Real-Time Adaptive Distributed State Estimation in Smart Grids

Soummya Kar* and José M. F. Moura*
* Department of Electrical and Computer Engineering, Carnegie Mellon University, Pittsburgh, PA, USA
E-mail: soummyak@andrew.cmu.edu, moura@ece.cmu.edu

***Abstract*—The paper presents a fully distributed framework for sequential recursive state estimation in inter-connected electrical power systems. Specifically, the setup considered involves a grid partitioned into multiple control areas that communicate over a *sparse* communication network. In the absence of a global sensor data fusion center (the conventional centralized SCADA) and with sensing model uncertainties, an adaptive distributed state estimation approach, the $\mathcal{DAE}$, is proposed in which the system control areas engage in a collaborative joint (model) learning and (state) estimation procedure through sequential information exchange over the pre-assigned communication network. The proposed distributed estimation methodology is recursive, in that, each system control area refines its state estimate at a given sampling instant by suitably combining its past estimate with the newly collected local measurement(s) and the information obtained from its communication neighbors. Under rather weak assumptions of global observability and connectivity of the control area communication network, the proposed distributed adaptive scheme is shown to yield consistent system state estimates (i.e., estimates that converge to the true system state in the large sample limit), the convergence rate being optimal in the Fisher information sense. As discussed, the proposed approach based on local communication and computation is suitable for real-time implementation as opposed to conventional centralized SCADA based estimation architectures with periodic data gathering and processing, thus being potentially more responsive and adaptive to sensed data generated by advanced non-conventional sensing resources like the PMUs with significantly higher system sampling rates.**

I. Introduction

A. Motivation

With the continued penetration of highly intermittent and vastly distributed wind and solar renewables, the power grid is rapidly evolving towards a complex interconnect of heterogenous distributed modules. While this offers huge potential to sustain our increasing energy needs, it necessitates a massive transformation of the existing highly centralized system to a *distributed* and *responsive* intelligent grid that will carefully exploit the opportunities offered by advances in power electronics, active load management, smart metering and communication-computation technologies ([1], [2]). Technically, enabled by advanced control, communication, and computation, wide area monitoring systems (WAMS) of the future will likely involve large numbers of fast information gathering and processing devices [3]. Institutionally, the power industry deregulation has led to the creation of multiple regional control areas (CAs), each only operating a fraction of the large interconnected power grid [4]. The technical and institutional changes suggest the need for more decentralized estimation and control in wide area power system operations [5].

The future operation paradigm will be largely based on distributed transactions and energy management, thus requiring pervasive system intelligence for efficient decision-making. The design of distributed state estimators is indeed key to providing such ubiquitous intelligence and sustaining the evolving demands and functionalities of the grid [6], [7], [8], [9], [10], [11], [12]. In particular, while large power interconnections such as the eastern/western interconnections are usually operated by several CAs, advanced applications such as WAMS and control require the state of the *entire* system to be available at all the CAs [13], [5]. Indeed, the conventional power system state estimation paradigm, in which centralized supervisory control and data acquisition (SCADA) systems periodically (roughly in intervals of about 5 minutes,see, for example, the New England ISO [14]) collect network-wide measurements through remote terminal units (RTUs), needs to be transformed into a centerless distributed platform where in real-time the different CAs exchange information to engage in a collaborative sequential state estimate update process. This is further motivated by the incorporation of non-conventional sensing resources in the grid, notably phasor measurement units (PMUs) [15] and intelligent electronic devices (IEDs) [16], [17], that are more accurate with considerably higher system sampling rate and that further provide real-time system information in terms of local protection and control data. Thus, for real-time system state monitoring, these heterogenous sensing data (with higher sampling rates) should be processed as and when they are captured, rather than communicating the huge volume of measured data to a center periodically. Clearly, a distributed approach is more appealing in this regard, as it is based on real-time information sharing among the CAs through peer-to-peer exchanges[1] leading to sequential estimate refinement as and when new information enters the system.

The traditional centralized SCADA based state estimation [19], [20] and bad data detection [21], [22] approaches

Work partially supported by NSF grants # CCF-1018509 and # CCF-1011903 and by AFOSR grant # FA9550101291.

[1]Note that inter-agent communication is typically much faster than synchronized data forwarding to a central SCADA and several real-time peer-to-peer type communication architectures are already in place, for example, the IEC 61850 standard, a protocol for inter-IED communication [18].

that are based on periodic snapshots of the system using conventional low sampling rate measurements are clearly unable to exploit the potentials offered by the non-conventional sensing resources with much faster sampling rates. For example, each PMU is capable of obtaining samples of the order of hundreds per second [23] rendering the transmission of all network measurements to a center periodically almost infeasible. Moreover, for efficient processing of the sensed measurements and real-time estimate update, a sequential information processing approach that fuses new data as soon as it is acquired is preferred over a periodic batch processing approach. Similarly, conventional approaches to mitigate the effect of corrupted measurements (resulting, for instance, from sensing anomalies) on the estimation performance, often called bad data detection and localization techniques, are essentially small-sample statistical tests based on normalized chi-squared residuals [21], [22], or more involved techniques such as combinatorial group testing [24] and hypothesis testing identification [25]. In general, the information content of a sensing resource may not be well predicted from a single sample and better results are likely to be obtained by studying their long term trend or large sample characteristics. Especially, given the abundance of sensing resources in the future grid with higher sampling rates, more general real-time sequential learning approaches should be adopted in order to reassess the contextual value of sensed information, so that the measurement data may be appropriately fused to yield more accurate state estimates.

Motivated by the challenges and requirements in the future smart grid, in this paper we consider adaptive and distributed state estimation procedures that: (1) exploits in real-time the sensing diversity offered by non-conventional sensing resources with higher accuracy and faster system sampling rate; (2) adapts to (stochastic) uncertainties in the information content of the sensing resources through online learning in conjunction with the estimation procedure, so as to effectively combine (fuse) the sensed data from different sensing methodologies by properly weighting their measurements; and (3) provides pervasive network-wide intelligence to improve the performance of the local protection units and controllers by appropriately distributing the information processing and communication overhead among the various network entities.

The rest of the paper is organized as follows. We start by introducing the distributed measurement model (with sensing uncertainties) resulting from a geographically distributed spread system of interest partitioned into several control areas (CAs) in Section II. As a benchmark for future comparison, in the same section, we further quantify the estimation performance of a hypothetical real-time centralized estimator (optimal) with access to all CA measurements at all times and perfect information about the physical and sensing parameters. In Section III we propose a generic distributed adaptive methodology, $\mathcal{DAE}$, for distributed state estimation with parametric uncertainties, in which the various CAs exchange information in real-time leading to recursive state estimates that instantaneously respond to the sensed data rather than waiting for a centralized SCADA to process the network-wide information periodically, as is done in existing centralized state estimation architectures. The main result of the paper characterizing the convergence of the $\mathcal{DAE}$ scheme is presented in Section IV. Finally, Section V concludes the paper.

B. Notation

We denote the k-dimensional Euclidean space by $\mathbb{R}^k$. The set of reals is denoted by $\mathbb{R}$, whereas $\mathbb{R}_+$ denotes the non-negative reals. For $a, b \in \mathbb{R}$, we will use the notations $a \vee b$ and $a \wedge b$ to denote the maximum and minimum of a and b respectively. The set of $k \times k$ real matrices is denoted by $\mathbb{R}^{k \times k}$. The corresponding subspace of symmetric matrices is denoted by $\mathbb{S}^k$. The cone of positive semidefinite matrices is denoted by $\mathbb{S}_+^k$, whereas $\mathbb{S}_{++}^k$ denotes the subset of positive definite matrices. The $k \times k$ identity matrix is denoted by I_k, while $\mathbf{1}_k, \mathbf{0}_k$ denote respectively the column vector of ones and zeros in $\mathbb{R}^k$. Often the symbol 0 is used to denote the $k \times p$ zero matrix, the dimensions being clear from the context. The operator $\|\cdot\|$ applied to a vector denotes the standard Euclidean $\mathcal{L}_2$ norm, while applied to matrices denotes the induced $\mathcal{L}_2$ norm, which is equivalent to the matrix spectral radius for symmetric matrices. The notation $A \otimes B$ is used to denote the Kronecker product of two matrices A and B.

Spectral graph theory: The inter-agent communication topology may be described by an *undirected* graph $G = (V, E)$, with $V = [1 \cdots N]$ and E the set of CAs (agents) and communication links (edges),respectively. The unordered pair $(n, l) \in E$ if there exists an edge between nodes n and l. We consider simple graphs, i.e., graphs devoid of self-loops and multiple edges. A graph is connected if there exists a path[2], between each pair of nodes. The neighborhood of node n is

$$\Omega_n = \{l \in V \mid (n, l) \in E\} \tag{1}$$

Node n has degree $d_n = |\Omega_n|$ (the number of edges with n as one end point.) The structure of the graph can be described by the symmetric $N \times N$ adjacency matrix, $A = [A_{nl}]$, $A_{nl} = 1$, if $(n, l) \in E$, $A_{nl} = 0$, otherwise. Let the degree matrix be the diagonal matrix $D = \operatorname{diag}(d_1 \cdots d_N)$. By definition, the positive semidefinite matrix $L = D - A$ is called the graph Laplacian matrix. The eigenvalues of L can be ordered as $0 = \lambda_1(L) \leq \lambda_2(L) \leq \cdots \leq \lambda_N(L)$, the eigenvector corresponding to $\lambda_1(L)$ being $(1/\sqrt{N})\mathbf{1}_N$. The multiplicity of the zero eigenvalue equals the number of connected components of the network; for a connected graph, $\lambda_2(L) > 0$. This second eigenvalue is the algebraic connectivity or the Fiedler value of the network; see [26] for detailed treatment of graphs and their spectral theory.

II. Problem Formulation

A. Sensing Model

For definiteness, in this paper, we address the problem of distributed and adaptive static state estimation in a power system comprising of several CAs, that are coupled by the

[2]A path between nodes n and l of length m is a sequence $(n = i_0, i_1, \cdots, i_m = l)$ of vertices, such that $(i_k, i_{k+1}) \in E\,\forall\, 0 \leq k \leq m-1$.

physical interactions. The state of the overall system is represented by a vector $\mathbf{x} \in \mathbb{R}^M$, which, for (linearized) DC state estimation, corresponds to the vector of bus angles. Let t denote time[3] and T the estimation horizon for static state estimation[4]. Assuming that the system state $\mathbf{x}$ stays constant over the estimation horizon $[0, T]$, the measurements at the n-th CA may be represented as a time series $\{\mathbf{y}_n(t)\}_{t=0}^{T}$, such that,

$$\mathbf{y}_n(t) = H_n\mathbf{x} + \mathbf{w}_n(t), \tag{2}$$

where: i) $\left\{\mathbf{y}_n(t) \in \mathbb{R}^{M_n}\right\}$ is the independent and identically distributed (i.i.d.) observation sequence for the n-th CA; and ii) for each n, $\{\mathbf{w}_n(t)\}$ is a zero-mean temporally i.i.d. noise sequence with covariance matrix R_n. Moreover, the sequences $\{\mathbf{w}_n(t)\}$ and $\{\mathbf{w}_l(t)\}$ are mutually uncorrelated for $n \neq l$. Generally each CA observes only a subset of M_n of the components of $\mathbf{x}$, with $M_n \ll M$. It is then necessary for the CAs to collaborate by means of occasional local message exchanges to achieve a reasonable estimate of the state $\mathbf{x}$. Moreover, due to inherent uncertainties in the deployment and the sensing environment, the statistics of the observation process (i.e., of the noise) is likely to be unknown apriori. For example, the exact observation noise variance depends on several factors beyond the control of the deployment process and should be learnt over time for reasonable estimation performance. In other words, prior knowledge of the spatial distribution of information content (i.e., which agent is more accurate than the others) may not be available, and the proposed estimation approach should be able to adaptively learn the true value of information leading to an accurate weighting of the various observation resources. We emphasize that, for meaningful state estimation, the sensed data obtained during the estimation interval of interest must be *simultaneously* used to: (1) learn the current values (states) of the uncertain parameters; and (2) estimate the system state $\mathbf{x}$ of interest. This necessitates the design of *adaptive* estimation schemes, i.e., schemes that adapt themselves to switching parameters and provide reasonable estimation performance by using the sensed data to learn the unknown parameters in conjunction to the estimation task.

Finally, keeping in mind the distributed nature of our sensing model and the estimation methodology (to be proposed soon), we make the following assumptions on (global) system observability and the inter-CA communication topology:

(A.1): The true observation noise covariance matrix R_n is positive definite for each n. We do not require observability at the local level, but impose the following global observability, i.e., the (normalized) Grammian matrix

$$\overline{\Sigma_{\mathrm{c}}} = \frac{1}{N}\sum_{n=1}^{N} H_n^T R_n^{-1} H_n \tag{3}$$

is invertible. Also, to begin with, each CA n has knowledge of its own local observation matrix H_n only, and the observation noise covariances R_n's are unknown apriori.

(A.2): In digital communications, packets may be lost at random times. To account for this, we let the links (or communication channels among the CAs) to fail, so that the edge set and the connectivity graph of the inter-CA network are time varying. Accordingly, the communication network at time t is modeled as an undirected graph, $G_t = (V, E_t)$ and the graph Laplacians as a sequence of i.i.d. Laplacian matrices $\{L_t\}$. We do not make any distributional assumptions on the link failure model. Although the link failures, and so the Laplacians, are independent at different times, during the same iteration, the link failures can be spatially dependent, i.e., correlated. This is more general and subsumes the erasure network model, where the link failures are independent over space *and* time. Wireless agent networks motivate this model since interference among the wireless communication channels correlates the link failures over space, while, over time, it is still reasonable to assume that the channels are memoryless or independent.

Connectedness of the graph is an important issue. We do not require that the random instantiations G_t of the graph be connected; in fact, it is possible to have all these instantiations to be disconnected. We only require that the graph stays connected on *average*. Denoting $\mathbb{E}[L_t]$ by $\overline{L}$, this is captured by assuming $\lambda_2\left(\overline{L}\right) > 0$. This weak connectivity requirement enables us to capture a broad class of asynchronous communication models; for example, the random asynchronous gossip protocol analyzed in [27] satisfies $\lambda_2\left(\overline{L}\right) > 0$ and hence falls under this framework. On the other hand, we assume that the inter-agent communication is noise-free and unquantized in the event of an active communication link; the problem of quantized data exchange in networked control systems (see, for example, [28], [29], [30] is an active research topic.

(A.3): The sequences $\{L_t\}$ and $\{\mathbf{w}_n(t)\}_{n\in V}$ are mutually independent.

B. Performance Benchmark: Centralized Sequential Estimator

A major drawback of the existing periodic SCADA estimator is that it is implemented only once at the end of the estimation interval $[0, T]$. The estimator, as such, is not responsive to instantaneously acquired data and given that T may be quite large it does not provide real time system monitoring. In practice, it is often desirable to detect abnormal system behavior (manifested by deviations of the system state from its nominal value) as early as possible, so that appropriate control action may be taken in real time thus mitigating critical system phenomena. In view of this, an idealized estimator should be updated in real time and should instantaneously reflect the information in the sensed data. The centralized real

[3]Note that time increments (t to $t+1$) in this paper correspond to the sampling instants of the sensing device with the highest sampling rate. In particular, if the PMU is the fastest measurement device in the system, we assume that samples are collected at each time instant t.

[4]Note that T is measured in terms of multiples of the time increments, the latter corresponding to the measurement frequency of the device with highest system sampling rate. In general, T corresponds to roughly the 5 minute estimation horizon (see [14]) in which the static state assumption holds. Given, that T is measured in terms of multiples of discrete time increments and the devices in question have quite high sensing rate (for example, the PMU has the potential to gather samples in the order of hundreds per second [23]), T is generally quite large. Noting that T also denotes the number of observation samples collected by a non-conventional sensing device in the static estimation horizon, the above implies that T is quite large generally.

time estimator may then be defined as a sequential process $\{\widehat{\mathbf{x}}_{\mathrm{R}}(t)\}_{t=0}^{T}$, where $\widehat{\mathbf{x}}_{\mathrm{R}}(t)$ corresponds to the (real-time) system state estimate at time t based on all sensed information till (and including) time t, i.e.,

$$\widehat{\mathbf{x}}_{\mathrm{R}}(t)=\mathrm{argmin}_{\mathbf{x}\in\mathbb{R}^{M}}\mathcal{Q}_{t}\left(\mathbf{x}\right), \tag{4}$$

where

$$\mathcal{Q}_{t}\left(\mathbf{x}\right)=\sum_{s=0}^{t}\sum_{n=1}^{N}\left(\mathbf{y}_{n}(s)-H_{n}\mathbf{x}\right)^{T}\left(R_{n}\right)^{-1}\left(\mathbf{y}_{n}(s)-H_{n}\mathbf{x}\right). \tag{5}$$

With perfect knowledge of the model parameters, H_n's and the R_n's, the real time estimator is indeed *optimal* as it provides the most up-to-date system state information at all time instants based on the sensed data gathered so far. The sequence $\{\widehat{\mathbf{x}}_{\mathrm{R}}(t)\}_{t=0}^{T}$ exhibits nice statistical properties, such as consistency and asymptotic normality [31], under reasonable structural and analytical assumptions on the sensing model. These are generally formulated as $T\to\infty$, i.e., as the interval of stationarity is relatively large or, alternatively, the number of sensed measurements per estimation period (the sampling rate) is high, which is typically the case with non-conventional sensing resources such as the PMUs. In fact, the following is well-known for the sequence $\{\widehat{\mathbf{x}}_{\mathrm{R}}(t)\}_{t=0}^{T}$ as $T\to\infty$ (see, for example, [32]):

Lemma 1: Let **(A.1)** hold. Then, the estimate sequence $\{\widehat{\mathbf{x}}_{\mathrm{R}}(t)\}$ is consistent, i.e., $\widehat{\mathbf{x}}_{\mathrm{R}}(t)\to\mathbf{x}$ as $t\to\infty$ almost surely (a.s.) Moreover, the estimate sequence $\{\widehat{\mathbf{x}}_{\mathrm{R}}(t)\}$ is asymptotically normal, i.e., as $t\to\infty$, the normalized residual $\sqrt{t}\left(\widehat{\mathbf{x}}_{\mathrm{R}}(t)-\mathbf{x}\right)$ converges in distribution to a zero-mean Gaussian random vector with positive definite covariance Σ_{c}^{-1}, with $\Sigma_{\mathrm{c}}=N\overline{\Sigma_{\mathrm{c}}}$, the latter quantity being defined in (3).

The asymptotic normality essentially means that the mean-squared estimation error (m.s.e.) decreases roughly as $(1/t)\Sigma_{\mathrm{c}}^{-1}$ with the number of samples t. In a sense, the asymptotic (co)variance Σ_{c}^{-1} may be intuitively viewed as the inverse of the signal-to-noise ratio (SNR). The covariance is lower bounded by the inverse of the Fisher information rate [31] of the measurement model and coincides with the latter if the measurement noise is Gaussian.

Although optimal, the usefulness of the centralized real-time estimate $\{\widehat{\mathbf{x}}_{\mathrm{R}}(t)\}$ is masked by the severe requirements it imposes on communication and computation. Implementation of the sequence $\{\widehat{\mathbf{x}}_{\mathrm{R}}(t)\}$ would require accumulating all sensed data at all time instants to a fusion center (or the centralized SCADA), which may not be feasible in real time, especially in the context of the future grid equipped with a plethora of sensing devices with high sampling rates but constrained by a bit limited communication medium. Moreover, the complexity of the optimization problem (4)-(5) blows up with the availability of more and more sensed data making it unscalable with growing network size. Finally, the above idealized estimate may only be constructed with prior knowledge of the sensing model parameters, which may not be the case as perturbations in the grid topology and the sensing process often cause them to switch from time to time.

This motivates us to consider more feasible estimation approaches that are scalable in terms of computation and communication, but nonetheless achieve the performance of the ideal centralized real time estimator $\{\widehat{\mathbf{x}}_{\mathrm{R}}(t)\}$. The estimation approach that we propose in this paper will largely resolve these limitations by being: (1) *distributed*, so that the communication burden imposed by real time transmission of all sensed data to a fusion center is shared by the different CAs through *local* low data rate sequential information exchanges among them; (2) *recursive*, so that the CA estimates are sequentially *updated* in real time as new data (innovation) enters the system, rather than overwhelming a fusion center (SCADA) with huge amounts of data to be processed all at once (4)-(5); and (3) *adaptive*, so that simultaneously with the distributed estimate update process, the uncertain sensing parameters are learnt online with a view to ascertaining the right measurement fusion (combination) rules for estimation.

III. $\mathcal{DAE}$: An Adaptive Distributed State Estimation Framework

Recall the distributed sensing model in (2). In this paper we study sensing uncertainties only, i.e., uncertainties in the measurement noise covariances R_n's, but, otherwise, assume that the local model matrix H_n is perfectly known at the n-th CA for each n. The uncertain noise covariances may correspond to faulty sensors due to natural wear and tear, or may even reflect useless sensor readings caused by unpredictable grid topology changes. Without prior knowledge of the measurement covariances (and hence the relative measurement qualities), implementing the least squares estimator of $\mathbf{x}$ is challenging (even in the centralized case) as the correct measurement fusion weights, being a function of the true covariances, are unavailable apriori. Thus, for successful state estimation, it is necessary to learn the covariances from sensed data, so that the estimation procedure may be updated over time with successive refinements of the covariance information. To this end, we present the following adaptive distributed estimation approach:

Estimate Update: The estimate update at agent n then proceeds as follows:

$$\mathbf{x}_{n}(t+1)=\mathbf{x}_{n}(t)-\beta_{t}\sum_{l\in\Omega_{n}(t)}\left(\mathbf{x}_{n}(t)-\mathbf{x}_{l}(t)\right)+\alpha_{t}K_{n}(t)\left(\mathbf{y}_{n}(t)-H_{n}\mathbf{x}_{n}(t)\right). \tag{6}$$

In the above, $\{\beta_t\}$ and $\{\alpha_t\}$ represent appropriate time-varying weighting factors for the agreement (consensus) and innovation (new observation) potentials, respectively, whereas, $\{K_n(t)\}$ is an adaptively chosen matrix gain process. Also, $\Omega_n(t)$ denotes the time-varying random neighborhood of agent n at time t.

Covariance Learning and Gain Update: The adaptive gain update at sensor n involves another $\{\mathcal{F}_t\}$ adapted distributed learning process that proceeds in parallel with the estimate update. In particular, we set

$$K_{n}(t)=\left(G_{n}(t)+\gamma_{t}I_{M}\right)^{-1}H_{n}^{T}\left(Q_{n}(t)+\gamma_{t}I_{M_{n}}\right)^{-1} \tag{7}$$

where $\{\gamma_t\}$ is a sequence of positive reals, such that $\gamma_t \to 0$ as $t \to \infty$. The $\{Q_n(t)\}$ and $\{G_n(t)\}$ evolve as:

$$Q_n(t+1) = \frac{1}{t}\sum_{s=0}^{t}\mathbf{y}_n(s)\mathbf{y}_n^T(s) - \left(\frac{1}{t}\sum_{s=0}^{t-1}\mathbf{y}_n(s)\right)\left(\frac{1}{t}\sum_{s=0}^{t-1}\mathbf{y}_n(s)\right)^T, \qquad (8)$$

$$G_n(t+1) = G_n(t) - \beta_t \sum_{l\in\Omega_n(t)}(G_n(t) - G_l(t)) + \alpha_t\left(H_n^T\left(Q_n(t)+\gamma_t I_N\right)^{-1}H_n - G_n(t)\right) \qquad (9)$$

These matrices are positive semidefinite with initial conditions $Q_n(0)$ and $G_n(0)$, respectively.

Note, in the above, the sequence $\{Q_n(t)\}$ is the sample covariance (unbiased) and serves as a consistent estimate of the local noise covariance R_n. In fact, the sample covariance estimates are not particularly necessary and any sequence $\{Q_n(t)\}$ such that $Q_n(t) \to R_n$ is sufficient for our purpose. Moreover, the following optional collaborative covariance refinement procedure may be performed at each agent n if it is of interest to obtain more efficient (faster convergence) local covariance estimates:

$$\widehat{R}_n(t) = \frac{1}{t}\sum_{s=0}^{t-1}\left(\mathbf{y}_n(s) - H_n\mathbf{x}_n(s)\right)\left(\mathbf{y}_n(s) - H_n\mathbf{x}_n(s)\right)^T \qquad (10)$$

In the following we introduce some additional assumptions on the observation noise process and the algorithm weight sequences to be in force unless otherwise stated.

(A.4): There exists $\varepsilon_1 > 0$, such that, for all n, $\mathbb{E}_{\theta^*}\left[\|\zeta_n(t)\|^{2+\varepsilon_1}\right] < \infty$, i.e., the measurement noise possesses moments of order greater than 2. Most of the reasonable noise models satisfy the above condition, the typically assumed Gaussian noise in fact possesses moments of all orders (and hence, satisfies the above condition).

(A.5): The weight sequences $\{\alpha_t\}$ and $\{\beta_t\}$ are given by

$$\alpha_t = \frac{a}{(t+1)^{\tau_1}} \text{ and } \beta_t = \frac{b}{(t+1)^{\tau_2}}, \qquad (11)$$

where $a, b > 0$, $0 < \tau_2 \leq \tau_1 \leq 1$ and $\tau_1 > \tau_2 + 1/(2+\varepsilon_1) + 1/2$.

Note that since $\varepsilon_1 > 0$, such a choice of the pair (τ_1, τ_2) is always possible, for example, by taking $\tau_1 = 1$ and $\tau_2 < 1/2 - 1/(2+\varepsilon_1)$. We comment on the choice of the weight sequences $\{\beta_t\}$ and $\{\alpha_t\}$ associated with the consensus and innovation potentials respectively (see (6)). From **(A.5)** we note that both the excitations for agent-collaboration (consensus) and local innovation are persistent, i.e., the sequences $\{\beta_t\}$ and $\{\alpha_t\}$ sum to ∞ - a standard requirement in stochastic approximation type algorithms to drive the updates to the desired from arbitrary initial conditions. Further, the square summability of $\{\alpha_t\}$ ($\tau_1 > 1/2$) is required to mitigate the effect of stochastic sensing noise perturbing the innovations. The requirement $\beta_t/\alpha_t \to \infty$ as $t \to \infty$ ($\tau_1 > \tau_2$), i.e., the asymptotic domination of the consensus potential over the local innovations ensures the right information mixing thus, as shown below, leading to optimal estimation performance. Technically, the different asymptotic decay rates of the two potentials lead to mixed time-scale stochastic recursions whose analyses require new techniques in stochastic approximation as developed in the paper.

We comment on the nature of the distributed learning process leading to online refinement of the estimator gains $K_n(t)$. First note that, in the absence of sensing uncertainties, i.e., with perfect knowledge of the covariances, a (non-adaptive) distributed estimation scheme for $\mathbf{x}$ may be designed as follows:

$$\mathbf{x}_n(t+1) = \mathbf{x}_n(t) - \beta_t\sum_{l\in\Omega_n(t)}(\mathbf{x}_n(t) - \mathbf{x}_l(t)) + \alpha_t\overline{\Sigma_{\mathrm{c}}}^{-1}H_n^T R_n^{-1}\left(\mathbf{y}_n(t) - H_n\mathbf{x}_n(t)\right) \qquad (12)$$

$$\overline{\Sigma_{\mathrm{c}}} = \frac{1}{N}\sum_{n=1}^{N}H_n^T R_n^{-1}H_n. \qquad (13)$$

The matrix $\overline{\Sigma_{\mathrm{c}}}$ denotes the (normalized) Gramian assumed to be positive definite[5]. The estimation methodology in (12) is open loop as no learning is required. Further, this estimation approach requires perfect knowledge of the entire network observation model at each agent, i.e., each network agent is fully aware of the model matrices $\{H_n\}_{n\in V}$ and the covariances $\{R_n\}_{n\in V}$ with V denoting the set of CAs $[1,\cdots,N]$. Under these requirements, it was shown in [33] that the distributed procedure (12) is asymptotically efficient as long as the inter-CA communication network is connected[6], i.e., achieves the asymptotic covariance Σ_{c}^{-1} of the optimal real time centralized SCADA estimator (the estimate $\{\widehat{\mathbf{x}}_{\mathrm{R}}(t)\}$ (4)) in the limit of large T, where $\Sigma_{\mathrm{c}} = N\overline{\Sigma_{\mathrm{c}}}$. In doing so, the scheme in [33] assumes each agent n has complete knowledge of the global parameters $\overline{\Sigma_{\mathrm{c}}}$ (and the local R_n), thus enabling the computation of the optimal local innovation gains at each agent leading to the best asymptotic covariance.

Now, considering the adaptive procedure (6)-(9), we note that the key departure from (12) is that, in the current example, the CAs are not aware of the global quantity $\overline{\Sigma_{\mathrm{c}}}$ and the local covariances R_n's and, hence, apriori are not able to compute and apply the optimal innovation gains $\overline{\Sigma_{\mathrm{c}}}^{-1}H_n^T R_n^{-1}$. This necessitates the additional gain update or learning process, in which over time the CAs try to refine their knowledge of the optimal gain matrices based on past data samples and mutual collaboration with the eventual goal of converging to the exact optimal gains. In particular, the update (8) corresponds to local learning of the unknown covariances R_n from sequentially sensed data, whereas the subsequent collaborative step (9)

[5]The positive definiteness of $\overline{\Sigma_{\mathrm{c}}}$ corresponds to global observability. Note, we do not require the individual CAs to be observable, i.e, each of the matrices H_n could be rank deficient.

[6]Note that the topology of the inter-CA network in terms of communicating information may be arbitrary as long as it is connected. In particular, the communication network may be significantly different and much sparser than the physical grid topology which is usually dense.

refines the knowledge of the (centralized) Gramian $\overline{\Sigma_{\mathrm{c}}}$ needed for choosing the optimal local innovation gains. The adaptive learning step incurs several additional complexities in the analysis of the $\mathcal{DAE}$ scheme (its linear counterpart) with respect to the non-adaptive open loop formulation in (12). Firstly, one needs to establish convergence of the adaptive gain sequence $\{K_n(t)\}$ to the exact optimal gains at each CA. More importantly, even in the event of convergence of the adaptive gains to the desired, the rate of convergence may be slow and apriori it is not clear whether the use of approximate gains (at least in the initial stages) will affect the convergence rate of the estimate update process or not. In other words, one needs to show that the usage of the convergent gain approximations entails no performance loss (in terms of asymptotic covariance) for the estimate update process. Another important observation is that, unlike (12), the estimates $\{\mathbf{x}_n(t), n \in V\}$ are no longer Markovian due to the dependence of the gains $K_n(t)$ on the past observations. From a technical viewpoint, this prevents the direct applicability of standard stochastic approximation techniques (see, for example, [34]) for convergence analysis. The need for non-standard technical approaches is further substantiated by the presence of mixed time scale potentials in the update processes, in particular, the different decay characteristics of the sequences $\{\alpha_t\}$ and $\{\beta_t\}$ for the consensus and innovation respectively.

IV. Main Results

We formally state the main results of the paper. The proofs follow from some general studies on distributed adaptive procedures recently undertaken in [35].

The first result concerns the asymptotic agreement or consensus among the various agent estimates.

Theorem 2: Let assumptions **(A.1)-(A.5)** hold. Then for each τ_0 such that

$$0 \leq \tau_0 < \tau_1 - \tau_2 - \frac{1}{2+\varepsilon_1}, \tag{14}$$

we have

$$\mathbb{P}\left(\lim_{t\to\infty}(t+1)^{\tau_0}\left\|\mathbf{x}_n(t)-\mathbf{x}_l(t)\right\|=0\right)=1 \tag{15}$$

for any pair of CAs n and l.

In words, Theorem 2 shows that the rate of agreement (at least the order) depends only on the difference $\tau_1 - \tau_2$ of the algorithm weight parameters, the latter quantifying the intensities of the *global agreement* and *local innovation* potentials relative to each other. Interestingly, the order of this convergence is independent of the network topology (as long as it is connected in the mean) and the distributed gain learning process (7)-(9).

Theorem 3: Let assumptions **(A.1)-(A.5)** hold with $\tau_1 = 1$ and $a \geq 1$. Then, for each n the estimate sequence $\{\mathbf{x}_n(t)\}$ is strongly consistent. In particular, we have

$$\mathbb{P}_{\theta^*}\left(\lim_{t\to\infty}(t+1)^{\tau}\left\|\mathbf{x}_n(t)-\theta^*\right\|=0\right)=1 \tag{16}$$

for each n and $\tau \in [0, 1/2)$.

The above convergence rate is optimal for pathwise convergence of estimates in the sense that (16) does not hold with $\tau = 1/2$ even for a centralized estimate sequence. This, in turn, is due to the asymptotic normality of the centralized estimator with a non-degenerate asymptotic covariance (see Theorem 4 for details). Again, the interesting and non-trivial fact to note here is that the distributed adaptive estimators retain the centralized convergence rate irrespective of the apparent information loss due to sparse inter-agent communication and lack of model information apriori.

The following result concerns the asymptotic normality of the estimates generated by the distributed procedure and establishes the asymptotic efficiency of the $\mathcal{DAE}$ scheme.

Theorem 4: Let assumptions **(A.1)-(A.5)** hold with $\tau_1 = 1$ and $a = 1$. Let Σ_{c} denote the matrix $\sum_{n=1}^{N} H_n^T R_n^{-1} H_n$. Then, for each n

$$\sqrt{(t+1)}\left(\mathbf{x}_n(t)-\theta^*\right) \Longrightarrow \mathcal{N}\left(\mathbf{0}, \Sigma_{\mathrm{c}}^{-1}\right), \tag{17}$$

where $\mathcal{N}(\cdot)$ and $\Longrightarrow$ denote the Gaussian distribution and weak convergence respectively.

Referring to the discussion in Section II-B (Lemma 1 in particular), we note that the $\mathcal{DAE}$ leads to the optimal error covariance decay attainable, in general, by any estimator (centralized) with information of the model parameters H_n's and R_n's only and no other specifics of the observation noise process. In particular, the distributed and adaptive $\mathcal{ADLE}$ is optimal in the class of linear centralized estimators when the noise distribution is arbitrary and is optimal in the Fisher information sense if the noise process is Gaussian. In a sense, Theorem 4 justifies the applicability and advantage of distributed estimation schemes. Apart from issues of robustness, implementing a centralized estimator is much more communication intensive as it requires transmitting all sensor data to a fusion center at all times. On the other hand, the distributed $\mathcal{DAE}$ algorithm requires only sparse local communication among the CAs at each step, and achieves the performance of a centralized estimator asymptotically as long as the communication network stays connected in the mean.

V. Conclusion

In this paper we presented a formalism for distributed state estimation in the electrical power grid. Under rather weak assumptions of global observability and connectivity of the control area communication network, the proposed distributed adaptive scheme is shown to yield consistent system state estimates, the convergence rate being optimal in the Fisher information sense. As such, the proposed approach (1) exploits in real-time the sensing diversity offered by non-conventional sensing resources with higher accuracy and faster system sampling rate; (2) adapts to (stochastic) uncertainties in the information content of the sensing resources through online learning in conjunction with the estimation procedure, so as to effectively combine (fuse) the sensed data from different sensing methodologies by properly weighting their measurements; and (3) provides pervasive network-wide intelligence to improve the performance of the local protection units and controllers by appropriately distributing the information processing and communication overhead among the various network entities. Future research in this direction include generalizing the above adaptive distributed framework to nonlinear (AC) models (see also [36]) and dynamic state estimation in the smart grid.

REFERENCES

[1] R. Anderson, "The distributed storage-generation smart electric grid of the future," *in Proc. Workshop, The 1050 Solution: Technol. Policies for a Low-Carbon Future*, 2004.

[2] A. Gomez-Exposito, A. Abur, A. de la Villa Jaen, and C. Gomez-Quiles, "A multilevel state estimation paradigm for smart grids," *Proceedings of the IEEE*, vol. 99, no. 6, pp. 952 – 976, June 2011.

[3] A. Bose, "Smart transmission grid applications and their supporting infrastructure," *IEEE Trans. Smart Grids*, vol. 1, no. 1, pp. 11–19, June 2010.

[4] F. F. Wu, K. Moslehi, and A. Bose, "Power system control centers: Past, present, and future," *Proceedings of the IEEE*, vol. 93, no. 11, pp. 1890–1908, November 2005.

[5] V. Terzija, G. Valverde, D. Cai, P. Regulski, V. Madani, J. Fitch, S. Skok, M. M. Begovic, and A. Phadke, "Power system control centers: Past, present, and future," *Proceedings of the IEEE*, vol. 99, no. 1, pp. 80–93, Jan. 2011.

[6] M. D. Ilić, L. Xie, U. A. Khan, and J. M. F. Moura, "Modeling of future cyberphysical energy systems for distributed sensing and control," *IEEE Transactions on Systems, Man and Cybernetics*, vol. 40, no. 4, pp. 825 – 838, July 2010.

[7] M. D. Ilić, J. W. Black, and M. Prica, "Distributed electric power systems of the future: Institutional and technological drivers for near-optimal performance," *Elect. Power Syst. Res.*, vol. 77, no. 9, pp. 1160–1177, July 2009.

[8] K. Tomsovic, D. Bakken, V. Venkatasubramanian, and A. Bose, "Designing the next generation of real-time control, communication, and computations for large power systems," *Proc. IEEE*, vol. 93, no. 5, pp. 965–979, May 2005.

[9] S. K. Mazumder, K. Acharya, and M. Tahir, "Joint optimization of control performance and network resource utilization in homogeneous power networks," *IEEE Trans. Ind. Electron*, vol. 56, no. 5, pp. 1736–1745, May 2009.

[10] S. M. Amin and B. F. Wollenberg, "Toward a smart grid: power delivery for the 21st century," *IEEE Power and Energy Magazine*, vol. 3, no. 5, pp. 34 – 41, Sept. 2005.

[11] A. Bose, "Smart transmission grid applications and their supporting infrastructure," *IEEE Transactions on Smart Grids*, vol. 1, no. 1, pp. 11–19, June 2010.

[12] F. F. Wu, K. Moslehi, and A. Bose, "Power system control centers: Past, present, and future," *Proceedings of the IEEE*, vol. 93, no. 11, pp. 1890–1908, November 2005.

[13] A. Bose, A. Abur, K. Y. K. Poon, and R. Emami, "Implementation issues for hierarchical state estimators," *PSERC Final Project Report*, August 2010.

[14] ISO New England Operating Procedure No. 18, "Metering and telemetering criteria," effective Date: June 1, 2010. Revision No. 10. http://www.iso-ne.com/rules_proceds/operating/isone/op18/op18_rto_final.pdf. [Online]. Available: http://www.iso-ne.com/rules_proceds/operating/isone/op18/op18_rto_final.pdf

[15] R. F. Nuqui and A. G. Phadke, "Phasor measurement unit placement techniques for complete and incomplete observability," *IEEE Transactions on Power Systems*, vol. 20, no. 4, pp. 2381 – 2388, 2005.

[16] *IEEE Standard Definition, Specification and Analysis of Systems Used for Supervisory Control, Data Acquisition, and Automatic Control*. IEEE Std. C37.1.1994, 1994.

[17] J. D. McDonald, *Electric Power Substation Engineering*. Boca Raton, FL: CRC Press, 2003.

[18] K. Brand, V. Lohmann, and W. Wimmer, *Substation Automation Handbook*. Bremgarten, Switzerland: Utility Automation Consulting Lohmann, 2003.

[19] F. C. Schweppe, J. Wildes, and A. Bose, "Power system static state estimation, parts I, II and III," *IEEE Trans. on Power Apparatus and Systems*, vol. 89, no. 1, pp. 120–135, January 1970.

[20] A. Monticelli, "Electric power system state estimation," *Proceedings of the IEEE*, vol. 88, no. 2, pp. 262 – 282, Feb. 2000.

[21] E. Handschin, F. C. Schweppe, J. Kohlas, and A. Fiechter, "Bad data analysis for power system state estimation," *IEEE Transactions on Power Apparatus and Systems*, vol. 94, no. 2, pp. 329–337, April 1975.

[22] A. Monticelli, F. F. Wu, and M. Yen, "Mutiple bad data identification for state estimation by combinatorial optimization," *IEEE Transactions on Power Delivery*, vol. 1, no. 3, pp. 361 – 369, July 1986.

[23] North American Electric Reliability Corporation, "Real-time application of synchrophasors for improving reliability," May 2010, http://www.nerc.com/docs/oc/rapirtf/Rapir_final-report-Aug-2010-lc.pdf. [Online]. Available: http://www.nerc.com/docs/oc/rapirtf/Rapir_final-report-Aug-2010-lc.pdf

[24] H. J. Koglin, T. Neisius, G. Beibler, and K. D. Schmitt, "Bad data detection and identification," *International Journal of Electrical Power and Energy Systems*, vol. 12, no. 2, pp. 94–103, April 1990.

[25] L. Mili, T. V. Cutsem, and M. Ribbens-Pavella, "Hypothesis testing identification: A new method for bad data analysis in power system state estimation," *IEEE Trans. Power App. Syst.*, vol. PAS-103, no. 11, p. 32393252, Nov. 1984.

[26] F. R. K. Chung, *Spectral Graph Theory*. Providence, RI : American Mathematical Society, 1997.

[27] S. Boyd, A. Ghosh, B. Prabhakar, and D. Shah, "Randomized gossip algorithms," *IEEE/ACM Trans. Netw.*, vol. 14, no. SI, pp. 2508–2530, 2006.

[28] S. Tatikonda and S. Mitter, "Control under communication constraints," *IEEE Transactions on Automatic Control*, vol. 49, no. 7, pp. 1056 – 1068, July 2004.

[29] A. Matveev and A. Savkin, "The problem of state estimation via asynchronous communication channels with irregular transmission times," *IEEE Transactions on Automatic Control*, vol. 48, no. 4, pp. 670–676, April 2006.

[30] K. Li and J. Baillieul, "Robust and efficient quantization and coding for control of multidimensional linear systems under data rate constraints," *International Journal of Robust and Nonlinear Control Special Issue: Communicating-Agent Networks*, vol. 17, no. 10-11, pp. 898–920, July 2007.

[31] L.M. Le Cam and G.L. Yang, *Asymptotics in statistics: some basic concepts*. New York: Springer-Verlag Inc., 2000.

[32] R. I. Jennrich, "Asymptotic properties of non-linear least squares estimators," *The Annals of Mathematical Statistics*, vol. 40, no. 2, pp. 633–643, 1969.

[33] S. Kar and J. M. F. Moura, "Convergence rate analysis of distributed gossip (linear parameter) estimation: Fundamental limits and tradeoffs," *IEEE Journal of Selected Topics in Signal Processing Signal Processing in Gossiping Algorithms Design and Applications*, vol. 5, no. 4, pp. 674–690, August 2011.

[34] M. Nevel'son and R. Has'minskii, *Stochastic Approximation and Recursive Estimation*. Providence, Rhode Island: American Mathematical Society, 1973.

[35] S. Kar, J. Moura, and H. Poor, "Distributed linear parameter estimation: asymptotically efficient adaptive strategies," 2011, submitted to the SIAM J. Control Optim., Initial Submission: Sept. 2011. http://arxiv.org/abs/1109.4960.

[36] S. Kar, J. Moura, and K. Ramanan, "Distributed parameter estimation in sensor networks: nonlinear observation models and imperfect communication," *IEEE Transactions on Information Theory*, vol. 58, no. 6, pp. 3575–3605, June 2012.