Decomposition of Covariance Matrix Using Cascade of Trees

Navid Tafaghodi Khajavi and Anthony Kuh
Deptartment of Electrical Engineering,
University of Hawaii at Manoa, Honolulu, HI 96822
Email: navidt@hawaii.edu, kuh@hawaii.edu

***Abstract*—We are looking at the statistical model approximation for jointly Gaussian random vectors. To do so, we are using a cascade of linear transformations that go beyond tree approximations. Here, we propose an algorithm which incorporates the Cholesky factorization method to compute the decomposition matrix and thus can approximate a simple graphical model using a cascade of the Cholesky factorization of the tree approximation transformations. The Cholesky decomposition keeps the sparsity pattern of the inverse decomposition and thus reduces computations for the tree structure linear transformation at each cascade stage of the algorithm. This is a different perspective on the approximation model, and algorithms such as Gaussian belief propagation can be used on this overall graph. We conclude with some simulation results.**

I. Introduction

Learning from high dimensional data in systems with complex models requires computational power which is not always available. In many cases, these complex systems and models are composed of multiple simpler models. Thus, a natural approach to understand the behavior of these systems is to decompose the system into simpler, yet informative model components. Doing so often translates into compromising the model accuracy versus its complexity. This paper specifically looks at Gaussian models and examines approximations based on the covariance matrix. The goal here is to approximate a decomposition of the covariance matrix using a cascade of simple and sparse models. The decomposition to simple components is useful for developing efficient estimation and inference algorithms for various applications. One such example is statistical model selection where given data, different algorithms impose different structure to model data [1].

Many engineering and computer science applications require using graphs to model dependencies between nodes of the graph. These applications include a diversity of areas from social networking to biomedical applications to transportation models to energy models. For these applications graphs must be approximated by simpler structures to reduce computational complexity. Here we briefly discuss an energy application. An electric distribution grid with measurements, distributed renewable energy sources, and decision making capabilities is referred to as the smart grid [2]. Smart grids often have large number of states (e.g. node voltages) making it computationally inefficient to gather the data and perform central state estimation in real-time fashion. Moreover, central state estimation requires many communication links between sensors on the distribution grid resulting in large costs. In contrast, distributed state estimators can give reasonably good estimates for large power grid systems in real-time while decreasing the number of necessary communications links. This fact causes a trade off between calculation time and accuracy of estimation. The distributed state estimation method over a factor graph based on loopy Gaussian BP proposed in [3] can perform in real-time fashion. To assure the convergence of loopy Gaussian BP algorithm, [3] considers simple models of the distributed renewable energy sources by approximating covariance matrices of the distributed renewable energy sources with simpler covariance matrices that have tree-like structures. However tree-like structures are poor approximates when the number of nodes is large [4] leading us to consider more complex graphical structures.

In this paper, we are generalizing a single-tree approximation algorithm using a sequence of tree approximations for sparse model approximation. The tree structure model is considered since this structure is simple and the optimal solution that minimizes the KL divergence can be easily computed using the Chow-Liu algorithm [5]. We use the Cholesky decomposition to factor the tree structured covariance matrix at each stage of the algorithm. We perform some simulations to confirm the results of this paper by looking at synthetic and real data and compare the performance of the proposed framework by comparing KL divergences. We also consider the singular value decomposition (SVD) and compare its performance to the Cholesky decomposition. Our simulation results also confirm the advantages of the cascade tree framework.

The rest of this paper is organized as follows. In section II we provide a summary of the Gaussian tree approximation. The Gaussian model approximation as a transformation is also discussed in this section. Section III presents the theory behind the proposed covariance matrix decomposition. This Section also outlines a greedy algorithm which is based on the symmetric CAM, the tree approximation algorithm and its Cholesky decomposition. Section IV provides some simulations over synthetic examples as well as a real solar data example from the island of Oahu obtained from NREL website data and investigates the quality of the proposed model approximation by using the KL divergence. Finally, Section V summarizes results of this paper.

Notation: Upper case and lower case letters denote random variables and their realizations, respectively; under-

lined letters stand for vectors; boldface upper case letters denote matrices; $(\cdot)^T$, $\text{tr}\{\cdot\}$ and $|\cdot|$ stand for transpose, trace and determinant of a matrix. In the rest of this paper, with shorthand notion, when we use the KL divergence between covariance matrices it means the KL divergence between their associated multivariate, zero-mean Gaussian distributions.

II. Gaussian tree approximation

In this section, we first review the tree approximation algorithm for Gaussian distributions. Then we explain the framework for the covariance transformation decomposition for any given model such as the tree model. The tree structure is a simple graphical model and can be computed efficiently. The loop-free structure of the tree structure also facilitates the implementation of distributed algorithms such as Gaussian BP. Later in the next section, we use the cascade of tree transformation decompositions to perform model approximation for graphical models.

A. Tree approximation for Gaussian distributions

In the tree approximation, we want to approximate a multivariate distribution by the product of lower order component distributions [6]. Let $\underline{X} \sim \mathcal{N}(\underline{0}, \mathbf{\Sigma})$ (i.e. jointly Gaussian with mean 0 and covariance matrix $\mathbf{\Sigma}$) where $\underline{X} \in \mathbb{R}^n$ have the graph representation $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ where sets $\mathcal{V}$ and $\mathcal{E}$ are the set of all vertices and edges of the graph representing $\underline{X}$.[1] Let $\underline{X}_\mathcal{T} \sim \mathcal{N}(\underline{0}, \widetilde{\mathbf{\Sigma}})$ have the graph representation $\mathcal{G}_\mathcal{T} = (\mathcal{V}, \mathcal{E}_\mathcal{T})$ where $\mathcal{E}_\mathcal{T}$ is a set of edges that represents a tree structure. The joint probability density function can be represented by joint pdfs of pairs of variables and marginal PDFs in the following convenient form

$$f_{\underline{X}_\mathcal{T}}(\underline{x}) = \prod_{(u,v)\in\mathcal{E}_\mathcal{T}} \frac{f_{\underline{X}^u,\underline{X}^v}(\underline{x}^u, \underline{x}^v)}{f_{\underline{X}^u}(\underline{x}^u) f_{\underline{X}^v}(\underline{x}^v)} \prod_{o\in\mathcal{V}} f_{\underline{X}^o}(\underline{x}^o). \quad (1)$$

Definition 1. *Let $\mathcal{T}_\mathbf{\Sigma}$ denote the set of all positive definite covariance matrices with following properties:*
1) These covariance matrices have tree structured Gaussian graphical models;
2) Picking any covariance matrix in this set, $\widetilde{\mathbf{\Sigma}} \in \mathcal{T}_\mathbf{\Sigma}$, the Gaussian distributions $\mathcal{N}(\underline{0}, \widetilde{\mathbf{\Sigma}})$ and $\mathcal{N}(\underline{0}, \mathbf{\Sigma})$, i) have the same marginal distributions, i.e. diagonal coefficients of $\widetilde{\mathbf{\Sigma}}$ and $\mathbf{\Sigma}$ are equal, and ii) have the same joint distribution of pairs of variables over the tree structured graph, $\mathcal{G}_\mathcal{T}$, i.e. covariances of $\widetilde{\mathbf{\Sigma}}$ and $\mathbf{\Sigma}$ over the tree are equal. ■

In the above definition $\mathcal{N}(\underline{0}, \widetilde{\mathbf{\Sigma}})$ obeys the product rule given in (1). Also, note that, the cardinality of the set $\mathcal{T}_\mathbf{\Sigma}$ is finite [7] since the number of all possible tree structured graphs with n nodes is finite.

Definition 2. *The KL divergence between two multivariate continuous distributions $p_{\underline{X}}(\underline{X})$ and $q_{\underline{X}}(\underline{X})$ is defined as*

$$\mathcal{D}\left(p_{\underline{X}}(\underline{x}) || q_{\underline{X}}(\underline{x})\right) = \int_\mathcal{X} p_{\underline{X}}(\underline{x}) \log \frac{p_{\underline{X}}(\underline{x})}{q_{\underline{X}}(\underline{x})}\, d\underline{x}$$

where $\mathcal{X}$ is the range of vector $\underline{x}$. ■

[1]Here, we assume a Markov random field where edges in the graph represents conditional dependencies given all other variables.

Chow-Liu MST method [5], was initially proposed for approximating the joint distribution of discrete variables by product of lower order distributions similar to (1) which involves no more than the pair of variables. The proposed KL divergence is used to quantify the distance between any distribution and its tree structure approximation.

The Chow-Liu MST algorithm for Gaussian distributions, minimizes the following optimization problem in order to find the optimal tree structured covariance matrix, $\mathbf{\Sigma}_\mathcal{T} \in \mathcal{T}_\mathbf{\Sigma}$

$$\mathbf{\Sigma}_\mathcal{T} = \underset{\widetilde{\mathbf{\Sigma}}\in\mathcal{T}_\mathbf{\Sigma}}{\arg\min}\ \mathcal{D}\left(f_{\underline{X}}(\underline{x}|\mathbf{\Sigma}) \middle|\middle| f_{\underline{X}_\mathcal{T}}(\underline{x}|\widetilde{\mathbf{\Sigma}})\right). \quad (2)$$

Here, $\mathcal{D}^\star \triangleq -\frac{1}{2}\log(|\mathbf{\Sigma}\mathbf{\Sigma}_\mathcal{T}^{-1}|)$ is the minimum possible KL divergence in (2) which is the distance between the given distribution and its optimal tree approximation. It is shown in [5] that the optimal solution for this problem 2 can be found efficiently using greedy algorithms [8], [9]. Their algorithm can be easily generalized for approximating the optimal tree structure of the joint distribution of Gaussian variables using equation (1) by adding edges one at a time [10]. In other words, given the knowledge of $\mathbf{\Sigma}$, the Chow-Liu algorithm can efficiently compute the optimal solution, i.e. $\mathbf{\Sigma}_\mathcal{T} = \text{chowliu}(\mathbf{\Sigma})$.

B. Gaussian model approximation as a transformation

Any zero-mean multivariate Gaussian distribution such as $\mathcal{N}(\underline{0}, \mathbf{\Sigma})$ where $\mathbf{\Sigma}$ is the covariance matrix, can be obtained through a linear transformation of the multivariate standard normal distribution, $\mathcal{N}(\underline{0}, \mathbf{I})$, where $\mathbf{I}$ is the identity matrix. Moreover, the decomposition matrix $\mathbf{C}$ is any square matrix that factors the covariance matrix $\mathbf{\Sigma}$ as $\mathbf{\Sigma} = \mathbf{C}\mathbf{C}^T$. In this scenario, the decomposition matrix $\mathbf{C}$ is also the transformation matrix. Let's assume that the desired model covariance matrix, $\mathbf{\Sigma}_\mathcal{M}$ and its decomposition matrix, $\mathbf{C}_\mathcal{M}$, i.e. $\mathbf{\Sigma}_\mathcal{M} = \mathbf{C}_\mathcal{M}\mathbf{C}_\mathcal{M}^T$, are given. Then, the model distribution $\mathcal{N}(\underline{0}, \mathbf{\Sigma}_\mathcal{M})$ is the transformation of the multivariate standard normal distribution. However, to generate the Gaussian distribution with covariance matrix, $\mathcal{N}(\underline{0}, \mathbf{\Sigma})$, using the model decomposition matrix, $\mathbf{C}_\mathcal{M}$, the input distribution has to have a certain covariance matrix, $\mathbf{\Delta}$. This covariance matrix is called the symmetric correlation approximation matrix and is defined as $\mathbf{\Delta} = \mathbf{C}_\mathcal{M}^{-1}\mathbf{\Sigma}\mathbf{C}_\mathcal{M}^{-T}$. We will give a formal definition for the symmetric CAM in section III where we consider cascade of tree approximation decompositions for graphical model approximation.

III. Cascade of Tree Decompositions Principle

In this section, we focus on the cascade of trees framework for model selection. We formulate the problem by considering the tree approximation as a transformation and we use multiple stages of these cascade trees to do model approximation.

Let $\mathbf{\Sigma} = \mathbf{C}\mathbf{C}^T$ and $\mathbf{\Sigma}_\mathcal{T} = \mathbf{C}_\mathcal{T}\mathbf{C}_\mathcal{T}^T$ be the tree approximation of $\mathbf{\Sigma}$ where decomposition matrices, $\mathbf{C}$ and $\mathbf{C}_\mathcal{T}$ are square-roots of the covariance matrices, $\mathbf{\Sigma}$ and $\mathbf{\Sigma}_\mathcal{T}$, respectively. Without loss of generality, in the rest of this paper, we look at the zero-mean Gaussian distributions with normalized

covariance matrix $\mathbf{\Sigma}$, i.e. covariance and correlation matrices are the same. Factoring covariances enable us to look at the problem as a transformation, as shown in figure 1.

Definition 3. *The symmetric correlation approximation matrix (CAM) for the tree approximation model is defined as* $\mathbf{\Delta} \triangleq \mathbf{C}_{\mathcal{T}}^{-1}\,\mathbf{\Sigma}\,\mathbf{C}_{\mathcal{T}}^{-T}$ *where* $\mathbf{C}_{\mathcal{T}}$ *is a square-root of* $\mathbf{\Sigma}_{\mathcal{T}}$. ■

The symmetric CAM for each step of the cascade tree algorithm is also defined using the transformation matrix $\mathbf{C}_{\mathcal{T}_i}$ and the previous step symmetric CAM.

Definition 4. *The symmetric correlation approximation matrix for the* i*-th step of the cascade tree approximation is defined as* $\mathbf{\Delta}_i \triangleq \mathbf{C}_{\mathcal{T}_i}^{-1}\mathbf{\Delta}_{i-1}\mathbf{C}_{\mathcal{T}_i}^{-T}$ *where* $\mathbf{\Delta}_0 \triangleq \mathbf{\Sigma}$, $\mathbf{\Sigma}_{\mathcal{T}_i} = \textit{chowliu}(\mathbf{\Delta}_{i-1})$ *and* $\mathbf{\Sigma}_{\mathcal{T}_i} \triangleq \mathbf{C}_{\mathcal{T}_i}\mathbf{C}_{\mathcal{T}_i}^{T}$ *where* $\mathbf{C}_{\mathcal{T}_i}$ *is a square-root of the* i*-the step covariance matrix,* $\mathbf{\Sigma}_{\mathcal{T}_i}$. ■

The construction is done by finding a tree approximation and then using a linear transformation applied to $\underline{X}$ to remove the tree approximation to get the residual vector $\underline{Z}_1$. This procedure is then repeated i times to get the residual vector $\underline{Z}_i$. Let $\underline{X}_{\mathcal{T}_i} \sim \mathcal{N}(\underline{0}, \mathbf{\Sigma}_{\mathcal{T}_i})$ be the tree approximation distribution for the residual random vector $\underline{Z}_{i-1} \sim \mathcal{N}(\underline{0}, \mathbf{\Delta}_{i-1})$[2] where $\mathbf{\Sigma}_{\mathcal{T}_i}$ is the tree approximation covariance matrix for $\mathbf{\Delta}_{i-1}$, i.e. $\mathbf{\Sigma}_{\mathcal{T}_i} = \text{chowliu}(\mathbf{\Delta}_{i-1})$.

Remark: For all $i \geq 1$, $\text{tr}\{\mathbf{\Delta}_i\} = n$. Trace of the CAM, $\mathbf{\Delta}_i$ is equal to n, since the covariance matrix $\mathbf{\Sigma}_{\mathcal{T}_i}$ at each iteration is obtained by the Chow-Liu algorithm and thus satisfies the covariance selection rules [11], i.e. $\text{tr}\{(\mathbf{\Delta}_{i-1} - \mathbf{\Sigma}_{\mathcal{T}_i})\mathbf{\Sigma}_{\mathcal{T}_i}^{-1}\} = 0$ and thus $\text{tr}\{\mathbf{\Delta}_{i-1}\mathbf{\Sigma}_{\mathcal{T}_i}^{-1}\} = n$.

If we just use one tree, $\mathbf{C}_{\mathcal{T}_1}$ we have a tree approximation. For a cascade of l trees the approximation model is constructed using a backwards algorithm via the following cascade of linear tree approximations; $(\mathbf{C}_{\mathcal{T}_1}\mathbf{C}_{\mathcal{T}_2}\ldots\mathbf{C}_{\mathcal{T}_l})(\mathbf{C}_{\mathcal{T}_1}\mathbf{C}_{\mathcal{T}_2}\ldots\mathbf{C}_{\mathcal{T}_l})^T = \mathbf{\Sigma}_{\mathcal{M}_l}$. We also have following properties in lemma 1 and lemma 2.

Lemma 1. *Let* $\underline{W} \sim \mathcal{N}(\underline{0}, \mathbf{I})$*, then*

(a) $\mathcal{D}(f_{\underline{Z}_i}(\underline{z})||f_{\underline{W}}(\underline{w})) = \mathcal{D}(f_{\underline{Z}_{i-1}}(\underline{z})||f_{\underline{X}_{\mathcal{T}_i}}(\underline{x}))$,

(b) $\mathcal{D}(f_{\underline{Z}_{i-1}}(\underline{z})||f_{\underline{X}_{\mathcal{T}_i}}(\underline{x})) \leq \mathcal{D}(f_{\underline{Z}_{i-1}}(\underline{z})||f_{\underline{W}}(\underline{w}))$,

where in (b) equality happens when $f_{\underline{Z}_{i-1}}(\underline{z}) = f_{\underline{W}}(\underline{w})$*, i.e* $\mathbf{\Delta}_{i-1} = \mathbf{I}$. ■

Lemma 1 states that the distribution of the i-th step residue random vector $\underline{Z}_i$ converges to the normal Gaussian random vector $\underline{W}$. Thus, in the cascade tree model approximation algorithm, we fix the number of cascade stages, l, and input the normal Gaussian random vector $\underline{W}$ to the cascade trees with l stages to do model approximation. The l-th step model covariance matrix approximation is

$$\mathbf{\Sigma}_{\mathcal{M}_l} = \mathbf{C}_{\mathcal{M}_l}\,\mathbf{C}_{\mathcal{M}_l}^{T}$$

where $\mathbf{C}_{\mathcal{M}_l} = \mathbf{C}_{\mathcal{T}_1}\mathbf{C}_{\mathcal{T}_2}...\mathbf{C}_{\mathcal{T}_i}...\mathbf{C}_{\mathcal{T}_l}$ is the model transformation. Note that, this is a backward construction.

Lemma 2. *KL divergence upper bound.*

$$\mathcal{D}(f_{\underline{X}}(\underline{x})||f_{\underline{X}_{\mathcal{M}_i}}(\underline{x})) \leq \mathcal{D}(f_{\underline{X}}(\underline{x})||f_{\underline{X}_{\mathcal{M}_{i-1}}}(\underline{x}))$$

with equality only happens if $\mathbf{\Delta}_{i-1} = \mathbf{I}$.

Proof.

$$\begin{aligned}\mathcal{D}(f_{\underline{X}}(\underline{x})||f_{\underline{X}_{\mathcal{M}_i}}(\underline{x})) &\overset{(a)}{=} \mathcal{D}(f_{\underline{Z}_i}(\underline{z})||f_{\underline{W}}(\underline{w}))\\ &\overset{(b)}{=} \mathcal{D}(f_{\underline{Z}_{i-1}}(\underline{z})||f_{\underline{X}_{\mathcal{T}_i}}(\underline{x}))\\ &\overset{(c)}{\leq} \mathcal{D}(f_{\underline{Z}_{i-1}}(\underline{z})||f_{\underline{W}}(\underline{w}))\\ &\overset{(d)}{=} \mathcal{D}(f_{\underline{X}}(\underline{x})||f_{\underline{X}_{\mathcal{M}_{i-1}}}(\underline{x}))\end{aligned}$$

where (a) and (d) are because of the invariance of KL divergence between Gaussian distributions to the transformation; (b) and (c) follow from lemma 1. Equality in (c) holds if $f_{\underline{Z}_{i-1}}(\underline{z}) = f_{\underline{W}}(\underline{w})$. ■

Theorem 3. The Cascade tree decomposition transformation. *As the number of cascade trees, i, increases, the KL divergence between the distribution of* $\underline{X}$ *and the model distribution monotonically decreases.* ■

In order to use the cascade tree transformation decomposition framework, we need to first, find the optimal Chow Liu tree, and then perform the Cholesky decomposition such that it preserves the tree graph structure. One can compute the symmetric CAM at each iteration of a greedy algorithm to efficiently find the optimal tree structure covariance matrix.

IV. Simulation Results and Discussion

In this Example, the covariance matrix is calculated based on datasets presented in [12]. The Oahu solar measurement grid dataset is obtained from the National Renewable Energy Laboratory (NREL) website [13]. This dataset consists of 19 sensors (17 horizontal sensors and two tilted sensors). For this dataset we normalized using standard normalization method and the zenith angle normalization method [12][3]. From the data obtained from these 19 solar sensors at the island of Oahu, we computed the spatial covariance matrix during the summer season at 12:00 PM averaged over a window of 5 minutes.

Figure 1 plots the gray scaled, sparsity pattern for the inverse of the Oahu solar measurement grid covariance matrix and various approximations of it. The top left plot shows the inverse of the original normalized covariance matrix while the bottom left plot indicates the first stage of the cascade tree approximation or the optimal Chow-Liu approximated model. The top middle plot shows the inverse of the second approximated model while the plot on the bottom middle indicates the inverse of the second stage tree approximation. The plot on the top right indicates the sparsity pattern of the inverse of third approximated model, while the bottom right plot shows the third stage tree approximation.

Figure 2 plots the log-scaled KL divergence between the distribution of the random vector $\underline{X}$ and the distribution of the model distribution after the i-th step of the cascade trees approximation with respect to the number of cascade trees transformation that are used in the approximation, i. This

[2] $\underline{Z}_0 \sim \mathcal{N}(\underline{0}, \mathbf{\Sigma})$.

[3] See [12] for more detailed description of dataset and other details about the normalization methods for the solar irradiation covariance matrix.

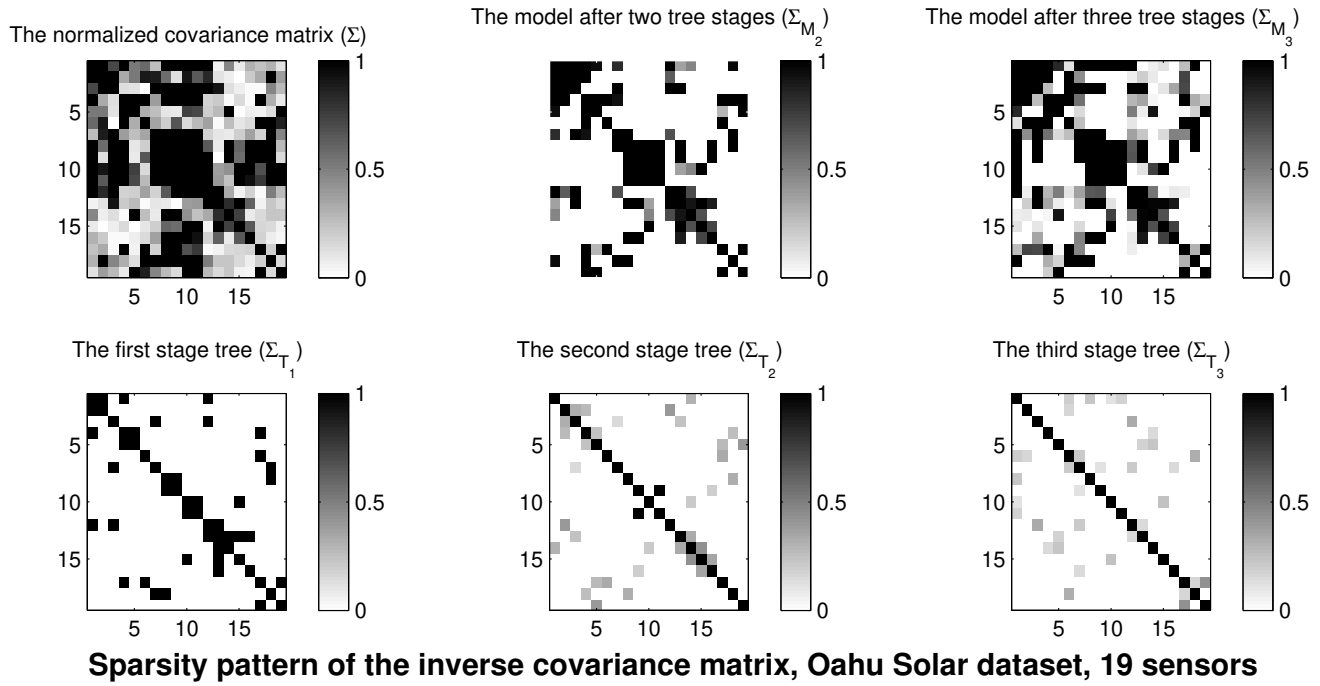

Fig. 1. Gray scaled, sparsity pattern for the inverse of the covariance matrix generated using the Oahu solar measurement grid dataset. **Top left:** Original normalized covariance matrix, **Bottom left:** first stage tree approximation and first model. **Top middle:** second approximated model, **Bottom middle:** second stage tree approximation. **Top right:** third approximated model, **Bottom right:** third stage tree approximation.

figure compares the performance of the proposed cascade trees approximation with different decomposition choices with the optimal star tree approximation. This figure also plots the result of the cascade trees framework with different decompositions such as the Cholesky $\mathbf{L}\mathbf{L}^T$ (keep the sparsity), the Cholesky $\mathbf{U}\mathbf{U}^T$ (does not keep the sparsity) and the SVD. From figure 2 we see that three of the decomposition transformations perform similarly with the star decomposition transformation performing the worse. If we compare the Chow-Liu tree to a cascade of two trees/ three trees the KL divergence decreases by respectively more than 50%/ 80%. By using the Chow-Liu algorithm to produce trees and then using the Cholesky factorization in general, this algorithm performs well as the KL divergence decreases relatively quickly. However, by using the star network systematically on all nodes except one we can guarantee that the cascade algorithm converges to the model after $n-1$ steps.

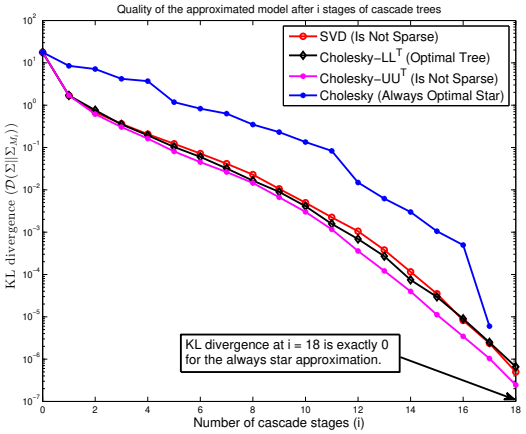

Fig. 2. KL divergence between the distribution of the random vector $\underline{X}$ and the model distribution after the i-th step of the cascade approximation v.s. the index of the cascade trees, i, for the island of Oahu solar data using different decompositions.

V. Conclusion

We look at the graphical model as a transformation and introduce a general framework to do model approximation for graphical models. We present results that guarantees the convergence of the proposed model approximation using the cascade of tree decompositions. We present a backward construction method combined with the Cholesky factorization. We confirm theoretical results using the examples provided in the simulation section where we look at synthetic and real data and compare the performance of the proposed framework by comparing KL divergences.

Acknowledgment

This work was supported in part by NSF grant ECCS-1310634, the Center for Science of Information (CSoI), an NSF Science and Technology Center, under grant agreement CCF-0939370, and the University of Hawaii REIS project.

References

[1] Justin HG Dauwels, *On graphical models for communications and machine learning: Algorithms, bounds, and analog implementation*, vol. 17, ETH Zurich, 2006.

[2] Janaka B Ekanayake, Nick Jenkins, Kithsiri Liyanage, Jianzhong Wu, and Akihiko Yokoyama, *Smart Grid: Technology and Applications*, John Wiley & Sons, 2012.
[3] N. T. Khajavi and A. Kuh, "First order Markov chain approximation of microgrid renewable generators covariance matrix," in *Proc. IEEE International Symposium on Information Theory, Istanbul, Turkey (ISIT' 13)*, July 2013, pp. 1207–1211.
[4] N. T. Khajavi and A. Kuh, "The quality of the covariance selection through detection problem and AUC bounds," *arXiv preprint arXiv:1605.05776*, 2016.
[5] C. K. Chow and C.N. Liu, "Approximating discrete probability distributions with dependence trees," *IEEE Transactions on Information Theory*, pp. 462–467, 1968.
[6] P. M. Lewis II, "Approximating probability distributions to reduce storage requirements," *Information and control*, vol. 2, no. 3, pp. 214–225, 1959.
[7] Curtis Greene and Glenn A Iba, "Cayley's formula for multidimensional trees," *Discrete Mathematics*, vol. 13, no. 1, pp. 1–11, 1975.
[8] Joseph B Kruskal, "On the shortest spanning subtree of a graph and the traveling salesman problem," *Proceedings of the American Mathematical society*, vol. 7, no. 1, pp. 48–50, 1956.
[9] R. C. Prim, "Shortest connection networks and some generalizations," *Bell system technical journal*, vol. 36, no. 6, pp. 1389–1401, 1957.
[10] A. Kavcic and J. M. F. Moura, "Matrices with banded inverses: Inversion algorithms and factorization of Gauss-Markov processes," *IEEE Transactions on Information Theory*, vol. 46, pp. 1495–1509, July 2000.
[11] A. P. Dempster, "Covariance selection," *Biometrics*, vol. 28, no. 1, pp. 157–175, March 1972.
[12] Navid Tafaghodi Khajavi, Anthony Kuh, and Narayana P Santhanam, "Spatial correlations for solar PV generation and its tree approximation analysis," in *Proceedings of the Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA ASC)*, Dec 2014, pp. 1–5.
[13] NREL. Solar irradiance website. [Online]. Available: http://www.nrel.gov/midc/, ," .