2015 European Control Conference (ECC)
July 15-17, 2015. Linz, Austria

A Distributed MAP Approach to Dynamic State Estimation with Applications in Power Networks

Yibing Sun[1], Minyue Fu[2], Bingchang Wang[1] and Huanshui Zhang[1]

Abstract— **This paper studies a state estimation problem for a networked dynamic system characterized by a communication graph. A new distributed state estimation method is based on a distributed MAP (maximum *a posteriori*) estimation algorithm for each node to update its local state. This distributed method is applied to the state estimation problem for a large power network and illustrated using the IEEE 118-bus system. It is shown that the performance of this method is close to that given by a centralized Kalman filtering approach and much better than that given by a local Kalman filtering approach, yet the computational complexity and communication load of the proposed method are low for each node, making the method scalable for large-sized networked systems.**

I. INTRODUCTION

Since the state estimation problem for power networks was introduced in [1] four decades ago, it has remained a fertile research area. Indeed, this area has received increasing attention from researchers in different fields in recent years, owing to the multidisciplinary nature of smart grid [2]. State estimators are broadly utilized to obtain an optimal estimation from redundant noisy measurements, and to estimate the state of a subnetwork which is not directly monitored for computational or economical reasons. The state estimation module is one of the key modules in the energy management system (EMS), playing a vital role in power dispatch, economic optimization, security analysis, voltage stability analysis, and fault detection, diagnosis and recovery [3].

Traditional state estimation algorithms are centralized, mostly based on the Kalman filtering technique [4]. However, distributed state estimation algorithms are necessary for large-sized networked systems, such as power networks, traffic networks and mobile sensor networks. State estimation algorithms can be divided into static and dynamic types. Under stationary operational conditions, power systems are usually treated as quasi-static systems. Many studies have been made on distributed static estimation approaches, and we refer the readers to [5-7]. Tai *et al.* [7] proposed a distributed weighted least-squares (WLS) estimation approach for static state estimation with the property that the local estimates converge to the same estimates obtained via the centralized estimator. At present, the static state estimators are widely used in the power networks under the reliable operation of the transmission and distribution systems. When transient dynamics are considered, power networks are typically modeled as dynamic systems. The dynamic change of loads gives rise to the adjustment of generations, which in turn leads to a change in the flows and injections at all buses. These dynamic changes can not be captured by the traditional static state estimation, thus promoting the development of dynamic state estimation [8-12]. Compared with the traditional static state estimation methods, these dynamic schemes have better accuracy and the ability to predict the future state, which is valuable for the operator to perform the security analysis and contingencies. Hence, the forecasting ability of dynamic state estimation plays an important role in the improvement of the overall EMS control and operation.

In this paper, we consider a dynamic state estimation problem of linear dynamic systems, with applications to power networks where the state vector consists of the voltage magnitude and angle at all buses in power networks. The objective of this paper is to develop a fully distributed dynamic estimation scheme for large-scale systems. Under the assumption that the communication graph of the network is acyclic, which is valid for many power networks, we provide a distributed MAP estimation algorithm to update the local state estimate at each time step and then apply local predictors to give the one-step-ahead prediction for the state vector which is treated as the priori to calculate the state estimate at the next sampling time. Compared to the centralized state estimation algorithm, this distributed algorithm offers suboptimal state estimation. But the major advantage of the distributed algorithm is that only local computation and communication are needed. We also show that the distributed algorithm at steady state converges in a finite number of iterations, which equals to the maximum path length of the acyclic graph. The main contribution of this paper is that it generalizes known results on static state estimation [7] to the dynamic case, and results in a much lighter communication load than that of [5] and [6]. Demonstrated via the IEEE-118 bus system, the performance of the proposed method is close to that given by a centralized Kalman filtering approach and much better than that given by a local Kalman filtering approach, yet the computational complexity and communication load of the proposed method are low for each node, making the method scalable for large-sized networked systems.

*This work is supported by the Taishan Scholar Construction Engineering by Shandong Government, the National Natural Science Foundation of China under Grants 61120106011, 61203029.

[1]Yibing Sun, Bingchang Wang and Huanshui Zhang are with School of Control Science and Engineering, Shandong University, Jinan 250061, P. R. China sun_yibing@126.com; bcwang@sdu.edu.cn; hszhang@sdu.edu.cn

[2]Minyue Fu is with School of Electrical Engineering and Computer Science, University of Newcastle, NSW 2308, Australia minyue.fu@newcastle.edu.au

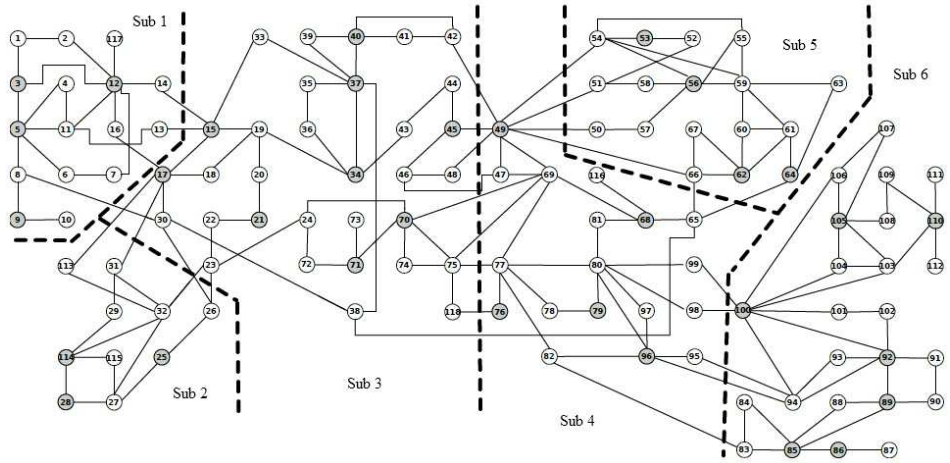

Fig. 1. Topological structure of the IEEE 118-bus system.

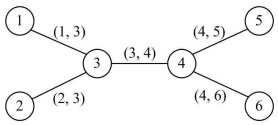

Fig. 2. The graph $\mathcal{G}$ depicting the partition of the 118-bus system.

II. PROBLEM STATEMENT

In this paper, we study a networked system with a connected graph $\mathcal{G} = (\mathcal{V},\ \mathcal{E})$, where each node $i \in \mathcal{V} = \{1,\ \cdots,\ n\}$ denotes a node corresponding to a control area and $\mathcal{E} \subset \mathcal{V} \times \mathcal{V}$ denotes the set of edges $(i,\ j)$ interconnecting nodes i and j. Let $\mathcal{N}_i = \{j :\ (i,\ j) \in \mathcal{E}\}$ be the set of neighbors of node i. It is assumed that graph is connected and undirected (i.e., there exists a two-way path between each pair of nodes) and that the graph is void of self-loops and multiple edges.

For a connected graph without loops, the length of a path is the number of edges forming it. The radius of node i is defined as the maximum length of a path between node i and any other node j in the graph, denoted by ε_i. The diameter of the graph is $\Gamma = \max\{\varepsilon_i,\ i \in \mathcal{V}\}$.

Consider the linear dynamic model for the power systems:

$$x_i(k+1) = A_i x_i(k) + \omega_i(k),\ i \in \mathcal{V}, \tag{1}$$

where $x_i(k) \in \mathbb{R}^{s_i}$ is the state of node i and $\omega_i(k) \in \mathbb{R}^{s_i}$ is the associated process noise.

Two types of measurements are available:

$$y_i(k) = C_i x_i(k) + \nu_i(k), \tag{2}$$

$$z_{i,j}(k) = B_{ij} x_i(k) + B_{ji} x_j(k) + \nu_{i,j}(k), \tag{3}$$

where $y_i(k) \in \mathbb{R}^{q_i}$ is the measurement of node i, $z_{i,j}(k) \in \mathbb{R}^{q_{ij}}$ is the measurement describing the interaction between node i and node j, $\nu_i(k) \in \mathbb{R}^{q_i}$ and $\nu_{i,j}(k) \in \mathbb{R}^{q_{ij}}$ are the associated measurement noises. We will call $y_i(k)$ the *local measurements* and $z_{i,j}(k)$ the *edge measurements*. The sample time k takes values of $0, 1, 2, \ldots$. The constant matrices A_i, C_i, B_{ij} and B_{ji} are the state transition matrix and the measurement matrices of appropriate dimensions. For simplicity, it is assumed that noises $\omega_i(k)$, $\nu_i(k)$ and $\nu_{i,j}(k)$ are independent white Gaussian with zero mean and covariances R_i, S_i and $T_{i,j}$, respectively, and the initial state $x_i(0)$ is also an independent Gaussian variable with mean $\bar{x}_i(0)$ and covariance $\Sigma_i(0)$.

The measurement model above is motivated by power networks where the state of each bus (or node) is measured by a local control center, and the edge measurement represents the so-called tie-line measurement between two subnetworks. For the partition made of the 118-bus system in Fig. 1, the graph $\mathcal{G}$ is described as Fig. 2, for which the edge (1,3) as an example means the existence of an edge measurement related to the subsystems 1 and 3. We note that edge measurements are natural for describing physical interactions in a large-scale dynamic system.

We have the following two assumptions:

Assumption 1: The graph $\mathcal{G}$ is acyclic.

Assumption 2: The measurement matrix C_i has full column rank, while the covariances of noises R_i, S_i and $T_{i,j}$, and the initial state $\Sigma_i(0)$ are invertible for $i \in \mathcal{V}$, $j \in \mathcal{N}_i$.

Assumption 1 means that the graph does not have a cycle. Note that this is not a severe requirement for power networks because many power networks can be partitioned into acyclic graphs, as shown in Figs. 1 and 2. Assumption 2 guarantees that state estimation error will be bounded and positive definite. Again, this is a common assumption for state estimation in power systems.

III. CENTRALIZED STATE ESTIMATION

In this section, we discuss the centralized state estimation approach. Firstly, we find that (1), (2) and (3) can be stacked. Denote the aggregated state $x(k) = (x_1^T(k),\ \cdots,\ x_n^T(k))^T$ and measurement $z(k) = (\cdots,\ y_i^T(k),\ \cdots,\ z_{i,j}^T(k),\ \cdots)^T$, so the state and measurement equations can be written

$$x(k+1) = Ax(k) + \omega(k), \tag{4}$$

$$z(k) = Hx(k) + \nu(k), \tag{5}$$

respectively, with $A = \text{diag}\,[A_1,\ \cdots,\ A_n]$,

$$H = \begin{bmatrix} & & & \cdots & & & \\ \cdots & 0 & C_i & 0 & \cdots & & \\ & & & \cdots & & & \\ \cdots & 0 & B_{ij} & 0 & B_{ji} & 0 & \cdots \\ & & & \cdots & & & \end{bmatrix},$$

$\omega(k) = (\omega_1^T(k),\ \cdots,\ \omega_n^T(k))^T$ and $\nu(k) = (\cdots,\ \nu_i^T(k),\ \cdots,\ \nu_{i,j}^T(k),\ \cdots)^T$, with covariance

$$R = \text{cov}(\omega(k)) = \text{diag}[R_1,\ \cdots,\ R_n],$$
$$R_* = \text{cov}(\nu(k)) = \text{diag}[\cdots,\ S_i,\ \cdots,\ T_{i,j},\ \cdots],$$

respectively. The initial state $x(0)$ is aggregated with mean $\bar{x}(0) = (\bar{x}_1^T(0),\ \cdots,\ \bar{x}_n^T(0))^T$ and covariance $\Sigma(0) = \text{diag}[\Sigma_1(0),\ \cdots,\ \Sigma_n(0)]$.

The centralized state estimator is a standard Kalman filter, which involves two parts: centralized maximum *a posteriori* (MAP) estimator and centralized one-step-ahead predictor. They are detailed below.

A. Centralized MAP Estimation

Let $\hat{x}(k|k)$ denote the estimate for the state vector at time instant k conditioned on the measurements available up to time k, and $\Sigma(k|k)$ be the associated estimation error covariance matrix. From [13], We get

$$\hat{x}(k|k) = \mathbb{E}\{x(k)|Z(k)\},$$
$$\Sigma(k|k) = \mathbb{E}\{(x(k) - \hat{x}(k|k))(x(k) - \hat{x}(k|k))^T|Z(k)\},$$

where $Z(k) = \{z(0),\ z(1),\ \cdots,\ z(k)\}$. The MAP estimation approach [14] is to compute

$$\hat{x}(k|k) = \arg\max_{x(k)} p(x(k)|Z(k)), \tag{6}$$

where $p(x(k)|Z(k))$ denotes the probability density function of $x(k)$ conditioned on the measurements $Z(k)$. Using the method in [15], the MAP estimator (6) is equivalent to

$$\hat{x}(k|k) = \arg\max_{x(k)} p(z(k)|x(k))p(x(k)|Z(k-1)). \tag{7}$$

From (5), (7) becomes

$$\hat{x}(k|k) = \arg\min_{x(k)} \Big[\big(z(k) - Hx(k)\big)^T R_*^{-1}\big(z(k) - Hx(k)\big) + \big(x(k) - \hat{x}(k|k-1)\big)^T \Sigma^{-1}(k|k-1)\big(x(k) - \hat{x}(k|k-1)\big)\Big], \tag{8}$$

where the variable $\hat{x}(k|k-1)$ denotes the prediction of the state at time instant k, based on the measurements available up to time $k-1$, and its error covariance matrix is $\Sigma(k|k-1)$. Then, it is easy to obtain the optimal estimation value as

$$\hat{x}(k|k) = \Sigma(k|k)\big(H^T R_*^{-1} z(k) + \Sigma^{-1}(k|k-1)\hat{x}(k|k-1)\big),$$

and the estimation error covariance is

$$\Sigma(k|k) = \big(H^T R_*^{-1} H + \Sigma^{-1}(k|k-1)\big)^{-1}, \tag{9}$$

which is initialized by $\hat{x}(0|-1) = \bar{x}(0)$ and $\Sigma(0|-1) = \Sigma(0)$. In this way we get the optimal estimate at each time instant in a recursive manner.

B. Centralized Prediction

This step is about optimal prediction of $x(k+1)$ using the available measurements $Z(k)$, i.e., we want to compute the conditional mean $\hat{x}(k+1|k)$. From (4) and the obtained estimation $\hat{x}(k|k)$ with its error covariance $\Sigma(k|k)$, we have

$$\hat{x}(k+1|k) = A\hat{x}(k|k), \quad \Sigma(k+1|k) = A\Sigma(k|k)A^T + R.$$

Once the measurement at time $k+1$ becomes available, $\hat{x}(k+1|k)$ and $\Sigma(k+1|k)$ will become the priori information for computing the centralized state estimate $\hat{x}(k+1|k+1)$ in the next time step, as described in the previous subsection. Hence, the recursion goes on.

Although this is the optimal solution to the estimation problem, it is necessary to install the centralized state estimator in a control center to collect all measurements over the entire network, and requires a powerful computer to do relatively heavy calculations. This will create heavy computation burden and communication bottleneck when the network size becomes large. In the next two sections, we will investigate the local state estimation and distributed state estimation methods, respectively.

IV. LOCAL STATE ESTIMATION

In this section, we describe a local state estimation method from the local viewpoint of node $i \in \mathcal{V}$, using only the local measurements. This is basically the same as the centralized state estimation method, except that the edge measurements are not employed. Naturally, this will lead to a sub-optimal estimator. We introduce this method for the purpose of comparison later.

For every $i \in \mathcal{V}$, the local Kalman filtering algorithm, which is obtained using (1) and (2), is given by the following recursive equations:

$$\hat{x}_i(k|k) = \hat{x}_i(k|k-1) + K_i(k)\big(y_i(k) - C_i\hat{x}_i(k|k-1)\big),$$
$$\Sigma_i(k|k) = \Sigma_i(k|k-1) - K_i(k)C_i\Sigma_i(k|k-1),$$
$$\hat{x}_i(k+1|k) = A_i\hat{x}_i(k|k),$$
$$\Sigma_i(k+1|k) = A_i\Sigma_i(k|k)A_i^T + R_i,$$

where

$$\Xi_i(k) = C_i\Sigma_i(k|k-1)C_i^T + S_i,$$
$$K_i(k) = \Sigma_i(k|k-1)C_i^T\Xi_i^{-1}(k).$$

V. DISTRIBUTED STATE ESTIMATION

This section bears the main contribution of this paper. We will describe the proposed distributed state estimation method. This method also involves two parts: a distributed MAP estimator at each subsystem replacing the centralized MAP estimator in the centralized center, and a local state predictor which is the same as in the local state estimator.

A. Distributed MAP

In the distributed MAP estimator, each node $i \in \mathcal{V}$ obtains an estimate $\hat{x}_i(k|k)$ for the local state $x_i(k)$ at time k, by using (2), (3) and the exchanged information from its

neighbors $j \in \mathcal{N}_i$. The purpose is to minimize the global objective function

$$J(x(k)) = \big(z(k) - Hx(k)\big)^T R_*^{-1} \big(z(k) - Hx(k)\big)$$
$$+\big(x(k) - \hat{x}(k|k-1)\big)^T \breve{\Sigma}^{-1}(k|k-1)\big(x(k) - \hat{x}(k|k-1)\big),$$

where $\breve{\Sigma}(k|k-1) = \text{diag}[\Sigma_1(k|k-1),\ \cdots,\ \Sigma_n(k|k-1)]$ with $\Sigma_i(k|k-1)$, $i = 1,\ \cdots,\ n$ being the block diagonal sub-matrix of $\Sigma(k|k-1)$.

From the problem description, we can see that although we assume that $\Sigma(0)$ is a block diagonal matrix in the global model, $\Sigma(k-1|k-1)$ is no longer diagonal at time $k = 1,\ 2,\ \cdots$, and then $\Sigma(k|k-1)$ is not diagonal. In order to achieve the distributed estimation scheme, in (9) we only use the diagonal elements of $\Sigma(k|k-1)$, i.e., we replace this matrix with $\breve{\Sigma}(k|k-1)$ to obtain $\Sigma(k|k)$. Hence, this leads to a suboptimal estimate.

Denote the suboptimal estimate by $\hat{x}^*(k|k)$ with its estimation error covariance $\Sigma^*(k|k)$, where $\hat{x}^*(k|k) = \arg\min_{x(k)} J(x(k))$. Also denote $\Sigma_i^*(k|k)$ $(i = 1,\ \cdots,\ n)$ as the block diagonal sub-matrix in $\Sigma^*(k|k)$ corresponding to the state $x_i(k)$ on node i. Furthermore, notice that the edge measurement $z_{i,j}(k)$ is usually of lower dimension, so every control center indeed only needs to transmit the reduced dimensional information $B_{ji}\hat{x}_j(k|k)$ instead of the complete local estimation $\hat{x}_j(k|k)$ to its neighbors and this indicates a very light communication load [7].

The construction of the distributed MAP estimator is split into three steps, which is developed in three lemmas.

Lemma 1 [7] *Consider a distributed state estimation scheme under Assumptions 1 and 2. Let $\hat{x}_i(k|k)$ and $\Sigma_i(k|k)$ be the resulting estimate and the associated estimation error covariance for $i \in \mathcal{V}$, respectively. If $\Sigma_i(k|k) = \Sigma_i^*(k|k)$, then $\hat{x}^*(k|k) = (\hat{x}_1^T(k|k)),\ \cdots,\ \hat{x}_n^T(k|k))^T$ minimizes $J(x(k))$ at time k.*

For each node $i \in \mathcal{V}$ and $j \in \mathcal{N}_i$, define

$$\bar{\alpha}_i(k) = C_i^T S_i^{-1} y_i(k) + \Sigma_i^{-1}(k|k-1)\hat{x}_i(k|k-1),$$
$$\bar{Q}_i(k) = C_i^T S_i^{-1} C_i + \Sigma_i^{-1}(k|k-1),$$
$$y_{ji}(k) = z_{i,j}(k) - \beta_j^i(k),\quad \beta_j^i(k) = B_{ji}\bar{Q}_j^{-1}(k)\bar{\alpha}_j(k),$$
$$S_{ji}(k) = T_{i,j} + \Phi_j^i(k),\quad \Phi_j^i(k) = B_{ji}\bar{Q}_j^{-1}(k)B_{ji}^T.$$

The following lemma studies the distributed MAP estimation of the local state in a graph with only two nodes.

Lemma 2 *Consider the system (1)-(3) with two nodes, i.e., nodes $i = 1,\ 2$. Under the distributed MAP estimation scheme, it can be obtained that*

$$\hat{x}_i^*(k|k) = \Sigma_i^*(k|k)(\bar{\alpha}_i(k) + B_{ij}^T S_{ji}^{-1}(k) y_{ji}(k)),$$
$$\Sigma_i^*(k|k) = (\bar{Q}_i(k) + B_{ij}^T S_{ji}^{-1}(k) B_{ij})^{-1},$$

where $y_{ji}(k)$ and $S_{ji}(k)$ are defined in the above. Furthermore,

$$\Sigma_i(k|k,\ \varepsilon_i + l) = \Sigma_i^*(k|k),\quad for\ all\ l \geq 0,$$

where $\varepsilon_i = 1$ and $\varepsilon_i + l$ is the step number of iteration.

As shown in Fig. 3, in an acyclic graph, if one and only one node's radius equals to 1 and the radii of all the other nodes is 2, then the graph is called a radial graph. The node with 1 radius indicated by i is known as the central node. The following lemma generalizes the result of two-node graph to the radial graph.

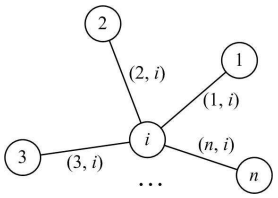

Fig. 3. Topological structure of a radial graph.

Lemma 3 *Suppose that Assumptions 1 and 2 hold. At time instant k, the distributed MAP estimation and the estimation error covariance of node $i \in \mathcal{V}$ are given by*

$$\hat{x}_i^*(k|k) = \Sigma_i^*(k|k)(\bar{\alpha}_i(k) + \sum_{j\in\mathcal{N}_i} B_{ij}^T S_{ji}^{-1}(k) y_{ji}(k)),$$
$$\Sigma_i^*(k|k) = (\bar{Q}_i(k) + \sum_{j\in\mathcal{N}_i} B_{ij}^T S_{ji}^{-1}(k) B_{ij})^{-1}.$$

A radial graph, as shown in Fig. 3, can be seen as a two-node graph, by combining nodes $1,\ \cdots,\ n$ into a single node. Thus, the proof is omitted due to space limitation.

Based on Lemma 3, we can design the main algorithm. Initially, every control center j calculates the local estimation $\hat{x}_j(k|k,\ 0)$ and the estimation error covariance $\Sigma_j(k|k,\ 0)$, using its local measurements and initial information of state. Then node i utilizes information $\beta_j^i(k,\ h)$ and $\Phi_j^i(k,\ h)$ received from its neighbor j to update the edge measurement, where h is the step number of iteration. Meanwhile, control center i computes $\alpha_i^j(k,\ h)$ and $Q_i^j(k,\ h)$ related to its most recently updated local estimation and estimation error covariance, and whereafter transmits them to its neighbors.

Algorithm 1 Distributed MAP estimation algorithm

For each $i \in \mathcal{V}$ and at time step $k = 0,\ 1,\ \cdots$:

1) Node i computes its local estimation and covariance:

$$\hat{x}_i(k|k,\ 0) = \bar{Q}_i^{-1}(k)\bar{\alpha}_i(k),\quad \Sigma_i(k|k,\ 0) = \bar{Q}_i^{-1}(k),$$

which are initialized by $\bar{x}_i(0)$ and $\Sigma_i(0)$.

2) Node i transmits the following information to $j \in \mathcal{N}_i$:

$$\beta_i^j(k,\ 0) = B_{ij}\hat{x}_i(k|k,\ 0),\quad \Phi_i^j(k,\ 0) = B_{ij}\Sigma_i(k|k,\ 0)B_{ij}^T.$$

3) Main loop ($h = 1,\ 2,\ \cdots$ is the step of iteration):

(a) Using the information transmitted from $j \in \mathcal{N}_i$, node i updates the edge information

$$y_{ji}(k,\ h) = z_{i,j}(k) - \beta_j^i(k,\ h-1),$$
$$S_{ji}(k,\ h) = T_{i,j} + \Phi_j^i(k,\ h-1).$$

(b) Node i computes the current state estimation and error covariance matrix:

$$\hat{x}_i(k|k,\ h) = Q_i^{-1}(k,h)\alpha_i(k,h),\ \ \Sigma_i(k|k,\ h) = Q_i^{-1}(k,h),$$

where

$$\alpha_i(k,\ h) = \bar{\alpha}_i(k) + \sum_{j\in\mathcal{N}_i} B_{ij}^T S_{ji}^{-1}(k,\ h) y_{ji}(k,\ h),$$
$$Q_i(k,\ h) = \bar{Q}_i(k) + \sum_{j\in\mathcal{N}_i} B_{ij}^T S_{ji}^{-1}(k,\ h) B_{ij}.$$

(c) Node i computes

$$\alpha_i^j(k,\ h) = \bar{\alpha}_i(k) + \sum_{m\in\mathcal{N}_i/\{j\}} B_{im}^T S_{mi}^{-1}(k,\ h) y_{mi}(k,\ h),$$
$$Q_i^j(k,\ h) = \bar{Q}_i(k) + \sum_{m\in\mathcal{N}_i/\{j\}} B_{im}^T S_{mi}^{-1}(k,\ h) B_{im},$$

and then transmit the following information to node j

$$\beta_i^j(k,\ h) = B_{ij}[Q_i^j(k,\ h)]^{-1}\alpha_i^j(k,\ h),$$
$$\Phi_i^j(k,\ h) = B_{ij}[Q_i^j(k,\ h)]^{-1}B_{ij}^T.$$

□

Theorem 1 *Suppose that Assumptions 1 and 2 hold. At time instant k, if Algorithm 1 is used, then for every $i \in \mathcal{V}$,*

$$\hat{x}_i(k|k,\ \varepsilon_i + l) = \hat{x}_i^*(k|k),$$
$$\Sigma_i(k|k,\ \varepsilon_i + l) = \Sigma_i^*(k|k),\quad for\ all\ l \geq 0,$$

where $\varepsilon_i + l$ is the step number of iteration.

Proof: The proof considers the state estimation of node i from its local point of view. At the first iteration $h = 1$, consider the radial sub-graph of $\mathcal{G}$, which has node i as the central node and all its neighbors. Apply Algorithm 1 together with Lemma 3 to this radial graph. After the first iteration, node i can obtain the desired suboptimal estimate corresponding to this sub-graph.

When the iteration $h = 2$, consider the radial sub-graph of $\mathcal{G}$, having i as the central node and all its neighbors, in which we combine each node $j \in \mathcal{N}_i$ with its neighbors but except node i into a single node as the neighbor of node i. Applying Algorithm 1, each node j computes the quantities $\beta_j^i(k,2)$ and $\Phi_j^i(k,2)$, and sends them to node i, in which they are builded by using the information $\beta_m^j(k,1)$ and $\Phi_m^j(k,1)$ previously received from node $m \in \mathcal{N}_j/\{i\}$. Then, the case of the second iteration is equivalent to using Lemma 3 to the above mentioned radial sub-graph, and node i is capable of computing the suboptimal estimate and estimation error covariance of this sub-graph formed by all nodes, which are less than or equal to two hops away from it.

At the iteration h, combine each node $j \in \mathcal{N}_i$ together with all the nodes, which are less than or equal to $h-1$ hops away from it, but without node i into a single node. Then consider the radial sub-graph which is formed by these nodes and has node i as the central one. Similar to the above argument, we get that node i is capable of computing the suboptimal estimate of this radial graph formed by all nodes, which are less than or equal to h hops away from node i. Since there is no more information transmitted from nodes after $h \geq \Gamma_i$ at each time k, node i will obtain the suboptimal estimate and its covariance corresponding to the whole graph $\mathcal{G}$, and this value will remain unchanged in the subsequent iterations. Then we achieve (14) for all $i \in \mathcal{V}$. □

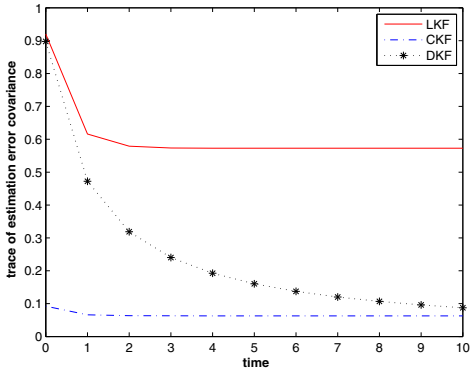

Fig. 4. The traces of the covariances of the three methods

It can be seen from (14) that the local estimates on all nodes converge after Γ steps.

B. Local Prediction

Based on (1) and the distributed state estimate $\hat{x}_i^*(k|k)$ with the associated error covariance $\Sigma_i^*(k|k)$, it is easy to obtain the state prediction with its covariance matrix

$$\hat{x}_i(k+1|k) = A_i\hat{x}_i^*(k|k), \Sigma_i(k+1|k) = A_i\Sigma_i^*(k|k)A_i^T + R_i.$$

Once the measurements at time $k+1$ become available, $\hat{x}_i(k+1|k)$ and $\Sigma_i(k+1|k)$ are used to build local state estimation $\hat{x}_i(k+1|k+1)$ and covariance $\Sigma_i(k+1|k+1)$, for each node $i \in \mathcal{V}$. Hence, Algorithm 1 goes on.

VI. SIMULATION RESULTS

In this section, we use the IEEE 118-bus system to test the performance of the distributed state estimator (or distributed Kalman filter (DKF)) by comparing with the centralized state estimator (or centralized Kalman filter (CKF)) and the local Kalman filter (LKF). As shown in Fig. 1, this 118-bus system is partitioned into 6 subsystems and the induced topology is acyclic. These subsystems are seen as nodes and connected by tie-lines, which is shown in Fig. 2. To simplify the simulations, we assume that phasor measurement units (PMUs) are used for measurements so that linear measurements are available.

The computational loads of the centralized and distributed MAP estimators are $O((\sum_{i=1}^n s_i)^3)$ and $\tilde{n}_i O(s_i^3)$, where $\tilde{n}_i$ denotes the cardinality of $\mathcal{N}_i$. The computational load of each distributed estimator relates to the number of its neighbors. Since each node only has a few neighbors, i.e., $\tilde{n}_i \ll n$, the computational load of the proposed distributed algorithm is smaller than that of the centralized method.

We run these three estimation algorithms under Assumptions 1 and 2. The sum of the trace of the estimation error covariance can be described as $\sum_{i=1}^n \mathrm{Tr}\{\Sigma_i(k|k)\}$, which can be used to describe all methods. We will use them to compare the performances of these estimation algorithms.

Fig. 4 shows that the trace of estimation error covariance of the DKF algorithm decreases faster than that of the LKF

scheme, which means that the performance of the DKF estimation algorithm is better than the LKF method. After converge to some constant, the estimation value of DKF scheme is only marginally larger than that of the CKF.

VII. CONCLUSIONS

A distributed method for state estimation is presented for large-scale networked systems with applications in multi-area interconnected power networks. The core of this new method is the distributed MAP estimator. Under the assumption that network graph is acyclic, the distributed MAP estimator converges to the suboptimal estimate at each local area in the steady state after a finite number of iterations, which equals to the diameter of the graph. The distributed estimation algorithm enjoys low computational complexity and communication load. Yet, when applied to power networks, its estimation accuracy is only marginally worse than the centralized MAP estimation algorithm, as demonstrated on the IEEE 118-bus system.

APPENDIX

Proof of Lemma 2: The proof is divided into two parts.

A. We consider the distributed MAP estimation scheme. Firstly, we discuss the state estimation on node 1 at time instant k, as the argument goes the same for the estimation on node 2. For node 2, it is clear that

$$\hat{x}_2(k|k) = \bar{Q}_2^{-1}(k)\bar{\alpha}_2(k), \quad \Sigma_2(k|k) = \bar{Q}_2^{-1}(k).$$

The information transmitted from node 2 to node 1 is

$$\beta_2^1(k) = B_{21}\hat{x}_2(k|k), \quad \Phi_2^1(k) = B_{21}\Sigma_2(k|k)B_{21}^T.$$

Node 1 updates the edge measurement using the received estimation from node 2:

$$y_{21}(k) = z_{1,2}(k) - \beta_2^1(k), \quad S_{21}(k) = T_{1,2} + \Phi_2^1(k).$$

Following by the distributed MAP estimator (8) with $z(k) = [y_1^T(k),\ y_{21}^T(k)]^T$, $H = [C_1^T,\ B_{12}^T]^T$, $R_* = \text{diag}[S_1,\ S_{21}(k)]$, the local estimation and its error covariance on node 1 are

$$\begin{aligned}\hat{x}_1(k|k) &= \Sigma_1(k|k)(\bar{\alpha}_1(k) + B_{12}^T S_{21}^{-1}(k)y_{21}(k)),\\ \Sigma_1(k|k) &= (\bar{Q}_1(k) + B_{12}^T S_{21}^{-1}(k)B_{12})^{-1}.\end{aligned}$$

With the same proof, we can also get that

$$\begin{aligned}\hat{x}_2(k|k) &= \Sigma_2(k|k)(\bar{\alpha}_2(k) + B_{21}^T S_{12}^{-1}(k)y_{12}(k)),\\ \Sigma_2(k|k) &= (\bar{Q}_2(k) + B_{21}^T S_{12}^{-1}(k)B_{21})^{-1}.\end{aligned}$$

B. The whole dynamic system can be written as (4) and (5), with $x(k) = (x_1^T(k),\ x_2^T(k))^T$, $\omega(k) = (\omega_1^T(k),\ \omega_2^T(k))^T$, $A = \text{diag}\,[A_1,\ A_2]$, $z(k) = (y_1^T(k),\ y_2^T(k),\ z_{1,2}^T(k))^T$, $\nu(k) = (\nu_1^T(k),\ \nu_2^T(k),\ \nu_{1,2}^T(k))^T$,

$$H = \begin{bmatrix} C_1 & 0 \\ 0 & C_2 \\ B_{12} & B_{21} \end{bmatrix}, \ R = \text{diag}\,[R_1,\ R_2]\,,$$

$R_* = \text{diag}[S_1,\ S_2,\ T_{1,2}]$. At time k, consider the centralized MAP estimation, in which $\Sigma(k|k-1)$ is replaced by $\tilde{\Sigma}(k|k-1)$. According to (9), we obtain

$$\Sigma^*(k|k) = \begin{bmatrix} \Gamma_{11} & \Gamma_{12} \\ \Gamma_{12}^T & \Gamma_{22} \end{bmatrix}^{-1},$$

where

$$\begin{aligned}\Gamma_{11} &= C_1^T S_1^{-1} C_1 + B_{12}^T T_{1,2}^{-1} B_{12} + \Sigma_1^{-1}(k|k-1),\\ \Gamma_{12} &= B_{12}^T T_{1,2}^{-1} B_{21},\\ \Gamma_{22} &= C_2^T S_2^{-1} C_2 + B_{21}^T T_{1,2}^{-1} B_{21} + \Sigma_2^{-1}(k|k-1).\end{aligned}$$

According to the matrix inverse lemma, we obtain that the first diagonal block of $\Sigma^*(k|k)$ is given by

$$\Sigma_1^*(k|k) = [C_1^T S_1^{-1} C_1 + \Sigma_1^{-1}(k|k-1) + B_{12}^T S_{21}^{-1}(k) B_{12}]^{-1}.$$

Following the same argument, we have that

$$\Sigma_2^*(k|k) = [C_2^T S_2^{-1} C_2 + \Sigma_2^{-1}(k|k-1) + B_{21}^T S_{12}^{-1}(0) B_{21}]^{-1}.$$

It is straightforward to get $\Sigma_i^*(k|k) = \Sigma_i(k|k)$, $i = 1,\ 2$.

Furthermore, on the basis of Lemma 1, we get $\hat{x}^*(k|k) = (\hat{x}_1^T(k|k)),\ \hat{x}_2^T(k|k))^T$. □

REFERENCES

[1] F. C. Schweppe and J. Wildes, Power system static-state estimation, part I, II, III, IEEE Trans. Power App. Syst., vol. 89, no. 1, pp. 120-135, Jan. 1970.

[2] F. F. Wu, Power system state estimation: a survey, Elect. Power Energy Syst., vol. 12, no. 1, pp. 80-87, Jan. 1990.

[3] F. F. Wu, K. Moslehi, and A. Bose, Power system control centers: past, present, and future, in Proc. IEEE, vol. 93, no. 11, 2005, pp. 1890-1908.

[4] M. B. Do Coutto Filho and J. C. S. de Souza, Forecasting-aided state estimation-part I: panorama, IEEE Trans. Power Sys., vol. 24, no. 4, pp. 1667-1677, Nov. 2009.

[5] L. Xie, D. H. Choi, and S. Kar, Cooperative distributed state estimation: local observability relaxed, in Proc. of IEEE Power and Energy Society General Meeting, San Diego, 2011, pp. 1-11.

[6] F. Pasqualetti, R. Carli, and F. Bullo, Distributed estimation via iterative projections with application to power network monitoring, Automatica, vol. 48, no. 5, pp. 747-758, May 2012.

[7] X. Tai, Z. Lin, M. Fu, and Y. Sun, A new distributed state estimation technique for power networks, in Proc. Amer. Control Conf. (ACC), Washington, 2013, pp. 3338-3343.

[8] K. Shih and S. Huang, Application of a robust algorithm for dynamic state estimation of a power system, IEEE Trans. Power Syst., vol. 17, no. 1, pp. 141-147, Feb. 2002.

[9] A. H. Sayed and C. G. Lopes, Distributed recursive least-squares strategies over adaptive networks, in Proc. 40th Asilomar Conf. Signals, Syst. Comput., Pacific Grove, 2006, pp. 233-237.

[10] F. S. Cattivelli, C. G. Lopes, and A. H. Sayed, Diffusion recursive least-squares for distributed estimation over adaptive networks, IEEE Trans. Signal Process., vol. 56, no. 5, pp. 1865-1877, May 2008.

[11] L. Yan, X. R. Li, Y. Xia, and M. Fu, Optimal sequential and distributed fusion for state estimation in cross-correlated noise, Automatica, vol. 49, no. 12, pp. 3607-3612, Dec. 2013.

[12] J. Feng and M. Zeng, Optimal distributed Kalman filtering fusion for a linear dynamic system with cross-correlated noises, Int. J. Syst. Sci., vol. 43, no. 2, pp. 385-398, Feb. 2012.

[13] B. D. O. Anderson and J. B. Moore, *Optimal Filtering*. Englewood Cliffs, NJ: Prentice-Hall, 1979.

[14] S. M. Kay, Fundamentals of Statistical Signal Processing. Englewood Cliffs, NJ: Prentice Hall, 1993.

[15] S. Arulampalam, S. Maskell, N. Gordon, and T. Clapp, A tutorial on particle filters for online nonlinear/non-gaussian Bayesian tracking, IEEE Trans. Signal Process., vol. 50, no. 2, pp. 174-188, Feb. 2002.