sparse polynomial representations for general state-space models[A/OL]. 2024. arXiv: 2411.15637. <https://arxiv.org/abs/2411.15637>.
- [23] CHOUZENOUX E, ELVIRA V. Sparse graphical linear dynamical systems[J]. Journal of Machine Learning Research, 2024, 25(223): 1-53.

- [24] TSAMPOURAKIS K, ELVIRA V. Approximating the likelihood ratio in linear-gaussian state-space models for change detection[C]//ICASSP 2022-2022 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). IEEE, 2022: 5912-5916.
- [25] SHARMA S, CHOUZENOUX É, ELVIRA V, et al. A dynamical model for stock forecasting via deep recurrent dictionary learning[Z]. 2023.
- [26] CUI C, DJURIĆ P M. Inference of time-varying graph topologies via gaussian processes [C]//ICASSP 2024-2024 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). IEEE, 2024: 13276-13280.
- [27] MOHANTY A, MOHAMED N, RAMANAN P, et al. Federated granger causality learning for interdependent clients with state space representation[A/OL]. 2025. arXiv: 2501.13890. <https://arxiv.org/abs/2501.13890>.

致谢

四年时光如白驹过隙，转眼间我的本科生活即将画上句点。回望这段时光，我不仅积累了丰富的知识与技能，更重要的是提升了我的分析和解决问题的能力，培养了良好的科学素养。虽然过程中走过一些弯路，但正是这些经历让我更加坚定了未来的方向，收获了许多宝贵的经验和教训。这一切的进步与成长，都离不开老师们的教诲和同学们的帮助，在此，我想向他们表达我最诚挚的谢意。

首先，我要特别感谢我的所有任课老师。在我还是一名本科生时，缺乏学术研究的经验，对于研究问题的重点、难点和热点往往难以把握，更难以评估自己的研究成果能够达到的层次。然而，老师们凭借他们渊博的知识和深刻的学术洞察力，给予了我无微不至的指导。这不仅在学术上给予了我明确的方向，帮助我理清思路，还在科研的过程中为我指引了正确的道路。他们严谨的治学态度、精益求精的工作作风，以及对学生的无私关怀，始终是我学习的榜样。在此，我想向老师们表达我最崇高的敬意和衷心的感谢，感谢他们在这段求学旅程中给予我的悉心指导和无私帮助。

其次，我要感谢我的同学们，特别是在研究过程中与我一同讨论、分享的伙伴们。大家在不同的学科领域中各具专长，在遇到问题时总能互相启发、共同解决。每一次的讨论和合作，不仅让我获得了新的知识，也让我深刻体会到团队合作的力量。你们的支持和鼓励是我不断进步的重要动力，在这里我要向你们表示最诚挚的感谢。

最后，我要深深感谢我的家人，是你们一直以来的无私奉献与默默支持，让我在面对困难和挑战时能够坚定信心、勇敢前行。你们给予我无限的爱与力量，使我能够在追求学术和个人成长的道路上不断超越自我，最终取得今天的成果。在此，我想对你们表达我最深沉的感激和爱。

四年的时光，虽然短暂，但每一步都充满了挑战与成长。感谢所有曾给予我帮助与支持的人，正是有了你们的陪伴，我才能走得更远、飞得更高。在未来的道路上，我将继续努力，不忘初心，勇敢追梦，回报这段珍贵的岁月。

邓肯

2025年5月26日

附录 A 其他图类型

小世界图 (Small World) 图 A-1 对比了小世界图的真实转移矩阵与估计矩阵。结果表明, GraphEM 的 Laplace+Gaussian 方法能够有效捕捉小世界网络所特有的局部聚类与远程连接特性, 其表现优于其他正则化策略。

无标度图 (Scale Free) 图 A-2 展示了无标度图的真实与估计转移矩阵。GraphEM Laplace+Gaussian 方法在识别枢纽节点和保持幂律度分布方面表现出色, 显示出其在处理无标度结构时的鲁棒性。

二部图 (Bipartite) 图 A-3 展示了二部图的真实与估计转移矩阵。GraphEM Laplace+Gaussian 方法能够准确捕捉二部结构, 展现了其建模不同节点集合之间关系的能力。

环图 (Cycle) 图 A-4 对比了环图的真实与估计转移矩阵。GraphEM Laplace+Gaussian 方法有效保留了环状结构, 展现了其建模周期性关系的能力。

星型图 (Star) 图 A-5 展示了星型图的真实与估计转移矩阵。GraphEM Laplace+Gaussian 方法能够准确识别中心节点及其与边缘节点的连接, 展现出其在建模星状网络中的有效性。

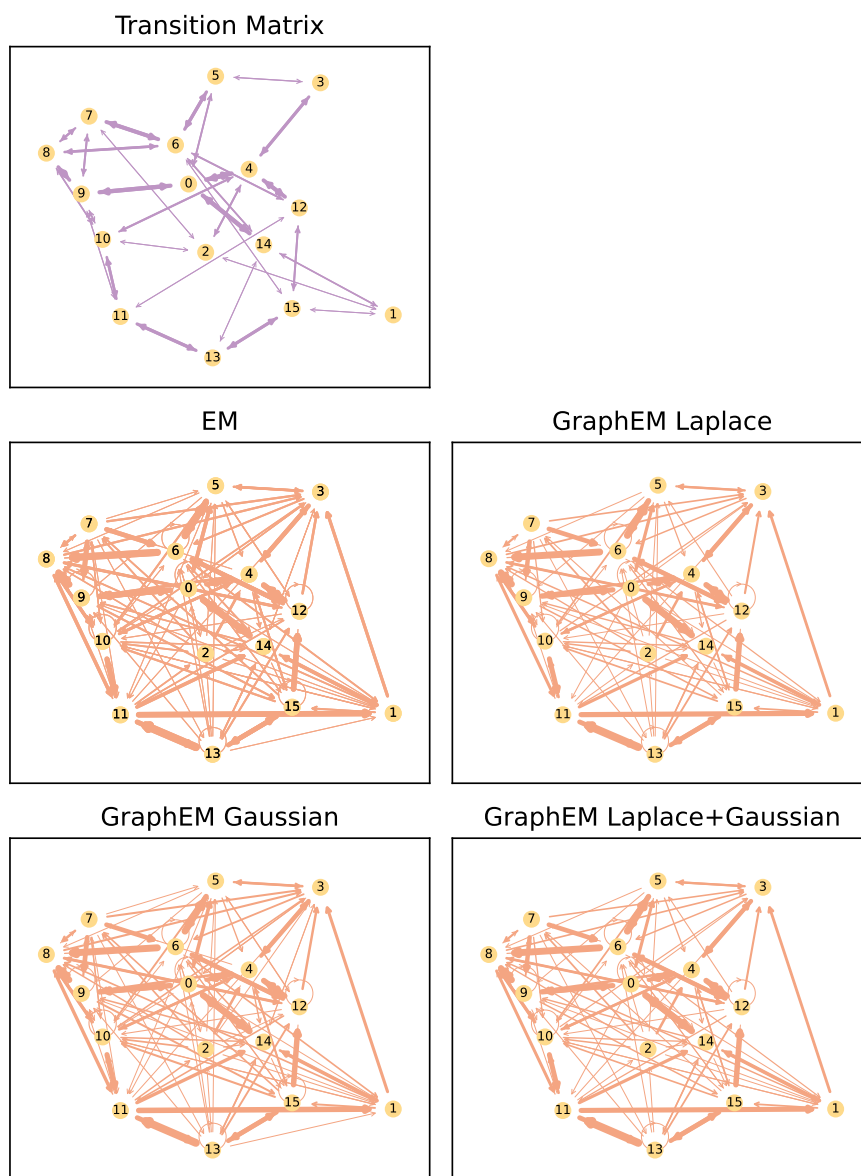


图 A-1 小世界图的真实转移矩阵与估计矩阵对比图。

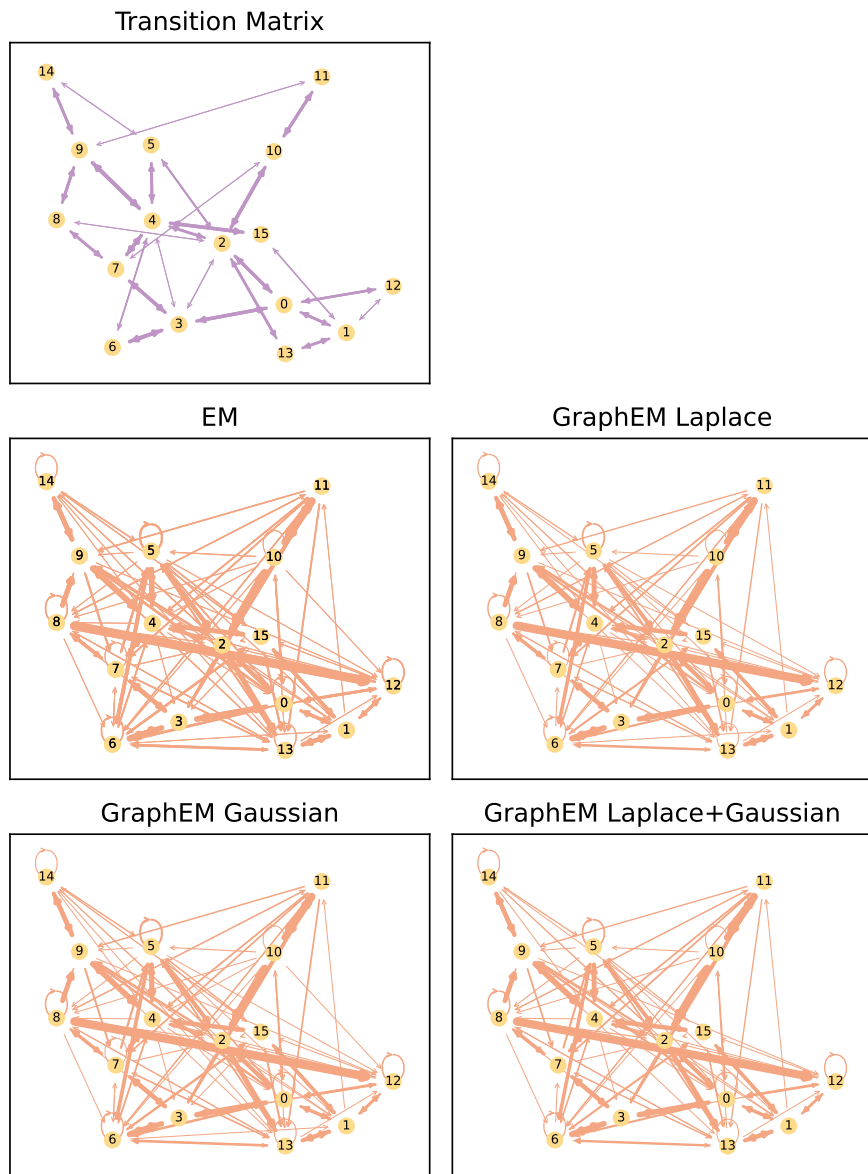


图 A-2 无标度图的真实转移矩阵与估计矩阵对比图。

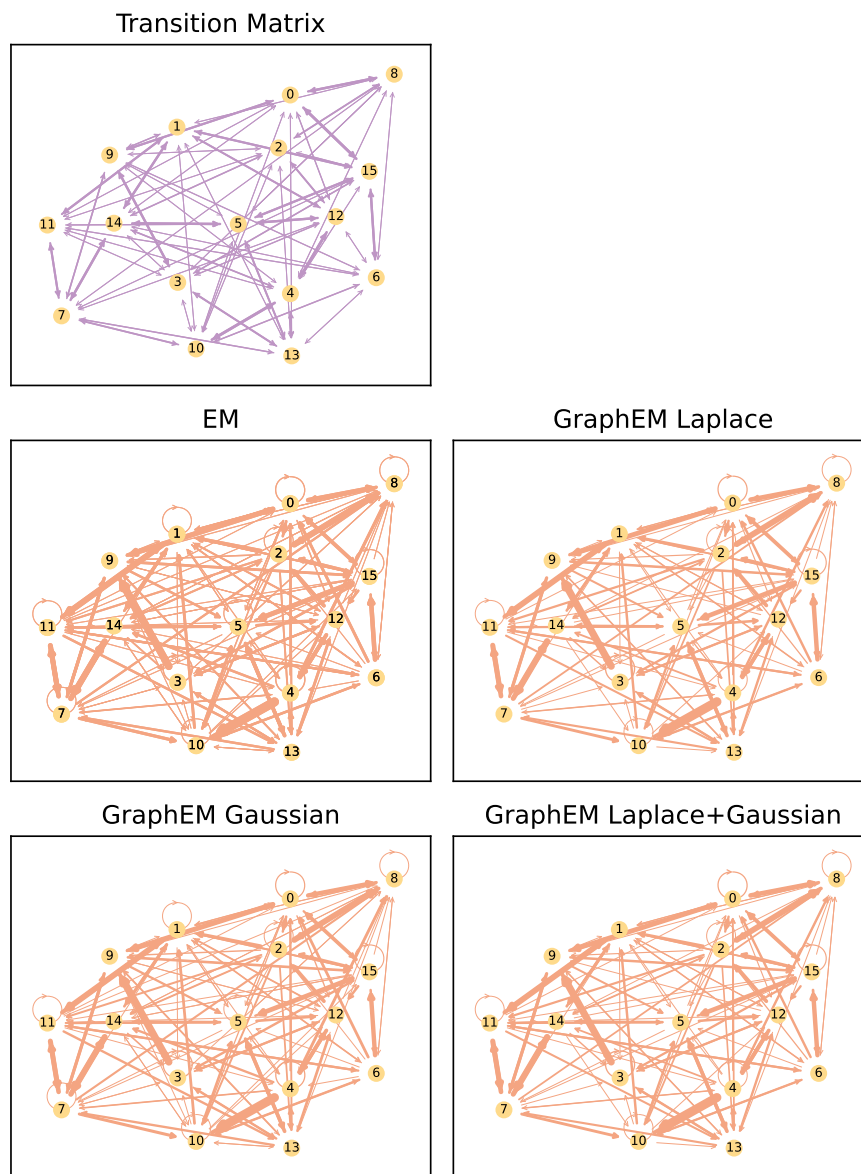


图 A-3 二部图的真实转移矩阵与估计矩阵对比图。

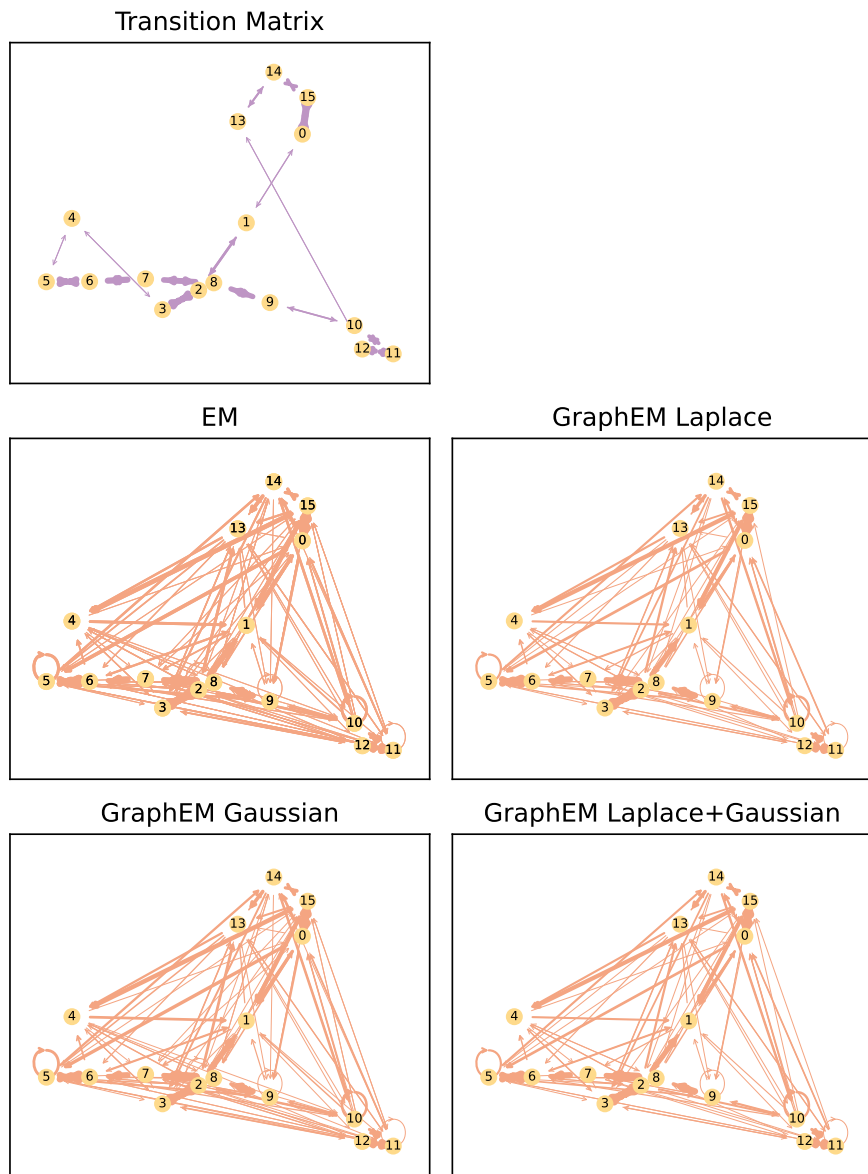


图 A-4 环图的真实转移矩阵与估计矩阵对比图。

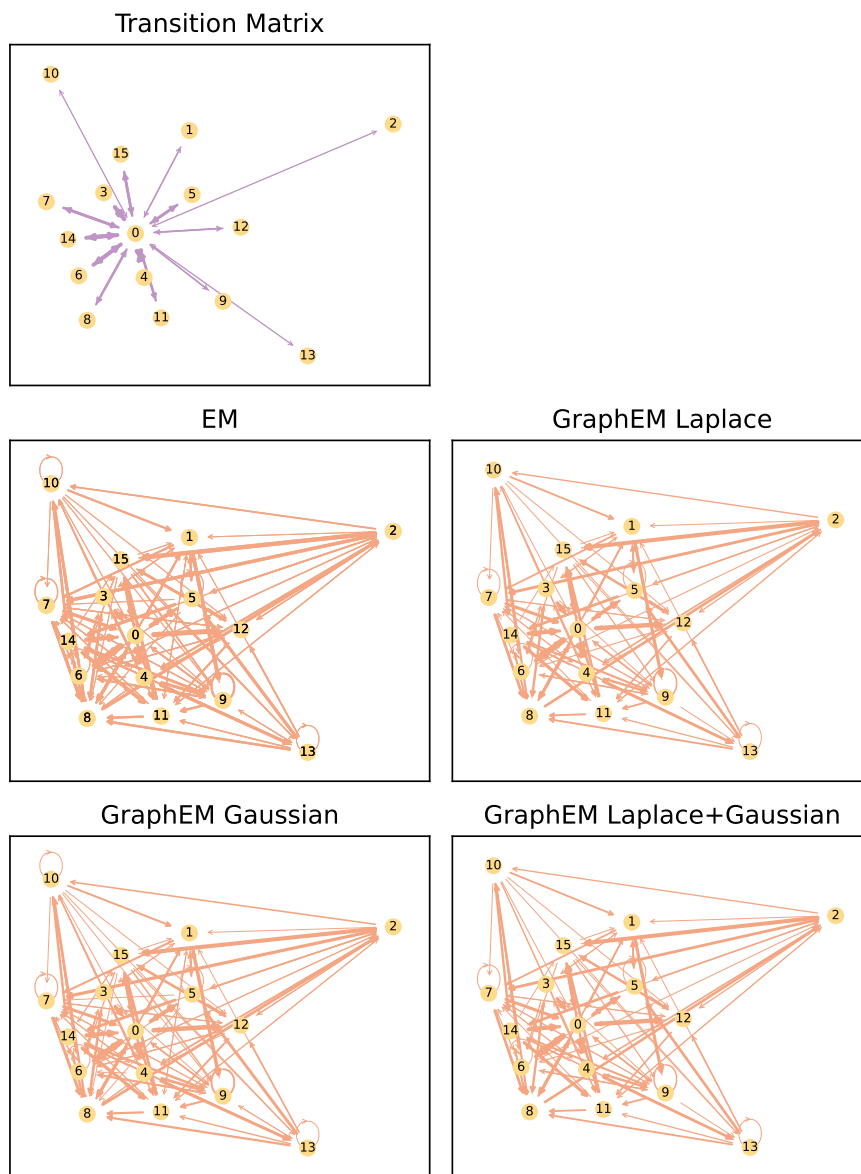


图 A-5 星型图的真实转移矩阵与估计矩阵对比图。