arXiv:2303.12569v1 [cs.CE] 22 Mar 2023

GraphIT：用于状态空间模型稀疏图推断的迭代重加权 ℓ_1 算法

´Emilie Chouzenoux[1] 和 V´ıctor Elvira[2]

[1] 巴黎-萨克雷大学、CentraleSup´elec、Inria、CVN，吉夫叙尔伊韦特，法国[2] 爱丁堡大学数学学院，英国

摘要

状态空间模型（SSM）是对多变量离散时间信号建模的常用工具。线性高斯状态空间模型（LG-SSM）应用广泛，因为在模型参数已知的情况下，它能够为推断提供闭式解。然而，在现实问题中，这些参数往往不可得，必须进行估计。促使这些参数具有稀疏性，既有助于提升可解释性，也能使推断易于处理。本文提出 GraphIT，这是一种主化-最小化（MM）算法，用于在稀疏先验下估计 LG-SSM 状态方程中的线性算子。我们提出了一类用途广泛的非凸正则化势函数。该 MM 方法借鉴了期望最大化方法和迭代重加权 l1 方法中的工具。具体而言，我们为目标函数推导出一个合适的凸上界，并使用近端分裂算法对其进行最小化。数值实验展示了所提出推断技术的优势。

索引词—状态空间模型、图推断、稀疏性、迭代重加权方案、MM 算法。

1. 引言

信号处理应用通常涉及处理离散时间多变量序列，以执行预测等统计任务。状态空间模型（SSM）将观测过程与随时间演化的隐状态联系起来。这种灵活的方法能够通过隐状态的动力学描述复杂现象。它还能够融入有关系统的物理知识，并提取处理数据时可推断出的相关结构。关于多变量时间序列的图模型已有大量文献，例如 [1–4], ，近年来也有不少研究将其应用于状态空间模型 [5, 6], ，其中还包括完全概率方法 [7, 8]。这类图表示有多种应用，例如生物学 [9, 10], 、社会网络分析 [11], 或神经科学 [12]。

状态空间模型中的图方法受益于稀疏性的促进作用，至少具有以下三项优势。首先，稀疏性增强了可解释性，例如，对于隐状态具有物理意义的模型而言，可解释性尤为重要。其次，稀疏性降低了参数的有效维数，从而缓解估计过程中的维数灾难。第三，还可以促进其他与稀疏性互补的性质，从而引入有关系统的有用先验知识，例如隐过程的稳定性和谱约束 [13]。

在 [5, 6], 中，我们介绍了 GraphEM，这是一种用于线性高斯（LG）状态空间模型（SSM）参数估计的期望最大化（EM）方法。具体而言，我们研究了如何估计转移矩阵这一具有挑战性的任务，该矩阵编码了多变量状态演化过程中的马尔可夫依赖。GraphEM 基于一种新颖的视角，将转移矩阵与有向图的邻接矩阵联系起来，从格兰杰因果关系的角度编码状态维度之间的因果关系。EM 过程中的 M 步采用近端分裂，并具有可靠的收敛保证。因此，GraphEM 将成熟的 EM 技术扩展到用于 LG-SSM 参数最大似然估计 [14–17] 的最大后验估计量，并能恰当地纳入复杂的先验项。不过，它仅适用于凸正则项。在稀疏图推断领域，SCAD [18], 、MCP [19], 和 CEL0[20] 等非凸惩罚已被证明比其凸替代项 ℓ_1 范数（即 Lasso）更适合重建极稀疏图 [21–23]。这些惩罚可视为准范数 ℓ_0的连续非凸近似。然而，非凸性使得标准优化工具无法使用。一种高效的策略是采用迭代重加权（IR）方案 [24, 25], ，这类方案基于主化最小化（MM）原理 [26] ，将复杂的目标函数重构为一系列易于处理的上界。IRℓ_1 算法已广泛应用于多个信号处理问题 [27–32], ，但据我们所知，尚未在状态空间模型中得到研究。

本文提出一种名为 GraphIT 的新方法

V.E.感谢法国国家研究署通过 PISCES（ANR-17-CE40-0031-01）项目、Leverhulme 研究奖学金（RF-2021-593）以及美国陆军研究实验室/美国陆军研究办公室通过资助项目 W911NF-22-1-0235 提供的支持。E.C.感谢欧洲研究委员会通过 MAJORIS ERC-2019-STG-850925 启动基金提供的支持。

该方法将 EM 技术与强大的 IRℓ_1 策略相结合，用于估计状态空间模型中的模型参数。具体而言，对于一大类能够增强稀疏性的非凸先验，GraphIT 能够估计 LG-状态空间模型的转移矩阵。得益于 IRℓ_1 框架，GraphIT 算法的更新简单且开销低。我们在多个合成数据集上的实验结果表明，该方法能够提高稀疏转移矩阵的恢复率。

本文结构如下。第 2 节介绍模型、滤波/平滑算法，并重新讨论 GraphEM 框架。第 3 节介绍新颖的 GraphIT 算法。第 4 节给出数值结果，第 5 节则作总结。

2. 背景

2.1. 线性状态空间模型中的推断

我们考虑由下式描述的线性高斯状态空间模型

$$\begin{aligned} \mathbf{x}_k &= \mathbf{A}\mathbf{x}_{k-1} + \mathbf{q}_k, \\ \mathbf{y}_k &= \mathbf{H}\mathbf{x}_k + \mathbf{r}_k, \end{aligned} \tag{1}$$

对于 $k = 1, \ldots, K$，其中 $\mathbf{x}_k \in \mathbb{R}^{N_x}$ 和 $\mathbf{y}_k \in \mathbb{R}^{N_y}$ 分别表示时刻 k的隐状态和观测，$\mathrm{A} = \mathbb{R}^{N_x \times N_x}$、$\mathrm{H} = \mathbb{R}^{N_y \times N_x}$ 以及 $\{\mathrm{q}_k\}_{k=1}^K \sim \mathcal{N}(0, \mathrm{Q})$ 和$\{\mathrm{r}_k\}_{k=1}^K \sim \mathcal{N}(0, \mathrm{R})$ 分别表示独立同分布的状态噪声和观测噪声过程。初始状态的分布为 $\mathrm{x}_0 \sim \mathcal{N}(\mathrm{x}_0; \boldsymbol{\mu}_0, \boldsymbol{\Sigma}_0)$，其中 $\boldsymbol{\mu}_0$和 $\boldsymbol{\Sigma}_0$已知。

在状态空间模型中，滤波问题是计算所谓滤波分布 $p(\mathrm{x}_k|\mathrm{y}_{1:k})$，其中 $\mathrm{y}_{1:k}$ ，$\{\mathrm{y}_j\}_{j=1}^k$。对于式（1）中的线性高斯状态空间模型，卡尔曼滤波器（卡尔曼滤波）能够以序贯且高效的方式精确计算滤波分布，得到$p(\mathrm{x}_k|\mathrm{y}_{1:k}) = \mathcal{N}(\mathrm{x}_k|\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k)$；也就是说，对于每个 $k = 1, \ldots, K$，它都会计算各高斯滤波分布 [33]的均值 $\boldsymbol{\mu}_k$ 和协方差 $\boldsymbol{\Sigma}_k$。平滑问题旨在计算 $p(\mathrm{x}_k|\mathrm{y}_{1:K})$，即对整组观测值进行条件化后的后验概率密度函数（pdf），适用于所有 $k = 1, \ldots, K$。在线性高斯状态空间模型中，借助 Rauch-Tung-Striebel（RTS）平滑器，也可以用闭式表达式（同样为高斯 pdf）得到平滑 pdf：$p(\mathrm{x}_k|\mathrm{y}_{1:K}) = \mathcal{N}(\mathrm{x}_k|\boldsymbol{\mu}_k^s, \boldsymbol{\Sigma}_k^s)$。卡尔曼滤波和 RTS 平滑器都需要已知模型参数 $\boldsymbol{\mu}_0$, $\boldsymbol{\Sigma}_0$, A, H, Q 和 R。上述模型中模型参数的估计一直是研究的热点，例如近期的研究[34, 35] (综述见 [36, 第 12] 章)。

2.2. 使用 GraphEM 估计转移矩阵

GraphEM [5, 6] 是我们最近提出的一种基于 EM 的算法，用于在对数凹先验分布 $p(\mathrm{A})$下对 A 进行最大后验（MAP）估计。GraphEM 旨在找到 $\mathcal{L}(\mathrm{A}) \ \mathcal{L}_0(\mathrm{A}) + \mathcal{L}_{1:K}(\mathrm{A})$的最小值，其中 $\mathcal{L}_0(\mathrm{A}) \ -\log\ p(\mathrm{A})$，且 $\mathcal{L}_{1:K}(\mathrm{A}) \ -\log p(\mathrm{y}_{1:K}|\mathrm{A})$。在 GraphEM 中，矩阵 A 被解释为对状态维度之间有向图的编码。先验 $p(\mathrm{A})$可以编码该图的结构信息（例如稀疏性）。该图在状态空间模型的解释中发挥着重要作用，并与格兰杰因果模型相关（参见 [6]中的讨论）。与式（1）中的模型对应的 A 的似然为

$$\mathcal{L}_{1:K}(\mathbf{A}) = \sum_{k=1}^{K} \frac{1}{2} \log|2\pi\mathbf{S}_k| + \frac{1}{2}\mathbf{z}_k^\top \mathbf{S}_k^{-1}\mathbf{z}_k, \tag{2}$$

其中，$\mathrm{z}_k = \mathrm{y}_k - \mathrm{HA}\boldsymbol{\mu}_{k-1}$，$\mathrm{S}_k$ 是预测分布 $p(\mathrm{y}_k|\mathrm{y}_{1:k-1}, \mathrm{A}) = \mathcal{N}(\mathrm{y}_k; \mathrm{HA}\boldsymbol{\mu}_{k-1}, \mathrm{S}_k)$ 的协方差矩阵；二者均由针对给定 A 运行的卡尔曼滤波得到（见 [36, 第 4.3] 节）。由于 (i) $\mathcal{L}_{1:K}$ 的形式是隐式的，且 (ii) $\mathcal{L}_0$ 可能不可微，直接最小化 $\mathcal{L}$ 较为困难。GraphEM 方法采用 EM 策略，通过依次构造 $\mathcal{L}_{1:K}$的上界来推广 [16], 中的方法，从而得到一系列易于处理的内层问题。

考虑给定 $\mathrm{A}' \in \mathbb{R}^{N_x \times N_x}$ 时卡尔曼滤波器/平滑器的输出。对每个 $k \in \{1, \ldots, K\}$，记 $\mathrm{G}_k = \boldsymbol{\Sigma}_k(\mathrm{A}')^\top(\mathrm{A}'\boldsymbol{\Sigma}_k(\mathrm{A}')^\top + \mathrm{Q})^{-1}$，并令 $\boldsymbol{\Psi} = \sum_{k=1}^K \boldsymbol{\Sigma}_k^s + \boldsymbol{\mu}_k^s(\boldsymbol{\mu}_k^s)^\top$，$\boldsymbol{\Phi} = \sum_{k=1}^K \boldsymbol{\Sigma}_{k-1}^s + \boldsymbol{\mu}_{k-1}^s(\boldsymbol{\mu}_{k-1}^s)^\top$，以及 $\boldsymbol{\Delta} = \sum_{k=1}^K \boldsymbol{\Sigma}_k^s \mathrm{G}_{k-1}^\top + \boldsymbol{\mu}_k^s(\boldsymbol{\mu}_{k-1}^s)^\top$。于是，根据 [36,第 12], 章

$$\begin{aligned} &(\forall \mathbf{A} \in \mathbb{R}^{N_x \times N_x}) \quad \mathcal{L}_{1:K}(\mathbf{A}) \leq \\ &-\int p(\mathbf{x}_{0:K}|\mathbf{y}_{1:K}, \mathbf{A}') \log p(\mathbf{x}_{0:K}, \mathbf{y}_{1:K}|\mathbf{A}) \mathrm{d}\mathbf{x}_{0:\mathrm{K}} + \mathrm{ct}_{/\mathbf{A}} \\ &= \frac{1}{2}\mathrm{tr}\left(\mathbf{Q}^{-1}(\boldsymbol{\Sigma} - \boldsymbol{\Delta}\mathbf{A}^\top - \mathbf{A}\boldsymbol{\Delta}^\top + \mathbf{A}\boldsymbol{\Phi}\mathbf{A}^\top)\right) + \mathrm{ct}_{/\mathbf{A}}, \end{aligned} \tag{3}$$

其中，tr 表示迹算子，$\mathrm{ct}_{/\mathbf{A}}$ 是一个与 A 无关的常数项，使得当 $\mathrm{A} = \mathrm{A}'$ 时等式成立。GraphEM 算法交替执行 E 步（构造式 (3) 中的右侧项）和 M 步（使用近端分裂方案最小化该项与 $\mathcal{L}_0$ 之和）。对于拉普拉斯先验这一简单情形（即 ℓ_1 范数惩罚项），M 步中的最小化可通过 Douglas-Rachford 算法完成，从而得到 [5]，中更简洁的 GraphEM 形式；而 [6] 提出了一种更复杂的原始-对偶分裂技术，用于处理一般的凸惩罚项。[6] 讨论了所得 GraphEM 迭代的收敛保证。

3. 提出的 Graphit 方法

3.1. 所考虑的惩罚项类别

我们的研究聚焦于非凸稀疏化惩罚项，这类惩罚项已在例如 [37] 中得到研究（另见 [38]，其研究背景为稳健估计）。我们定义

$$\mathcal{L}_0(\mathbf{A}) = \sum_{1 \leq i,j \leq N_x} \rho(|A_{i,j}|), \tag{4}$$

Name	$\rho(\lvert u\rvert)$	$\rho'(\lvert u\rvert)$
对数和	$\gamma\lambda(\log(\lvert u\rvert+\lambda)\ -\log(\lambda))$，其中 $\lambda>0$	$\gamma\lambda(\lvert u\rvert+\lambda)^{-1}$
atan	$\frac{2}{\lambda}$反正切$(\lambda\lvert u\rvert)$，其中 $\lambda>0$	$\frac{\gamma}{1+\lambda^2u^2}$
Mangasarian	$\frac{2}{\lambda}(1-\exp(-\lambda\lvert u\rvert))$	$\gamma\exp(-\lambda\lvert u\rvert)$
MCP	$\begin{cases}\gamma\lvert u\rvert-\frac{u^2}{2\lambda} & \text{如果 } \lvert u\rvert\le\lambda\gamma \\ \frac{\lambda\gamma^2}{2} & \text{否则}\end{cases}$，其中 $\lambda>0$	$\begin{cases}\gamma-\frac{\lvert u\rvert}{\lambda} & \text{如果 } \lvert u\rvert\le\lambda\gamma \\ 0 & \text{否则}\end{cases}$
SCAD	$\begin{cases}\gamma\lvert u\rvert & \text{如果 } \lvert u\rvert\le\gamma \\ \frac{-2a\gamma\lvert u\rvert+u^2+\gamma^2}{2(a-1)} & \text{如果 } \gamma<\lvert u\rvert\le a\gamma\text{，其中 } a>2 \\ \frac{(a+1)\gamma^2}{2} & \text{否则}\end{cases}$	$\begin{cases}\gamma & \text{如果 } \lvert u\rvert\le\gamma \\ \frac{a\gamma-\lvert u\rvert}{a-1} & \text{如果 } \gamma<\lvert u\rvert\le a\gamma \\ 0 & \text{否则}\end{cases}$

表1：势函数及其导数的示例，其中超参数为 $\gamma>0$，使得 $\rho'(0^+)=\gamma$。

其中 $A_{i,j}$ 表示 A 的第 (i,j) 个元素，而 $\rho:\ \mathbb{R}\to[0,+\infty)$ 是一个满足以下条件的势函数：

1. ρ 连续，且在 $[0,+\infty)$上递增，其中$\rho(0)=0$；2. ρ 在$(0,+\infty)$上可微，其导数（有定义时）记为 ρ'；
3. ρ' 在$(0,+\infty)$上递减，且 $\lim_{u\to0^+}\rho'(u)=\gamma\in(0,+\infty)$。

一大类势函数满足上述要求。表1列出了一些相关示例。除 SCAD 外，在我们的示例中，已证明当 $\lambda\to0^+$时，$u\mapsto\rho(\lvert u\rvert)$上图收敛于离散势函数 $u\mapsto\delta_{u\neq0}$（输入值为零时等于0；否则等于1）。因此，式(4)中的先验可视为 $\gamma\ell_0$的近似。图1展示了部分势函数的图形，其中 $\gamma=1$。我们还展示了 ℓ_1和 ℓ_0 惩罚项以供比较。所考虑的先验类别是连续的，但仍是 ℓ_0的非凸近似。

3.2. IRℓ_1方法

对势函数 ρ 所作的假设带来了一个关键性质，该性质是所谓 IRℓ_1 优化方法 [39]的核心。设 $v\in\mathbb{R}$。那么，

$$(\forall u\in\mathbb{R})\quad \rho(\lvert u\rvert)\le\rho'(\lvert v\rvert)(\lvert u\rvert-\lvert v\rvert)+\rho(\lvert v\rvert),\tag{5}$$

其中等号在 $u=v$ 处成立。该结果的证明见 [29]。因此，对于式 (4)，在某个 $\mathrm{A}'\in\mathbb{R}^{N_x\times N_x}$ 处可得到如下凸上界：

$$\mathcal{L}_0(\mathbf{A})\le\|\mathbf{\Omega}(\mathbf{A}')\odot\mathbf{A}\|_1+\mathrm{ct}_{/\mathbf{A}},\tag{6}$$

其中 $\mathrm{ct}_{/\mathbf{A}}$ 满足在 $\mathrm{A}=\mathrm{A}'$处等号成立。上文中，$\Omega(\mathrm{A}')\in[0,+\infty)^{N_x\times N_x}$ 是一个权重矩阵，定义为 $\Omega_{i,j}=\rho'(\lvert A_{i,j}\rvert)$，适用于每个 (i,j)。此外，$\odot$ 是 Hadamard（即逐元素）乘积。表1给出了我们各个示例中 $\rho'(\lvert\cdot\rvert)$ 的表达式。IRℓ_1 技术旨在最小化 $\mathcal{L}_0+f$，其中f 是一个简单函数（通常为二次项）；它通过迭代求解可处理的最小化问题来实现

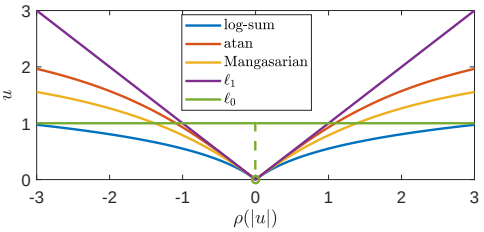

图1：函数 $\rho(\lvert\cdot\rvert)$的示例，其中 $\lambda=0.5$ 和 $\gamma=1$。

$\|\Omega(\mathrm{A}')\odot\cdot\|_1+f$。当 f 是最小二乘失配时，IRℓ_1 中的内层问题可以采用文献中针对 Lasso 惩罚 [40–42]下的惩罚最小二乘所提出的多种高效方法之一来求解。

3.3. GraphIT 算法

下面我们介绍 GraphIT 算法，其概要见算法 1。GraphIT 是一种 MM 算法，用于在我们的状态空间模型中搜索 A 的最大后验估计（即 $\mathcal{L}$ 的最小化解），其中采用式 (4) 所定义的先验。该算法采用迭代结构，交替执行主化和最小化步骤。难点在于构造主化近似，以及所需的用于对其进行最小化的优化算法。

算法 1 GraphIT 算法

初始 $\mathrm{A}^{(0)}$ 和惩罚参数。精度 $\varepsilon>0$。

对于 $i=1,\ 2,\ \ldots$

主化：

使用 $\mathrm{A}^{(i-1)}$ 运行卡尔曼滤波和 RTS 平滑器，并构造

$$(\forall\mathrm{A}\in\mathbb{R}^{N_x\times N_x})\quad \mathcal{Q}\left(\mathrm{A};\ \mathrm{A}^{(i-1)}\right)=\frac{1}{2}\mathrm{tr}\left(\mathrm{Q}^{-1}(\mathbf{\Sigma}-\mathbf{\Delta}\mathrm{A}^\top-\mathrm{A}\mathbf{\Delta}^\top+\mathrm{A}\mathbf{\Phi}\mathrm{A}^\top)\right)+\|\Omega(\mathrm{A}^{(i-1)})\odot\mathrm{A}\|_1.\tag{7}$$

最小化：

使用 DR 求解器求解 $\mathrm{A}^{(i)}=\mathrm{argmin}_{\mathrm{A}}\left(\mathcal{Q}(\mathrm{A};\mathrm{A}^{(i-1)})\right)$。

停止条件：

如果 $\|\mathrm{A}^{(i)}-\mathrm{A}^{(i-1)}\|_F\le\varepsilon\|\mathrm{A}^{(i-1)}\|_F$，则停止递归。

给定来自上一次迭代的估计 $\mathrm{A}^{(i-1)}$，我们构造式（7）中定义的特定主化函数 $\mathcal{Q}(\mathrm{A};\mathrm{A}^{(i-1)})$，该函数结合了（i）基于 EM 的主化，其中涉及使用转移矩阵 $\mathrm{A}^{(i-1)}$运行卡尔曼滤波器/平滑器所得的输出；以及（ii）式（4）中先验项的一个凸上界（来自 IRℓ_1方法）。利用（3）和（6），可以证明，对于每个 $\mathrm{A}\in\mathbb{R}^{N_x\times N_x}$ 和每次迭代 $i\in\mathbb{N}$，式（7）中的函数满足 $\mathcal{L}(\mathrm{A})\le\mathcal{Q}(\mathrm{A};\mathrm{A}^{(i-1)})+\mathrm{ct}_{/\mathbf{A}}$（再次说明，常数项的设定使得在 $\mathrm{A}=\mathrm{A}^{(i-1)}$处两者相等）。实际中，通过 Douglas-Rachford（DR）算法 [43]来最小化上述上界，并以当前迭代值 $\mathrm{A}^{(i-1)}$作为初始值。作为 MM 构造的优点，序列 $\{\mathcal{L}(\mathrm{A}^{(i)})\}_{i\in\mathbb{N}}$ 单调递减

收敛到有限极限的序列。

4. 实验评估

我们在根据模型（1）模拟生成的时间序列上展示所提出的 GraphIT 方法的性能，其中 $N_x = N_y \in \{8, 16\}$。随机生成稀疏矩阵 A，其支撑大小（即相关图中的边数）为 $S \in \{4, 8, 16\}$。对于六种情形中的每一种，我们随机选择 S 个索引，将 A 中对应元素设为独立同分布标准正态分布的一个实现值，并将 $N_x^2 - S$ 其他元素设为零。对 A 的非零元素进行缩放，使 A 的谱范数小于 1，从而保证马尔可夫过程的稳定性。最后，我们设 $\mathbf{H} = \mathrm{Id}, K = 10^3$，$\mathbf{Q} = \sigma_{\mathbf{Q}}^2\mathrm{Id}$，$\mathbf{R} = \sigma_{\mathbf{R}}^2\mathrm{Id}$，$\mathbf{\Sigma}_0 = \sigma_0^2\mathrm{Id}$，其中 $(\sigma_{\mathbf{Q}}, \sigma_{\mathbf{R}}, \sigma_0) = (10^{-1}, 10^{-1}, 10^{-4})$。

对于六个数据集中的每一个，我们都使用对数和惩罚项（表1，第一行）和精度 $\xi = 10^{-3}$ 运行 GraphIT（最多进行 50 次迭代）。我们还展示了使用 ℓ_1 惩罚项的 GraphEM，以及通过 EM 算法计算的最大似然估计量（MLEM） [15, 16]的结果。所有方法均以一个元素为 $A_{i,j} = (10^{-1})^{|i-j|}$ 的矩阵作为初始值，并将其重新缩放，使谱范数等于 0.99。我们使用矩阵 A 的相对均方误差（RMSE）、图边检测的准确率和 F1 分数来比较这些方法，其中对矩阵 A 的元素采用阈值 10^{-10}。所有结果均为 50 次实验的平均值。通过网格搜索策略调整 GraphIT 和 GraphEM 的正则化参数，以最小化单次实验中的 RMSE。

结果汇总于表 2。MLEM 未在 A 上加入稀疏先验，这解释了其表现不佳，尤其是在边可检测性方面。GraphEM 和 GraphIT 的得分都优于 MLEM。值得注意的是，图越稀疏（即 S 越小），GraphIT 的结果就越好。无论是均方根误差还是检测指标，GraphIT 在所有数据集上的表现都显著优于竞争方法。特别是，尽管 GraphEM 的估计器中集成了稀疏化 ℓ_1 惩罚项，但在极稀疏图上仍表现不佳，导致均方根误差上升、准确率下降。相比之下，GraphIT 在所有指标上都表现出色。当稀疏程度降低（即 S 增大）时，GraphEM 和 GraphIT 的表现趋于相似。表 2 的最后一列还列出了各方法在 50 次实验中的平均计算时间。所有方法的运行时间相近（代码在 Matlab 2021a 中运行，使用配备 32 GB 内存的第 11 代英特尔(R) Core(TM) i7-1185G7 3.00GHz 处理器）。

最后，我们在图 2 中展示两个图重建示例。我们比较两种设置下的真实图（左）、GraphEM（中）和 GraphIT（右）估计结果，即 $(N_x, S) = (8, 16)$（上）和 $(N_x, S) = (16, 16)$（下）。可以看出，在两个稀疏图的情况下，GraphIT 在恢复图的结构和权重方面表现出更强的能力。

表 2：GraphIT、GraphEM 和 MLEM 的结果与计算时间。

(N_x, S)	方法	均方根误差	准确率	F1	时间（秒）
(8,4)	GraphIT	**0.185**	**0.964**	**0.776**	1.523
	GraphEM	0.244	0.920	0.635	1.186
	MLEM	0.401	0.063	0.118	1.862
(8,8)	GraphIT	**0.161**	**0.976**	**0.901**	1.907
	GraphEM	0.214	0.936	0.784	3.467
	MLEM	0.345	0.125	0.222	1.774
(8,16)	GraphIT	**0.190**	**0.931**	**0.861**	1.663
	GraphEM	0.194	0.865	0.772	3.506
	MLEM	0.248	0.250	0.400	2.040
(16,4)	GraphIT	**0.234**	**0.990**	**0.749**	4.218
	GraphEM	0.325	0.928	0.302	4.185
	MLEM	0.792	0.016	0.031	4.741
(16,8)	GraphIT	**0.257**	**0.987**	**0.808**	4.005
	GraphEM	0.322	0.907	0.403	3.916
	MLEM	0.650	0.031	0.061	3.964
(16,16)	GraphIT	**0.350**	**0.959**	**0.606**	4.393
	GraphEM	0.362	0.905	0.494	5.572
	MLEM	0.710	0.063	0.118	3.129

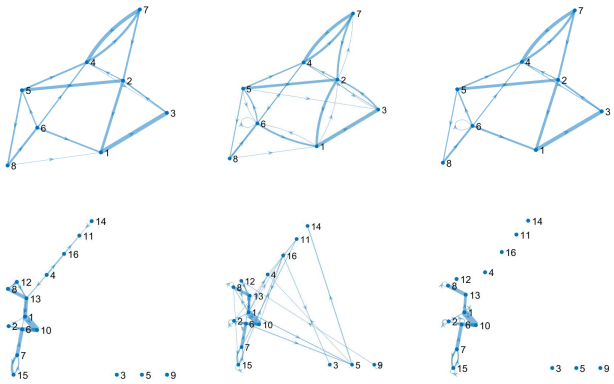

图 2：真实图（左）、GraphEM（中）和 GraphIT（右）对 $(N_x, S) = (8, 16)$（上）及 $(N_x, S) = (16, 16)$（下）的估计。

5. 结论

本文提出了 GraphIT，这是一种 MM 算法，用于估计线性高斯状态空间模型中的稀疏转移矩阵，以描述隐状态之间的相互作用。我们考虑了一类非凸正则化项，以促使所求矩阵具有稀疏性；这里将其解释为与有向图的邻接矩阵相关。该新方法利用了 EM框架和迭代重加权方法所固有的主化性质。由此得到的凸上界可通过近端分裂求解器高效地最小化。通过受控场景下的数值结果，我们展示了该方法在准确恢复稀疏转移矩阵方面的强大能力。

参考文献

[1] M. Eichler，《多元时间序列的图模型》，《概率论及相关领域》，第 153 卷，第 1 期，第 233–268 页，2012年 6 月。

[2] F. R. Bach 和 M. I. Jordan，“平稳时间序列的图模型学习”，IEEE信号处理汇刊，第 52 卷，第 8 期，第 2189–2199 页，2004 年 8 月。

[3] D. Barber 和 A. T. Cemgil，“时间序列的图模型”，IEEE 信号处理杂志，第 27 卷，第 6 期，第 18–28 页，2010 年 11 月。

[4] V.N. Ioannidis、Y. Shen 和 G. B. Giannakis，“动态图上的拓扑结构与动态过程的半盲推断”，IEEE信号处理汇刊，第 67 卷，第 9 期，第 2263–2274 页，2019 年。

[5] E. Chouzenoux 和 V. Elvira，“GraphEM：盲卡尔曼滤波器的 EM 算法 - 在图稀疏性约束下进行”，载于第 45 届 IEEE 国际声学、语音与信号处理会议论文集（ICASSP 2020），2020 年 5 月 4–8 日，第 5840–5844 页。

[6] V. Elvira 和 E. Chouzenoux，“线性高斯状态空间模型中的图推断”，IEEE信号处理汇刊，第 70 卷，第 4757–4771 页，2022 年。

[7] B. Cox 和 V. Elvira，“稀疏线性模型中的参数估计 -高斯状态 - 空间模型中的可逆跳跃马尔可夫链蒙特卡罗方法，”载于第30届欧洲信号处理会议（EUSIPCO 2022）论文集。IEEE，2022年，第797–801页。

[8] B. Cox 和 V. Elvira，“线性模型参数的稀疏贝叶斯估计 - 高斯状态空间模型，”IEEE信号处理汇刊（即将发表），2023年。

[9] A. Pirayre、C. Couprie、L. Duval 和 J.-C. Pesquet，《BRANE Clust：基于聚类的基因调控网络推断细化》，IEEE/ACM计算生物学与生物信息学汇刊，第15卷，第3期，第850–860页，2018年5月。

[10] D. Luengo, G. Rios-Munoz, V. Elvira, C. Sanchez, and A. Artes-Rodriguez，“Hi - 用于心房颤动心腔内电图因果关系检索的分层算法》，《IEEE 生物医学与健康信息学汇刊》，第 12 卷，第 1 期，第 143–155 页，2019 年 1 月。

[11] C. Ravazzi、R. Tempo 和 F. Dabbene，《在稀疏社交网络中学习影响结构》，《IEEE 网络系统控制汇刊》，第 PP 卷，第 1–1 页，2017年12月。

[12] J. Richiardi、S. Achard、B. Horst、和 D. V. D. Ville，《利用脑图进行机器学习》，《IEEE 信号处理杂志》，第 30 卷，第 3 期，第 58–70 页，2013。

[13] J. Chiu、Y. Deng 和 A. M. Rush，《结构化模型中快速推断的低秩约束》，载于《神经信息处理系统进展》，A. Beygelzimer、Y. Dauphin、P. Liang 和 J. W. Vaughan 编，2021。

[14] S. Roweis 和 Z. Ghahramani，《线性高斯模型的统一综述》，《Neural Computation》，第 11 卷，第 305–345 页，1999。

[15] Z. Ghahramani 和 G. Hinton，《线性模型的参数估计 - 耳动态系统”，技术报告，多伦多大学，1996，http://mlg.eng.cam.ac.uk/zoubin/course04/tr-96-2.pdf。

[16] R. H. Shumway 和 D. S. Stoffer，“使用 EM 算法进行时间序列平滑和预测的方法”，《时间序列分析杂志》，第 3 卷，第 4 期，第 253–264 页，1982。

[17] O. Cappe、E. Moulines 和 T. Ridden，《隐马尔可夫模型中的推断》，Springer统计学丛书。Springer，纽约，NY，第 1 版，2005。

[18] J. Fan 和 R. Li，“通过非凹惩罚似然进行变量选择及其预言机性质”，美国统计学会期刊，第 96 卷，第 1348–1360 页，2001。

[19] C.-H. Zhang，《极小极大凹惩罚下近乎无偏的变量选择》，《统计年鉴》，第 894–942 页，2010。

[20] E.Soubies、L. Blanc-F´eraud 和 G. Aubert，《用于最小二乘正则化问题的连续精确 l0 惩罚（CEL0）》，SIAM 成像科学期刊，第 3 期，第 1574–1606 页，2015年。

[21] J. Ying、J. de Miranda Cardoso 和 D. P. Palomar，《拉普拉斯约束图模型下的非凸稀疏图学习》，载于《神经信息处理系统进展》，2020 年。

[22] J. de Miranda Cardoso, J. Ying 和 D. Palomar，《重尾图模型 - 尾部市场”，载于《神经信息处理系统进展》。2021，卷34，第19989–20001页，Curran Associates, Inc.[23] D. Williams，“超越 Lasso：高斯图模型中非凸正则化综述”，技术报告，2020，https://psyarxiv.com/ad57p/.[24] D. Wipf 和 S. Nagarajan，“用于寻找稀疏解的迭代重加权 ℓ_1 和 ℓ_2 方法”，IEEE 信号处理精选主题期刊，卷4，第2期，第317–329页，2010.[25] P. Huber，《稳健统计》，Wiley，第2版，2009。

[26] Y. Sun, P. Babu, and D. Palomar，“通信和机器学习中的最小化算法”，IEEE信号处理汇刊，卷65，第3期，第794–816页，2017年。

[27] I. Daubechies、R. DeVore、M. Fornasier 和 C. Gunturk，《用于稀疏恢复的迭代重加权最小二乘最小化》，《纯粹与应用数学通讯》，第 63 卷，第 1 期，第 1–38 页，2009。

[28] R. Chartrand 和 V. Wotao，《用于压缩感知的迭代重加权算法》，载于 IEEE声学、语音与信号处理国际会议论文集（ICASSP 2008），2008年，第 3869–3872 页。

[29] P. Ochs、A. Dosovitskiy、T. Brox 和 T. Pock，《关于计算机视觉中用于非光滑非凸优化的迭代重加权算法》，SIAM 影像科学杂志，第 8 卷，第 1 期，第 331–372 页，2015年。

[30] R. E. Carrillo、J. D. McEwen 和 Y. Wiaux，《稀疏性平均重加权分析（SARA）：一种用于射电干涉成像的新算法》，《皇家天文学会月刊》，第 426 卷，2012年10月，第 1223–1234 页。

[31] Q. Sun, K. M. Tan, H. Liu, and T. Zhang，《通过自适应凸松弛进行图形非凸优化》，载于《国际机器学习大会论文集》（ICML 2018），2018年。

[32] J. Ying 和 J. Pesquet，《在稀疏的非凸图模型下学习稀疏图》，技术报告，2020，https://arxiv.org/abs/2006.14925。

[33] R. E. Kalman，《线性滤波与预测问题的新方法》，《基础工程学杂志》，第 - 82卷，第35–45页，1960年。

[34] S. Sharma、V. Elvira、E. Chouzenoux 和 A. Majumdar，《面向状态空间模型的递归字典学习及其在股票预测中的应用》，《神经计算》，第450卷，第1–13页，2021。

[35] S. Sharma, A. Majumdar, V. Elvira, E. Chouzenoux，“用于短期负荷预测的盲卡尔曼滤波”，IEEE 电力系统汇刊，第 35 卷，第 6 期，第 4916–4919 页，2020。

[36] S. Sarkka，《贝叶斯滤波与平滑》，第 3 版，2013.

[37] G. Gasso, A. Rakotomamonjy 和 S. Canu，“利用一类非凸惩罚和直流规划恢复稀疏信号”，IEEE信号处理汇刊，第 57 卷，第 12 期，第 4686–4698 页，2009。

[38] X. Pan、Q. Sun 和 W.-X. Zhou，“迭代重加权 ℓ_1 惩罚稳健回归”，电子统计学期刊，第 15 卷，第 1 期，第 3287–3348 页，2021。

[39] H. Zou 和 R. Li，“非凹惩罚似然模型中的单步稀疏估计”，统计学年鉴，第 36 卷，第 1509–1533 页，2008年。

[40] P. L. Combettes 和 J.-C. Pesquet，“数据科学中的不动点策略”，IEEE信号处理汇刊，第 69 卷，第 3878–3905 页，2021。

[41] I. Daubechies, M. Defrise 和 C. De Mol，“带稀疏约束的线性逆问题迭代阈值算法”，《纯粹与应用数学通讯》，第57卷第11期，第1413–1457页，2004年。

[42] A. Beck 和 M. Teboulle，“一种快速迭代影像科学杂志，第2卷，第1期，第183–202页，2009年。

[43] P. Combettes 和 J. Pesquet，“一种 Douglas -拉赫福德分裂方法，用于求解非光滑 - 平滑凸变分信号恢复》，IEEE 信号处理精选主题期刊，第 1 卷，第 4 期，第 564–574 页，2007年12月。