ISIT'2013 1569720491

First Order Markov Chain Approximation of Microgrid Renewable Generators Covariance Matrix

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

***Abstract*—Smart grids present interesting challenges as we integrate renewable energy sources and allow for customer participation in the decision making. A key to smart grids is getting information from the grid in real time and especially at the distribution level beyond the substation, which we refer to as the microgrid. This paper presents a distributed state estimation algorithm for microgrids with distributed renewable energy generation. We use a factor graph approach to model the microgrid network with renewable generators. The renewable generators are correlated resulting in many loops in the factor graph. This presents a problem when using distributed algorithms such as belief propagation. To limit the number of loops in the factor graph, we approximate the correlation among the renewable generators using a Markov chain approach. The algorithm is sub-optimal, but has low complexity using a greedy approach and Cholesky factorization. We present a simple microgrid example with renewable generators and show through simulations that our approximate solution gives performance close to the optimal solution.**

I. Introduction

Given real-time information obtained from measurements, a power grid operator's goal is to maintain a system that can reliably deliver energy to consumers, even if the generation of power changes stochastically. Thus, local awareness is an inseparable, crucial part of the smart grid. Due to facility costs and availability conditions for a large complex system like the smart grid, direct error-free measurement of all or even a large number of system state variables is practically infeasible, which means that power grid operators need computationally efficient tools like signal processing approaches to estimate system state variables. Signal processing approaches and state estimation methods are powerful methods for system monitoring. It is shown in the literature that given a set of noisy measurements, a state estimator can provide the best complete estimate of the current system state by processing data collected at measurement units throughout the system [1]. Consequently, a state estimator is considered as an essential tool for smart grid system monitoring.

There are two broad types of available state estimator algorithms. Centralized state estimators collect all available data from all the sensors and perform accurate state estimation, but there are computational complexity concerns. [2] and [3] have recently addressed this issue by distributing messages between local areas. In distributed state estimators, grid elements pass local messages to their neighbors and perform state estimation using belief propagation (BP) algorithm [1]. Since microgrid electrical connections have tree structures, their factor graph representations also have tree structures and BP algorithms converge to the maximum likelihood solution.

Due to a large number of states in a smart grid, it is often difficult to perform state estimation in real-time fashion. In contrast, distributed state estimators can give reasonably good estimates for large grid systems in real-time. This fact provides a trade off between calculation time and accuracy of estimation. In this paper, we seek ways to increase the accuracy of distributed state estimators that pass local messages. We modify the distributed state estimation method proposed in [4] such that it can perform in real-time fashion. Since smart grid operators need real-time state estimation, distributed state estimators that achieve good accuracy are desirable but present major challenges. Note that, with increasing penetration of distributed correlated renewable energy generators (REGs) this adds complexity to the smart grid. A factor graph representation with REGs introduces many loops to the original graph. This can create problems for the convergence of the BP algorithm.

To assure the convergence of distributed algorithm, we need to decrease the number of loops in the microgrid factor graph representation. Here we assume REGs are jointly Gaussian random variables and approximate these random variables using a first order Markov chain. Our goal is to find an approximate covariance matrix that represents the REG subgraph as a chain structure. A first approach, is the well-known covariance selection problem solved initially by Burg [5]. A more generalized solution was presented by Dempster [6]. Recently, it was also shown in [7] that the Dempster solution is a sub optimal solution for the minimum mean square estimation (MMSE) problem. Here we define an optimization problem based on an MMSE *regret* function [8] to take into account observation noise and the microgrid structure. We propose a greedy ad-hoc algorithm to find an upper bound for the optimal solution, since finding the exact solution is not straight-forward. Note, in this paper we apply this method to a static state estimation problem, but these method are applicable to a wide range of dynamic signal processing problems. Our simulation results show that the greedy algorithm using the BP algorithm can achieve good performance using a MMSE cost criterion.

The rest of this paper is organized as follow. Section II gives a formulation of the problem. Since for large number of REGs,

the optimal solution for the regret function is computationally difficult to obtain, we present a greedy ad-hoc solution in Section III. In Section IV we present some simulation results. Finally, Section V summarizes results of this paper.

Notation: Upper case and lower case letters denote random variables and their realizations, respectively; underlined letters stand for vectors; boldface upper case letters denote matrices; $(\cdot)^T$, $(\cdot)^+$ and $\mathrm{E}(\cdot)$ stand for transposition, pseudo inverse and expectation, respectively.

II. Problem statement

We are interested in estimating the microgrid bus voltages, $\underline{V} \in \mathbb{R}^l$, and branch currents, $\underline{I} \in \mathbb{R}^{n-l}$, while integrating REG sources. We consider REGs which are modeled as correlated jointly Gaussian current sources. Let $\underline{X} \triangleq [\underline{V}\ \underline{I}]^T \in \mathbb{R}^n$ be the state vector and $\underline{G} \in \mathbb{R}^p$ be the vector of REGs where $p < n$. Let $\underline{Y} \in \mathbb{R}^m$ be the observation vector where $m < n$. Also, assume independent and identically distributed (i.i.d.) noisy sensors with noise variance ν^2. The microgrid is described as[1]:

$$\underline{X} = \mathbf{A}\underline{G} \tag{1}$$

and

$$\underline{Y} = \mathbf{B}\underline{X} + \underline{W} \tag{2}$$

where $\mathbf{A}$ is microgrid characteristic matrix. $\underline{G}$ is REGs' random vector, $\underline{G} \sim \mathcal{N}(\underline{0}, \mathbf{\Sigma})$. $\underline{W}$ is a zero-mean i.i.d. Gaussian vector with covariance matrix $\mathbf{D} = \nu^2\mathbf{I}_m$ with $\mathbf{I}_m$ being an identity matrix of dimension m. $\mathbf{B}$ is an $m \times n$ binary matrix with orthonormal rows, where each row has one '1', i.e. $\mathbf{B}\mathbf{B}^T = \mathbf{I}_m$. Note that, we propose a linear model in (1) and (2). Since REGs random vector, $\underline{G}$, and noisy measurements, $\underline{W}$, are Gaussian, the states, i.e., voltages and currents and observations are jointly Gaussian. Let $\mathbf{H} \triangleq \mathbf{\Sigma}_{\underline{XY}}\mathbf{\Sigma}_{\underline{Y}}^{-1}$, where $\mathbf{\Sigma}_{\underline{XY}} = \mathrm{E}(\underline{X}\underline{Y}^T)$ and $\mathbf{\Sigma}_{\underline{Y}} = \mathrm{E}(\underline{Y}\underline{Y}^T)$. The MMSE estimator is given by [9]

$$\hat{\underline{X}}(\mathbf{\Sigma}, \underline{Y}) = E(\underline{X}|\underline{Y}) = \mathbf{H}(\mathbf{\Sigma})\underline{Y} \tag{3}$$

Let $\mathbf{Q} \triangleq \mathbf{B}\mathbf{A}$. Considering (1) and (2), we have:

$$\mathbf{H}(\mathbf{\Sigma}) = \mathbf{A}\mathbf{\Sigma}\mathbf{Q}^T\left(\mathbf{Q}\mathbf{\Sigma}\mathbf{Q}^T + \mathbf{D}\right)^{-1}, \tag{4}$$

Here our goal is to find an approximate covariance matrix $\widetilde{\mathbf{\Sigma}}$, such that the overall microgrid graph (that incorporates the scheme of $\widetilde{\mathbf{\Sigma}}$ in place of $\mathbf{\Sigma}$ in the graph) has few loops. Here we propose to achieve this goal by looking for the matrix $\widetilde{\mathbf{\Sigma}}$ that corresponds to a first order Markov chain process. This means that there exist a permutation $\mathbf{P}$, such that $(\mathbf{P}\widetilde{\mathbf{\Sigma}}\mathbf{P}^T)^{-1}$ is a tridiagonal matrix.

Definition 1. *Let $\mathcal{M}_1$ denote the set of all positive definite covariance matrices of first-order Markov process. That is, $\mathcal{M}_1$ consists of all matrices $\widetilde{\mathbf{\Sigma}}$, for which there exist a permutation $\mathbf{P}$, such that $\mathbf{T} = (\mathbf{P}\widetilde{\mathbf{\Sigma}}\mathbf{P}^T)^{-1}$ is a tridiagonal matrix.* ■

[1]Without loss of generality, we may assume that $\underline{X}$, $\underline{G}$ and $\underline{Y}$ are zero-mean vectors.

Let $\hat{\underline{X}}(\widetilde{\mathbf{\Sigma}}, \underline{Y})$ be the linear state estimate in (3) when $\widetilde{\mathbf{\Sigma}}$ is used instead of $\mathbf{\Sigma}$. We define error by $\underline{\mathcal{E}} = \hat{\underline{X}}(\widetilde{\mathbf{\Sigma}}, \underline{Y}) - \hat{\underline{X}}(\mathbf{\Sigma}, \underline{Y})$. Using $\Sigma_{\underline{Y}}$, the error covariance matrix is given by

$$\mathrm{E}(\underline{\mathcal{E}}\underline{\mathcal{E}}^T) = (\widetilde{\mathbf{H}}(\widetilde{\mathbf{\Sigma}}) - \mathbf{H}(\mathbf{\Sigma}))\Sigma_{\underline{Y}}(\widetilde{\mathbf{H}}(\widetilde{\mathbf{\Sigma}}) - \mathbf{H}(\mathbf{\Sigma}))^T. \tag{5}$$

Our task is to find the matrix $\widetilde{\mathbf{\Sigma}} = \widetilde{\mathbf{\Sigma}}^*$ that minimizes the total error $\mathrm{tr}\ \mathrm{E}(\underline{\mathcal{E}}\underline{\mathcal{E}}^T)$.

$$\widetilde{\mathbf{\Sigma}}^* = \underset{\widetilde{\mathbf{\Sigma}} \in \mathcal{M}_1}{\arg\min}\, \mathrm{tr}\ \mathrm{E}(\underline{\mathcal{E}}\underline{\mathcal{E}}^T) \tag{6}$$

Definition 2. *Let ρ be the* **regret** *function between $\mathbf{\Omega}$ and $\mathbf{\Delta}$ with respect to the parameters $\mathbf{\Gamma}$ and $\mathbf{\Theta}$ with the following definition:*

$$\rho_{<\mathbf{\Gamma},\mathbf{\Theta}>}(\mathbf{\Omega}, \mathbf{\Delta}) = \mathrm{tr}\{\mathbf{\Gamma}[\mathbf{\Omega} - \mathbf{\Delta}]\mathbf{\Theta}[\mathbf{\Omega} - \mathbf{\Delta}]^T\mathbf{\Gamma}^T\} \tag{7}$$

■

Note that, the regret definition above is similar to the regret definition proposed in [8].

Definition 3. *Let the* **discrepancy** *of the matrix $\widetilde{\mathbf{\Sigma}}$ (the desired approximation of covariance matrix $\mathbf{\Sigma}$) be defined as:*

$$\mathcal{D}(\widetilde{\mathbf{\Sigma}}) = \rho_{<\mathbf{A}\ ,\ \mathbf{Q}\mathbf{\Sigma}\mathbf{Q}^T+\mathbf{D}>} (\widetilde{\mathbf{\Sigma}}\mathbf{Q}^T(\mathbf{Q}\widetilde{\mathbf{\Sigma}}\mathbf{Q}^T + \mathbf{D})^{-1}, \mathbf{\Sigma}\mathbf{Q}^T(\mathbf{Q}\mathbf{\Sigma}\mathbf{Q}^T + \mathbf{D})^{-1}) \tag{8}$$

where ρ is the regret function. ■

The optimization problem then is given by

$$\widetilde{\mathbf{\Sigma}}^* = \underset{\widetilde{\mathbf{\Sigma}} \in \mathcal{M}_1}{\arg\min}\, \mathcal{D}(\widetilde{\mathbf{\Sigma}}). \tag{9}$$

Since $\widetilde{\mathbf{\Sigma}}$ and $\mathbf{D}$ are non-singular square matrices, applying the Matrix Inversion Lemma [10], (4) can be written as:

$$\mathbf{H}(\mathbf{\Sigma}) = \mathbf{A}(\mathbf{\Sigma}^{-1} + \mathbf{Q}^T\mathbf{D}^{-1}\mathbf{Q})^{-1}\mathbf{Q}^T\mathbf{D}^{-1}. \tag{10}$$

Let $\mathbf{S} \triangleq \mathbf{Q}^T\mathbf{D}^{-1}\mathbf{Q}$. Replacing (10) in (5) , the discrepancy (8) can be written as:

$$\mathcal{D}(\widetilde{\mathbf{\Sigma}}) = \rho_{<\mathbf{A}\ ,\ \mathbf{S}\mathbf{\Sigma}\mathbf{S}+\mathbf{S}>}((\widetilde{\mathbf{\Sigma}}^{-1} + \mathbf{S})^{-1}, (\mathbf{\Sigma}^{-1} + \mathbf{S})^{-1}). \tag{11}$$

The optimization problem in (9) can be rewritten as the following optimization problem with an inner and an outer minimization:

$$(\mathbf{P}^*, \mathbf{T}^*) = \underset{\mathbf{P}}{\arg\min}\ \min_{\mathbf{T} \in \mathcal{T}}\ \mathcal{D}(\mathbf{P}\mathbf{T}^{-1}\mathbf{P}^T). \tag{12}$$

where $\mathcal{T}$ is the set of all positive definite tridiagonal matrices, $\mathbf{P}$ is a permutation matrix and $\widetilde{\mathbf{\Sigma}}^* = \mathbf{P}^*\mathbf{T}^{*-1}\mathbf{P}^{*T}$. The outer optimization in (12) is an exhaustive search whose complexity is $\mathcal{O}(p!)$. The inner optimization finds the best tridiagonal matrix $\mathbf{T}$ for a given permutation matrix $\mathbf{P}$. Due to the difficulty of the optimization problem (12), we propose an ad-hoc solution to avoid exhaustive search. If $\widetilde{\mathbf{\Sigma}} = \mathbf{P}\mathbf{T}^{-1}\mathbf{P}^T$ is an ad-hoc solution to (12), then it provides an upper bound on the optimal discrepancy $\mathcal{D}(\widetilde{\mathbf{\Sigma}}^*)$, i.e., $\mathcal{D}(\widetilde{\mathbf{\Sigma}}) \geq \mathcal{D}(\widetilde{\mathbf{\Sigma}}^*)$. Therefore, the search for good (suboptimal) solutions to (12) is equivalent to constructing tight upper bounds on $\mathcal{D}(\widetilde{\mathbf{\Sigma}}^*)$. In the following sections, we focus on jointly solving the

inner and the outer optimization problems (12) by proposing a greedy algorithm. In what follows, without lost of generality, we assume that the $\widetilde{\mathbf{\Sigma}}^{-1}$ has a tridiagonal structure. The result of the outer optimization is $\mathbf{P}=\mathbf{I}$ and we search for the best matrix $\widetilde{\mathbf{\Sigma}}$ whose inverse is tridiagonal, i.e. , $\widetilde{\mathbf{\Sigma}}^{-1}=\mathbf{T}$.

III. Greedy Ad-hoc Solution

In this Section we discuss the details of our greedy algorithm. We initially choose a 2×2 matrix and iteratively construct a $p\times p$ matrix. Subsection III-A describes basics of the iteration step. Subsection III-B discusses the numerical algorithm of solving for the optimal values of the Cholesky factorization. Finally, subsection III-C discusses the overall run time of the algorithm and implementing different permutations.

A. Basic algorithm

As mentioned earlier, without loss of generality, we assume that $\mathbf{P}=\mathbf{I}$, so that we only need to worry about the inner optimization problem in (12). We design an iterative greedy algorithm. Assume that in iteration $l-1$, we have an $(l-1)\times(l-1)$ tridiagonal matrix $\mathbf{T}_{l-1}\in\mathcal{T}_{l-1}$. In iteration l, we find an $l\times l$ tridiagonal matrix $\mathbf{T}_l\in\mathcal{T}_l$ such that $\mathbf{T}_{l-1}$ is its principle submatrix, i.e. the following constraint holds

$$\mathbf{T}_{l-1}=[\mathbf{I}_{l-1}\ \underline{0}]\mathbf{T}_l[\mathbf{I}_{l-1}\ \underline{0}]^T. \tag{13}$$

Thereby, we find $\mathbf{T}_l$ to optimize the size l version of the optimization problem in (11). To that end, we define the l-th step discrepancy as

$$\mathcal{D}_l(\mathbf{T}_l^{-1})\triangleq\rho_{<\mathbf{A}_l,\mathbf{S}_l\mathbf{\Pi}_l^{-1}\mathbf{S}_l+\mathbf{S}_l>}((\mathbf{T}_l+\mathbf{S}_l)^{-1},(\mathbf{\Pi}_l+\mathbf{S}_l)^{-1}) \tag{14}$$

where $A_l\triangleq A[\mathbf{I}_{l-1}\ \mathbf{0}]^T$, $\mathbf{S}_l\triangleq[\mathbf{I}_{l-1}\ \mathbf{0}]\mathbf{S}[\mathbf{I}_{l-1}\ \mathbf{0}]^T$ and $\mathbf{\Pi}_l\triangleq([\mathbf{I}_{l-1}\ \mathbf{0}]\mathbf{\Sigma}[\mathbf{I}_{l-1}\ \mathbf{0}]^T)^{-1}$. Now, we construct $\mathbf{T}_l$ as the tridiagonal matrix that minimizes the l-th step discrepency (14) under the constraint in (13). Precisely, in the l-th iteration, we set

$$\mathbf{T}_l\triangleq\underset{\substack{\mathbf{T}\in\mathcal{T}_l:\\ \mathbf{T}_{l-1}=[\mathbf{I}_{l-1}\ \underline{0}]\mathbf{T}[\mathbf{I}_{l-1}\ \underline{0}]^T}}{\arg\min}\mathcal{D}_l(\mathbf{T}^{-1}) \tag{15}$$

After the p-th iteration, we set $\widetilde{\mathbf{\Sigma}}=\mathbf{T}_p^{-1}$.

B. Numerical solution

We do not have an analytical method for solving (15), but we can easily construct a semi-analytical (partially analytical and partially numerical) method for solving (15) in each step of the greedy algorithm. Given $\mathbf{T}_{l-1}$, the $l\times l$ tridiagonal matrix $\mathbf{T}\in\mathcal{T}_l$ that satisfies the constraint of optimization problem (15) can be decomposed by applying backward Cholesky decomposition as [11]:

$$\mathbf{T}=\mathbf{T}(w)+\underline{e}\underline{e}^T\sigma^{-2} \tag{16}$$

where σ^2 and w are backward Cholesky parameters, $\underline{e}$ is a unit vector orthogonal to $[\mathbf{I}_{l-1}\ \underline{0}]$, i.e. $[\mathbf{I}_{l-1}\ \underline{0}]\underline{e}=\underline{0}$, and

$$\begin{aligned}\mathbf{T}(w)=&[\mathbf{I}_{l-1}\ \underline{0}]^T\mathbf{T}_{l-1}[\mathbf{I}_{l-1}\ \underline{0}]\\&+\left(\begin{bmatrix}\underline{0}\\1\\w\end{bmatrix}\begin{bmatrix}\underline{0}\\1\\w\end{bmatrix}^T-\begin{bmatrix}\underline{0}\\1\\0\end{bmatrix}\begin{bmatrix}\underline{0}\\1\\0\end{bmatrix}^T\right)\begin{bmatrix}\underline{0}\\1\end{bmatrix}\mathbf{T}_{l-1}\begin{bmatrix}\underline{0}\\1\end{bmatrix}^T.\end{aligned} \tag{17}$$

We then solve (15) in the l-th iteration of the greedy algorithm. Substituting (16), (15) can be rewritten as

$$(\sigma_l^2,w_l)=\arg\min_w\min_{\sigma^2>0}\mathcal{D}_l\left((\mathbf{T}(w)+\underline{e}\underline{e}^T\sigma^{-2})^{-1}\right) \tag{18}$$

To solve the optimization problem in (18), first we introduce the following Lemma and Theorem[2].

Lemma 1. *Let $\mathbf{\Omega}=\underline{e}\underline{e}^T\eta$ where η is a scalar parameter, then regret* (7) *can be written as:*

$$\rho_{<\mathbf{\Gamma},\mathbf{\Theta}>}(\underline{e}\underline{e}^T\eta,\mathbf{\Delta})=\rho_{<\mathbf{\Gamma}\underline{e},\underline{e}^T\mathbf{\Theta}\underline{e}>}(\eta,\delta)+c(\delta) \tag{19}$$

where $\delta=(\mathbf{\Gamma}\underline{e})^+\mathbf{\Gamma}\mathbf{\Delta}\mathbf{\Theta}\underline{e}(\underline{e}^T\mathbf{\Theta}\underline{e})^{-1}$ is a scalar parameter and $c(\delta)=\mathrm{tr}\{\mathbf{\Gamma}\mathbf{\Delta}\mathbf{\Theta}\mathbf{\Delta}^T\mathbf{\Gamma}^T\}-\mathrm{tr}\{\mathbf{\Gamma}\underline{e}\delta\underline{e}^T\mathbf{\Theta}\underline{e}\delta^T(\mathbf{\Gamma}\underline{e})^T\}$. ■

Given $\mathbf{T}(w)$, i.e. given $\mathbf{T}_{l-1}$ and w, define $\mathbf{R}_l\triangleq\mathbf{T}(w)+\mathbf{S}_l=\mathbf{U}\mathbf{\Lambda}\mathbf{U}^T$ where $\mathbf{U}$ and $\mathbf{\Lambda}$ are unitary matrix and diagonal matrix, respectively. Then, the inner optimization in (18) can be rewritten as

$$\begin{aligned}\sigma_l^2=\arg\min_{\sigma^2>0}&\rho_{<\mathbf{A}_l\mathbf{U},\mathbf{U}^T(\mathbf{S}_l\mathbf{\Pi}_l^{-1}\mathbf{S}_l+\mathbf{S}_l)\mathbf{U}>}\\&((\mathbf{\Lambda}+\underline{u}\underline{u}^T\sigma^{-2})^{-1},\mathbf{U}^T(\mathbf{\Pi}_l+\mathbf{S}_l)^{-1}\mathbf{U})\end{aligned} \tag{20}$$

where $\underline{u}=\mathbf{U}^T\underline{e}$. Note that, rank of matrix $\mathbf{R}_l$ is either l or $l-1$. Lemma 1 and (20) lead to the following Theorem.

Theorem 2. *a) If rank of $\mathbf{R}_l$ is l, then the l-th step solution of inner optimization problem in* (18) *is identical to the solution of the following optimization problem:*

$$\sigma_l^2=\arg\min_{\sigma^2>0}((\sigma^2+\underline{e}^T\mathbf{R}_l^{-1}\underline{e})^{-1}-f_l)^2 \tag{21}$$

where

$$\begin{aligned}f_l=&(\mathbf{A}_l\mathbf{R}_l^{-1}\underline{e})^+\mathbf{A}_l(\mathbf{R}_l^{-1}-(\mathbf{\Pi}_l+\mathbf{S}_l)^{-1})(\mathbf{S}_l\mathbf{\Pi}_l^{-1}\mathbf{S}_l+\mathbf{S}_l)\\&\mathbf{R}_l^{-1}\underline{e}(\underline{e}^T\mathbf{R}_l^{-1}(\mathbf{S}_l\mathbf{\Pi}_l^{-1}\mathbf{S}_l+\mathbf{S}_l)\mathbf{R}_l^{-1}\underline{e})^{-1}.\end{aligned} \tag{22}$$

b) If rank of $\mathbf{R}_l$ is $l-1$, i.e. $\mathbf{\Lambda}=[\mathbf{I}\ \underline{0}]^T\mathbf{\Lambda}'[\mathbf{I}\ \underline{0}]$ where $\mathbf{\Lambda}'$ is an $(l-1)\times(l-1)$ diagonal submatrix of $\mathbf{\Lambda}$, then the l-th step solution of inner optimization problem in (18) *is identical to the solution of the following optimization problem:*

$$\sigma_l^2=\arg\min_{\sigma^2>0}(\underline{e}^T\underline{u}\sigma^2-g_l)^2 \tag{23}$$

where

$$\begin{aligned}g_l=&(\mathbf{A}_l\mathbf{U}\underline{e})^+\mathbf{A}_l(\mathbf{U}\mathbf{K}\mathbf{U}^T-(\mathbf{\Pi}_l+\mathbf{S}_l)^{-1})(\mathbf{S}_l\mathbf{\Pi}_l^{-1}\mathbf{S}_l+\mathbf{S}_l)\\&\mathbf{U}\underline{e}(\underline{e}^T\mathbf{U}^T(\mathbf{S}_l\mathbf{\Pi}_l^{-1}\mathbf{S}_l+\mathbf{S}_l)\mathbf{U}\underline{e})^{-1}\end{aligned} \tag{24}$$

where $\mathbf{K}=[\mathbf{I}\quad-(\underline{e}^T\underline{u})^{-1}[\mathbf{I}\ \underline{0}]\underline{u}]^T\mathbf{\Lambda}'[\mathbf{I}\quad-(\underline{e}^T\underline{u})^{-1}[\mathbf{I}\ \underline{0}]\underline{u}]$. ■

To achieve our numerical solution, we choose several values of w and find σ_l^2 for each of them. Then, we choose the best pair of σ_l^2 and w, i.e. σ_l^2 and w_l, as the solution for the optimization problem in (18). Having the pair of (σ_l^2,w_l), we can construct $\mathbf{T}_l$ using (16). Continuing the procedure, we achieve $\mathbf{T}_p^{-1}\in\mathcal{M}_1$ as the covariance matrix of the approximated first-order Markov process.

[2]Proofs of Lemma and Theorem are written up in a subsequent paper and are not presented here due to space constraints.

TABLE I
SIMULATION PARAMETERS

Parameter	V	r_V	r	R	$\mathrm{E}(\mathbf{G})$
Value	110	1	1	20	$[10\ 7\ 13\ 6\ 2\ 5]^T$

C. Incorporating the search for the permutation into the greedy algorithm

Initialization step: Set $l = 2$. Setting $\mathbf{T}_2$ equal to inverse of any 2×2 principal submatrix of covariance matrix $\mathbf{\Sigma}$, implies that $\mathcal{D}_2(\mathbf{T}_2^{-1}) = 0$. So, inverse of each 2×2 principal submatrix is a solution for $\mathbf{T}_2$ at the initialization step of the greedy algorithm. Since, there are $\frac{p(p-1)}{2}$ 2×2 principal submatrix, the greedy algorithm must run $\frac{p(p-1)}{2}$ times to find the approximate covariance matrix.

Iteration step: Let $\mathcal{P}_l$ denote the set of matrices $\mathbf{P}_l$ of size $p \times l$ whose l columns consist of l different unit vectors, i.e. $\mathbf{P}_l$ is a $p \times l$ submatrix of a $p \times p$ permutation matrix $\mathbf{P}$. Assume that in the $(l-1)$-th iteration of the greedy algorithm, we have $\widetilde{\mathbf{P}}_{l-1} \in \mathcal{P}_{l-1}$ and $\widetilde{\mathbf{T}}_{l-1} \in \mathcal{T}_{l-1}$. In the l-th iteration of the greedy algorithm, we seek to find new matrices $\widetilde{\mathbf{P}}_l \in \mathcal{P}_l$ and $\widetilde{\mathbf{T}}_l \in \mathcal{T}_l$ that minimize a properly defined l-th order discrepency $\mathcal{D}'_l(\widetilde{\mathbf{T}}_l)$ and satisfy the following constraints:

a) $\widetilde{\mathbf{P}}_l$ equals either $[\widetilde{\mathbf{P}}_{l-1}\ \widetilde{e}]$ or $[\widetilde{e}\ \widetilde{\mathbf{P}}_{l-1}]$ where $\widetilde{e}$ is a unit vector orthogonal to $\widetilde{\mathbf{P}}_{l-1}$, i.e. $\widetilde{\mathbf{P}}_{l-1}\widetilde{e} = \underline{0}$,

b) $\widetilde{\mathbf{T}}_l$ is a tridiagonal matrix such that $\widetilde{\mathbf{T}}_{l-1}$ is its submatrix, i.e. either $[\mathbf{I}_{l-1}\ \underline{0}]\widetilde{\mathbf{T}}_l[\mathbf{I}_{l-1}\ \underline{0}]^T = \widetilde{\mathbf{T}}_{l-1}$ or $[\underline{0}\ \mathbf{I}_{l-1}]\widetilde{\mathbf{T}}_l[\underline{0}\ \mathbf{I}_{l-1}]^T = \widetilde{\mathbf{T}}_{l-1}$

Constraints a) and b) can be compactly written as

$$\widetilde{\mathbf{P}}_{l-1}^T\widetilde{\mathbf{P}}_l \in \{[\mathbf{I}_{l-1}\ \underline{0}], [\underline{0}\ \mathbf{I}_{l-1}]\}, \tag{25}$$

and

$$\widetilde{\mathbf{P}}_{l-1}^T\widetilde{\mathbf{P}}_l\widetilde{\mathbf{T}}_l\widetilde{\mathbf{P}}_l^T\widetilde{\mathbf{P}}_{l-1} = \widetilde{\mathbf{T}}_{l-1}. \tag{26}$$

It is now only left to properly define the l-th order discrepancy $\mathcal{D}'_l(\widetilde{\mathbf{T}}_l^{-1})$ and formulate the l-th step of the greedy algorithm as a minimization of $\mathcal{D}'_l(\widetilde{\mathbf{T}}_l^{-1})$ with constraints (25) and (26). Specifically, we find $\widetilde{\mathbf{P}}_l$ and $\widetilde{\mathbf{T}}_l$ as

$$(\widetilde{\mathbf{P}}_l, \widetilde{\mathbf{T}}_l) = \underset{\substack{(\mathbf{P}_l,\mathbf{T}_l)\in\mathcal{P}_l\times\mathcal{T}_l:\\ \widetilde{\mathbf{P}}_{l-1}^T\mathbf{P}_l\in\{[\mathbf{I}_{l-1}\ \underline{0}],[\underline{0}\ \mathbf{I}_{l-1}]\},\\ \widetilde{\mathbf{P}}_{l-1}^T\mathbf{P}_l\mathbf{T}_l\mathbf{P}_l^T\widetilde{\mathbf{P}}_{l-1}=\widetilde{\mathbf{T}}_{l-1}}}{\arg\min} \mathcal{D}'_l(\mathbf{T}_l^{-1}), \tag{27}$$

where

$$\mathcal{D}'_l(\mathbf{T}_l^{-1}) \triangleq \rho_{<\mathbf{A}'_l,\mathbf{S}'_l\mathbf{\Pi}'^{-1}_l\mathbf{S}'_l+\mathbf{S}'_l>}((\mathbf{T}_l+\mathbf{S}'_l)^{-1},(\mathbf{\Pi}'_l+\mathbf{S}'_l)^{-1}) \tag{28}$$

where $\mathbf{A}'_l \triangleq \mathbf{A}\mathbf{P}_l$, $\mathbf{S}'_l \triangleq \mathbf{P}_l^T\mathbf{S}\mathbf{P}_l$ and $\mathbf{\Pi}'_l \triangleq (\mathbf{P}_l^T\mathbf{\Sigma}\mathbf{P}_l)^{-1}$. Notice that if $\mathbf{P}_l = \mathbf{I}_l$, we have $\mathcal{D}'_l(\mathbf{T}_l^{-1}) = \mathcal{D}_l(\mathbf{T}_l^{-1})$. Hence, (27) is the equivalent of the greedy algorithm in (15), with the caveat that (27) updates both the permutation matrix $\widetilde{\mathbf{P}}_l$ as well as the tridiagonal matrix $\widetilde{\mathbf{T}}_l$. The algorithm stops after p-th iteration by setting $\widetilde{\mathbf{\Sigma}} = (\widetilde{\mathbf{P}}\widetilde{\mathbf{T}}\widetilde{\mathbf{P}}^T)^{-1}$ where $\widetilde{\mathbf{P}} = \widetilde{\mathbf{P}}_p$ and $\widetilde{\mathbf{T}} = \widetilde{\mathbf{T}}_p$. The construction of the algorithm guarantees that $(\widetilde{\mathbf{P}}^T\widetilde{\mathbf{\Sigma}}\widetilde{\mathbf{P}})^{-1} = \widetilde{\mathbf{T}}$ is tridiagonal.

Definition 4. *Let $\mathcal{C}$ denotes the index set of all $\frac{p(p-1)}{2}$ possible initializations.*

The greedy algorithm is as follow:

Greedy Algorithm

$\forall c \in \mathcal{C}$,

- Initialization Step ($l = 2$):
  - $\widetilde{\mathbf{P}}_2^c = \mathbf{P}_2^c$ where $\mathbf{P}_2^c \in \mathcal{P}_2$
  - $\widetilde{\mathbf{T}}_2^c = (\widetilde{\mathbf{P}}_2^{cT}\mathbf{\Sigma}\widetilde{\mathbf{P}}_2^c)^{-1}$
- l-th Step, $l \leftarrow l + 1$, until $l = p$:
  - Find σ_l^{c2} as a function of w using Theorem 2
  - Find $\widetilde{\mathbf{P}}_l^c$ by exhaustive search and w_l^c numerically from

$$(\widetilde{\mathbf{P}}_l^c, w_l^c) = \underset{\substack{\mathbf{P}_l\in\mathcal{P}_l:\\ \widetilde{\mathbf{P}}_{l-1}^{cT}\mathbf{P}_l\in\{[\mathbf{I}_{l-1}\ \underline{0}],[\underline{0}\ \mathbf{I}_{l-1}]\},}}{\arg\min} \min_w \mathcal{D}'_l\left((\mathbf{T}^c(w) + \underline{e}_l\sigma_l^{c-2}\underline{e}_l^T)^{-1}\right)$$

  - $\widetilde{\mathbf{T}}_l^c = \mathbf{T}^c(w_l) + \underline{e}_l\sigma_l^{c-2}\underline{e}_l^T$

After the p-th step of greedy algorithm, find c^* as:

$$c^* = \underset{c\in\mathcal{C}}{\arg\min}\ \mathcal{D}\left(\widetilde{\mathbf{P}}_p^c\widetilde{\mathbf{T}}_p^{c\,-1}\widetilde{\mathbf{P}}_p^{c\,T}\right)$$

The greedy solution is: $\widetilde{\mathbf{\Sigma}}_p^* = \widetilde{\mathbf{P}}_p^{c^*}\widetilde{\mathbf{T}}_p^{c^*\,-1}\widetilde{\mathbf{P}}_p^{c^*\,T}$

Here, $\widetilde{\mathbf{\Sigma}}_p^*$ is the greedy solution and is an upper bound for the optimal solution, i.e., $\mathcal{D}(\widetilde{\mathbf{\Sigma}}_p^*) \geq \mathcal{D}(\widetilde{\mathbf{\Sigma}}^*)$. Ignoring the computation of finding numerical solution for w_l at the l-th step (for $l > 2$) of greedy algorithm, number of search is $2(p - l + 1)$. Since size of set $\mathcal{C}$ is equal to $\frac{p(p-1)}{2}$, number of total search to find the greedy solution is $\frac{p(p-1)}{2}\sum_{l=3}^{p} 2(p-l+1) = \frac{p(p-1)^2(p-2)}{2}$ which is polynomial in p, $\mathcal{O}(p^4)$, and is better than $\mathcal{O}(p!)$. The computation of this algorithm is done off-line to find a greedy solution for the first-order Markov process approximation. After this algorithm is completed we can then run the BP algorithm replacing the original covariance matrix of REGs with the approximated covariance matrix resulting in the overall graph having less loops.

IV. SIMULATION RESULTS

In this section, we perform some simulations to validate our results. To do so, we generate the circuit presented in Fig. 1 which consist of a substation, 11 Buses, 30 branches, 6 correlated REGs and 4 observation sensors. Table I presents value of parameters in our simulations. To generate the correlation matrix of REGs, we assume two dimensional location for REGs as follow: $L(x, y) \in \{(3, 2), (4, 1), (1, -2), (3, -1), (3, -3), (4, -3)\}$. Let d_{ij} be the distance between the i-th and the j-th REG and $\zeta > 0$ and $\kappa > 0$ be constants, then the correlation model can be defined as $\varphi_{ij} \triangleq e^{\frac{-d_{ij}^2}{2\zeta}}$ and so, the covariance matrix is $\mathbf{\Phi} = [\varphi_{ij}] + \kappa\mathbf{I}$. Note that large value of ζ resulting in a more correlated covariance matrix [4].

Fig. 2 plots the discrepancy function vs. the observation noise variance for modified Burg method [5] and greedy method while $\kappa = 1$ and $\zeta = 10$. This figure indicates that the proposed greedy solution performs well comparing the

modified Burg solution essentially for large values of noise variance since based on our intuition, in each step of greedy solution we jointly find σ_l^2 and w_l considering the system model and the observation noise. Note that complexity of both algorithms is polynomial in p. Complexity of greedy algorithm is $\mathcal{O}(p^4)$ while it is $\mathcal{O}(p^3)$ for the modified Burg algorithm.

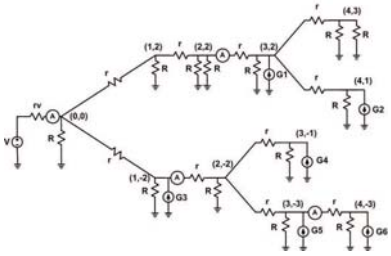

Fig. 1. Microgrid example with renewable generators.

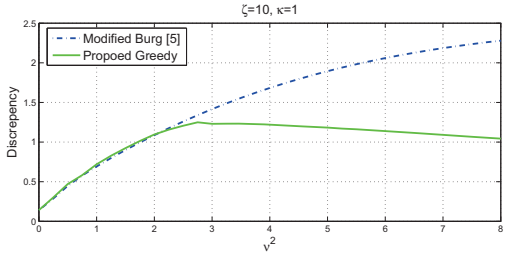

Fig. 2. Discrepancy function vs. the observation noise variance for modified Burg method and proposed greedy method.

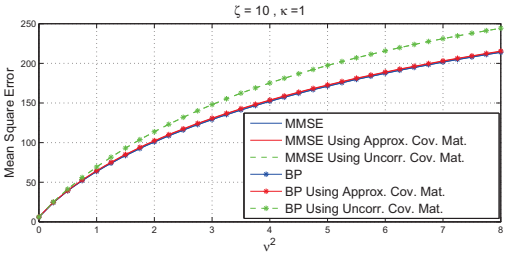

Fig. 3. Mean square error behavior for different centralized (MMSE) and distributed (BP) algorithms.

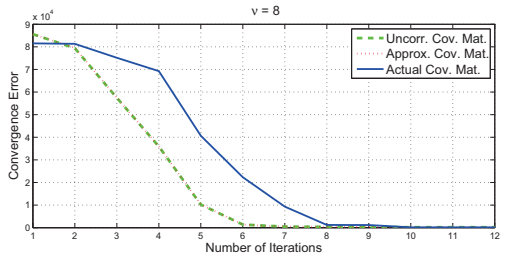

Fig. 4. Convergence error behavior of BP algorithm using uncorrelated, approximated first-order Markov process and actual REGs' covariance matrix.

Fig. 3 indicates the behavior of the centralized (MMSE) and the distributed (BP) estimation algorithms. Fig. 4 plots the convergence behavior of the BP algorithm for different iterations when we use uncorrelated, approximated and original covariance matrix of REGs. As it is shown in the figure, the BP algorithm using original covariance matrix takes more iterations to converge compared to the BP algorithm which uses the approximated covariance matrix while their performance in terms of MMSE is almost the same. The BP algorithm using the uncorrelated covariance matrix takes less iterations to converge than the other two methods, but it loses the estimation accuracy in terms of MMSE compared to other methods.

V. Conclusion

The optimal MMSE solution involves inverting large matrices and so is infeasible due to complexity. This paper presents a distributed state estimators by performing loopy GBP algorithm, since states and sensor observations has jointly Gaussian distribution. To assure the convergence of loopy GBP algorithm, we approximate the covariance matrix of inputs using greedy algorithm which gives us a simple distributed state estimator with good performance.

Acknowledgment

This work was supported in part by NSF grants ECCS-098344, D0E grant, DE-0E0000394, and the University of Hawaii REIS project. The authors acknowledge the help of Prof. Aleksander Kavcic in formulation and solving this problem.

References

[1] H.-A. Loeliger, J. Dauwels, J. Hu, S. Korl, L. Ping, and F. R. Kschischang, "The factor graph approach to model-based signal processing," *Proceedings of the IEEE*, vol. 95, no. 6, pp. 1295–1322, June 2007.

[2] L. Xie, D.H. Choi, S. Kar, , and H.V. Poor, "Fully distributed state estimation for wide-area monitoring systems," *IEEE Transactions on Smart Grid*, vol. 3, no. 3, pp. 1154–1169, Sept. 2012.

[3] V. Kekatos and G. B. Giannakis, "Distributed robust power system state estimation," *IEEE Trans. on Power Systems*, vol. 28, pp. 1617 – 1626, May 2013.

[4] Y. Hu, A. Kuh, T. Yang, and A. Kavcic, "A belief propagation based power distribution system state estimator," *IEEE Computational Intelligence Magazine*, vol. 6, no. 3, pp. 36–46, 2011.

[5] J. P. Burg, "Maximum entropy spectral analysis," *in Proc.37th Meet.Society of Exploration Geophysicists, 1967. Reprinted in Modern Spectrum Analysis, D. G. Childers, Ed. New York: IEEE Press*, pp. 34–41, 1978.

[6] A. P. Dempster, "Covariance selection," *Biometrics*, vol. 28, no. 1, pp. 157–175, March 1972.

[7] S. Vichik and Y. Oshman, "Optimal covariance selection for estimation using graphical models," *American Control Conference (ACC), 2011*, pp. 5049 –5054, June 29 - July 01 2011.

[8] Y.C. Eldar and N. Merhav, "A competitive minimax approach to robust estimation and random parameters," *IEEE Transactions on Signal Process.*, vol. 52, pp. 1931–1946, 2004.

[9] S. M. Kay, *Fundamentals of Statistical Signal Processing: Estimation Theory*, Englewood Cliffs, NJ: PTR Prentice Hall, 1993.

[10] K. B. Petersen and M. S. Pedersen, *The Matrix Cookbook*, Technical University of Denmark, http://www2.imm.dtu.dk/pubdb/p.php?3274, October 2008.

[11] 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.