GRAPHIT: Iterative reweighted l1 algorithm for sparse graph inference in state-space models

Emilie Chouzenoux, Víctor Elvira

GRAPHIT: ITERATIVE REWEIGHTED ℓ_1 ALGORITHM FOR SPARSE GRAPH INFERENCE IN STATE-SPACE MODELS

Émilie Chouzenoux[1] *and Víctor Elvira*[2]

[1] Université Paris-Saclay, CentraleSupélec, Inria, CVN, Gif-sur-Yvette, France
[2] School of Mathematics, University of Edinburgh, United Kingdom

ABSTRACT

State-space models (SSMs) are a common tool for modeling multi-variate discrete-time signals. The linear-Gaussian (LG) SSM is widely applied as it allows for a closed-form solution at inference, if the model parameters are known. However, they are rarely available in real-world problems and must be estimated. Promoting sparsity of these parameters favours both interpretability and tractable inference. In this work, we propose GraphIT, a majorization-minimization (MM) algorithm for estimating the linear operator in the state equation of an LG-SSM under sparse prior. A versatile family of non-convex regularization potentials is proposed. The MM method relies on tools inherited from the expectation-maximization methodology and the iterated reweighted-l1 approach. In particular, we derive a suitable convex upper bound for the objective function, that we then minimize using a proximal splitting algorithm. Numerical experiments illustrate the benefits of the proposed inference technique.

***Index Terms*—** State-space model, graphical inference, sparsity, iterative reweighted scheme, MM algorithm.

1. INTRODUCTION

Signal processing applications often involve dealing with discrete-time multi-variate series that need to be processed to perform statistical tasks such as prediction. State-space models (SSMs) relate the observation process with a hidden state that evolves over time. This versatile approach allows to describe complex phenomena through the dynamics of the hidden state. It also allows for the incorporation physical knowledge about the system and to retrieve relevant structure that can be inferred when processing the data. There exists abundant literature in graphical models for multi-variate time series, e.g., [1–4], and more recently also within SSMs [5, 6], including also fully probabilistic approaches [7, 8]. There are multiple applications of such graphical representations, e.g., in biology [9, 10], social network analysis [11], or neuroscience [12].

The graphical approach in SSMs benefits from the promotion of sparsity with, at least, the following three advantages. First, sparsity enhances interpretability, which is highly desired for instance in models where the hidden state has a physical interpretation. Second, it reduces the effective dimension of the parameters, alleviating the curse of dimensionality during estimation. Third, it is possible to promote other properties that are complementary with sparsity, allowing to introduce useful prior knowledge on the system, e.g., stability of the hidden process and spectral constraints [13].

In [5, 6], we introduced GraphEM, an expectation-maximization (EM) method for parameter estimation in linear Gaussian (LG) SSMs. Specifically, we addressed the challenging task of estimating the transition matrix that encodes the Markovian dependencies in the evolution of the multi-variate state. GraphEM is based on a novel perspective that relates the transition matrix to the adjacency matrix of a directed graph, encoding causal relationship among state dimensions in the Granger-causality sense. The M-step in the EM procedure relies on a proximal splitting, benefiting from sound convergence guarantees. GraphEM thus extends the well-established EM technique for maximum likelihood estimation of LG-SSM parameters [14–17] into a maximum a posteriori estimator that suitably encompasses sophisticated prior terms. However, it is limited to convex regularizers. In the literature of sparse graph inference, non-convex penalties such as SCAD [18], MCP [19], CEL0 [20] have shown to be more suited than their convex surrogate ℓ_1 norm (i.e., Lasso) to properly reconstruct very sparse graphs [21–23]. These penalties can be viewed as continuous non-convex approximations to the quasi-norm ℓ_0. However, non-convexity prevents the use of standard optimization tools. An efficient strategy is to resort to iterative reweighted (IR) schemes [24, 25], which are based on the majorization-minimization (MM) principle [26] to recast the complicated objective function into a sequence of tractable upper bounds. IRℓ_1 algorithms have been widely applied to several signal processing problems [27–32], but, to our knowledge, remain unstudied within SSMs.

In this paper, we propose a novel method called GraphIT

V.E. acknowledges support from the Agence Nationale de la Recherche of France under PISCES (ANR-17-CE40-0031-01) project, the Leverhulme Research Fellowship (RF-2021-593), and by ARL/ARO under grant W911NF-22-1-0235. E.C. acknowledges support from the European Research Council under Starting Grant MAJORIS ERC-2019-STG-850925.

that combines the EM technique with the powerful IRℓ_1 strategy for the estimation of model parameters in SSMs. More precisely, GraphIT is able to provide an estimation of the LG-SSM transition matrix for a wide class of non-convex sparsity enhancing priors. Thanks to the IRℓ_1 framework, GraphIT algorithm benefits from simple and not expensive updates. Experimental results on several synthetic datasets allow us to assess the ability of our method for improving the recovery rate for sparse transition matrices.

The paper is structured as follows. In Section 2, we describe the model, the filtering/smoothing algorithms, and revisit the GraphEM framework. The novel GraphIT algorithm is presented in Section 3. The paper ends with numerical results in Section 4 and some concluding remarks in Section 5.

2. BACKGROUND

2.1. Inference in linear state-space model

We consider the LG-SSM described by

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

for $k = 1, \ldots, K$, where $\mathbf{x}_k \in \mathbb{R}^{N_x}$ and $\mathbf{y}_k \in \mathbb{R}^{N_y}$ are the hidden state and observation at time k, respectively, $\mathbf{A} = \mathbb{R}^{N_x \times N_x}$, $\mathbf{H} = \mathbb{R}^{N_y \times N_x}$, and $\{\mathbf{q}_k\}_{k=1}^K \sim \mathcal{N}(0, \mathbf{Q})$ and $\{\mathbf{r}_k\}_{k=1}^K \sim \mathcal{N}(0, \mathbf{R})$ the i.i.d. state and observation noise processes, respectively. The distribution of the initial state is $\mathbf{x}_0 \sim \mathcal{N}(\mathbf{x}_0; \boldsymbol{\mu}_0, \boldsymbol{\Sigma}_0)$ with known $\boldsymbol{\mu}_0$ and $\boldsymbol{\Sigma}_0$.

In SSMs, the filtering problem consists in the computation of the so-called filtering distributions, $p(\mathbf{x}_k|\mathbf{y}_{1:k})$, where $\mathbf{y}_{1:k} \triangleq \{\mathbf{y}_j\}_{j=1}^k$. In the LG-SSM of Eq. (1), the Kalman filter (KF) exactly computes the filtering distributions in a sequential and efficient manner, yielding $p(\mathbf{x}_k|\mathbf{y}_{1:k}) = \mathcal{N}(\mathbf{x}_k|\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k)$, i.e., for every $k = 1, \ldots, K$, it computes the mean $\boldsymbol{\mu}_k$ and covariance $\boldsymbol{\Sigma}_k$ of each Gaussian filtering distribution [33]. The smoothing problem aims at computing $p(\mathbf{x}_k|\mathbf{y}_{1:K})$, i.e., the posterior probability density functions (pdfs) conditioning on the whole set of observations instead, for all $k = 1, \ldots, K$. In the LG-SSM, the smoothing pdfs are also available in a closed-form expression (again, Gaussian pdfs) by the Rauch-Tung-Striebel (RTS) smoother, $p(\mathbf{x}_k|\mathbf{y}_{1:K}) = \mathcal{N}(\mathbf{x}_k|\boldsymbol{\mu}_k^s, \boldsymbol{\Sigma}_k^s)$. Both KF and RTS smoother require to know the model parameters $\boldsymbol{\mu}_0$, $\boldsymbol{\Sigma}_0$, $\mathbf{A}$, $\mathbf{H}$, $\mathbf{Q}$, and $\mathbf{R}$. The estimation of the model parameters in the model above has been a subject of intense research, e.g., recently [34, 35] (see [36, Ch. 12] for a review).

2.2. GraphEM to estimate the transition matrix

GraphEM [5, 6] is an EM-based algorithm that we recently introduced for the maximum a posteriori (MAP) estimation of $\mathbf{A}$ under log-concave prior distribution $p(\mathbf{A})$. GraphEM aims at finding the minimum of $\mathcal{L}(\mathbf{A}) \triangleq \mathcal{L}_0(\mathbf{A}) + \mathcal{L}_{1:K}(\mathbf{A})$ with $\mathcal{L}_0(\mathbf{A}) \triangleq -\log p(\mathbf{A})$ and $\mathcal{L}_{1:K}(\mathbf{A}) \triangleq -\log p(\mathbf{y}_{1:K}|\mathbf{A})$. The matrix $\mathbf{A}$ is interpreted in GraphEM to encode a directed graph among the state dimensions. Prior $p(\mathbf{A})$ can encode structural information on such graph (e.g., sparsity). This graph plays an important role in the SSM interpretation, with links to Granger-causality models (see a discussion in [6]). The likelihood of $\mathbf{A}$ associated to the model in Eq. (1) is

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

where $\mathbf{z}_k = \mathbf{y}_k - \mathbf{H}\mathbf{A}\boldsymbol{\mu}_{k-1}$ and $\mathbf{S}_k$ is the covariance matrix of the predictive distribution $p(\mathbf{y}_k|\mathbf{y}_{1:k-1}, \mathbf{A}) = \mathcal{N}(\mathbf{y}_k; \mathbf{H}\mathbf{A}\boldsymbol{\mu}_{k-1}, \mathbf{S}_k)$, both being obtained by the KF, run for a given $\mathbf{A}$ (see [36, Section 4.3]). The direct minimization of $\mathcal{L}$ is difficult due to (i) the implicit form of $\mathcal{L}_{1:K}$, and (ii) the possibly non-differentiability of $\mathcal{L}_0$. The GraphEM approach implements an EM strategy, generalizing the one in [16], by relying on successive upper bounds of $\mathcal{L}_{1:K}$, that leads to obtain a sequence of tractable inner problems.

Let us consider KF/RTS outputs for a given $\mathbf{A}' \in \mathbb{R}^{N_x \times N_x}$. Denote, for every $k \in \{1, \ldots, K\}$, $\mathbf{G}_k = \boldsymbol{\Sigma}_k(\mathbf{A}')^\top(\mathbf{A}'\boldsymbol{\Sigma}_k(\mathbf{A}')^\top + \mathbf{Q})^{-1}$, and set $\boldsymbol{\Psi} = \sum_{k=1}^K \boldsymbol{\Sigma}_k^s + \boldsymbol{\mu}_k^s(\boldsymbol{\mu}_k^s)^\top$, $\boldsymbol{\Phi} = \sum_{k=1}^K \boldsymbol{\Sigma}_{k-1}^s + \boldsymbol{\mu}_{k-1}^s(\boldsymbol{\mu}_{k-1}^s)^\top$, and $\boldsymbol{\Delta} = \sum_{k=1}^K \boldsymbol{\Sigma}_k^s\mathbf{G}_{k-1}^\top + \boldsymbol{\mu}_k^s(\boldsymbol{\mu}_{k-1}^s)^\top$. Then, according to [36, Chap. 12],

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

where tr is the trace operator, and $\mathrm{ct}_{/\mathbf{A}}$ is a constant term not depending on $\mathbf{A}$, such that equality holds at $\mathbf{A} = \mathbf{A}'$. The GraphEM algorithm iterates alternating between an E-step, building the right term in (3), and an M-step minimizing the sum of this term with $\mathcal{L}_0$, using a proximal splitting scheme. In the simple case of a Laplace prior (i.e., ℓ_1 norm penalty), the minimization in the M-step can be performed by a Douglas-Rachford algorithm, leading to the simpler form of GraphEM in [5], while a more sophisticated primal-dual splitting technique is proposed in [6] for dealing with a generic convex penalty. Convergence guarantees for the resulting GraphEM iteration are discussed in [6].

3. PROPOSED GRAPHIT APPROACH

3.1. Considered class of penalties

Our study focuses on the class of non-convex sparsifying penalties that have been studied for instance in [37] (see also [38] in the context of robust estimation). We define

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

Name	$\rho(\lvert u\rvert)$	$\rho'(\lvert u\rvert)$
log-sum	$\gamma\lambda(\log(\lvert u\rvert+\lambda)-\log(\lambda))$, with $\lambda>0$	$\gamma\lambda(\lvert u\rvert+\lambda)^{-1}$
atan	$\frac{\gamma}{\lambda}\operatorname{atan}(\lambda\lvert u\rvert)$, with $\lambda>0$	$\frac{\gamma}{1+\lambda^2u^2}$
Mangasarian	$\frac{\gamma}{\lambda}(1-\exp(-\lambda\lvert u\rvert))$	$\gamma\exp(-\lambda\lvert u\rvert)$
MCP	$\begin{cases}\gamma\lvert u\rvert-\frac{u^2}{2\lambda} & \text{if } \lvert u\rvert\leq\lambda\gamma \\ \frac{\lambda\gamma^2}{2} & \text{otherwise}\end{cases}$, with $\lambda>0$	$\begin{cases}\gamma-\frac{\lvert u\rvert}{\lambda} & \text{if } \lvert u\rvert\leq\lambda\gamma \\ 0 & \text{otherwise}\end{cases}$
SCAD	$\begin{cases}\gamma\lvert u\rvert & \text{if } \lvert u\rvert\leq\gamma \\ -\frac{\gamma^2-2a\gamma\lvert u\rvert+u^2}{2(a-1)} & \text{if } \gamma<\lvert u\rvert\leq a\gamma \text{, with } a>2 \\ \frac{(a+1)\gamma^2}{2} & \text{otherwise}\end{cases}$	$\begin{cases}\gamma & \text{if } \lvert u\rvert\leq\gamma \\ \gamma-\frac{2\lvert u\rvert-(2a-1)\gamma}{2(a-1)} & \text{if } \gamma<\lvert u\rvert\leq a\gamma \\ 0 & \text{otherwise}\end{cases}$

Table 1: Example of potential functions and their derivatives with hyper-parameter $\gamma>0$ such that $\rho'(0^+)=\gamma$.

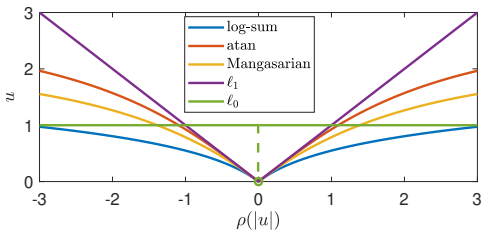

Fig. 1: Example of functions $\rho(|\cdot|)$, with $\lambda=0.5$ and $\gamma=1$.

where $A_{i,j}$ denotes the (i,j) entry of $\mathbf{A}$, and $\rho:\mathbb{R}\to[0,+\infty)$ is a potential satisfying:

1. ρ is continuous, and it is increasing on $[0,+\infty)$ with $\rho(0)=0$;
2. ρ is differentiable on $(0,+\infty)$, with derivative (when defined) denoted ρ';
3. ρ' is decreasing on $(0,+\infty)$ and $\lim_{u\to0^+}\rho'(u)=\gamma\in(0,+\infty)$.

A large class of potential functions satisfy the above requirements. Some relevant examples are listed in Table 1. Except for SCAD, in our examples, epi-convergence of $u\mapsto\rho(|u|)$ to the discrete potential $u\mapsto\delta_{u\neq0}$ (equals 0 for a zero-valued input; 1 otherwise) is established for $\lambda\to0^+$. Then, the prior in (4) can be seen as an approximation of $\gamma\ell_0$. Fig. 1 shows the plots for a subset of potentials, with $\gamma=1$. We also show ℓ_1 and ℓ_0 penalties for comparison. The considered class of priors is continuous, but still non-convex approximations of ℓ_0.

3.2. IRℓ_1 approach

The properties assumed on the potential ρ yield a key property, that is at the core of the so-called IRℓ_1 optimization approach [39]. Let $v\in\mathbb{R}$. Then,

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

where equality holds at $u=v$. A proof for this result can be found in [29]. The following convex upper bound is then obtained, for (4), at some $\mathbf{A}'\in\mathbb{R}^{N_x\times N_x}$:

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

where $\mathrm{ct}_{/\mathbf{A}}$ is such that equality holds at $\mathbf{A}=\mathbf{A}'$. Hereabove, $\mathbf{\Omega}(\mathbf{A}')\in[0,+\infty)^{N_x\times N_x}$ is a matrix of weights defined as $\Omega_{i,j}=\rho'(|A_{i,j}|)$ for every (i,j). Moreover, $\odot$ is the Hadamard (i.e., element-wise) product. We provide in Table 1 the expression for $\rho'(|\cdot|)$ for our examples. The IRℓ_1 technique amounts at minimizing $\mathcal{L}_0+f$, where f a simple function (typically a quadratic term), by iteratively solving the tractable problem of the minimization of $\|\mathbf{\Omega}(\mathbf{A}')\odot\cdot\|_1+f$. When f is a least square discrepancy, the inner problem in IRℓ_1 can be solved using one of the several efficient approaches proposed in the literature for penalized least squares under Lasso penalty [40–42].

3.3. GraphIT algorithm

In the following, we describe the GraphIT algorithm, which is summarized in Algorithm 1. GraphIT is an MM algorithm, to search for the MAP estimate of $\mathbf{A}$ within our SSM (i.e., the minimizer of $\mathcal{L}$), under the prior defined in (4). It has an iterative structure, which alternates over majorization and minimization steps. The difficulty lies in the construction of the majorization approximation and in the required optimization algorithm to minimize it.

Algorithm 1 GraphIT algorithm

Initial $\mathbf{A}^{(0)}$ and penalty parameters. Precision $\varepsilon>0$.
For $i=1,2,\ldots$
Majorization:
Run KF and RTS smoother using $\mathbf{A}^{(i-1)}$ and construct

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

Minimization:
Solve $\mathbf{A}^{(i)}=\operatorname{argmin}_{\mathbf{A}}\left(\mathcal{Q}(\mathbf{A};\mathbf{A}^{(i-1)})\right)$ using DR solver.
Stopping condition:
If $\|\mathbf{A}^{(i)}-\mathbf{A}^{(i-1)}\|_F\leq\varepsilon\|\mathbf{A}^{(i-1)}\|_F$, stop the recursion.

Given an estimation, $\mathbf{A}^{(i-1)}$, coming from the previous iteration, we construct a specific majorization function $\mathcal{Q}(\mathbf{A};\mathbf{A}^{(i-1)})$ defined in Eq. (7) that combines (i) an EM-based majorization involving outputs from running the KF/RTS with transition matrix $\mathbf{A}^{(i-1)}$, and (ii) a convex upper bound for our prior term in Eq. (4) (coming from IRℓ_1 approach). Using (3) and (6), we can show that the function in Eq. (7) is such that, for every $\mathbf{A}\in\mathbb{R}^{N_x\times N_x}$ and every iteration $i\in\mathbb{N}$, $\mathcal{L}(\mathbf{A})\leq\mathcal{Q}(\mathbf{A};\mathbf{A}^{(i-1)})+\mathrm{ct}_{/\mathbf{A}}$ (heragain, the constant term is set so as to have equality at $\mathbf{A}=\mathbf{A}^{(i-1)}$). In practice, the minimization of the above upper bound is done through the Douglas-Rachford (DR) algorithm [43] initialized at the current iterate $\mathbf{A}^{(i-1)}$. As a virtue of the MM construction, the sequence $\{\mathcal{L}(\mathbf{A}^{(i)})\}_{i\in\mathbb{N}}$ is a decreasing sequence converging to a finite limit.

4. EXPERIMENTAL EVALUATION

We illustrate the performance of the proposed GraphIT method on a time series simulated according to model (1), with $N_x = N_y \in \{8, 16\}$. Sparse matrices $\mathbf{A}$ are randomly generated, with support size (i.e., number of edges in the associated graph) $S \in \{4, 8, 16\}$. For each of the six scenarios, we randomly select S indexes for which we set the associated entry in $\mathbf{A}$ equals to a realization of an i.i.d. standard normal distribution and with the $N_x^2 - S$ other entries being set to zero. The non-zero entries of $\mathbf{A}$ are scaled so that the spectral norm of $\mathbf{A}$ is less than one to guarantee the stability of the Markov process. We finally set $\mathbf{H} = \mathbf{Id}$, $K = 10^3$ and $\mathbf{Q} = \sigma_{\mathbf{Q}}^2\mathbf{Id}$, $\mathbf{R} = \sigma_{\mathbf{R}}^2\mathbf{Id}$, $\mathbf{\Sigma}_0 = \sigma_0^2\mathbf{Id}$ with $(\sigma_{\mathbf{Q}}, \sigma_{\mathbf{R}}, \sigma_0) = (10^{-1}, 10^{-1}, 10^{-4})$.

For each of the six datasets, we ran GraphIT using log-sum penalty (Table 1, first row) and precision $\xi = 10^{-3}$ (with a maximum of 50 iterations). We also display the results of GraphEM using ℓ_1 penalty, and of the maximum likelihood estimator computed by an EM algorithm (MLEM) [15, 16]. All methods are initialized with a matrix with entries $A_{i,j} = (10^{-1})^{|i-j|}$, rescaled to have spectral norm equals to 0.99. The methods are compared in terms of relative mean square error (RMSE) on $\mathbf{A}$, accuracy and F1 scores for the graph edge detection, using a threshold of 10^{-10} on the entries of matrix $\mathbf{A}$. All results are averaged on 50 realizations. The regularization parameters of GraphIT and GraphEM are fine-tuned through a grid-search strategy to minimize the RMSE on a single realization.

The results are summarized in Table 2. MLEM does not include sparsity prior on $\mathbf{A}$, which explains its poor performance, especially in terms of edge detectability. GraphEM and GraphIT both reach better scores than MLEM. It is noticeable that, the sparser the graph (i.e., the smaller S), the better results for GraphIT. The superiority of GraphIT over its competitors on all datasets is remarkable both in terms of RMSE and detection metrics. In particular, despite integrating a sparsifying ℓ_1 penalty in its estimator, GraphEM struggles for very sparse graphs, leading to an increase in RMSE and a lowered accuracy. In contrast, GraphIT performs well for all metrics. When the level of sparsity decreases (i.e., S increases), both GraphEM and GraphIT tend to perform similarly. We also include in the last column of Table 2, the averaged computing times over 50 realizations for each methods. All methods take a similar running time (code is run in Matlab 2021a on an 11th Gen Intel(R) Core(TM) i7-1185G7 3.00GHz with 32 Go RAM).

We finally display two examples of graph reconstruction in Fig. 2. We compare the true graph (left), with GraphEM (middle) and GraphIT (right) estimates in both setups, i.e., $(N_x, S) = (8, 16)$ (top) and $(N_x, S) = (16, 16)$ (bottom). We can see the superior ability of GraphIT to recover the graphs shape and weights in the case of two sparse graphs.

Table 2: Results and computing times for GraphIT, GraphEM, and MLEM.

(N_x, S)	method	RMSE	accur.	F1	Time (s.)
(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

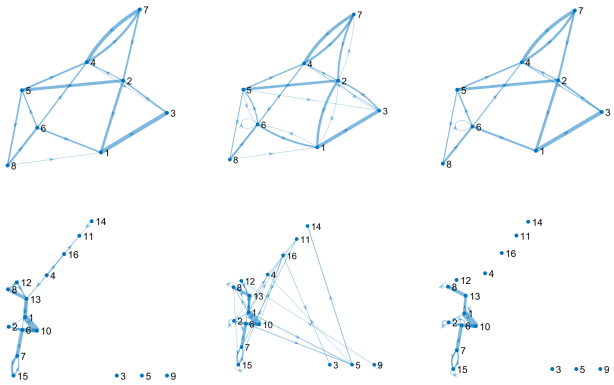

Fig. 2: True graph (left), GraphEM (middle), and GraphIT (right) estimates for $(N_x, S) = (8, 16)$ (top) and $(N_x, S) = (16, 16)$ (bottom).

5. CONCLUSION

In this paper, we have proposed GraphIT, an MM algorithm for the estimation of a sparse transition matrix describing hidden state interactions in an LG-SSM. A family of non-convex regularization term is considered to enforce the sparsity of the sought matrix, here interpreted as related to the adjacency matrix of a directed graph. The novel method exploits majorization properties inherited from both the EM framework and the iterated reweighted ℓ_1 methodology. The resulting convex upper bounds can be efficiently minimized through a proximal splitting solver. Through numerical results on controlled scenarios, we illustrate the great ability of the method to properly retrieve sparse transition matrices.

References

[1] M. Eichler, "Graphical modelling of multivariate time series," *Probability Theory and Related Fields*, vol. 153, no. 1, pp. 233–268, Jun. 2012.

[2] F. R. Bach and M. I. Jordan, "Learning graphical models for stationary time series," *IEEE Transactions on Signal Processing*, vol. 52, no. 8, pp. 2189–2199, Aug. 2004.

[3] D. Barber and A. T. Cemgil, "Graphical models for time-series," *IEEE Signal Processing Magazine*, vol. 27, no. 6, pp. 18–28, Nov 2010.

[4] V. N. Ioannidis, Y. Shen, and G. B. Giannakis, "Semi-blind inference of topologies and dynamical processes over dynamic graphs," *IEEE Transactions on Signal Processing*, vol. 67, no. 9, pp. 2263–2274, 2019.

[5] E. Chouzenoux and V. Elvira, "GraphEM: EM algorithm for blind Kalman filtering under graphical sparsity constraints," in *Proceedings of the 45th IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP 2020)*, 4–8 May 2020, pp. 5840–5844.

[6] V. Elvira and E. Chouzenoux, "Graphical inference in linear-Gaussian state-space models," *IEEE Transactions on Signal Processing*, vol. 70, pp. 4757–4771, 2022.

[7] B. Cox and V. Elvira, "Parameter estimation in sparse linear-Gaussian state-space models via reversible jump Markov Chain Monte Carlo," in *Proceedings of the 30th European Signal Processing Conference (EUSIPCO 2022)*. IEEE, 2022, pp. 797–801.

[8] B. Cox and V. Elvira, "Sparse Bayesian estimation of parameters in linear-gaussian state-space models," *IEEE Transactions on Signal Processing (to appear in)*, 2023.

[9] A. Pirayre, C. Couprie, L. Duval, and J.-C. Pesquet, "BRANE Clust: Cluster-assisted gene regulatory network inference refinement," *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, vol. 15, no. 3, pp. 850–860, May 2018.

[10] D. Luengo, G. Rios-Munoz, V. Elvira, C. Sanchez, and A. Artes-Rodriguez, "Hierarchical algorithms for causality retrieval in atrial fibrillation intracavitary electrograms," *IEEE Journal of Biomedical and Health Informatics*, vol. 12, no. 1, pp. 143–155, Jan. 2019.

[11] C. Ravazzi, R. Tempo, and F. Dabbene, "Learning influence structure in sparse social networks," *IEEE Transactions on Control of Network Systems*, vol. PP, pp. 1–1, 12 2017.

[12] J. Richiardi, S. Achard, B. Horst, , and D. V. D. Ville, "Machine learning with brain graphs," *IEEE Signal Processing Magazine*, vol. 30, no. 3, pp. 58–70, 2013.

[13] J. Chiu, Y. Deng, and A. M. Rush, "Low-rank constraints for fast inference in structured models," in *Advances in Neural Information Processing Systems*, A. Beygelzimer, Y. Dauphin, P. Liang, and J. W. Vaughan, Eds., 2021.

[14] S. Roweis and Z. Ghahramani, "A unifying review of linear Gaussian models," *Neural Computation*, vol. 11, pp. 305–345, 1999.

[15] Z. Ghahramani and G. Hinton, "Parameter estimation for linear dynamic systems," Tech. Rep., University of Toronto, 1996, http://mlg.eng.cam.ac.uk/zoubin/course04/tr-96-2.pdf.

[16] R. H. Shumway and D. S. Stoffer, "An approach to time series smoothing and forecasting using the EM algorithm," *Journal of Time Series Analysis*, vol. 3, no. 4, pp. 253–264, 1982.

[17] O. Cappe, E. Moulines, and T. Ridden, *Inference in Hidden Markov Models*, Springer Series in Statistics. Springer New York, NY, 1st edition, 2005.

[18] J. Fan and R. Li, "Variable selection via nonconcave penalized likelihood and its oracle properties," *Journal of the American Statistical Association*, vol. 96, pp. 1348–1360, 2001.

[19] C.-H. Zhang, "Nearly unbiased variable selection under minimax concave penalty," *The Annals of Statistics*, pp. 894–942, 2010.

[20] E. Soubies, L. Blanc-Féraud, and G. Aubert, "A continuous exact l0 penalty (cel0) for least-squares regularized problem," *SIAM Journal on Imaging Science*, , no. 3, pp. 1574–1606, 2015.

[21] J. Ying, J. de Miranda Cardoso, and D. P. Palomar, "Nonconvex sparse graph learning under Laplacian constrained graphical model," in *Advances in Neural Information Processing Systems*, 2020.

[22] J. de Miranda Cardoso, J. Ying, and D. Palomar, "Graphical models in heavy-tailed markets," in *Advances in Neural Information Processing Systems*. 2021, vol. 34, pp. 19989–20001, Curran Associates, Inc.

[23] D. Williams, "Beyond Lasso: a survey of nonconvex regularization in Gaussian graphical models," Tech. Rep., 2020, `https://psyarxiv.com/ad57p/`.

[24] D. Wipf and S. Nagarajan, "Iterative reweighted ℓ_1 and ℓ_2 methods for finding sparse solutions," *IEEE Journal of Selected Topics in Signal Processing*, vol. 4, no. 2, pp. 317–329, 2010.

[25] P. Huber, *Robust Statistics*, Wiley, 2nd edition, 2009.

[26] Y. Sun, P. Babu, and D. P. Palomar, "Majorization-minimization algorithms in signal processing, communications, and machine learning," *IEEE Transactions on Signal Processing*, vol. 65, no. 3, pp. 794–816, 2017.

[27] I. Daubechies, R. DeVore, M. Fornasier, and C. Gunturk, "Iteratively reweighted least squares minimization for sparse recovery," *Communications on Pure and Applied Mathematics*, vol. 63, no. 1, pp. 1–38, 2009.

[28] R. Chartrand and Y. Wotao, "Iteratively reweighted algorithms for compressive sensing," in *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP 2008)*, 2008, pp. 3869–3872.

[29] P. Ochs, A. Dosovitskiy, T. Brox, and T. Pock, "On iteratively reweighted algorithms for nonsmooth nonconvex optimization in computer vision," *SIAM Journal on Imaging Sciences*, vol. 8, no. 1, pp. 331–372, 2015.

[30] R. E. Carrillo, J. D. McEwen, and Y. Wiaux, "Sparsity Averaging Reweighted Analysis (SARA): a novel algorithm for radio-interferometric imaging," *Monthly Notices of the Royal Astronomical Society*, vol. 426, no. 2, pp. 1223–1234, 10 2012.

[31] Q. Sun, K. M. Tan, H. Liu, and T. Zhang, "Graphical nonconvex optimization via an adaptive convex relaxation," in *Proceedings of International Conference on Machine Learning (ICML 2018)*, 2018.

[32] J. Ying, J. V. d. M. Cardoso, and D. P. Palomar, "Does the ℓ_1-norm learn a sparse graph under Laplacian constrained graphical models?," Tech. Rep., 2020, https://arxiv.org/abs/2006.14925.

[33] R. E. Kalman, "A new approach to linear filtering and prediction problems," *Journal of Basic Engineering*, vol. 82, pp. 35–45, 1960.

[34] S. Sharma, V. Elvira, E. Chouzenoux, and A. Majumdar, "Recurrent dictionary learning for state-space models with an application in stock forecasting," *Neurocomputing*, vol. 450, pp. 1–13, 2021.

[35] S. Sharma, A. Majumdar, V. Elvira, and E. Chouzenoux, "Blind kalman filtering for short-term load forecasting," *IEEE Transactions on Power Systems*, vol. 35, no. 6, pp. 4916–4919, 2020.

[36] S. Sarkka, *Bayesian Filtering and Smoothing*, 3 edition, 2013.

[37] G. Gasso, A. Rakotomamonjy, and S. Canu, "Recovering sparse signals with a certain family of nonconvex penalties and dc programming," *IEEE Transactions on Signal Processing*, vol. 57, no. 12, pp. 4686–4698, 2009.

[38] X. Pan, Q. Sun, and W.-X. Zhou, "Iteratively reweighted ℓ_1-penalized robust regression," *Electronic Journal of Statistics*, vol. 15, no. 1, pp. 3287 – 3348, 2021.

[39] H. Zou and R. Li, "One-step sparse estimates in nonconcave penalized likelihood models," *Annals of Statistics*, vol. 36, pp. 1509–1533, 2008.

[40] P. L. Combettes and J.-C. Pesquet, "Fixed point strategies in data science," *IEEE Transactions on Signal Processing*, vol. 69, pp. 3878–3905, 2021.

[41] I. Daubechies, M. Defrise, and C. De Mol, "An iterative thresholding algorithm for linear inverse problems with a sparsity constraint," *Communications on Pure and Applied Mathematics*, vol. 57, no. 11, pp. 1413–1457, 2004.

[42] A. Beck and M. Teboulle, "A fast iterative shrinkage-thresholding algorithm for linear inverse problems," *SIAM Journal on Imaging Sciences*, vol. 2, no. 1, pp. 183–202, 2009.

[43] P. Combettes and J. Pesquet, "A Douglas-Rachford splitting approach to nonsmooth convex variational signal recovery," *IEEE Journal of Selected Topics in Signal Processing*, vol. 1, no. 4, pp. 564–574, Dec. 2007.