Proceedings of the 34th Chinese Control Conference
July 28-30, 2015, Hangzhou, China

# Dynamic State Estimation in Power Systems Using A Distributed MAP Method

SUN Yibing[1], FU Minyue[2,3], WANG Bingchang[1], ZHANG Huanshui[1]

1. School of Control Science and Engineering, Shandong University, Jinan 250061, P. R. China
E-mail: sun_yibing@126.com, bcwang@sdu.edu.cn, hszhang@sdu.edu.cn

2. School of Control Science and Engineering, Zhejiang University, Hangzhou 310027, P. R. China

3. School of Electrical Engineering and Computer Science, University of Newcastle, NSW 2308, Australia
E-mail: minyue.fu@newcastle.edu.au

**Abstract:** This paper studies a dynamic state estimation problem for power systems, which can be seen as the quasi-static systems. The state vector of each subsystem (called node) in power networks is expressed by measurements. Based on a distributed maximum *a posteriori* (MAP) estimation technique, a fully distributed state estimation method is presented to update the local state at each time instant. Also, the assumption of local observability of every node is no longer needed. Tests on the IEEE 118-bus system are used to show the performance of the proposed approach and compare its results with a centralized state estimation method providing the optimal state estimate for the entire power networks, a local state estimation method using the edge measurements as the local measurements to estimate its local state, and a distributed static state estimation algorithm.

**Key Words:** Distributed state estimation, MAP estimation, Kalman filter, Power networks

## 1 Introduction

As our society emphasizes the importance of smarter electricity networks to support sustainable energy utilization, power networks are currently undergoing tremendous changes, and then a great variety of techniques are built into smart grids. Since Schweppe *et al.* [1] introduced the idea of (static) state estimation in power networks forty years ago, state estimation has remained an attractive and contentious research field, owing to the multidisciplinary nature of smart grids [2]. The state estimation module is a key module in the energy management system (EMS), playing a vital role in power dispatch, economic optimization, security analysis, voltage stability analysis, diagnosis and recovery [3].

Normally, distributed state estimation algorithms can be divided into static and dynamic ones. For the last few decades, there has been much research activity focused on static state estimation [4-7]. In fact, smart grids are typically dynamic systems due to the dynamic nature of system's loads. When the requirement of real-time and accurate monitoring in power networks becomes urgent, the dynamic features are considered to improve state estimation methods and the static state estimators can not efficiently and accurately capture the dynamic behaviors. In this case, the dynamic state estimation techniques were developed [8-11], in which the state vectors are estimated based on a prediction-correction process. These dynamic methods obtained in literature mainly related to the Kalman filtering technique; see [10, 11]. Compared with the static state estimation methods, these dynamic schemes have better accuracy and the ability to predict the future state, which is reasonably valuable for the operator to perform the security analysis and predict contingencies, and then take appropriate steps to counter them one time step ahead. Hence, the forecasting ability of the dynamic state estimation techniques plays an important role in the improvement of the overall EMS control and operation.

This work is supported by the Taishan Scholar Construction Engineering by Shandong Government, the National Natural Science Foundation of China under Grants 61120106011, 61203029, 61403233.

In this paper, a *fully distributed* dynamic estimation algorithm with relaxed local state observability for large-scale networks is exploited, i.e., the local measurements in one area may not be captured or exploited to estimate the local state, and the local state estimate can also be obtained by exchanging the boundary information with neighboring areas. Under the assumption that the communication graph of power networks is acyclic, which is a common assumption for a great deal of practical systems, we present an iterative distributed MAP estimation technique to update the local state estimation of every subsystem at each time instant. We find that at each time instant, the distributed MAP estimator at steady state converges after a finite number of iteration, which equals to the maximum path length of the acyclic graph. The main contribution of this paper generalizes the static state estimation method considered in [7] to the dynamic case, and the accuracy of our proposed estimation method in dynamic case is better than that in [7], since the result in [7] does not depend on the previous state prediction, and the static method is incapable of forecasting the system state. We can see from the simulation results that the distributed state estimation algorithm is slightly less accurate than the centralized MAP method but much better than the local and static static state estimation algorithms in [7]. Nevertheless, the computational complexity and communication load of the proposed distributed algorithm are lower than the centralized one, and this method is scalable to wide-area networked power networks.

## 2 System modeling and problem setup

We use a communication graph $\mathcal{G} = (\mathcal{V},\ \mathcal{E})$ to represent the multi-area power networks, where $\mathcal{V} = \{1,\ \cdots,\ N\}$ is the set of nodes with each node representing a control area and $\mathcal{E} \subset \mathcal{V} \times \mathcal{V}$ is the set of edges $(i,\ j) \in \mathcal{E}$ expressing that there exists an edge between nodes $i$ and $j$. Let $\mathcal{N}_i = \{j : (i,\ j) \in \mathcal{E}\}$ be the set of neighbors of node $i$. Denote $\mathcal{N}_i/\{j\}$ as the set of nodes which do not include the node $j$. We also assume that the graph $\mathcal{G}$ is connected and undirected, i.e.,

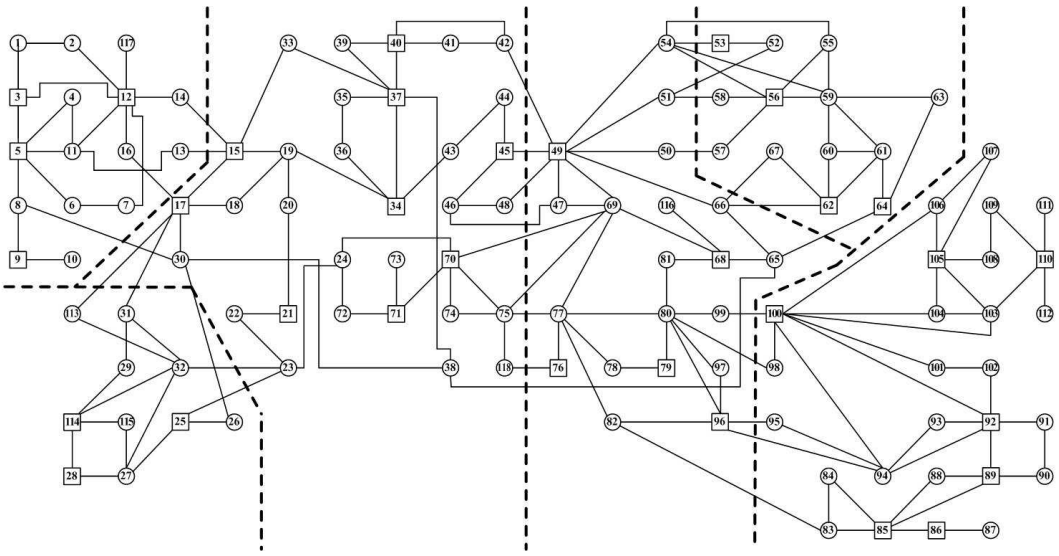

Fig. 1: Topological structure of the IEEE 118-bus system.

there exists a two-way path between each pair of nodes, and the graph is void of self-loops and multiple edges.

*Notation:* We denote the set of all $l-$dimensional real column vectors and $l \times q$ real matrices by $\mathbb{R}^l$ and $\mathbb{R}^{l\times q}$, respectively. $M^T$ denotes the transpose of $M$. $\mathbb{E}\{x\}$ represents the expectation of the stochastic variable $x$. The shorthand $\text{diag}[A_1,\ A_2,\ \cdots,\ A_n]$ denotes a block diagonal matrix with the diagonal blocks being the matrices $A_1,\ \cdots,\ A_n$.

In this paper, we consider the quasi-static systems, determined by slow dynamic changes of the load. For the $i$th control center, consider the following linear dynamic system

$$x_i(k+1) = A_i x_i(k) + G_i + \omega_i(k), \tag{1}$$

where $k$ is the time sample taking values of $0, 1, 2, \ldots$, $x_i(k) \in \mathbb{R}^{r_i}$ represents the local state of node $i$, matrix $A_i$ and vector $G_i$ describe the transition process of the states, $\omega_i(k) \in \mathbb{R}^{r_i}$ is the process noise assumed to be independent white Gaussian with zero mean and covariance $R_i$.

The measurements can be classified into two types, i.e., *local measurements* which are only functions of the states of one control center, and *edge measurements* representing the tie-line measurements related to the states of two neighboring nodes. Therefore, for nodes $i$ and $j$ which are neighbors, the two types of measurements can be represented as

$$y_i(k) = C_i x_i(k) + \nu_i(k), \tag{2}$$
$$z_{i,j}(k) = B_{i,j} x_i(k) + B_{j,i} x_j(k) + \nu_{i,j}(k), \tag{3}$$

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 nodes $i$ and $j$, $\nu_i(k)$ and $\nu_{i,j}(k)$ are the measurement noises assumed to be independent white Gaussian with zero mean and covariances $S_i$ and $T_{i,j}$. $C_i$, $B_{i,j}$ and $B_{j,i}$ are the measurement matrices. Assume that the initial state $x_i(0)$ is independent Gaussian with mean $\bar{x}_i(0)$ and covariance $P_i(0)$. Since (3) is shared by nodes $i$ and $j$, $z_{i,j}(k)$ and $z_{j,i}(k)$ are identical and the same goes to $\nu_{i,j}(k)$ and $\nu_{j,i}(k)$.

For the IEEE 118-bus system as in Fig. 1, Fig. 2 is described to abstract the partition, where the edge (1,3) for an

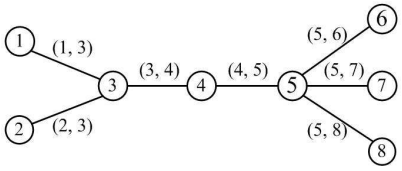


Fig. 2: The graph $\mathcal{G}$ depicting the partition of the 118-bus system.

example indicates that there exists edge measurements related to nodes 1 and 3. In large-scale systems, edge measurements are natural for describing physical interactions.

Denote $x(k) = (x_1^T(k),\ \cdots,\ x_N^T(k))^T$ and $z(k) = (\cdots,\ y_i^T(k),\ \cdots,\ z_{i,j}^T(k),\ \cdots)^T$, for (1), (2) and (3). The state and measurement equations can be written as

$$x(k+1) = Ax(k) + G + \omega(k), \tag{4}$$
$$z(k) = Hx(k) + \nu(k), \tag{5}$$

where $A = \text{diag}\,[A_1,\ \cdots,\ A_N]$, $G = (G_1^T,\ \cdots,\ G_N^T)^T$,

$$H = \begin{bmatrix} & & & \cdots & & & \\ \cdots & 0 & C_i & 0 & \cdots & & \\ & & & \cdots & & & \\ \cdots & 0 & B_{i,j} & 0 & B_{j,i} & 0 & \cdots \\ & & & \cdots & & & \end{bmatrix},$$

the covariances of noises $\omega(k) = (\omega_1^T(k),\ \cdots,\ \omega_N^T(k))^T$ and $\nu(k) = (\cdots,\ \nu_i^T(k),\ \cdots,\ \nu_{i,j}^T(k),\ \cdots)^T$ are $R = \text{cov}(\omega(k)) = \text{diag}[R_1,\ \cdots,\ R_N]$ and $R_* = \text{cov}(\nu(k)) = \text{diag}[\cdots,\ S_i,\ \cdots,\ T_{i,j},\ \cdots]$, respectively. $x(0)$ is the initial state with mean $\bar{x}(0) = (\bar{x}_1^T(0),\ \cdots,\ \bar{x}_N^T(0))^T$ and covariance $P(0) = \text{diag}[P_1(0),\ \cdots,\ P_N(0)]$.

We have the following two assumptions:

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

**Assumption 2**: The matrix $H$ has full column rank and the covariances of noises $R \geq 0$, $R_*$ and the initial state $P(0)$ are positive definite.

Assumption 1 means that the graph corresponding to a partition of power networks does not have a cycle. A great

number of power networks in practice can be abstracted as acyclic graphs, such as the power systems constructed along the coastline. Assumption 2 implies that local topological observability is waived for each node, and we only need to guarantee the globally topological observability of all nodes in the graph, which is equivalent to the matrix $H$ having full column rank. In other words, the local measurements in one area may not be sufficient to give an state estimate with a bounded estimation error covariance. Thus, Assumptions 1 and 2 are not severe requirements for state estimation problems in power networks.

## 3 Centralized MAP estimation algorithm

In this section, we propose the centralized state estimation method based on the MAP technique. The centralized MAP estimator is a standard Kalman filter. Let $Z(k) = \{z(0),\ z(1),\ \cdots,\ z(k)\}$. The estimate $\hat{x}(k|k)$ and the error covariance $\Sigma(k|k)$ at time instant $k$ can be viewed as ([12])

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

The one-step ahead prediction $\hat{x}(k|k-1)$ and its error covariance $\Sigma(k|k-1)$ of the state can be similarly defined.

The centralized state estimator contains two parts: centralized MAP estimator and centralized predictor. The main purpose is to calculate the optimal estimate $\hat{x}^{opt}(k|k)$.

### 3.1 Centralized MAP Estimation

The MAP estimation method can be written as (see [13])

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

where $p(x(k)|Z(k))$ denotes the probability density function of $x(k)$ conditioned on $Z(k)$. According to Bayesian rule in [14], the MAP estimator (6) is equivalent to

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

From (5) and (7), we obtain

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

Therefore, the centralized MAP estimate and its estimation error covariance take the following forms:

$$\hat{x}^{opt}(k|k) = Q^{-1}(k)\alpha(k), \tag{8}$$
$$\Sigma^{opt}(k|k) = Q^{-1}(k), \tag{9}$$

where

$$\alpha(k) = H^T R_*^{-1} z(k) + \Sigma^{-1}(k|k-1)\hat{x}(k|k-1), \tag{10}$$
$$Q(k) = H^T R_*^{-1} H + \Sigma^{-1}(k|k-1). \tag{11}$$

### 3.2 Centralized Prediction

This part focuses on the state prediction step to calculate the prediction $\hat{x}(k+1|k)$ by using the measurements $Z(k)$. Utilizing (4), the obtained $\hat{x}^{opt}(k|k)$ and $\Sigma^{opt}(k|k)$, we get

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

which will become the priori to compute the state estimation at time $k+1$.

The state estimate obtained from the centralized MAP estimator is optimal. The centralized state estimator is typically installed in a control center gathering all measurements over the network, which requires a powerful computer to do relatively heavy calculations and creates a bottleneck for communication, when we consider a large-scale network. In order to solve the problem resulted from the centralized state estimation scheme, we will investigate local and distributed state estimation methods in the next two sections.

## 4 Local state estimation method

In this sequel, treating the edge measurements as the local measurements of each node $i \in \mathcal{V}$, we describe a local state estimation method of node $i$.

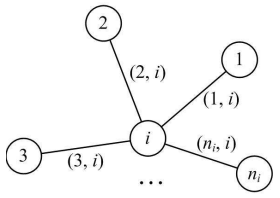


Fig. 3: Topological structure of a radial graph.

As shown in Fig. 3, a *radial graph* is a tree with $n_i + 1$ nodes, where one node called the central node whose radius equals to 1 and the other nodes' radii are 2.

For an interconnected system depicted as a *radial graph*, its aggregated state and measurement equations at time instant $k$, take the form as (4) and (5), where $x(k) = (x_i^T(k),\ x_1^T(k),\ \cdots,\ x_{n_i}^T(k))^T$, $z(k) = (y_i^T(k),\ y_1^T(k),\ \cdots,\ y_{n_i}^T(k),\ z_{i,1}^T(k),\ \cdots,\ z_{i,n_i}^T(k))^T$, $R^i = \text{diag}[R_i,\ R_1,\ \cdots,\ R_{n_i}]$ and $R_*^i = \text{diag}[S_i,\ S_1,\ \cdots,\ S_{n_i},\ T_{i,1},\ \cdots,\ T_{i,n_i}]$ are the covariances of the stacked noises.

Consider the modified measurement equation of node $i$:

$$\begin{bmatrix} y_i(k) \\ z_{i,1}(k) \\ \vdots \\ z_{i,n_i}(k) \end{bmatrix} = \begin{bmatrix} C_i \\ B_{i,1} \\ \vdots \\ B_{i,n_i} \end{bmatrix} x_i(k) + \begin{bmatrix} \nu_i(k) \\ \nu_{i,1}(k) \\ \vdots \\ \nu_{i,n_i}(k) \end{bmatrix}, \tag{12}$$

which is obtained from (5) by getting rid of the measurements and states unrelated to $x_i(k)$. The rest measurements can be considered as the local measurements of node $i$ to estimate its state. Then using the local MAP estimator as (8) and (9), the *pseudo information vector* $\breve{\alpha}_i(k)$ and the *pseudo information matrix* $\breve{Q}_i(k)$ for (12) are given as

$$\tilde{\alpha}_i(k) = C_i^T S_i^{-1} y_i(k) + \Sigma_i^{-1}(k|k-1)\hat{x}_i(k|k-1),$$
$$\tilde{Q}_i(k) = C_i^T S_i^{-1} C_i + \Sigma_i^{-1}(k|k-1),$$
$$\breve{\alpha}_i(k) = \tilde{\alpha}_i(k) + \sum_{j\in\mathcal{N}_i} B_{i,j}^T T_{i,j}^{-1} z_{i,j}(k) \tag{13}$$
$$\breve{Q}_i(k) = \tilde{Q}_i(k) + \sum_{j\in\mathcal{N}_i} B_{i,j}^T T_{i,j}^{-1} B_{i,j}. \tag{14}$$

For node $j \in \mathcal{N}_i$, only using the local measurements (2)

and the priori, we obtain the measurement-update equations:

$$\bar{x}_j(k|k) = \hat{x}_j(k|k-1) + K_j(k)\big(y_j(k) - C_j\hat{x}_j(k|k-1)\big),$$
$$\bar{\Sigma}_j(k|k) = \Sigma_j(k|k-1) - K_j(k)C_j\Sigma_j(k|k-1),$$

where $K_j(k) = \Sigma_j(k|k-1)C_j^T\big(C_j\Sigma_j(k|k-1)C_j^T + S_j\big)^{-1}$ is the Kalman gain. Next, node $j$ computes $\theta_{i,j}^0(k) = B_{j,i}\bar{x}_j(k|k)$, $\Theta_{i,j}^0(k) = B_{j,i}\bar{\Sigma}_j(k|k)B_{j,i}^T$, and sends them to node $i$. Using the information transmitted from neighbors, node $i$ can calculate its state estimation and covariance.

**Theorem 1.** *Consider an interconnected system represented by a radial graph $\mathcal{G}$ in Fig. 3. At time instant $k$, the local state estimation and the error covariance of node $i \in \mathcal{V}$ are*

$$\hat{x}_i(k|k) = \Sigma_i(k|k)\Big(\breve{\alpha}_i(k) - \sum_{j=1}^{n_i}\beta_{i,j}^0(k)\Big), \qquad (15)$$
$$\Sigma_i(k|k) = \Big(\breve{Q}_i(k) - \sum_{j=1}^{n_i}\Phi_{i,j}^0(k)\Big)^{-1},$$

*where $\eta_{i,j} = T_{i,j}^{-1}B_{i,j}$, $\beta_{i,j}^0(k) = \eta_{i,j}^T\theta_{i,j}^0(k)$, $\Phi_{i,j}^0(k) = \eta_{i,j}^T\Theta_{i,j}^0(k)\eta_{i,j}$.*

*In the next section, we will give a result similarly to Theorem 1, and we will prove this important result in detail. Hence, we omit the proof of Theorem 1 here.*

We can obtain a similar result of node $j$, i.e.,

$$\breve{\alpha}_j(k) = \tilde{\alpha}_j(k) + B_{j,i}^T T_{i,j}^{-1} z_{i,j}(k),$$
$$\breve{Q}_j(k) = \tilde{Q}_j(k) + B_{j,i}^T T_{i,j}^{-1} B_{j,i}.$$

In the prediction step, we get the time-update equations:

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

Compared with the centralized state estimation method, the local state estimate obtained from (15) is suboptimal. But it plays a role in the performance comparison with centralized and distributed MAP estimation algorithms later.

## 5 Distributed MAP estimation algorithm

We will propose a distributed state estimation algorithm including two parts: distributed MAP estimators and local state predictors, which are the same as that in Section 4.

In the distributed MAP estimation method, each node $i \in \mathcal{V}$ launches a local state estimator to build $\hat{x}_i(k|k)$ at time instant $k$, based on (2), (3) and the information transmitted from node $j \in \mathcal{N}_i$. The purpose is to compute the stacked estimate $\hat{x}(k|k)$ minimizing the cost function $J$, i.e.,

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

where $\hat{x}(k|k-1) = [\hat{x}_1^T(k|k-1),\ \cdots,\ \hat{x}_N^T(k|k-1)]^T$, $\tilde{\Sigma}(k|k-1) = \text{diag}[\Sigma_1(k|k-1),\ \cdots,\ \Sigma_N(k|k-1)]$ and $\Sigma_i(k|k-1)$, $i = 1,\ \cdots,\ N$ are the diagonal elements of $\Sigma(k|k-1)$.

From the problem description, although the initial covariance matrix $P(0)$ is block diagonal in the global model, we can see that using the centralized MAP estimator and centralized predictor in Section 3, $\Sigma(k|k-1)$ is no longer a block diagonal matrix. In order to achieve the distributed state estimation approach, we use $\tilde{\Sigma}(k|k-1)$ replacing $\Sigma(k|k-1)$ as the priori to update $\hat{x}(k|k)$ and $\Sigma(k|k)$. In addition, we expect that the distributed state estimation algorithm could converge in a finite number of iterations at each sample time.

Let $\hat{x}^*(k|k)$ and $\Sigma^*(k|k)$ denote the state estimate and the estimation error covariance, obtained by the distributed MAP estimation algorithm, where $\hat{x}^*(k|k) = \arg\min_{x(k)} J(x(k))$ is the aggregation of $\hat{x}_i^*(k|k)$, $i = 1,\ \cdots,\ N$. Corresponding to the local state $x_i(k)$ on each node $i$, denote $\Sigma_i^*(k|k)$ as the block diagonal sub-matrix of $\Sigma^*(k|k)$. 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 $\hat{x}_j(k|k)$ to its neighbors and this indicates a very light communication load.

Consider the *radial graph* as shown in Fig. 3. From (13) and (14), the *pseudo estimate* and *pseudo estimation error covariance* of node $i$ are

$$\breve{x}_i(k|k) = \breve{Q}_i^{-1}(k)\breve{\alpha}_i(k),\ \ \breve{\Sigma}_i(k|k) = \breve{Q}_i^{-1}(k).$$

Similarly, the *pseudo estimate* and *pseudo estimation error covariance* for the local state $x_j(k)$ are

$$\breve{x}_j(k|k) = \breve{Q}_j^{-1}(k)\breve{\alpha}_j(k),\ \ \breve{\Sigma}_j(k|k) = \breve{Q}_j^{-1}(k).$$

In order to obtain the local estimation and error covariance for $x_i(k)$, node $i$ needs to combine $\breve{x}_i(k|k)$ and $\breve{\Sigma}_i(k|k)$ with the information transmitted from its neighbors. Define

$$\beta_{i,j}(k) = B_{i,j}^T T_{i,j}^{-1} B_{j,i}\breve{x}_j(k|k),$$
$$\Phi_{i,j}(k) = B_{i,j}^T T_{i,j}^{-1} B_{j,i}\breve{\Sigma}_j(k|k)B_{j,i}^T T_{i,j}^{-1} B_{i,j},$$

which denote the updates transmitted from node $j \in \mathcal{N}_i$.

**Lemma 1.** *Suppose that Assumption 2 holds. Consider the radial graph $\mathcal{G}$ in Fig. 3. At time instant $k$, the distributed MAP estimate and the error covariance of node $i \in \mathcal{V}$ are*

$$\hat{x}_i^*(k|k) = \Sigma_i^*(k|k)\Big(\breve{\alpha}_i(k) - \sum_{j=1}^{n_i}\beta_{i,j}(k)\Big),$$
$$\Sigma_i^*(k|k) = \Big(\breve{Q}_i(k) - \sum_{j=1}^{n_i}\Phi_{i,j}(k)\Big)^{-1}.$$

Before proving Lemma 1, we first consider the two-node graph of nodes 1 and 2.

**Lemma 2.** *Consider an interconnected graph of two nodes 1 and 2, where $x(k) = (x_1^T(k),\ x_2^T(k))^T$, $z(k) = (y_1^T(k),\ y_2^T(k),\ z_{1,2}^T(k))^T$, the stacked noises can be similarly defined, $A^1 = diag[A_1,\ A_2]$, $R^1 = diag[R_1,\ R_2]$ and $R_*^1 = diag[S_1,\ S_2,\ T_{1,2}]$. At time instant $k$, the distributed MAP estimate and its error covariance of node $i = 1,\ 2$ are*

$$\hat{x}_i^*(k|k) = \Sigma_i^*(k|k)\Big(\breve{\alpha}_i(k) - \beta_{i,j}(k)\Big),$$
$$\Sigma_i^*(k|k) = \Big(\breve{Q}_i(k) - \Phi_{i,j}(k)\Big)^{-1},\ \ j \in \mathcal{N}_i,$$

*which is initialized by $\hat{x}_i(0|-1) = \bar{x}_i(0)$, $\Sigma_i(0|-1) = P_i(0)$ and where $\breve{\alpha}_i(k) = \tilde{\alpha}_i(k) + B_{i,j}^T T_{i,j}^{-1} z_{i,j}(k)$, $\breve{Q}_i(k) = \tilde{Q}_i(k) + B_{i,j}^T T_{i,j}^{-1} B_{i,j}$.*

**Proof:** At time instant $k$, consider the centralized MAP estimation scheme, in which $\Sigma(k|k-1)$ is replaced by $\tilde{\Sigma}(k|k-1)$, and from (10) and (11), we get

$$\alpha(k) = \left[\begin{array}{c} \alpha_1(k) \\ \alpha_2(k) \end{array}\right], \ Q(k) = \left[\begin{array}{cc} \Gamma_{11}(k) & \Gamma_{12}(k) \\ \Gamma_{12}^T(k) & \Gamma_{22}(k) \end{array}\right],$$

where

$$\alpha_i(k) = \breve{\alpha}_i(k), \ \Gamma_{ii}(k) = \breve{Q}_i(k), \ \Gamma_{ij}(k) = B_{i,j}^T T_{i,j}^{-1} B_{j,i}.$$

Based on (8)-(9), and using the matrix inverse lemma, we get

$$\hat{x}^*(k|k) = Q^{-1}(k)\alpha(k) = \left[\begin{array}{c} \hat{x}_1^*(k|k) \\ \hat{x}_2^*(k|k) \end{array}\right],$$
$$\Sigma^*(k|k) = Q^{-1}(k) = \left[\begin{array}{cc} \Sigma_1^*(k|k) & \Delta_{12}(k) \\ \Delta_{21}(k) & \Sigma_2^*(k|k) \end{array}\right],$$

where

$$\begin{aligned}
\Delta_{12}(k) &= -\Sigma_1^*(k|k)\Gamma_{12}(k)\Gamma_{22}^{-1}(k), \\
\Delta_{21}(k) &= -\Sigma_2^*(k|k)\Gamma_{12}^T(k)\Gamma_{11}^{-1}(k), \\
\hat{x}_1^*(k|k) &= \Sigma_1^*(k|k)\breve{\alpha}_1(k) - \Sigma_1^*(k|k)\Gamma_{12}(k)\breve{Q}_2^{-1}(k)\breve{\alpha}_2(k) \\
&= \Sigma_1^*(k|k)\Big(\breve{\alpha}_1(k) - \beta_{1,2}(k)\Big), \\
\Sigma_1^*(k|k) &= \Big(\Gamma_{11}(k) - \Gamma_{12}(k)\Gamma_{22}^{-1}(k)\Gamma_{12}^T(k)\Big)^{-1} \\
&= \Big(\breve{Q}_1(k) - \Phi_{1,2}(k)\Big)^{-1}.
\end{aligned}$$

The case of node 2 can be similarly obtained. □

*The proof of Lemma 1 can be directly deduced from Lemma 2 by replacing central node 1 by $i$ and node 2 by nodes $1, 2, \cdots, n_i$ in Fig. 3. So we omit the proof here.*

Based on Lemma 1, we can design the distributed MAP estimation algorithm. Initially, we treat every node $i$ as the central node of the *radial graph* to calculate $\breve{x}_i(k|k, 0)$ and $\breve{\Sigma}_i(k|k, 0)$, using its initial information of state and the modified measurement equation. Then at iteration $h$, node $i$ updates its local state estimation $\hat{x}_i(k|k, h)$ and the associated covariance $\Sigma_i(k|k, h)$ by combining its own $\breve{x}_i(k|k, 0)$ and $\breve{\Sigma}_i(k|k, 0)$ with the information $(\theta_{i,j}(k, h-1), \Theta_{i,j}(k, h-1))$ received from its neighbor $j \in \mathcal{N}_i$. Meanwhile, each node $i$ computes $(\theta_{j,i}(k, h), \Theta_{j,i}(k, h))$ related to its most recently updated local state estimate and error covariance, and then transmits them to its neighboring node $j$. A key point in Algorithm 1 is that the information $(\theta_{j,i}(k, h), \Theta_{j,i}(k, h))$ do not contain the information that node $i$ previously received from node $j$.

**Algorithm 1 Distributed MAP estimation algorithm**

1) At $k = 1, 2, \cdots$, node $i \in \mathcal{V}$ computes the *pseudo estimation* and the *pseudo estimation error covariance*:

$$\begin{aligned}
\breve{x}_i(k|k, 0) &= \breve{\Sigma}_i(k|k, 0)\breve{\alpha}_i(k, 0), \\
\breve{\Sigma}_i(k|k, 0) &= \breve{Q}_i^{-1}(k, 0),
\end{aligned}$$

where $\breve{\alpha}_i(k, 0)$ and $\breve{Q}_i(k, 0)$ are defined as (13) and (14). If $k = 0$, the above priors are replaced by $\bar{x}_i(0)$ and $P_i(0)$.

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

$$\begin{aligned}
\theta_{j,i}(k, 0) &= B_{i,j}\breve{x}_i(k|k, 0), \\
\Theta_{j,i}(k, 0) &= B_{i,j}\breve{\Sigma}_i(k|k, 0)B_{i,j}^T.
\end{aligned}$$

**Main loop:** $h = 1, 2, \cdots$ is the step number of iteration.

1) Let $\theta_{i,j}(k, h-1)$ and $\Theta_{i,j}(k, h-1)$ be the information received from node $j$. Node $i$ updates the local estimation and its covariance based on the received information:

$$\begin{aligned}
\hat{x}_i(k|k, h) &= \Sigma_i(k|k, h)\Big(\breve{\alpha}_i(k, 0) - \sum_{j\in\mathcal{N}_i} \beta_{i,j}(k, h-1)\Big), \\
\Sigma_i(k|k, h) &= \Big(\breve{Q}_i(k, 0) - \sum_{j\in\mathcal{N}_i} \Phi_{i,j}(k, h-1)\Big)^{-1},
\end{aligned}$$

where

$$\begin{aligned}
\beta_{i,j}(k, h-1) &= B_{i,j}^T T_{i,j}^{-1}\theta_{i,j}(k, h-1), \\
\Phi_{i,j}(k, h-1) &= B_{i,j}^T T_{i,j}^{-1}\Theta_{i,j}(k, h-1)T_{i,j}^{-1}B_{i,j}.
\end{aligned}$$

2) Transmit information to node $j \in \mathcal{N}_i$. Node $i$ computes

$$\begin{aligned}
\breve{x}_{j,i}(k|k, h) &= \breve{\Sigma}_{j,i}(k|k, h)\Big(\breve{\alpha}_i(k, 0) \\
&\qquad - \sum_{m\in\mathcal{N}_i/\{j\}} \beta_{i,m}(k, h-1)\Big), \\
\breve{\Sigma}_{j,i}(k|k, h) &= \Big(\breve{Q}_i(k, 0) - \sum_{m\in\mathcal{N}_i/\{j\}} \Phi_{i,m}(k, h-1)\Big)^{-1},
\end{aligned}$$

and then transmits the following information to node $j$:

$$\begin{aligned}
\theta_{j,i}(k, h) &= B_{i,j}\breve{x}_{j,i}(k|k, h), \\
\Theta_{j,i}(k, h) &= B_{i,j}\breve{\Sigma}_{j,i}(k|k, h)B_{i,j}^T.
\end{aligned}$$

□

For a interconnected graph, 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 $\varepsilon_i$. The diameter $\Gamma$ of the graph is the maximum radius among all nodes.

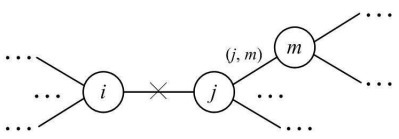


Fig. 4: Illustration for the proof of Theorem 2.

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

$$\begin{aligned}
\hat{x}_i(k|k, \varepsilon_i + l) &= \hat{x}_i^*(k|k), \\
\Sigma_i(k|k, \varepsilon_i + l) &= \Sigma_i^*(k|k), \quad for\ all\ l \geq 0,
\end{aligned} \tag{16}$$

*where $\varepsilon_i + l$ is the step number of iteration, $\hat{x}_i^*(k|k)$ and $\Sigma_i^*(k|k)$ are obtained as in Lemma 1.*

**Proof:** Consider the acyclic graph $\mathcal{G}$ as in Fig. 4. At the first iteration $h = 1$, we also consider the radial sub-graphs of $\mathcal{G}$, having each $i$ as its central node. Obviously, the two-node graph and *radial graph* considered in Lemmas 1 and 2 are this case. From Lemmas 1 or 2, since we regard the neighbors as single nodes to transmit data to node $i$, which never changes after step 1. Hence, the local state estimate of node $i$ reaches the suboptimal values after $h = 1$. Meanwhile, similar computation is done at node $j \in \mathcal{N}_i$.

For the second iteration, consider the radial sub-graph formed by all nodes which are two hops away from node $i$. It has $i$ as central node, in which we combine each node $j \in \mathcal{N}_i$ with its neighbors but except $i$ into a single node, i.e., we combine nodes $j$ with $m$ as the neighbor of node $i$. We still call this node $j$, which already contains the information transmitted from node $m$ at iteration $h = 1$. According to Algorithm 1, node $j \in \mathcal{N}_i$ builds $\beta_j^i(k,2)$ and $\Phi_j^i(k,2)$ to be transmitted to node $i$, without using the information $\beta_i^j(k,1)$ and $\Phi_i^j(k,1)$ previously received from node $i$. Hence, the second iteration is equivalent to applying Lemma 1 to the aforementioned radial graph. So node $i$ is able to compute the suboptimal estimate of the sub-graph.

At iteration $h$, consider the radial sub-graph formed by all nodes which are $h$ hops away from node $i$. Combine each node $j \in \mathcal{N}_i$ with these nodes, which are less than or equal to $h-1$ hops away from it, but except $i$ into a single node. Following the above argument, node $j$ already contains the information transmitted from these nodes, and node $i$ is able to compute the suboptimal estimate of this radial sub-graph.

Since there is no more information transmitted from nodes after $h \geq \Gamma_i$, node $i$ gets the suboptimal estimate of the whole graph $\mathcal{G}$, and this value will remain unchanged in the subsequent iterations. Then we achieve (16) for all $i \in \mathcal{V}$. □

It can be seen from (16) that the local estimates on all nodes converge after $\Gamma$ steps at each time instant.

**Remark 1.** *Since the local predictors are the same in Section 4, we omit this part here.*

## 6 Simulation result

To compare the performance of the distributed state estimator (DSE) presented in this paper with the centralized state estimator (CSE), the local state estimator (LSE) and the distributed static state estimator (DSSE), the IEEE 118-bus system is used, which is split into eight areas. We can see that the resulting network shown in Fig. 2 is acyclic, so Assumption 1 is satisfied.

We run these four state estimation algorithms under Assumptions 1 and 2. We will use the sum of the trace of the estimation error covariance $\sum_{i=1}^{8} \mathrm{Tr}\{\Sigma_i(k|k)\}$ to compare the performances of these algorithms.

The time step of tests in the simulation is chosen as 1s. Fig. 5 shows that the CSE is the optimal, followed by the DSE, the LSE and the DSSE are the worst in the sense of these cases, which means that the performance of the DSE is only slightly worse than that of the CSE, while it is largely better than that of the LSE and the DSSE after converging to some values.

## 7 Conclusion

Based on the MAP estimation technique, we provide an iterative distributed method to update the local state estimate of every node in large sized power networks at each time. We relax the assumption that not every control center must be locally observable and the state estimate of each node can still be recovered by edge measurements and boundary information transmitted from neighbors. An illustrative example in the IEEE 118-bus system confirms the effectiveness of the proposed scheme, especially when we consider the computational complexity and communication load.

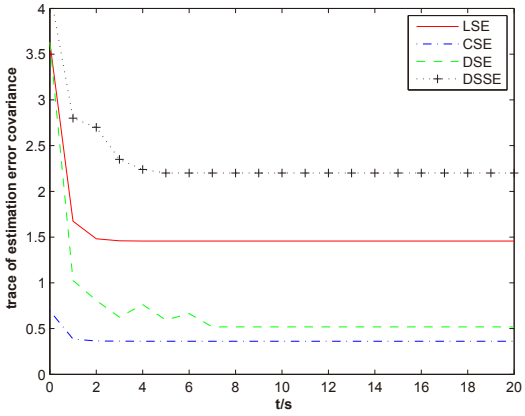


Fig. 5: The traces of the estimation error covariances of the CSE, LSE, DSE and DSSE estimates

## References


[1] F. C. Schweppe and J. Wildes, Power system static-state estimation, part I, II, III, *IEEE Trans. on Power Apparatus and Systems*, 89(1): 120–135, 1970.

[2] F. F. Wu, Power system state estimation: a survey, *International Journal of Electrical Power and Energy Systems*, 12(1): 80–87, 1990.

[3] F. F. Wu, K. Moslehi, and A. Bose, Power system control centers: past, present, and future, in *Proceedings of the IEEE*, 2005: 1890–1908.

[4] W. Jiang, V. Vittal, and G. T. Heydt, A distributed state estimator utilizing synchronized phasor measurements, *IEEE Trans. on Power Systems*, 22(2) 563–571, 2007.

[5] L. Xie, D. H. Choi, S. Kar, and H. V. Poor, Fully distributed state estimation for wide-area monitoring systems, *IEEE Trans. on Smart Grid*, 3(3): 1154–1169, 2012.

[6] F. Pasqualetti, R. Carli, and F. Bullo, Distributed estimation via iterative projections with application to power network monitoring, *Automatica*, 48(5): 747–758, 2012.

[7] X. Tai, Z. Lin, M. Fu, and Y. Sun, A new distributed state estimation technique for power networks, *2013 American Control Conference*, 2013: 3338–3343.

[8] F. S. Cattivelli, C. G. Lopes, and A. H. Sayed, Diffusion recursive least-squares for distributed estimation over adaptive networks, *IEEE Trans. on Signal Processing*, 56(5): 2069–2084, 2008.

[9] H. Xue, Q. Jia, N. Wang, Z. Bo, H. Wang, and H. Ma, A dynamic state estimation method with PMU and SCADA measurement for power systems, In *the 8th International Power Engineering Conference*, 2007: 848–853.

[10] G. Valverde and V. Terzija, Unscented Kalman filter for power system dynamic state estimation, *IET Generation, Transmission and Distribution*, 5(1): 29–37, 2011.

[11] S. B. Wang, W. Z. Gao, and A. P. S. Meliopoulos, An alternative method for power system dynamic state estimation based on unscented transform, *IEEE Trans. Power Systems*, 27(2): 942–950, 2012.

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

[13] S. M. Kay, *Fundamentals of Statistical Signal Processing*. Upper Saddle River, NJ: Prentice Hall, 1993.

[14] S. Arulampalam, S. Maskell, N. Gordon, and T. Clapp, A tutorial on particle filters for online nonlinear/non-gaussian Bayesian tracking, *IEEE Trans. on Signal Processing*, 50(2): 174–188, 2002.