IEEE Signal Processing Society | IEEE Transactions on Signal Processing

Consensus on State and Time: Decentralized Regression with Asynchronous Sampling

Journal:	*Transactions on Signal Processing*
Manuscript ID:	T-SP-17750-2014.R1
Manuscript Type:	Regular Paper
Date Submitted by the Author:	n/a
Complete List of Authors:	Wai, Hoi-To; University of California, Davis, Electrical and Computer Engineering Scaglione, Anna; UC Davis, ECE
EDICS:	142. SPE-DETC Detection and estimation in power grid < SPE SIGNAL PROCESSING FOR ENGERGY AND POWER SYSTEMS, 143. SPE-DP Distributed processing and protocols in energy systems < SPE SIGNAL PROCESSING FOR ENGERGY AND POWER SYSTEMS, 105. SAM-APPL Applications of sensor & array multichannel processing < SAM SENSOR ARRAY AND MULTICHANNEL PROCESSING

Consensus on State and Time: Decentralized Regression with Asynchronous Sampling

Hoi-To Wai and Anna Scaglione

***Abstract*—An implicit assumption made in several studies on sensor systems is that the time and frequency at which sensor measurements are taken is consistent across all the distributed sensing sites. In reality, the times of measurement often lack consistency and integrity, and this is an intrinsic vulnerability of wide area sensor system. Data logs coming from different Analog to Digital Converters (ADCs) are not in phase and may differ also in the sampling rate, in some cases because heterogeneity in the sensors and in others because the data are simply not refreshed in the data historians with the same frequency. Lack of good synchronization in sensing may be the result of a malfunction or also due to intentional delay attacks.**

This premise motivates our work, where we advance the area of decentralized signal processing and consider explicitly timing errors and non-homogenous sampling rates in least square estimation problems with distributed sensing. For linear observations models, we provide a necessary and sufficient condition for identifiability of the time offsets. We propose an algorithm for the joint regression on the state vector and time offsets. The algorithm also exploits the asynchrony and redundancy in the spatial sampling to attain sub-Nyquist sampling resolution of the slow sensor feeds. Importantly, this also leads to the development of a novel decentralized algorithm. The efficacies of the proposed decentralized algorithm are shown by both convergence analysis and numerical simulations.

Index terms— decentralized state estimation, sampling offsets, sub-Nyquist sampling, smart grid.

I. Introduction

TODAY there is significant interest in developing *decentralized* signal processing techniques for solving regression problems that arises in array processing and control applications (e.g., see [3]–[12] and the references therein). These decentralized algorithms overcome the lack of *observability* in individual sensors by merging communication with computations in a resilient fashion, relying on (possibly randomized) near-neighbors communications.

One of the implicit assumptions made in the vast majority of related literatures is that the measurements are sampled in a synchronous manner. Such an assumption is valid only when: i) the system state evolution is sufficiently slow that the lack of synchrony in sampling is negligible; or ii) the timing information is sufficiently accurate (with the aid of for instance of a GPS receiver) to calibrate the ADCs; and iii) the sensors employed follow the same sampling rate. These assumption are quite limiting. For example, in power grid, local clocks in the measurement devices are prone to malicious attacks [13]; in sensor array processing, the desire of processing signals over high frequency carriers and wide bandwidths has made the design of hardwares for synchronization more challenging.

This paper attacks the problem of non-ideal sampling by addressing these issues in a unified manner, while considering the case of power grid as an immediate application example. Our formulation imposes very mild restrictions. Specifically, assuming that the relationship between the measurements and state is *memoryless*, we model the system state variables as band-limited continuous-time signals and the measurements as samples taken at different sampling frequencies and with an unknown time offset. The analysis and algorithms we propose rely only on the sampling expansion (i.e. *smoothness* in the state signal). On the practical side, by representing the down-sampled and time-shifted signals in the frequency domain, we tackle the new regression problem on state and time offsets using a novel decentralized algorithm. The decentralized algorithm is proven to converge both analytically and empirically.

The remainder of this paper is organized as follows. In Section II, we introduce the system model with linear measurement and non-ideal sampling. Specifically, we derive an equivalent frequency-domain representation where our analysis and algorithms are based upon. In Section III, we derive conditions under which the accurate state and times offsets are recoverable via solving the proposed regression problem. A decentralized algorithm for tackling the proposed regression problem will be discussed and analyzed in Section IV. In Section V, we discuss the extension to non-linear measurement models and its applications on the power grids. Finally, the paper is concluded by the simulation results that show the efficiencies of the proposed method in Section VI.

Notations: We follow the standard notations used in signal processing literature. The operator $\otimes$ denotes the Kronecker product, $(\cdot)^*$ denotes complex conjugate, $\mathrm{DTFT}\{\cdot\}$ denotes the standard DTFT transformation $\sum_{n=-\infty}^{\infty} \mathbf{x}[n] e^{-j\omega n}$.

A. Related works

Techniques for mitigating timing errors have long been considered in the control theory literatures, e.g., [14]–[17]. A common feature among these works is that they adopt a Kalman filtering approach and are often combined with a Taylor approximation to the system dynamics. As a result, applying these techniques requires an a-priori knowledge of

This material is based upon work supported by the Department of Energy under Award Number DE-OE0000097. Preliminary versions of this work were presented at IEEE SAM 2014 [1] and Asilomar 2014 [2].

H.-T. Wai and A. Scaglione are with Department of Electrical and Computer Engineering, University of California, Davis, CA 95616, USA. E-mails: {htwai,ascaglione}@ucdavis.edu.

the dynamic equations that represent the evolution of system state, and that the timing offset is sufficiently small such that the Taylor approximation is accurate. A good example of prior works that is relevant to ours is [18]. In this paper, Yang et. al. considered a static system where the effects of non-ideal sampling is resulted from down-converting signals with high-frequency carriers (e.g., in power system state estimation (PSSE) with linear measurements) and tackled the joint time offsets and state regression using a Taylor approximation.

What differentiates our work is the direct manner in which we model the effects of non-ideal sampling. In fact, we do not require any a-priori knowledge on the state evolution dynamics nor assumptions on the magnitude of time offsets. Our work is in the same spirit with sampling using time-interleaved ADCs [19], [20]. Furthermore, it overlaps with the recent works on sub-Nyquist recovery in [21]–[23], reviewed in [24]. In these works, it is usually required that the signal being sampled satisfies certain properties known a-priori to the system, while ours focuses on the blind calibration problem and fusion of measurements.

Decentralized algorithms over sensor network have been considered in [11], [12], [25]–[27]. For example, [11], [12] combines gradient (or quasi-Newton) descent and consensus protocols to develop the diffusion-based LMS (or RLS) algorithm; [28] proposes a Gossip-based Gauss-Newton algorithm. While most of these algorithms applies to general optimization problems, they are guaranteed to converge only when the optimization is convex. However, our regression problem with non-ideal sampling is non-convex and the aforementioned algorithms cannot be directly applied. We have proposed a new decentralized algorithm that exploits structures in the said regression problem. More importantly, the algorithm is proven to converge under certain conditions.

II. System model

Consider a sensor network equipped with P sensors. The sensor network monitors the system state that can be modelled as a continuous-time signal $\mathbf{x}_c(t) \in \mathbb{C}^N$. Specifically, we study the case when the measurement is linear in the system state such that the pth sensor observes the following at time t:

$$\boldsymbol{\zeta}_p(t) = \mathbf{H}_p \mathbf{x}_c(t) + \boldsymbol{v}_p(t), \tag{1}$$

where $\mathbf{H}_p \in \mathbb{C}^{M_p \times N}$ is the measurement matrix with $M_p \leq N$ and $\boldsymbol{v}_p(t) \sim \mathcal{CN}(0, \sigma_w^2 \mathbf{I})$ is an additive white noise. For the applications of (1) on sensor array processing, we refer our readers to [6]–[8]; in addition, some recent applications on power system state estimation can be found in [3]–[5]. Our model can also be extended to the general case with non-linear measurement, i.e., when $\boldsymbol{\zeta}_p(t)$ is non-linear in $\mathbf{x}_c(t)$; see Section V.

The objective is to estimate $\mathbf{x}_c(t)$ using $\boldsymbol{\zeta}_p(t)$. In the control theory literatures [15]–[18], [29], the system state $\mathbf{x}_c(t)$ is often modelled by a linear/non-linear dynamical system. We consider the scenario when such a knowledge on the underlying dynamical system is not available. Instead, we study the state estimation problem under the model implied by the following assumptions:

Assumption 1. *The system state $\mathbf{x}_c(t)$ is band-limited by $W/2$ Hz, i.e., $\mathbf{x}_c(t)$ is a smooth signal.*

Assumption 2. *The measurements are collected from the continuous-time signal $\boldsymbol{\zeta}_p(t)$ under a non-ideal sampling model — the pth sensor samples $\boldsymbol{\zeta}_p(t)$ at time:*

$$t_p^n = (nA_p - b_p)T_s, \ n = 0, 1, 2, ..., \tag{2}$$

where $T_s = 1/W$ second is the Nyquist sampling period, $b_p \in \mathbb{R}$ is the normalized time offset at sensor p and the sampling factor $A_p \geq 1$ is an integer. In addition, we set $b_1 = 0$ to avoid ambiguity.

Assumption 3. *The down-sampling factor $A_p \in \mathbb{Z}_+$ is known while the time offset $b_p \in \mathcal{B}_p$ is unknown. The interval $\mathcal{B}_p \subseteq \mathbb{R}$ is convex and known.*

The consequences of Assumption 1 to 3 are discussed as follows. Assumption 1 is the key enabling assumption for estimating $\mathbf{x}_c(t)$ from samples of measurements in (1). Under the assumption, it suffices to obtain $\mathbf{x}[n] = \mathbf{x}_c(nT_s)$ in order to estimate $\mathbf{x}_c(t)$. In fact, by observing that

$$\boldsymbol{\zeta}_p(nT_s) = \mathbf{H}_p \mathbf{x}[n] + \boldsymbol{v}_p(nT_s), \tag{3}$$

the system state $\mathbf{x}[n]$ can be readily estimated by solving a least square optimization with the data $\{\boldsymbol{\zeta}_p(nT_s)\}_p$. In fact, Assumption 1 with the synchronous sampling model is one of the implicit assumptions made in most literatures on sensor array processing [3]–[9].

In reality, obtaining the set of data $\{\boldsymbol{\zeta}_p(nT_s)\}_p$ is impossible since it requires the physical system to sample at time $t_p^n = nT_s$ for *all* sensors. This requires synchronization, and it is hard to enforce for sensors placed over a physical system that occupies a wide area. In this case, often measurement samples are collected at *heterogenous* sampling rate and with sampling offsets. This is the reason why Assumption 2 was imposed. Under Assumption 2, the *sampled* version of (1) can be expressed as:

$$\boldsymbol{\zeta}_p[n] \triangleq \boldsymbol{\zeta}_p(t_p^n) = \mathbf{H}_p \mathbf{x}_c((nA_p - b_p)T_s) + \boldsymbol{v}_p[n]. \tag{4}$$

Lastly, Assumption 3 is justified by the fact that the sampling rate of a sensor is usually known a-priori, while the time offsets are unpredictable.

The aim of this paper is to study the estimation problem of $\mathbf{x}[n]$ and b_p. We first observe that estimating $\mathbf{x}[n]$ and b_p on a sample-by-sample basis from $\{\boldsymbol{\zeta}_p[n]\}_{p=1}^K$ alone is impossible as the former terms do not appear in the right hand side of (4). As a remedy, we consider a frequency domain representation for (4) and leverage on the following observation:

Observation 1. *Let $\mathbf{x}_c(t)$ be a bandlimited signal with bandwidth $W/2$ Hz. We denote $\mathbf{x}[n] = \mathbf{x}_c(nT_s)$ with $T_s = 1/W$ as its discrete time equivalent and $\mathbf{X}(e^{j\omega}) = \mathrm{DTFT}\{\mathbf{x}[n]\}$ is the discrete time Fourier transform (DTFT) spectrum. Then:*

$$\mathrm{DTFT}\{\mathbf{x}_c((nA - b_p)T_s)\} = \frac{1}{A} \sum_{a=0}^{A-1} e^{-jb_p \Omega_A^a(\omega)} \mathbf{X}\left(e^{j\Omega_A^a(\omega)}\right), \tag{5}$$

where

$$\Omega_A^a(\omega) \triangleq \left(\frac{\omega}{A} - \frac{a}{A} 2\pi\right) \bmod (-\pi, \pi]. \tag{6}$$

The proof of Observation 1 is relegated to Appendix A. In fact, the DTFT spectrum of $\mathbf{x}((nA-b_p)T_s)$ is a weighted combination of the stretched and shifted versions of $\mathbf{X}(e^{j\omega})$.

For ease of exploration, it will be useful to consider that the K sensors are sampling at the same (sub-Nyquist) rate. This can be done by creating further decimated samples from $\boldsymbol{\zeta}_p[n]$. In particular, we define the constants:

$$A \triangleq \mathsf{LCM}\{A_1, ..., A_P\} \text{ and } Q_p \triangleq A/A_p \tag{7}$$

and decompose $\boldsymbol{\zeta}_p[n]$ into Q_p streams of samples:

$$\boldsymbol{\zeta}_p^q[n] \triangleq \boldsymbol{\zeta}_p[Q_p n - q],\ q = 0, 1, ..., Q_p - 1. \tag{8}$$

Each of $\boldsymbol{\zeta}_p^q[n]$ is a sequence of samples of $\boldsymbol{\zeta}_p(t)$ downsampled by A and offsetted by $qA_p + b_p$ unit of time. Applying Observation 1 to (4) and (8) gives:

$$\mathbf{Z}_p^q(e^{j\omega}) = \frac{1}{A}\times \sum_{a=0}^{A-1} e^{-j(b_p+qA_p)\Omega_A^a(\omega)}\mathbf{H}_p\mathbf{X}\left(e^{j\Omega_A^a(\omega)}\right) + \mathbf{V}_p(e^{j\omega}) \tag{9}$$

for $\omega \in (-\pi, \pi]$. Here, $\mathbf{X}(e^{j\omega}) = \sum_n \mathbf{x}[n]e^{-j\omega n}$ and $\mathbf{V}_p^q(e^{j\omega}) = \sum_n \boldsymbol{v}_p^q[n]e^{-j\omega n}$ are the DTFT spectrum of $\mathbf{x}[n]$ and $\boldsymbol{v}_p^q[n]$, respectively.

To simplify (9), we define the following extended state spectrum:

$$\mathbf{X}_A(\omega) \triangleq \left[\mathbf{X}\left(e^{j\Omega_A^0(\omega)}\right)^T\ \mathbf{X}\left(e^{j\Omega_A^1(\omega)}\right)^T\ \ldots\ \mathbf{X}\left(e^{j\Omega_A^{A-1}(\omega)}\right)^T\right]^T \tag{10}$$

and the extended measurement matrix, i.e.,

$$\boldsymbol{\mathcal{H}}_p(b_p, \omega) \triangleq \boldsymbol{\Theta}_p(b_p, \omega) \otimes \mathbf{H}_p \tag{11}$$

where

$$\boldsymbol{\Theta}_p(b_p, \omega) \triangleq \frac{1}{A}\times \begin{bmatrix} e^{-j(b_p)\Omega_A^0(\omega)} & \ldots & e^{-j(b_p)\Omega_A^{A-1}(\omega)} \\ e^{-j(b_p+A_p)\Omega_A^0(\omega)} & \ldots & e^{-j(b_p+A_p)\Omega_A^{A-1}(\omega)} \\ \ldots & \ldots & \ldots \\ e^{-j(b_p+(Q_p-1)A_p)\Omega_A^0(\omega)} & \ldots & e^{-j(b_p+(Q_p-1)A_p)\Omega_A^{A-1}(\omega)} \end{bmatrix}.$$

Then, (9) can be conveniently expressed as:

$$\boldsymbol{\mathcal{Z}}_p(e^{j\omega}) = \boldsymbol{\mathcal{H}}_p(b_p, \omega)\mathbf{X}_A(\omega) + \boldsymbol{\mathcal{V}}_p(e^{j\omega}), \tag{12}$$

for $\omega \in (-\pi, \pi]$, where $\boldsymbol{\mathcal{Z}}_p(e^{j\omega})$ and $\boldsymbol{\mathcal{V}}_p(e^{j\omega})$ can be formed by vertically concatenating the vectors $\{\mathbf{Z}_p^q(e^{j\omega})\}_q$ and $\{\mathbf{V}_p^q(e^{j\omega})\}_q$, respectively.

Observe that there is a one-to-one correspondence between $\mathbf{X}_A(\omega)$ and $\mathbf{X}(e^{j\omega})$ since each entry in the extended spectrum $\mathbf{X}_A(\omega)$ is non-repeating as the intervals $\Omega_A^a((-\pi, \pi])$ and $\Omega_A^b((-\pi, \pi])$ are disjoint whenever $a \neq b$. Consequently, estimating $\mathbf{X}_A(\omega)$ is equivalent to estimating the time domain sequence $\{\mathbf{x}[n]\}_n$. The latter can be obtained by first converting $\mathbf{X}_A(\omega)$ to $\mathbf{X}(e^{j\omega})$, and then performing an inverse DTFT.

To conclude, we observe that the measured spectrum $\boldsymbol{\mathcal{Z}}_p(e^{j\omega})$ can be expressed as a linear transformation of $\mathbf{X}_A(\omega)$. In the sequel, we will study the model (12) from two different aspects — i) to derive a set of identifiability conditions such that we can uniquely identify b_p and $\mathbf{X}_A(\omega)$ from $\boldsymbol{\mathcal{Z}}_p(e^{j\omega})$; ii) to propose a tractable, decentralized algorithm for retrieving b_p and $\mathbf{X}_A(\omega)$.

III. Identifiability Condition

This section derives an identifiability condition for (12). Recall that our intention is to estimate *jointly* the time offsets and the state spectrum, i.e., the tuple $(\mathbf{X}_A(\omega), \mathbf{b})$ with $\mathbf{b} = (b_1, b_2, ..., b_K)$, from the linear system (12). Under such context, we define:

Definition 1. *The sensing system $\{\mathbf{H}_p\}_{p=1}^P$ is said to be identifiable*[1] *under non-ideal sampling if and only if for any $\{\boldsymbol{\mathcal{Z}}_p(e^{j\omega})\}_p$ that is generated by $(\mathbf{X}_A(\omega), \mathbf{b})$, the tuple $(\mathbf{X}_A(\omega), \mathbf{b})$ is the only one satisfying* (12).

In other words, if the system is *identifiable*, then one can recover the tuple $(\mathbf{X}_A(\omega), \mathbf{b})$ unambiguously from $\{\boldsymbol{\mathcal{Z}}_p(e^{j\omega})\}_p$. As a comment, blind identification conditions are explored in the absence of noise, see e.g., [30], [31]. Our main result is summarized by the following proposition:

Proposition 1. *Consider the following matrix:*

$$\hat{\boldsymbol{\mathcal{H}}}(\hat{\mathbf{b}}, \mathbf{b}, \omega) \triangleq \left[\overline{\boldsymbol{\mathcal{H}}}(\hat{\mathbf{b}}, \omega)\ \ -\overline{\boldsymbol{\mathcal{H}}}(\mathbf{b}, \omega)\right], \tag{13}$$

where $\mathbf{b} = [b_1\ b_2\ \ldots\ b_P]$ and

$$\overline{\boldsymbol{\mathcal{H}}}(\mathbf{b}, \omega) \triangleq \begin{bmatrix} \boldsymbol{\mathcal{H}}_1(b_1, \omega) \\ \cdots \\ \boldsymbol{\mathcal{H}}_P(b_P, \omega) \end{bmatrix}. \tag{14}$$

Assuming that the measurement is noiseless, i.e., $\boldsymbol{\mathcal{V}}_p(e^{j\omega}) = \mathbf{0}$, $\mathbf{X}_A(\omega) \neq \mathbf{0}$ and $\mathrm{rank}(\overline{\boldsymbol{\mathcal{H}}}(\mathbf{b}, \omega)) = AN$ for all ω. Then, the sensing system $\{\mathbf{H}_p\}_p$ is identifiable if and only if

$$\mathrm{rank}(\hat{\boldsymbol{\mathcal{H}}}(\hat{\mathbf{b}}, \mathbf{b}, \omega)) = 2AN. \tag{15}$$

for all ω whenever $\hat{\mathbf{b}} \neq \mathbf{b}$ (with $b_1 = \hat{b}_1 = 0$).

Proof. Notice that any tuple $(\hat{\mathbf{X}}_A(\omega), \hat{\mathbf{b}})$ satisfying (12) with $\boldsymbol{\mathcal{V}}_p(e^{j\omega}) = \mathbf{0}$ must fulfill the homogeneous equation:

$$\hat{\boldsymbol{\mathcal{H}}}(\hat{\mathbf{b}}, \mathbf{b}, \omega)\begin{bmatrix} \hat{\mathbf{X}}_A(\omega) \\ \mathbf{X}_A(\omega) \end{bmatrix} = \mathbf{0}, \tag{16}$$

since $\boldsymbol{\mathcal{Z}}_p(e^{j\omega}) = \boldsymbol{\mathcal{H}}_p(b_p, \omega)\mathbf{X}_A(e^{j\omega})$ for all p and ω.

We first prove the sufficient condition. If $\hat{\boldsymbol{\mathcal{H}}}(\hat{\mathbf{b}}, \mathbf{b}, \omega)$ is full column rank whenever $\hat{\mathbf{b}} \neq \mathbf{b}$, then (16) implies that whenever $\mathbf{b} \neq \hat{\mathbf{b}}$, we have $\mathbf{X}_A(\omega) = \hat{\mathbf{X}}_A(\omega) = \mathbf{0}$, leading to a contradiction. As such, $\mathbf{b} = \hat{\mathbf{b}}$ and the assumption $\mathrm{rank}(\overline{\boldsymbol{\mathcal{H}}}(\mathbf{b}, \omega)) = AN$ guarantees that $\hat{\mathbf{X}}_A(\omega) = \mathbf{X}_A(\omega)$.

For the necessary condition, let us assume that $\hat{\boldsymbol{\mathcal{H}}}(\hat{\mathbf{b}}, \mathbf{b}, \omega)$ is not full rank for some $\hat{\mathbf{b}} \neq \mathbf{b}$. In this case, for some instances of $\mathbf{X}_A(\omega)$ there exists a tuple $(\hat{\mathbf{X}}_A(\omega), \hat{\mathbf{b}}) \neq (\mathbf{X}_A(\omega), \mathbf{b})$ such that (16) is satisfied. This contradicts the uniqueness of $(\mathbf{X}_A(\omega), \mathbf{b})$. **Q.E.D.**

Proposition 1 provides the identifiability condition for which the joint recovery of both sampling offsets and state

[1] Identifiability is interchangeable here with the notion of observability that is prevalent in the control theory terminology, e.g., [3].

spectrum are possible. However, verifying the condition that $\text{rank}(\hat{\boldsymbol{\mathcal{H}}}(\hat{\mathbf{b}}, \mathbf{b}, \omega)) = 2AN$ for all ω is a non-trivial problem. In the following, we derive several insightful conditions for identifiability which are easy to verify.

We begin by examining an assumption made in Proposition 1 that $\text{rank}(\overline{\boldsymbol{\mathcal{H}}}(\mathbf{b}, \omega)) = AN$. This assumption is equivalent to enforcing identifiability on the system with *known* time offsets, which is necessitated by:

$$\textstyle\sum_{p=1}^{P} \min\{Q_p \text{rank}(\mathbf{H}_p), AN_c(\mathbf{H}_p)\} \geq AN, \tag{17}$$

where $N_c(\mathbf{H}_p) \leq N$ is the number of non-zero columns in $\mathbf{H}_p$. The above condition is obtained by upper bounding $\overline{\boldsymbol{\mathcal{H}}}(\mathbf{b}, \omega)$ by the sum of rank of its sub-matrices, $\sum_p \text{rank}(\boldsymbol{\mathcal{H}}_p(b_p, \omega))$. Furthermore, we have:

$$\begin{aligned} \text{rank}(\boldsymbol{\mathcal{H}}_p(b_p, \omega)) &= \text{rank}(\boldsymbol{\Theta}_p(b_p, \omega) \otimes \mathbf{H}_p) \\ &= \min\{Q_p \text{rank}(\mathbf{H}_p), AN_c(\mathbf{H}_p)\}, \end{aligned} \tag{18}$$

where the second equality is due to the Vandermonde structure in $\boldsymbol{\Theta}_p(b_p, \omega)$ [32].

Another interesting observation is that if b_p's are not distinct, then the upper bound (17) is loose. To see why, we suppose that $b_p = b_q$, then the rank of the submatrix $[\boldsymbol{\mathcal{H}}_p(b_p, \omega)^T \ \boldsymbol{\mathcal{H}}_q(b_q, \omega)^T]^T = \boldsymbol{\Theta}_p(b_p, \omega) \otimes [\mathbf{H}_p^T \ \mathbf{H}_q^T]^T$ is upper bounded by

$$\text{rank}(\boldsymbol{\Theta}_p(b_p, \omega) \otimes [\mathbf{H}_p^T \ \mathbf{H}_q^T]^T) \leq Q_p N.$$

This is because the rank of $[\mathbf{H}_p^T \ \mathbf{H}_q^T]^T$ is upper bounded by N. Such observation suggests that the existence of time offsets may be beneficial, especially when the sensors are down-sampling.

Using the similar reasoning as in (17), we conclude that the identifiability condition in Proposition 1 is necessitated by:

Corollary 1. *The condition* $\text{rank}(\hat{\boldsymbol{\mathcal{H}}}(\hat{\mathbf{b}}, \mathbf{b}, \omega)) = 2AN$ *is fulfilled only if*

$$\textstyle\sum_{p=1}^{P} \min\{Q_p \text{rank}(\mathbf{H}_p), AN_c(\mathbf{H}_p)\} \geq 2AN \tag{19}$$

provided that the sampling offsets b_p *are distinct.*

Eq. (19) provides a guideline for the deployment of sensors in a robust sensing system. In particular, it suggests that

$$\textstyle\sum_{p=1}^{P} \min\{Q_p M_p, AN_c(\mathbf{H}_p)\} \geq 2AN. \tag{20}$$

We remark that (19) is only a necessary condition. However, our numerical experiments suggest that when $\mathbf{H}_p$ is a Gaussian random matrix and (20) is satisfied, then the identifiability condition in Proposition 1 holds with high probability.

IV. Consensus on State and Time

This section studies a method to recover the state spectrum and time offset in a decentralized fashion. We treat each sensor, indexed by p as in the previous discussions, as a computing node that processes the knowledge of $\mathbf{H}_p$ and $\{\boldsymbol{\zeta}_p[n]\}_{n=0}^{Q_p L-1}$, where L is a designated frame size. The sensors are connected through a communication network described by $G = (\mathcal{V}, \mathcal{E})$. The goal is to estimate the unknowns (the state spectrum and time offset) while the sensors perform local computations. The sensors communicate with their neighbors only occasionally.

Our first step is to formulate the regression problem for estimating $\{b_p\}_p$ and $\mathbf{X}_A(\omega)$ in (12). As the sensors possesses only finite-length data $\{\boldsymbol{\zeta}_p[n]\}_{n=0}^{Q_p L-1}$, we resort to taking approximation and discretizing the both sides of Eq. (12) by a K-point discrete Fourier transform (DFT), i.e.,

$$\boldsymbol{\mathcal{Z}}_p[k] \approx \boldsymbol{\mathcal{H}}_p(b_p, \omega_k)\mathbf{X}_A(\omega_k) + \boldsymbol{\mathcal{V}}_p[k], \tag{21}$$

where $\omega_k \triangleq 2\pi(k - K + 1)/(K)$ is the frequency that the kth DFT point is related to. The approximation above is exact when $\boldsymbol{z}_p[n]$ is a periodic sequence with length L. For general signals, the approximation error decays as $\mathcal{O}(L^{-1})$ [33][2]. The K-point DFT spectrum of $\boldsymbol{\mathcal{Z}}_p[k]$ is obtained from L samples of $\boldsymbol{z}_p[n]$ as:

$$\boldsymbol{\mathcal{Z}}_p[k] = \textstyle\sum_{n=0}^{L-1} \boldsymbol{z}_p[n] e^{-j\omega_k n}, \tag{22}$$

where $\boldsymbol{z}_p[n] \triangleq [\boldsymbol{\zeta}_p^0[n]^T \ \ldots \ \boldsymbol{\zeta}_p^{Q_p-1}[n]^T]^T$ (cf. (8)) is a concatenation of the Q_p streams of data and K satisfies $L \leq K$. Notice that $\{\mathbf{X}_A(\omega_k)\}_{k=0}^{K-1}$ is a re-ordering of $\{\mathbf{X}(e^{j\tilde{\omega}_k})\}_k$ where $\tilde{\omega}_k \triangleq 2\pi(k - AK + 1)/(AK)$. The latter corresponds to the state sequence $\{\mathbf{x}[n]\}_{n=0}^{AL-1}$ over the finite length's frame of interest.

The discretized noise spectrum $\boldsymbol{\mathcal{V}}_p[k]$ is Gaussian and white. Furthermore, if $K = L$, then $\boldsymbol{\mathcal{V}}_p[k]$ are independent. As such, the maximum likelihood estimation problem of $\mathbf{X}_A(\omega_k)$ and b_p can be given as:

$$\begin{aligned} \min_{\substack{\hat{\mathbf{X}}_A(\omega_k), \forall k, \\ \hat{b}_p, \forall p}} \ & \sum_{p=1}^{P} \sum_{k=0}^{K-1} \left\| \boldsymbol{\mathcal{Z}}_p[k] - \boldsymbol{\mathcal{H}}_p(\hat{b}_p, \omega_k)\hat{\mathbf{X}}_A(\omega_k) \right\|_2^2 \\ \text{s.t.} \ & \hat{b}_p \in \mathcal{B}_p, \ \forall \ p. \end{aligned} \tag{23}$$

The next step is to develop a decentralized algorithm for (23).

A. Decentralized optimization with nuisance parameter

Note that (23) is a non-convex optimization problem. In particular, it can be regarded as a regression problem with local, nuisance parameter $\{\hat{b}_p\}_p$. To this end, a natural approach for tackling the problem is to apply an alternating optimization (AO) strategy, which works by alternating between the updates of $\{\hat{\mathbf{X}}_A(\omega_k)\}_k$ and $\hat{\mathbf{b}}$. To fix ideas, we let $t \in \mathbb{N}$ be the iteration index and define the objective function in (23) as $f(\{\hat{\mathbf{X}}_A(\omega_k)\}_k, \hat{\mathbf{b}})$. A practical AO strategy can be described by the following recursion:

$$\begin{aligned} \{\mathbf{X}_A^{(t+1)}(\omega_k)\}_k &\leftarrow \arg\min_{\hat{\mathbf{X}}_A(\cdot)} f(\{\hat{\mathbf{X}}_A(\omega_k)\}_k, \mathbf{b}^{(t)}), \\ \mathbf{b}^{(t+1)} &\leftarrow \mathcal{P}_{\mathcal{B}}\big(\mathbf{b}^{(t)} - \beta \nabla_{\mathbf{b}} f(\{\mathbf{X}_A^{(t)}(\omega_k)\}_k, \mathbf{b}^{(t)})\big), \end{aligned} \tag{24}$$

where $\mathcal{P}_{\mathcal{B}}(\cdot)$ is the projection onto the convex set $\mathcal{B}_1 \times \cdots \times \mathcal{B}_P$ and $\beta > 0$ is a step size. Notice that the optimization of $\mathbf{b}$ is taken care by a projected gradient descent update instead of an exact minimization.

There are several motivations for us to apply AO to (23). First, we observe that update of the time offsets $\mathbf{b}^{(t+1)}$ can be computed locally, which is due to the fact that

[2] In [33, Lemma 3.4.1], it was shown that $\|\boldsymbol{\mathcal{Z}}_p[k] - \boldsymbol{\mathcal{H}}_p(b_p, \omega_k)\mathbf{X}_A(e^{j\omega_k}) - \boldsymbol{\mathcal{V}}_p[k]\|$ is finitely bounded, while the squared magnitude of $\boldsymbol{\mathcal{Z}}_p[k]$ grows as $\mathcal{O}(L)$. The error is thus $\mathcal{O}(L^{-1})$.

$f(\{\mathbf{X}_A(\omega_k)\}_k, \mathbf{b})$ can be written as a summation of P functions, each of them depending only on b_p and is known to the pth sensor. Second, the optimization of state spectrum $\{\mathbf{X}_A(\omega_k)\}_k$ is equivalent to solving a standard linear least square problem. The latter admits a closed form solution. Lastly, the recursion (24) can be analyzed as a special case of the Block Successive Minimization Method in [34]. In particular, the recursion is shown to converge to a stationary point of (23).

In the interest of tackling (23) distributively, we see that the first update in (24) involves data from all sensors. In particular,

$$\begin{aligned}\mathbf{X}_A^{(t+1)}(\omega_k) &= \mathbf{X}_{A,k}^{\star}(\mathbf{b}^{(t)}) \\ &= \Big(\overline{\boldsymbol{\mathcal{H}}\boldsymbol{\mathcal{H}}}(\mathbf{b}^{(t)}, \omega_k)\Big)^{-1}\overline{\boldsymbol{\mathcal{H}}\boldsymbol{\mathcal{Z}}}(\mathbf{b}^{(t)}, \omega_k),\end{aligned} \tag{25}$$

with

$$\overline{\boldsymbol{\mathcal{H}}\boldsymbol{\mathcal{H}}}(\mathbf{b}^{(t)}, \omega_k) \triangleq (1/P)\sum_{p=1}^{P} \boldsymbol{\mathcal{H}}_p(b_p^{(t)}, \omega_k)^H \boldsymbol{\mathcal{H}}_p(b_p^{(t)}, \omega_k), \tag{26}$$

$$\overline{\boldsymbol{\mathcal{H}}\boldsymbol{\mathcal{Z}}}(\mathbf{b}^{(t)}, \omega_k) \triangleq (1/P)\sum_{p=1}^{P} \boldsymbol{\mathcal{H}}_p(b_p^{(t)}, \omega_k)^H \boldsymbol{\mathcal{Z}}_p[k], \tag{27}$$

where $(\cdot)^H$ denotes Hermitian transpose. As seen, (25) requires the knowledge of $\overline{\boldsymbol{\mathcal{H}}\boldsymbol{\mathcal{H}}}(\cdot)$, $\overline{\boldsymbol{\mathcal{H}}\boldsymbol{\mathcal{Z}}}(\cdot)$, which is not available to the individual sensors.

There are a number of decentralized algorithm that can be applied to solve the linear least square problem in (24). To list a few options, the diffusion-based LMS and RLS algorithm are proposed in [11], [12], respectively; the ADMM method is applied in [4], etc.

It is preferred to apply a decentralized algorithm with fast convergence. In this regard, we propose to directly compute (25) in a decentralized fashion. The idea is to leverage on the fact that $\overline{\boldsymbol{\mathcal{H}}\boldsymbol{\mathcal{H}}}(\cdot)$ and $\overline{\boldsymbol{\mathcal{H}}\boldsymbol{\mathcal{Z}}}(\cdot)$ can be replaced by the averages of the collection $\{\boldsymbol{\mathcal{H}}_p(\cdot)^H\boldsymbol{\mathcal{H}}_p(\cdot)\}_p$ and $\{\boldsymbol{\mathcal{H}}_p(\cdot)^H\boldsymbol{\mathcal{Z}}_p[k]\}_p$, respectively. As a result, the computation of (25) can be treated as a decentralized averaging problem. To compute the averages, we apply the Gossip-based average consensus (G-AC) protocol in [35], which is described in the following.

Let us take

$$\boldsymbol{y}_{p,k}^0 \triangleq \begin{bmatrix} \boldsymbol{\mathcal{H}}\boldsymbol{\mathcal{Z}}_p^0[k] \\ \boldsymbol{\mathcal{H}}\boldsymbol{\mathcal{H}}_p^0[k] \end{bmatrix} \triangleq \begin{bmatrix} \mathbf{H}_p(b_p^{(t)}, \omega_k)^H \boldsymbol{\mathcal{Z}}_p[k] \\ \mathrm{vec}(\mathbf{H}_p(b_p^{(t)}, \omega_k)^H \mathbf{H}_p(b_p^{(t)}, \omega_k)) \end{bmatrix},$$

which is a $((AN)^2 + AN)$-dimensional complex vector. It suffices to compute (25) by obtaining the average of $\{\boldsymbol{y}_{p,k}^0\}_p$, i.e.,

$$\overline{\boldsymbol{y}}_k = (1/P)\sum_{p=1}^{P} \boldsymbol{y}_{p,k}^0. \tag{28}$$

The G-AC protocol achieves $\bar{\boldsymbol{y}}_k$ by performing the following recursions:

$$\boldsymbol{y}_{p,k}^{\ell} = \textstyle\sum_{q\in\mathcal{N}_p} W_{pq}^{\ell}\boldsymbol{y}_{q,k}^{\ell-1}, \tag{29}$$

where $\mathcal{N}_p \subseteq \mathcal{V}$ denotes the set of neighbors of sensor p. To guarantee convergence, the mixing matrix $\mathbf{W}^{\ell} = [W_{pq}^{\ell}]_{p,q}$ satisfies a certain set of mild conditions, e.g., it is required to be *doubly stochastic*, i.e., $\mathbf{W}^{\ell}\mathbf{1} = \mathbf{1}$ and $\mathbf{1}^T\mathbf{W}^{\ell} = \mathbf{1}^T$. For a more detailed discussion, see [35].

As seen in (29), at each G-AC step $\ell \in \mathbb{N}$, the sensor p only obtains information from its immediate neighbors, i.e., $q \in \mathcal{N}_p$. Moreover, as $\mathbf{W}^{\ell}$ can be time-varying, only a subset of links $\mathcal{E}^{\ell} \subseteq \mathcal{E}$ are required to be active at each G-AC step. The G-AC method requires only *local computation* and it allows *random* communication between the agents. Finally, the variable $\mathbf{X}_{p,A,k}^{\ell_t}(\mathbf{b}^{(t)})$ is computed using the *approximate* averages stored at the pth agent after ℓ_t G-AC steps, i.e.,

$$\mathbf{X}_{p,A,k}^{\ell_t}(\mathbf{b}^{(t)}) = \big(\mathrm{vec}^{-1}(\boldsymbol{\mathcal{H}}\boldsymbol{\mathcal{H}}_p^{\ell_t}[k])\big)^{-1}\big(\boldsymbol{\mathcal{H}}\boldsymbol{\mathcal{Z}}_p^{\ell_t}[k]\big). \tag{30}$$

Notice that we apply the G-AC protocol to compute approximates to all the K points in DFT spectrum $\mathbf{X}_{A,k}^{\star}(\mathbf{b}^{(t)})$.

Combining AO and G-AC results in a decentralized algorithm for (23). We call this algorithm the Gossip-based alternating optimization (G-AO) algorithm, as summarized in Algorithm 1.

Algorithm 1 The G-AO algorithm for (23).

1: **Initialize:** $\{\{\mathbf{X}_{p,A}^{(0)}(\omega_k)\}_k\}_{p=1}^P$, $\{b_p^{(0)}\}_{p=1}^P$;
2: **for** $t = 0, 1, ...$ **do**
3: The network computes $\mathbf{X}_{p,A,k}^{\ell_t}(\mathbf{b}^{(t)})$ for each (p,k) using ℓ_t G-AC steps (cf. (30)).
4: **for** $p = 1, 2, ..., P$ **do**
5: Agent p updates its copies of $\mathbf{X}_{p,A}(\cdot)$ and b_p as:

$$\mathbf{X}_{p,A}^{(t+1)}(\omega_k) \leftarrow \mathbf{X}_{p,A,k}^{\ell_t}(\mathbf{b}^{(t)}),\ k = 0, ..., K-1,$$

$$b_p^{(t+1)} \leftarrow \mathcal{P}_{\mathcal{B}_p}\big(b_p^{(t)} - \beta\nabla_{b_p} f(\{\mathbf{X}_{p,A}^{(t+1)}(\omega_k)\}_k, \mathbf{b}^{(t)})\big)$$

As mentioned, the update of b_p can be performed distributively since f is separable.
6: **end for**
7: **end for**
8: **Return:** $\{\{\mathbf{X}_{p,A}^{(t+1)}(\omega_k)\}_k\}_{p=1}^P$, $\{b_p^{(t+1)}\}_{p=1}^P$.

B. Convergence analysis

To study the convergence of the G-AO algorithm, we first need to study the convergence rate of G-AC. As shown in [35], the recursion (29) converges to the true average vector $\bar{\boldsymbol{y}}_k$ under several assumptions on $\{\mathbf{W}^{\ell}\}_{\ell}$. In fact, the rate of convergence is exponential, i.e., $\|\boldsymbol{y}_{p,k}^{\ell} - \overline{\boldsymbol{y}}_k\|_2 = \mathcal{O}(\lambda_{\bar{W}}^{\ell})$ where $0 < \lambda_{\bar{W}} < 1$ is the second largest eigenvalue of the matrix $\bar{\mathbf{W}} = \mathbb{E}_{\ell}\{\mathbf{W}^{\ell}\}$. Consequently, the accuracy on the approximation $\mathbf{X}_{p,A,k}^{\ell_t}(\mathbf{b}^{(t)})$ improves exponentially with ℓ_t:

Proposition 2. *Suppose that*

$$C_0 \cdot C_1 \cdot \lambda_{\bar{W}}^{\ell_t} < 1, \tag{31}$$

where $C_0 = \max_{\mathbf{b},k} \|\sum_p \mathbf{H}_p(b_p, \omega_k)^H\mathbf{H}_p(b_p, \omega_k)\|$ *and* $C_1 = \max_{\mathbf{b},k} \|(\sum_p \mathbf{H}_p(b_p, \omega_k)^H\mathbf{H}_p(b_p, \omega_k))^{-1}\|$ *are finite constants, then the spectrum* $\mathbf{X}_{p,A,k}^{\ell_t}(\mathbf{b}^{(t)})$ *computed in* (30) *using* ℓ_t *G-AC steps satisfies:*

$$\big\|\mathbf{X}_{A,k}^{\star}(\mathbf{b}^{(t)}) - \mathbf{X}_{p,A,k}^{\ell_t}(\mathbf{b}^{(t)})\big\|_2 = \mathcal{O}(\lambda_{\bar{W}}^{\ell_t}),\ \forall\ p, k, \tag{32}$$

where ℓ_t *is the number of G-AC steps at the* t*th iteration.*

The proof is relegated to Appendix B. An important implication is that the approximation improves exponentially with ℓ_t, i.e., the number of G-AC steps per iteration.

We are ready to state the following theorem regarding the convergence of the G-AO algorithm.

Theorem 1. *Let $(\{\mathbf{X}^\star_{A,k}\}_k, \mathbf{b}^\star)$ be a local minimum to (23). Suppose that f is (m_o, M_o)-strongly convex in the neighborhood $\mathcal{N}_{R^\star}(\{\mathbf{X}^\star_{A,k}\}_k, \mathbf{b}^\star)$ and it is Lipschitz continuous with constant L_o. Suppose that (31) holds, $\beta \leq 1/M_o$, $B = \max_t \|(\{\hat{\mathbf{X}}^{(t)}_{A,k}\}_k, \mathbf{b}^{(t)}) - (\{\mathbf{X}^\star_{A,k}\}_k, \mathbf{b}^\star)\|_2 < \infty$ and $B \leq R^\star$, then we have:*

$$\lim_{t\to\infty} \left\|(\{\hat{\mathbf{X}}^{(t)}_{A,k}\}_k, \mathbf{b}^{(t)}) - (\{\mathbf{X}^\star_{A,k}\}_k, \mathbf{b}^\star)\right\|_2^2 = \sqrt{\mathcal{O}(\lambda_W^{\ell_{min}})}, \tag{33}$$

where $\ell_{min} = \min_t \ell_t$ is the minimum number of G-AC steps taken and $\hat{\mathbf{X}}^{(t)}_{A,k} \triangleq (1/P)\sum_{p=1}^P \mathbf{X}^{(t)}_{p,A}(e^{j\omega_k})$.

The proof is provided in Appendix C, which is based on studying the error dynamics of the G-AO algorithm as a second order dynamical system. In fact, Theorem 1 implies that if the G-AO algorithm stays close enough to a local minimum, then the algorithm converges to an approximate of that local minimum, where the approximation accuracy improves exponentially with ℓ_{min}.

The strong convexity assumption on f around a local minimum may appear restrictive at first. However, our numerical results indicate that Theorem 1 can accurately predict the performance of G-AO algorithm applied on (23).

V. Extensions & Applications

In this section, we extend the previous formulations/algorithms to the case with general measurement (e.g., non-linear). This model pertains to the case when the measured signal at the sensor is non-linearly related to the underlying system state. As an application example, we demonstrate that the formulation can be employed to robustify power system state estimation (PSSE) systems.

Considering Assumption 1 to 3 in Section II and the sampling architecture in Fig. 1, we can express the measured sample at sensor p as:

$$\begin{aligned}\boldsymbol{\zeta}_p[n] &= \overline{(\boldsymbol{h}_p \circ \mathbf{x}_c)}(T_s(nA_p - b_p)) + \boldsymbol{v}_p[n],\\ \overline{(\boldsymbol{h}_p \circ \mathbf{x}_c)}(t) &= \mathrm{LPF}\{(\boldsymbol{h}_p(\mathbf{x}_c(t))\},\end{aligned} \tag{34}$$

where $\boldsymbol{h}_p : \mathbb{C}^N \to \mathbb{C}^{M_p}$ is a general measurement function and $\overline{(\boldsymbol{h}_p \circ \mathbf{x}_c)}(t)$ is the low-pass filtered version of $\boldsymbol{h}_p(\mathbf{x}_c(t))$ with a cutoff frequency at $W/2$ Hz. The reason for introducing the low-pass-filtered is explained in Remark 1.

Under the same spirit as in our previous developments, the next step is to consider the frequency domain equivalent to (34) via Observation 1. To simplify notations, we consider $A_p = A$ for all p in the following. The K-point DFT of $\boldsymbol{\zeta}_p[n]$ is given by (cf. (22)):

$$\begin{aligned}&\mathbf{Z}_p[k] \approx \frac{1}{A}\times\\ &\sum_{a=0}^{A-1} e^{-jb_p\Omega_A^a(\omega_k)} \sum_{n=0}^{AL-1} \boldsymbol{h}_p(\mathbf{x}[n]) e^{-j\Omega_A^a(\omega_k)n} + \mathbf{V}_p[k],\end{aligned} \tag{35}$$

where we recall that $\omega_k = 2\pi(k-K+1)/K$. Analogous to the previous derivations, it follows that we can formulate the nonlinear regression problem as:

$$\min_{\{\hat{\mathbf{x}}[n]\}_{n=0}^{AL-1}, \hat{b}_p, \forall p} \quad \sum_{p=1}^P \sum_{k=0}^{K-1} \left\|\boldsymbol{g}_{p,k}\big(\{\hat{\mathbf{x}}[n]\}_{n=0}^{AL-1}, \hat{b}_p\big)\right\|_2^2, \tag{36}$$

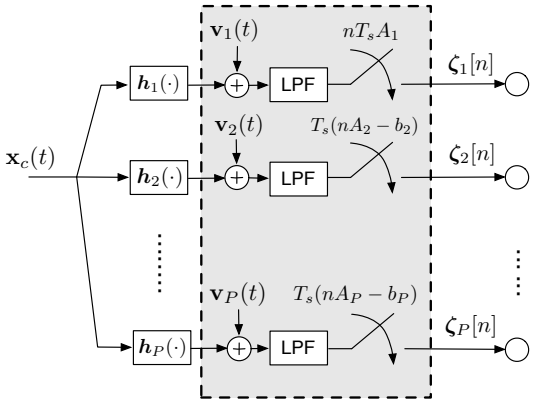

Fig. 1. The architecture of the sample & estimation system (cf. (4)). The low pass filters (LPF) have a cutoff frequency at $W/2$ Hz.

where

$$\begin{aligned}&\boldsymbol{g}_{p,k}\big(\{\hat{\mathbf{x}}[n]\}_{n=0}^{AL-1}, \hat{b}_p\big) \triangleq \mathbf{Z}_p[k] -\\ &\frac{1}{A}\sum_{a=0}^{A-1} e^{-j\hat{b}_p\Omega_A^a(\omega_k)} \sum_{n=0}^{AL-1} \boldsymbol{h}_p(\hat{\mathbf{x}}[n]) e^{-jn\Omega_A^a(\omega_k)}.\end{aligned} \tag{37}$$

Notice that the above formulation allows us to consider hybrid measurement, i.e., $\boldsymbol{h}_p$ can be linear for some p.

Remark 1. We observe that in general $\boldsymbol{h}_p(\mathbf{x}_c(t))$ has a bandwidth higher than $W/2$ Hz due to the nonlinearity of $\boldsymbol{h}_p$. In particular, the measurement function cannot be analyzed separately in $\boldsymbol{h}_p(\mathbf{x}_c(t))$ as in the linear case. The low pass filter is introduced to remedy this such that Observation 1 can be applied. Moreover, it can be verified that

$$\overline{(\boldsymbol{h}_p \circ \mathbf{x}_c)}(nT_s) = \boldsymbol{h}_p(\mathbf{x}[n]),\ \forall\ n, \tag{38}$$

where the equality can be established by studying the spectrum of the both sides.

A. Extending the G-AO algorithm

Similar to (23), Problem (36) is also a non-convex problem. To develop a decentralized algorithm for (36), we apply a modified version of the G-AO algorithm. The modified algorithm is based on the Gossip-based Gauss Newton (GGN) method in [28] and the AO strategy.

We observe that the G-AO algorithm cannot be applied directly to (36). In fact, even when $\mathbf{b}$ is fixed in (36), the optimization problem remains non-convex and does not admit a closed form solution. We develop our algorithm by borrowing insights from the GGN method.

Let $t \in \mathbb{N}$ be the iteration index and assume that $\mathbf{b}^{(t)}$ is fixed. The damped GN direction at $\{\mathbf{x}^{(t)}[m]\}_m$ for $\mathbf{x}[n]$ is given as:

$$\begin{aligned}&\mathbf{d}_n(\{\mathbf{x}_p^{(t)}[m]\}_m, \mathbf{b}^{(t)}) =\\ &\big(\lambda_{GN}\mathbf{I} + \overline{\boldsymbol{\mathcal{G}}\boldsymbol{\mathcal{G}}}(\{\mathbf{x}_p^{(t)}[m]\}_m, \mathbf{b}^{(t)})\big)^{-1}\overline{\boldsymbol{\mathcal{G}}\boldsymbol{\mathcal{Z}}}(\{\mathbf{x}_p^{(t)}[m]\}_m, \mathbf{b}^{(t)}),\end{aligned} \tag{39}$$

where $\lambda_{GN} \geq 0$ is the damped GN parameter, $\overline{\boldsymbol{\mathcal{G}}\boldsymbol{\mathcal{G}}}$ and $\overline{\boldsymbol{\mathcal{G}}\boldsymbol{\mathcal{Z}}}$ are both defined in terms of the Jacobian matrix of $\boldsymbol{g}_{p,k}(\cdot)$:

$$\overline{\boldsymbol{\mathcal{G}}\boldsymbol{\mathcal{G}}}(\{\mathbf{x}_p[m]\}_m, \mathbf{b}) = \sum_{p=1}^P \sum_{k=0}^{K-1} \boldsymbol{G}_{p,k}^n(\cdot,\cdot)^H \boldsymbol{G}_{p,k}^n(\{\mathbf{x}[m]\}_m, b_p)$$

Algorithm 2 The GGN-AO algorithm for (36).

Initialize: $\{\{\mathbf{x}_p^{(0)}[n]\}_n\}_{p=1}^P$, $\{b_p^{(0)}\}_{p=1}^P$;
for $t = 0, 1, ...$ **do**
  Update: $\mathbf{x}_p^{(t+1)}[n] \leftarrow \mathbf{x}_p^{(t)}[n]$ for all $n = 0, ..., AL-1$.
  for $n = 0, 1, ..., AL-1$ **do**
    The network computes the (approximate) GN direction $\mathbf{d}_{p,n}^{\ell_t}(\{\mathbf{x}_p^{(t+1)}[m]\}_m, \mathbf{b}^{(t)})$ for each p using ℓ_t G-AC steps (cf. (39)).
    for $p = 1, 2, ..., P$ **do**
      Agent p updates its copies of $\mathbf{x}_p[\cdot]$ and b_p as:

$$\mathbf{x}_p^{(t+1)}[n] \leftarrow \mathbf{x}_p^{(t)}[n] - \mathbf{d}_{p,n}^{\ell_t}(\{\mathbf{x}_p^{(t)}[m]\}_m, \mathbf{b}^{(t)}).$$

    end for
  end for
  for $p = 1, 2, ..., P$ **do**
    Agent p updates b_p by:

$$b_p^{(t+1)} \leftarrow \mathcal{P}_{\mathcal{B}_p}\big(b_p^{(t)} - \beta \nabla_{b_p} f_{NL}(\{\mathbf{x}_p^{(t+1)}[n]\}_n, \mathbf{b}^{(t)})\big)$$

  end for
end for
Return: $\{\{\mathbf{x}_p^{(t+1)}[n]\}_n\}_{p=1}^P$, $\{b_p^{(t+1)}\}_{p=1}^P$.

$$\overline{\boldsymbol{\mathcal{G}}\boldsymbol{\mathcal{Z}}}(\{\mathbf{x}_p[m]\}_m, \mathbf{b}) = \sum_{p=1}^{P}\sum_{k=0}^{K-1} \boldsymbol{G}_{p,k}^n(\cdot,\cdot)^H \boldsymbol{g}_{p,k}(\{\mathbf{x}[m]\}_m, b_p)$$

such that $\boldsymbol{G}_{p,k}^n(\{\mathbf{x}[m]\}_m, b_p)$ is the $N \times N$ Jacobian matrix of $\boldsymbol{g}_{p,k}(\cdot,\cdot)$ taken at $(\{\mathbf{x}[m]\}_m, b_p)$ with respect to the vector $\mathbf{x}[n]$. Notice that the state vector $\mathbf{x}[m]$ is updated in a Gauss-Seidal like manner [36].

Importantly, we observe that the GN direction can be calculated distributively using the G-AC protocol. This suggests us to combine the AO algorithm and the Gossip-GN algorithm to develop a decentralized algorithm for (36). In particular, the pseudo code for the GGN-AO algorithm for (36) is now summarized in Algorithm 2, where we have denoted the objective function in (36) as $f_{NL}(\{\mathbf{x}_p[n]\}_n, \mathbf{b})$.

In contrast to the G-AO algorithm studied in Section IV, the GGN-AO algorithm applied to (36) entails a higher complexity in general. For instance, the update of state variable $\mathbf{x}[n]$ in the GGN-AO algorithm is based on the GN method, which is an iterative method by nature. In contrast, the G-AO's counterpart of the update relies on a closed form solution. Intuitively, the G-AO algorithm will exhibit a faster convergence rate.

B. Application: Robustifying the PSSE systems

This subsection applies the models developed in this paper to robustify the power system state estimation (PSSE) systems. In particular, we consider the problem of PSSE in wide area measurement systems, where data from Phasor Measurement Units (PMUs) and the legacy Supervisory Control And Data Acquisition (SCADA) systems are combined to provide the state estimates.

The system's state of interest is the complex envelope $\mathbf{x}_c^e(t)$ of the voltage on each bus, which is related to the actual voltage on bus i by $[\mathbf{x}_c(t)]_i = \Re\{[\mathbf{x}_c^e(t)]_i e^{j\omega_0 t}\}$, where ω_0 is the operating frequency (typically 60/50 Hz) of the power grid. Typically $\mathbf{x}_c^e(t)$ has a smaller bandwidth than 60 Hz.

There are two types of sensors/systems that are used in the power grid:

PMU — The PMU installed on bus i takes samples of the voltage phasor $[\mathbf{x}_c^e(t)]_i$ and the current flow phasors $I_{ij}^e(t)$ on branches that are connected to bus i. Typically, the PMU relies on a local oscillator, synchronized using GPS clocks, to take samples of $[\mathbf{x}_c^e(t)]_i$ and $I_{ij}^e(t)$ [37]. The GPS clocks may be tampered in the case of an attack. As such, we model the voltage measurements obtained at the PMU on bus i as:

$$V_i[n] = [\mathbf{x}_c^e[n]]_i = [\mathbf{x}_c^e(nT_s - b_p)]_i e^{-j\omega_0 b_p} + v_i[n], \quad (40)$$

where $v_i[n]$ is the measurement noise, b_p represents the sampling offset, and the current measurements:

$$\begin{aligned} &I_{ij}^e[n] = e^{-j\omega_0 b_p} Y_{ij} \times \\ &\big([\mathbf{x}_c^e(nT_s - b_p)]_i - [\mathbf{x}_c^e(nT_s - b_p)]_j\big) + v_{ij}[n], \end{aligned} \quad (41)$$

where Y_{ij} is the admittance of branch (i,j). Notice that we have assumed Nyquist sampling, i.e., $A_p = 1$, as the PMU has a sampling rate of 10-30 Hz [37].

Let $\mathcal{I}_p \subseteq \{1, ..., N\}$ be the buses included in the pth sensing site. By stacking $\{V_i[n]\}_{i\in\mathcal{I}_p}$ and $\{I_{ij}[n]\}_{j\in\mathcal{N}_i, i\in\mathcal{I}_p}$ vertically as $\boldsymbol{\zeta}_p[n]$, where $\mathcal{N}_i$ denotes the buses connected to bus i, the samples obtained at the pth PMUs' site is modelled as:

$$\boldsymbol{\zeta}_p[n] = \mathbf{H}_p \mathbf{x}_c^e(nT_s - b_p) e^{-j\omega_0 b_p} + \boldsymbol{v}_p[n]. \quad (42)$$

SCADAs — The SCADA installed on bus i samples on the injected complex power $S_i(t)$ as well as the complex power $S_{ij}(t)$ that flows to/from bus i. For example, the power flow phasor through branch (i,j) is given by

$$S_{ij}(t) = [\mathbf{x}_c^e(t)]_i \big([\mathbf{x}_c^e(t)]_i - [\mathbf{x}_c^e(t)]_j\big)^* Y_{ij}^*. \quad (43)$$

Similarly, the injected power is

$$S_i(t) = |[\mathbf{x}_c^e(t)]_i|^2 Y_i^S + \textstyle\sum_{j\in\mathcal{N}_i} S_{ij}(t) \quad (44)$$

where Y_i^S is the shunt admittance from bus i to the ground. In fact, these quantities are obtained by measuring the active power and reactive power. Besides the measurement models, another factor that differentiates SCADA systems from PMUs is the sample rate used. In fact, the sampling rate is only 0.2–0.5 Hz [38] for SCADA. In light of this, we stack the relevant $\{S_i(t)\}_{i\in\mathcal{I}_p}$ and $\{S_{ij}(t)\}_{j\in\mathcal{N}_i, i\in\mathcal{I}_p}$ to form the $\boldsymbol{h}_p(\mathbf{x}_c(t))$ in (34). Under the assumption of an LPF and using an appropriate choice of A_p, we see that the model described in (35) applies to the sampled SCADA data.

We observe that both of the models for PMUs and SCADA fit into the descriptions in (34). Subsequently, the *robust* PSSE problem can be formulated in a similar manner as (36).

Remark 2. Assume the case with only asynchronous PMUs. If $b_p \ll T_s$, we can assume that the power system state is stable relative to the time offsets in PMUs, i.e., we have $\mathbf{x}_c^e(nT_s - b_p) \approx \mathbf{x}_c^e(nT_s)$, the regression problem for $\mathbf{x}_c^e(nT_s)$ and b_p can be formulated as a special case of (23):

$$\min_{\hat{\mathbf{x}}_c^e(nT_s), \{\hat{b}_p\}} \sum_{p=1}^{K} \big\|\boldsymbol{\zeta}_p[n] - \mathbf{H}_p \hat{\mathbf{x}}_c^e(nT_s) e^{-j\omega_0 \hat{b}_p}\big\|_2^2. \quad (45)$$

The G-AO algorithm can then be applied to tackle (45).

VI. Numerical results

To highlight different aspects affecting the performance of the techniques for joint state and sampling offset estimation, this section will be divided into two parts — i) the first part includes simulations pertaining the linear model in Section II with sensing matrices that are complex Gaussian zero mean i.i.d. coefficients (e.g. Rayleigh fading); ii) the second part is focused on the application to PSSE problem with a non-linear measurement model.

We first explain how the non-ideally sampled measurements are generated. Notice that in (4), the discrete-time measurement $\boldsymbol{\zeta}_p[n]$ is equivalent to the infinite sum $\sum_{m=-\infty}^{\infty} \boldsymbol{\zeta}_p((n-m)T_s)\sin(\pi(A_p m - b_p))/(\pi(A_p m - b_p))$. It is impossible to evaluate the infinite sum, we thus truncate the latter by a finite sum from $m = 0$ to $m = AL-1$. In this way, note that $\{\boldsymbol{\zeta}_p[n]\}_{n=0}^{Q_p L-1}$ retains the information from $\{\mathbf{x}[n]\}_{n=0}^{AL-1}$.

Moreover, our numerical experience shows that it is necessary to apply a pre-processing window to obtain the measured spectrum in (22), so as to reduce the modelling error introduced by discrete approximation. We instead take $\boldsymbol{Z}_p[k] = \sum_{n=0}^{L-1} w[n]\boldsymbol{z}_p[n]e^{-j\omega_k n}$, where $w[n]$ is the Blackman window [39]. Due to the windowing operation, the estimated state at the boundaries can be unreliable. As such, unless otherwise specified, we evaluate only the mean squared error (MSE) for the state $\mathbf{x}[n]$ in the middle of the frame. Specifically, the MSE is computed as the per sample error $\mathbb{E}[(1/(2G+1))\sum_{n=AL/2-G}^{AL/2+G}\|\mathbf{x}[n]-\hat{\mathbf{x}}[n]\|_2^2/N]$, where $\hat{\mathbf{x}}[n]$ is the estimated state and $G = \lfloor 0.15AL \rfloor$. The squared error for time offsets is calculated as $\mathbb{E}[\sum_{p=1}^{P}(b_p - \hat{b}_p)^2]$.

For the other simulation parameters, we fix $K = 192$ as the DFT size and $\sigma_w^2 = 10^{-2}$ as the noise variance. We perform 100 Monte-Carlo simulation trials to get the averages. The G-AO/GGN-AO algorithm is terminated when the relative decrease in objective value is less than 0.1%. The G-AO algorithm is initialized with $b_p = 0$ for all p. The communication network is generated as an Erdos-Renyi graph with parameter $p = 0.5$. We assume that the mixing matrix $\mathbf{W}$ is static with Metropolis-Hastings weight [40]. The error is evaluated as the maximum MSEs evaluated for each sensor.

A. Example: Rayleigh fading

In the following examples, we focus on the performance of proposeds method under the linear model specified in Section II. In particular, the states $\mathbf{x}[n]$ and measurement matrices $\mathbf{H}_p$ are generated as random vectors/matrices with unit variance i.i.d. complex Gaussian random entries.

Our first example considers a system with sub-Nyquist sampling, i.e., we set $A_p = 2$ for all p. The system dimensions are set as $M = 4, N = 8, P = 12$. The time offsets are uniformly drawn from $\mathcal{B} = [-0.5, 0.5]$. Notice that under sub-Nyquist sampling, without exploiting the time offsets between the sensors, it is impossible to estimate the state vector $\mathbf{x}[n]$ for all n. Therefore, as a benchmark, we provide the MSE evaluated by comparing $\{\mathbf{x}[n]\}_n$ with an interpolated state sequence estimated from the sub-Nyquist measurements.

The simulation result from this example is depicted in Fig. 2, where we compare the MSE in state and in $\{b_p\}_p$ against the frame size L. From the figure, we see that the error metrics of the proposed algorithm decrease as L increases. It is due to the improved approximation to the true DTFT spectrum. In fact, the MSE in state decays as $\mathcal{O}(L^{-1})$, coinciding with the discussions that follows (21). On the other hand, the G-AO algorithm achieves a similar performance with its centralized counterpart. Especially, as ℓ_{min} increases, the performance of the former approaches that of the latter. This observation is in line with the analysis results on G-AO from Theorem 1.

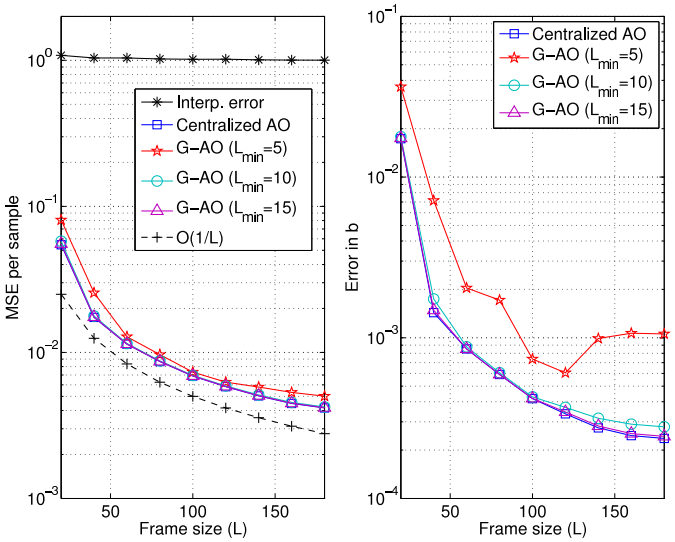

Fig. 2. Comparing the MSE performance against the frame size L. (Left) On estimating $\{\boldsymbol{x}[n]\}_n$. (Right) On estimating $\{b_p\}_p$.

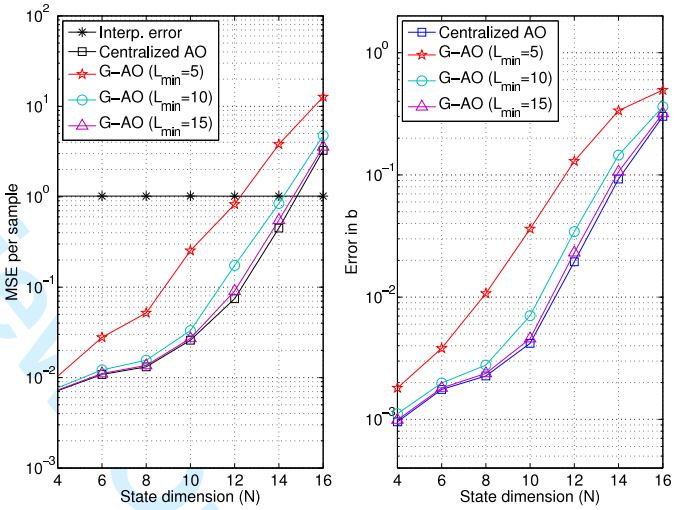

Fig. 3. Comparing the MSE performance against the state dimension N. (Left) On estimating $\{\boldsymbol{x}[n]\}_n$. (Right) On estimating $\{b_p\}_p$.

In the second example, we examine the identifiability condition in Section III. Specifically, the system parameters are $M = 3, P = 12, L = 120$ and the MSEs are compared with different state dimension N. Fig. 3 shows the result from this example. Recall that from Corollary 1, the identifiability condition is likely to be satisfied if $N \leq 9$. This is evident from the figure that the MSE increases significantly when $N \geq 10$. The discrepancy is due to the fact that Corollary 1 is derived based on a noiseless model and the G-AO algorithm may have

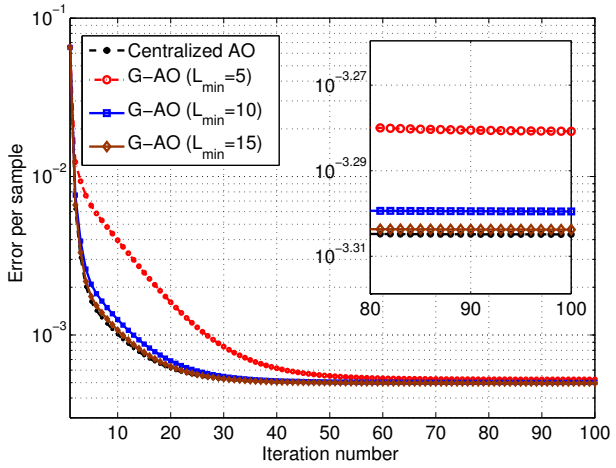

Fig. 4. Comparing the state estimation error against iteration number of the proposed algorithms. $M = 4, P = 12, N = 8, L = 120$.

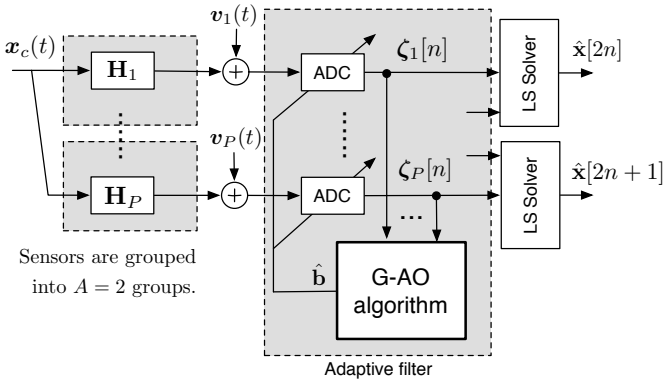

Fig. 5. Incorporating the proposed regression method into standard least squares state estimators. (Consider the special case when $A = A_p = 2$.)

been initialized close enough to the true optimum.

The next example, shown in Fig. 4, examines the convergence speed of the G-AO algorithm, for which we track the state estimation error as G-AO algorithm proceeds. In this example, we set $M = 4, P = 12, N = 8, L = 120$ and consider solving a randomly generated instance of (23). We observe that the error is gradually decreasing as the algorithm progresses and converges in about 30-40 iterations. Importantly, we see that the G-AO algorithm follows closely with its centralized counterpart, suggesting that they both achieve a similar performance.

Estimating the state via adaptive filtering — As an extension, we study a practical scheme for incorporating the proposed regression method into standard state estimators.

Assuming that the time offsets $\{\hat{b}_p\}_p$ are estimated with the G-AO algorithm in an earlier stage, our idea is to consider a *sub-optimal* adaptive filter, which reverts only the effects of non-ideal sampling by adding delays to the ADCs. For example, if $A = A_p = 2$, the sensors can be divided into two groups, one with a delay tap of $\tau_p = 1 - \hat{b}_p$ and the other with $\tau_p = 2 - \hat{b}_p$, etc. Notice that we have created two groups of sensors, one sampling for $\mathbf{x}[2n]$, and the other one sampling for $\mathbf{x}[2n+1]$. The main advantage for applying the adaptive filter is to reduce complexity. In fact, the regression problem on the post-filtered data becomes a standard least-square optimization that admits a closed form solution; see Fig. 5.

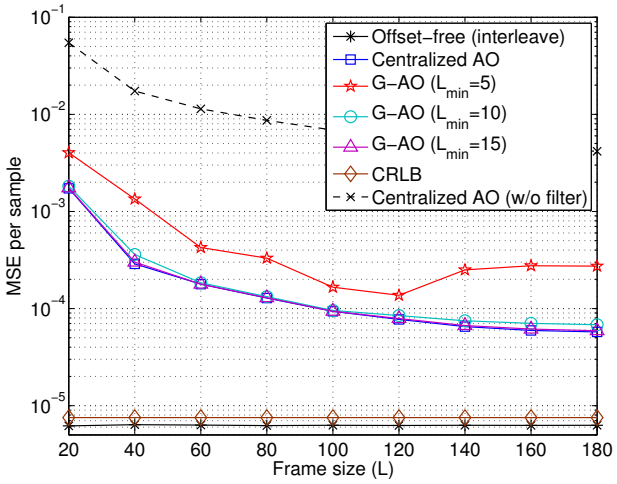

Fig. 6. Comparing the MSE in state estimate against L upon applying the adaptive filter. System settings are $M = 4, P = 12, N = 8$.

Fig. 6 shows the MSE in state estimate against L using the inverse filter designed from the time offset estimate $\hat{b}_p$. The simulation settings are the same as in Fig. 2. The sensors grouping is done by grouping the first $P/2$ sensors into one group; and the rest of the sensors into another. As a benchmark, the 'offset-free (interleave)' refers to the system with *perfectly aligned*, time-interleaved sensors; and the 'CRLB' is the Cramer-Rao's lower bound in the error performance by assuming that the signal is periodic, i.e., the DFT approximation is exact. The 'offset-free (interleave)' benchmark requires perfect knowledge of $\{b_p\}_p$. From the simulation results, the MSE performance using the proposed algorithm is comparable to the benchmark when L is large.

B. Example: Power System State Estimation

In the following examples, we evaluate the performance of the proposed algorithms applied to PSSE. We consider the IEEE-30 bus test case in Fig. 7. To simulate the power system dynamics, the state vector is generated as $\mathbf{x}^e[n] = \mathbf{x}_c^e(nT_s) = \mathbf{x}_t + \mathbf{d}[n]$, where $\mathbf{x}_t$ is the voltage vector in the IEEE-30 test case in MATPOWER [41] and $\mathbf{d}[n] \sim \mathcal{CN}(0, \sigma_d^2\mathbf{I})$ models the fluctuation of voltages. Bus 1 is assumed to be the slack bus with zero phase angle.

We first consider the case where only PMUs are employed. We assume that $b_p \ll T_s$ such that $\mathbf{x}_c^e(nT_s - b_p) \approx \mathbf{x}_c^e(nT_s)$. As noted by Remark 2, the corresponding joint regression problem (45) can be tackled using the G-AO algorithm. To ensure identifiability for both state and time offsets, the PMUs collect measurements from the buses $\{1, 2, 4, 5, 6, 9, 13, 15, 17, 19, 20, 21, 23, 25, 27, 29\}$.

The simulation result, as shown in Fig. 8, is performed with 1000 Monte-Carlo trials. In particular, we compare the MSE in state estimation against different range of sampling offset σ_b, where the sampling offsets b_p is generated as uniformly

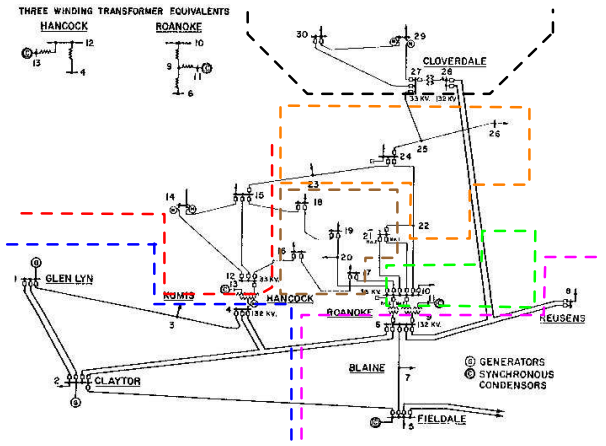

Fig. 7. The IEEE-30 bus system. We divide the PMUs into $P = 7$ sensing sites, as indicated by the colored dotted lines. The IEEE-30 bus system is partitioned as $\mathcal{A}_1 = \{1, 2, 3, 4\}, \mathcal{A}_2 = \{5, 6, 7, 8\}, \mathcal{A}_3 = \{9, 10, 11\}$, $\mathcal{A}_4 = \{12, 13, 14, 15\}, \mathcal{A}_5 = \{16, 17, 18, 19, 20, 21\}, \mathcal{A}_6 = \{22, 23, 24, 25, 26\}, \mathcal{A}_7 = \{27, 28, 29, 30\}$.

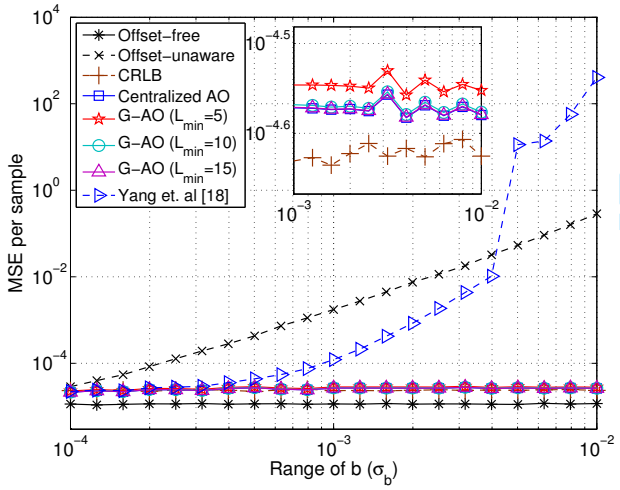

Fig. 8. Comparing the MSE per sample of an IEEE-30 bus system with only PMUs against the variance in sampling offset in b_p and we fix $\sigma_w = 10^{-2}$. We also have $\sigma_d = 10^{-2}$ and $\sum_p M_p = 62, N = 30$.

distributed over $[-\sigma_b, \sigma_b]$. As observed, the proposed method achieves a performance on par with the method (Yang et. al.) from [18] when σ_b is small. The latter is outperformed by ours when $\sigma_b \geq 10^{-3}$. Moreover, our proposed algorithm achieves a MSE that is close to the Cramer Rao's lower bound (CRLB). This demonstrates the benefit of modelling non-ideal sampling directly.

The next simulation example considers the case where we combine both PMU and SCADA data. In particular, we consider the setting when only 8 PMUs are installed on bus $\{1, 4, 5, 9, 11, 15, 20, 24, 27\}$ in an IEEE-30 system. The number of installed PMUs is insufficient to provide identifiability if we only rely on PMU data. To remedy, we deploy 30 SCADA systems (each installed on a bus) on the power grid to monitor every bus and branch in the grid. We consider $L = 40$ and apply the centralized version of the GGN-AO algorithm, which is initialized by $\mathbf{x}_p^{(0)}[n] = \mathbf{x}_t$ for all n and $b_p = 0$ for all p.

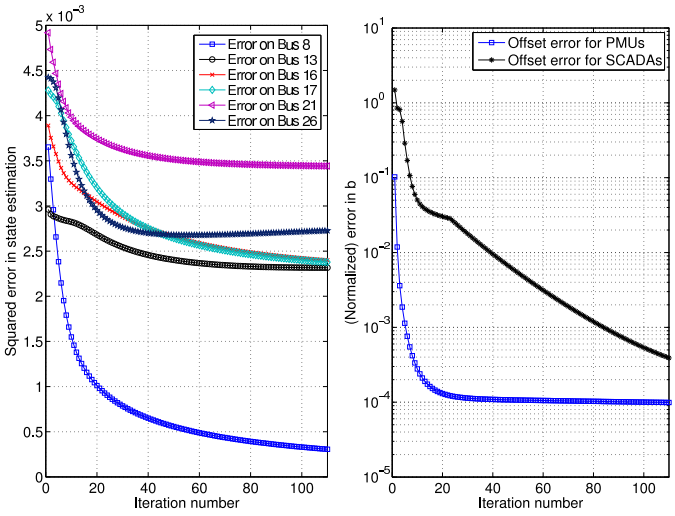

Fig. 9. The state estimation error performance against the iteration number over an IEEE-30 bus system. We have set $\lambda_{GN} = 10^2$ for the GGN-AO algorithm. (Left) On estimating the power system state at different bus. (Right) On estimating $\{b_p\}_p$ for PMUs and SCADAs. The latter error is normalized according to the range of time offsets b_p.

Our aim is to demonstrate that the full power grid's state in *transient* can be revealed using the sub-Nyquist SCADA data and insufficient amount of PMU data. We assume that the complex envelope $\mathbf{x}_c^e(t)$ has a bandwidth of 0.5 Hz such that the Nyquist sampling period is $T_s = 1$ second. For simplicity, the SCADA systems has a sampling period of $A_p T_s = 2$ second, while the PMUs are able to capture samples with period shorter than $A_p T_s = 1$ second. A snapshot of the estimation result is shown in Fig. 9, in which we compare the state/time offset estimation error against iteration number. We have only shown the errors in estimating the voltage on buses $\{8, 13, 16, 17, 21, 26\}$. Notice that these buses are unobservable using the PMU data alone. That said, from the figure, we observe that the state estimation error from these buses is fairly low. More importantly, the error in estimating $\{b_p\}_p$ is at the order of 10^{-4} to 10^{-3}. This allows us to design the adaptive filter and treat the SCADA systems as the *time-interleaved* sensors in a similar fashion as the last subsection.

C. Discussions

There are only ∼1,000 installed PMUs in total in the North America Power Grid as of 2014, while the number of buses exceeds 10,000 [42]. Although more PMUs are being installed, the number of PMUs employed is still insufficient to provide full identifiability of the entire power grid in the near future.

As demonstrated in the last simulation example, we see that deploying PMUs with SCADAs on the power grid is beneficial in the sense that the power system state in *transient* can be captured. In fact, using the result from Corollary 1, the fundamental limit on the maximum bandwidth for the power system state can be estimated. We assume that the PMUs are sampling at 10 Hz while the SCADAs are sampling at 0.5 Hz, each sensor takes $M_p = 3$ measurements which depends on the state of $N_c(\mathbf{H}_p) = 5$ buses. Now, set $Q_p = 1$ for the

SCADAs, $Q_p = 20$ for the PMUs and using Eq. (20) yield:

$$\#\text{ PMUs} \cdot \min\{60, 5A\} + \#\text{ SCADAs} \cdot 3 \geq 20000A \quad (46)$$

In particular, when the number of PMUs and SCADAs is 1,000 and 10,000, respectively, the maximum possible A is 2. That is, the maximum possible bandwidth for system transient signal is 0.5 Hz, which is a two-fold improvement over the case with SCADAs alone. Naturally, the individual sensors renders transients variations at 10 Hz visible, but due to their relative scarcity the gain in resolving transients of nearby locations is degraded. To what extent aliasing corrupts the reconstruction of faster transients is a topic of future research.

VII. Conclusions

In this paper, we have addressed the issue of non-ideal sampling in decentralized regression. Our contributions are multi-fold: i) we propose a new joint regression problem that estimates the system state and unknown time offsets under sub-Nyquist sampling; ii) we derive a set of identifiability conditions that guarantee perfect state and time offset estimation; iii) we develop a decentralized algorithm that solves the joint regression problem and prove its convergence. The efficacy of the proposed algorithm is demonstrated through numerical examples on both synthetic data and realistic power systems.

Appendix A
Proof of Observation 1

The observation is the consequence of a several properties from the Fourier transform. We first derive the DTFT spectrum of $\mathbf{x}_c((n-b_p)T_s)$. Under the Assumption 1 and by the Shannon's interpolation formula, we observe that:

$$\mathbf{x}_c((n-b_p)T_s) = \sum_{m=-\infty}^{\infty} \mathbf{x}[n-m]\frac{\sin(\pi(m-b_p))}{\pi(m-b_p)}, \quad (47)$$

which implies

$$\text{DTFT}\{\mathbf{x}_c((n-b_p)T_s)\} = \mathbf{X}(e^{j\omega})e^{-jb_p\left(\omega \bmod (-\pi,\pi]\right)}. \quad (48)$$

Now, by decomposing the DTFT spectrum of $\mathbf{x}((nA-b_p)T_s)$ into its polyphase components and using results from Chapter 4 in [43], we can obtain (5).

Appendix B
Proof of Proposition 2

In the following proof, we shall abbreviate $\boldsymbol{\mathcal{H}}\boldsymbol{\mathcal{H}}_p^{\ell_t}$ as $\boldsymbol{\mathcal{H}}_p^{\ell_t}$, $\boldsymbol{\mathcal{H}}\boldsymbol{\mathcal{Z}}_p^{\ell_t}$ as $\boldsymbol{\mathcal{Z}}_p^{\ell_t}$, $\overline{\boldsymbol{\mathcal{H}}\boldsymbol{\mathcal{H}}}$ as $\overline{\boldsymbol{\mathcal{H}}}$ and $\overline{\boldsymbol{\mathcal{H}}\boldsymbol{\mathcal{Z}}}$ as $\overline{\boldsymbol{\mathcal{Z}}}$. We also drop the dependence on $\mathbf{b}$ and k of the vectors/matrices as they are irrelevant in the proof. Our goal is to show the following:

$$\|(\boldsymbol{\mathcal{H}}_p^{\ell_t})^{-1}\boldsymbol{\mathcal{Z}}_p^{\ell_t} - (\overline{\boldsymbol{\mathcal{H}}})^{-1}\overline{\boldsymbol{\mathcal{Z}}}\| = \mathcal{O}(\lambda_{\bar{W}}^{\ell_t}). \quad (49)$$

We now observe that:

$$\boldsymbol{\mathcal{H}}_p^{\ell_t} = \overline{\boldsymbol{\mathcal{H}}} + \delta\boldsymbol{\mathcal{H}}_p^{\ell_t}, \;\; \boldsymbol{\mathcal{Z}}_p^{\ell_t} = \overline{\boldsymbol{\mathcal{Z}}} + \delta\boldsymbol{\mathcal{Z}}_p^{\ell_t} \quad (50)$$

where the error terms satisfy

$$\|\delta\boldsymbol{\mathcal{H}}_p^{\ell_t}\| \leq C_0\lambda_{\bar{W}}^{\ell_t}, \;\; \|\delta\boldsymbol{\mathcal{Z}}_p^{\ell_t}\| \leq C_z\lambda_{\bar{W}}^{\ell_t}, \quad (51)$$

for some $C_z < \infty$. The error bounds in (51) are due to exponential convergence of the G-AC protocol [35]. Under assumption (31), the matrix inverse admits a series expansion [32]:

$$(\overline{\boldsymbol{\mathcal{H}}} + \delta\boldsymbol{\mathcal{H}}_p^{\ell_t})^{-1} = \sum_{q=0}^{\infty}(-1)^q(\overline{\boldsymbol{\mathcal{H}}}^{-1}\delta\boldsymbol{\mathcal{H}}_p^{\ell_t})^q(\overline{\boldsymbol{\mathcal{H}}})^{-1}. \quad (52)$$

Consequently, the left hand side in (49) can be upper bounded by:

$$\begin{aligned}
&\|(\boldsymbol{\mathcal{H}}_p^{\ell_t})^{-1}\boldsymbol{\mathcal{Z}}_p^{\ell_t} - (\overline{\boldsymbol{\mathcal{H}}})^{-1}\overline{\boldsymbol{\mathcal{Z}}}\| \\
&= \left\|\overline{\boldsymbol{\mathcal{H}}}^{-1}\delta\boldsymbol{\mathcal{Z}}_p^{\ell_t} + \sum_{q=1}^{\infty}(-1)^q(\overline{\boldsymbol{\mathcal{H}}}^{-1}\delta\boldsymbol{\mathcal{H}}_p^{\ell_t})^q\overline{\boldsymbol{\mathcal{H}}}^{-1}(\overline{\boldsymbol{\mathcal{Z}}} + \delta\boldsymbol{\mathcal{Z}}_p^{\ell_t})\right\| \\
&\leq \|\overline{\boldsymbol{\mathcal{H}}}^{-1}\|\left(C_z\lambda_{\bar{W}}^{\ell_t} + \sum_{q=1}^{\infty}\|(\delta\boldsymbol{\mathcal{H}}_p^{\ell_t}\overline{\boldsymbol{\mathcal{H}}}^{-1})^q\|\|\overline{\boldsymbol{\mathcal{Z}}} + \delta\boldsymbol{\mathcal{Z}}_p^{\ell_t}\|\right).
\end{aligned} \quad (53)$$

Assuming that $\|\overline{\boldsymbol{\mathcal{Z}}} + \delta\boldsymbol{\mathcal{Z}}_p^{\ell_t}\| \leq C_Z$, we have

$$\begin{aligned}
&\|(\boldsymbol{\mathcal{H}}_p^{\ell_t})^{-1}\boldsymbol{\mathcal{Z}}_p^{\ell_t} - (\overline{\boldsymbol{\mathcal{H}}})^{-1}\overline{\boldsymbol{\mathcal{Z}}}\| \\
&\quad\leq C_1\left(C_z\lambda_{\bar{W}}^{\ell_t} + C_Z\sum_{q=1}^{\infty}\|(\delta\boldsymbol{\mathcal{H}}_p^{\ell_t}\overline{\boldsymbol{\mathcal{H}}}^{-1})^q\|\right) \\
&\quad\leq C_1\left(C_z\lambda_{\bar{W}}^{\ell_t} + C_Z\sum_{q=1}^{\infty}(C_0C_1\lambda_{\bar{W}}^{\ell_t})^q\right)
\end{aligned} \quad (54)$$

Since $C_0C_1\lambda_{\bar{W}}^{\ell_t} < 1$, we have

$$\begin{aligned}
&\|(\boldsymbol{\mathcal{H}}_p^{\ell_t})^{-1}\boldsymbol{\mathcal{Z}}_p^{\ell_t} - (\overline{\boldsymbol{\mathcal{H}}})^{-1}\overline{\boldsymbol{\mathcal{Z}}}\| \\
&\quad\leq C_1\left(C_z\lambda_{\bar{W}}^{\ell_t} + \frac{C_ZC_0C_1}{1-C_0C_1\lambda_{\bar{W}}^{\ell_t}}\lambda_{\bar{W}}^{\ell_t}\right) = \mathcal{O}(\lambda_{\bar{W}}^{\ell_t}).
\end{aligned} \quad (55)$$

The proof is completed.

Appendix C
Proof of Theorem 1

To facilitate our discussions, we introduce the shorthand notations: $\boldsymbol{x}_p^{(t)} \triangleq \{\mathbf{X}_{p,A}^{(t)}(\omega_k)\}_k$, $\boldsymbol{x}_p^{\ell_t}(\mathbf{b}^{(t)}) \triangleq \{\mathbf{X}_{p,A,k}(\mathbf{b}^{(t)})\}_k$, $\boldsymbol{x}_t^\star \triangleq \mathbf{X}_{A,k}(\mathbf{b}^{(t-1)})$, $\boldsymbol{x}^{(t)} \triangleq (\boldsymbol{x}_1^{(t)}, ..., \boldsymbol{x}_P^{(t)})$ and $\hat{\boldsymbol{x}}^{(t)} \triangleq (1/P)\sum_{p=1}^P \boldsymbol{x}_p^{(t)}$. With a slight abuse of the notations, we let $f(\boldsymbol{x}^{(t)}, \mathbf{b}) \triangleq \sum_{p=1}^P f_p(\boldsymbol{x}_p^{(t)}, b_p)$. We also set $\beta = 1/M_o$.

We can assign upper bounds to the following norms of the differences:

$$\textstyle\sum_{p=1}^P \|\boldsymbol{x}_p^{(t)} - \boldsymbol{x}_t^\star\|_2 \leq \theta(\lambda_{\bar{W}}^{\ell_t}) \triangleq C_\theta\lambda_{\bar{W}}^{\ell_t}, \quad (56)$$

$$\textstyle\sum_{p=1}^P \|\hat{\boldsymbol{x}}^{(t)} - \boldsymbol{x}_t^\star\|_2 \leq \phi(\lambda_{\bar{W}}^{\ell_t}) \triangleq C_\phi\lambda_{\bar{W}}^{\ell_t}, \quad (57)$$

$$\textstyle\sum_{p=1}^P \|\boldsymbol{x}_p^{(t)} - \hat{\boldsymbol{x}}^{(t)}\|_2 \leq \psi(\lambda_{\bar{W}}^{\ell_t}) \triangleq C_\psi\lambda_{\bar{W}}^{\ell_t}. \quad (58)$$

The first inequality is due to Proposition 2. The latter two inequalities can be derived using triangular inequalities on (56).

Under the assumption in Theorem 1, for all t, $(\hat{\boldsymbol{x}}^{(t)}, \mathbf{b}^{(t)})$ stays in the neighborhood $\mathcal{N}_{R^\star}(\boldsymbol{x}^\star, \mathbf{b}^\star)$ such that f is strongly convex. Our idea is to study the dynamics of the following:

$$\Delta^{(t)} = f(\hat{\boldsymbol{x}}^{(t)}, \mathbf{b}^{(t)}) - f(\boldsymbol{x}^\star, \mathbf{b}^\star) \quad (59)$$

Observe that:

$$\begin{aligned}&\Delta^{(t)}-\Delta^{(t-1)}\\&=f(\hat{\boldsymbol{x}}^{(t)},\mathbf{b}^{(t)})-f(\hat{\boldsymbol{x}}^{(t-1)},\mathbf{b}^{(t-1)})\\&=f(\boldsymbol{x}^{(t)},\mathbf{b}^{(t)})-f(\boldsymbol{x}^{(t)},\mathbf{b}^{(t)})+f(\hat{\boldsymbol{x}}^{(t)},\mathbf{b}^{(t)})\\&\quad-f(\hat{\boldsymbol{x}}^{(t-1)},\mathbf{b}^{(t-1)})\\&\leq L_o\psi(\lambda_{\bar{W}}^{\ell_t})+f(\boldsymbol{x}^{(t)},\mathbf{b}^{(t)})-f(\hat{\boldsymbol{x}}^{(t-1)},\mathbf{b}^{(t-1)})\\&\leq L_o\psi(\lambda_{\bar{W}}^{\ell_t})+f(\boldsymbol{x}^{(t)},\mathbf{b}^{(t-1)})-\\&\quad f(\hat{\boldsymbol{x}}^{(t-1)},\mathbf{b}^{(t-1)})-\frac{M_o}{2}\|\mathbf{b}^{(t)}-\mathbf{b}^{(t-1)}\|_2^2\end{aligned}\tag{60}$$

where the first inequality is due to Lipschitz continuity of f and (58); the second inequality is due to the descent lemma [36]. Moreover, we have:

$$\begin{aligned}f(\boldsymbol{x}^{(t)},\mathbf{b}^{(t-1)})&\leq f(\boldsymbol{x}_t^\star,\mathbf{b}^{(t-1)})+L_o\theta(\lambda_{\bar{W}}^{\ell_t})\\&\leq f(\hat{\boldsymbol{x}}^{(t-1)},\mathbf{b}^{(t-1)})+L_o\theta(\lambda_{\bar{W}}^{\ell_t}),\end{aligned}\tag{61}$$

where the first inequality is due to (56) and the second inequality is due to the optimality of $\boldsymbol{x}_t^\star$ with $\mathbf{b}^{(t-1)}$ fixed. Therefore,

$$\Delta^{(t)}-\Delta^{(t-1)}\leq L_o(\theta(\lambda_{\bar{W}}^{\ell_t})+\psi(\lambda_{\bar{W}}^{\ell_t}))-\frac{M_o}{2}\|\mathbf{b}^{(t)}-\mathbf{b}^{(t-1)}\|_2^2.\tag{62}$$

Our next task is to lower bound $\|\mathbf{b}^{(t)}-\mathbf{b}^{(t-1)}\|_2^2$. To this end, we proceed by:

$$\begin{aligned}\Delta^{(t)}&=f(\hat{\boldsymbol{x}}^{(t)},\mathbf{b}^{(t)})-f(\boldsymbol{x}^\star,\mathbf{b}^\star)\\&=f(\boldsymbol{x}_t^\star,\mathbf{b}^{(t)})-f(\boldsymbol{x}_t^\star,\mathbf{b}^{(t)})+f(\hat{\boldsymbol{x}}^{(t)},\mathbf{b}^{(t)})-f(\boldsymbol{x}^\star,\mathbf{b}^\star)\\&\leq L_o\phi(\lambda_{\bar{W}}^{\ell_t})+\langle\nabla_b f(\boldsymbol{x}_t^\star,\mathbf{b}^{(t)}),\mathbf{b}^{(t)}-\mathbf{b}^\star\rangle+\\&\qquad BM_o\|\mathbf{b}^{(t)}-\mathbf{b}^{(t-1)}\|_2\end{aligned}\tag{63}$$

where in the last inequality, we have used i) f is Lipschitz continuous and (57), ii) f is locally convex and iii)

$$\begin{aligned}&\langle\nabla_x f(\boldsymbol{x}_t^\star,\mathbf{b}^{(t)}),\boldsymbol{x}_t^\star-\boldsymbol{x}^\star\rangle\\&\quad=\langle\nabla_x f(\boldsymbol{x}_t^\star,\mathbf{b}^{(t)})-\nabla_x f(\boldsymbol{x}_t^\star,\mathbf{b}^{(t-1)}),\boldsymbol{x}_t^\star-\boldsymbol{x}^\star\rangle\\&\quad\leq BM_o\|\mathbf{b}^{(t)}-\mathbf{b}^{(t-1)}\|_2.\end{aligned}\tag{64}$$

The equality is due to $\nabla_x f(\boldsymbol{x}_t^\star,\mathbf{b}^{(t-1)})=\mathbf{0}$.

Our next endeavor is to upper bound $\langle\nabla_b f(\boldsymbol{x}_t^\star,\mathbf{b}^{(t)}),\mathbf{b}^{(t)}-\mathbf{b}^\star\rangle$. To this end, we observe

$$\begin{aligned}&\nabla_b f(\boldsymbol{x}_t^\star,\mathbf{b}^{(t)})=\nabla_b f(\boldsymbol{x}_t^\star,\mathbf{b}^{(t)})-\nabla_b f(\boldsymbol{x}^{(t)},\mathbf{b}^{(t-1)})+\\&\tfrac{1}{\beta}\big((\mathbf{b}^{(t-1)}-\mathbf{b}^{(t)})+\mathbf{b}^{(t)}-(\mathbf{b}^{(t-1)}-\beta\nabla_b f(\boldsymbol{x}^{(t)},\mathbf{b}^{(t-1)})\big),\end{aligned}\tag{65}$$

since the latter terms cancel each other. Together with the following inequalities:

$$\begin{aligned}\langle\nabla_b f(\boldsymbol{x}_t^\star,\mathbf{b}^{(t)})-\nabla_b f(\boldsymbol{x}^{(t)},\mathbf{b}^{(t-1)}),\mathbf{b}^{(t)}-\mathbf{b}^\star\rangle\leq\\BM_o\big(\theta(\lambda_{\bar{W}}^{\ell_t})+\|\mathbf{b}^{(t)}-\mathbf{b}^{(t-1)}\|_2\big),\end{aligned}\tag{66}$$

which is a consequence of Cauchy-Schwarz and (56). Moreover, we have

$$\langle\mathbf{b}^{(t)}-(\mathbf{b}^{(t-1)}-\beta\nabla_b f(\boldsymbol{x}^{(t)},\mathbf{b}^{(t-1)})),\mathbf{b}^{(t)}-\mathbf{b}^\star\rangle\leq 0,\tag{67}$$

since $\mathbf{b}^{(t)}$ is the projection of $\mathbf{b}^{(t-1)}-\beta\nabla_b f(\boldsymbol{x}^{(t)},\mathbf{b}^{(t-1)})$ onto $\mathcal{B}$ and $\mathbf{b}^\star\in\mathcal{B}$.

Consequently,

$$\begin{aligned}\langle\nabla_b f(\boldsymbol{x}_t^\star,\mathbf{b}^{(t)}),\mathbf{b}^{(t)}-\mathbf{b}^\star\rangle\leq BM_o\big(\theta(\lambda_{\bar{W}}^{\ell_t})+\\\|\mathbf{b}^{(t)}-\mathbf{b}^{(t-1)}\|_2\big)+(B/\beta)\|\mathbf{b}^{(t)}-\mathbf{b}^{(t-1)}\|_2\end{aligned}\tag{68}$$

and $\Delta^{(t)}$ is upper bounded by:

$$\Delta^{(t)}\leq L_o\phi(\lambda_{\bar{W}}^{\ell_t})+BM_o\theta(\lambda_{\bar{W}}^{\ell_t})+3BM_o\|\mathbf{b}^{(t)}-\mathbf{b}^{(t-1)}\|_2\tag{69}$$

Plugging the above results back to (62) yields the following:

$$\begin{aligned}\Delta^{(t)}-\Delta^{(t-1)}\leq L_o(\theta(\lambda_{\bar{W}}^{\ell_t})+\psi(\lambda_{\bar{W}}^{\ell_t}))-\\\gamma_o\max\{\Delta^{(t)}-(L_o\phi(\lambda_{\bar{W}}^{\ell_t})+BM_o\theta(\lambda_{\bar{W}}^{\ell_t})),0\}^2\end{aligned}\tag{70}$$

where $\gamma_o=1/(18B^2M_o)$.

Since $\Delta^{(t)}$ is non-negative, we can simplify (70) by considering the upper bound $\xi^{(t)}$ such that $\Delta^{(t)}\leq\xi^{(t)}$ and $\mathcal{O}(\lambda_{\bar{W}}^{\ell_t})\leq\mathcal{O}(\lambda_{\bar{W}}^{\ell_{min}})$ for all t:

$$\begin{aligned}\xi^{(t)}-\xi^{(t-1)}=L_o(\theta(\lambda_{\bar{W}}^{\ell_{min}})+\psi(\lambda_{\bar{W}}^{\ell_{min}}))-\\\gamma_o\max\{\xi^{(t)}-(L_o\phi(\lambda_{\bar{W}}^{\ell_{min}})+BM_o\theta(\lambda_{\bar{W}}^{\ell_{min}})),0\}^2\end{aligned}\tag{71}$$

A fixed point $\bar{\xi}$ to the above system must satisfy:

$$\begin{aligned}&L_o(\theta(\lambda_{\bar{W}}^{\ell_{min}})+\psi(\lambda_{\bar{W}}^{\ell_{min}}))=\\&\gamma_o\max\{\bar{\xi}-(L_o\phi(\lambda_{\bar{W}}^{\ell_{min}})+BM_o\theta(\lambda_{\bar{W}}^{\ell_{min}})),0\}^2\end{aligned}\tag{72}$$

which implies

$$\begin{aligned}\bar{\xi}=(L_o\phi(\lambda_{\bar{W}}^{\ell_{min}})+BM_o\theta(\lambda_{\bar{W}}^{\ell_{min}}))+\\\sqrt{L_o(\theta(\lambda_{\bar{W}}^{\ell_{min}})+\psi(\lambda_{\bar{W}}^{\ell_{min}}))/\gamma_o}=\sqrt{\mathcal{O}(\lambda_{\bar{W}}^{\ell_{min}})}.\end{aligned}\tag{73}$$

It can be verified that the above fixed point is stable. In fact, it is the only fixed point for the upper bound system (71).

Finally, from (73) and the local strong convexity of f, we have the following chain

$$\begin{aligned}\lim_{t\to\infty}\|(\hat{\boldsymbol{x}}^{(t)},\mathbf{b}^{(t)})-(\boldsymbol{x}^\star,\mathbf{b}^\star)\|^2&\leq\frac{2}{m_o}\lim_{t\to\infty}\Delta^{(t)}\\&\leq\sqrt{\mathcal{O}(\lambda_{\bar{W}}^{\ell_{min}})},\end{aligned}\tag{74}$$

which completes the proof.

ACKNOWLEDGMENT

The authors would like to thank the anonymous reviewers for their valuable comments to improve the manuscript.

REFERENCES

[1] H.-T. Wai and A. Scaglione, "State Estimation with Sampling Offsets in Wide Area Measurement Systems," in *Proc' IEEE SAM 2014*, Jun. 2014, pp. 49–52.

[2] ——, "Decentralized Regression with Asynchronous Sub-Nyquist Sampling," to appear in *Proc' Asilomar 2014*.

[3] L. Xie, D.-h. Choi, S. Kar, and H. V. Poor, "Fully Distributed State Estimation for Wide-Area Monitoring Systems," *IEEE Trans. Smart Grid*, vol. 3, no. 3, pp. 1154–1169, Sep. 2012.

[4] V. Kekatos and G. B. Giannakis, "Distributed Robust Power System State Estimation," *IEEE Trans. Power Syst.*, vol. 28, no. 2, pp. 1617–1626, May 2013.

[5] X. Li and A. Scaglione, "Robust Decentralized State Estimation and Tracking for Power Systems via Network Gossiping," *IEEE J. Sel. Areas Commun.*, vol. 31, no. 7, pp. 1184–1194, Jul. 2013.

[6] A. G. Dimakis, S. Kar, J. M. F. Moura, M. G. Rabbat, and A. Scaglione, "Gossip Algorithms for Distributed Signal Processing," *Proc. IEEE*, vol. 98, no. 11, pp. 1847–1864, Nov. 2010.

[7] S. Kar, J. M. F. Moura, and K. Ramanan, "Distributed Parameter Estimation in Sensor Networks: Nonlinear Observation Models and Imperfect Communication," *IEEE Trans. Inform. Theory*, vol. 58, no. 6, pp. 3575–3605, Jun. 2012.

[8] I. D. Schizas, A. Ribeiro, and G. B. Giannakis, "Consensus in Ad Hoc WSNs With Noisy Links Part I: Distributed Estimation of Deterministic Signals," *IEEE Trans. Signal Process.*, vol. 56, no. 1, pp. 350–364, Jan. 2008.

[9] M. Rabbat and R. Nowak, "Distributed optimization in sensor networks," in *Proc' IPSN'04*, 2004.

[10] R. Olfati-Saber, J. A. Fax, and R. M. Murray, "Consensus and Cooperation in Networked Multi-Agent Systems," *Proc. IEEE*, vol. 95, no. 1, pp. 215–233, Jan. 2007.

[11] A. H. Sayed, S.-y. Tu, J. Chen, X. Zhao, and Z. J. Towfic, "Diffusion strategies for adaptation and learning over networks: an examination of distributed strategies and network behavior," *IEEE Signal Process. Mag.*, vol. 30, no. 3, pp. 155–171, May 2013.

[12] 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.

[13] D. P. Shepard, T. E. Humphreys, and A. Fansler, "Evaluation of the vulnerability of phasor measurement units to GPS spoofing attacks," *International Journal of Critical Infrastructure Protection*, vol. 5, no. 3-4, pp. 146–153, Dec. 2012.

[14] A. Ferreira and J. Fernandes, "A survey on time delay system estimation," *Proc' ECC 1997*.

[15] S. Julier and J. Uhlmann, "Fusion of time delayed measurements with uncertain time delays," in *Proc' ACC 2005*, pp. 4028–4033.

[16] J.-O. Nilsson, I. Skog, and P. Handel, "Joint state and measurement time-delay estimation of nonlinear state space systems," in *Proc' ISSPA 2010*, May 2010, pp. 324–328.

[17] J. Nilsson and P. Händel, "Timing estimation in distributed sensor and control systems with central processing," *arXiv preprint arXiv:1309.1864*, vol. 46, no. 0, 2013. [Online]. Available: http://arxiv.org/abs/1309.1864

[18] P. Yang, Z. Tan, A. Wiesel, and A. Nehorai, "Power System State Estimation Using PMUs With Imperfect Synchronization," *IEEE Trans. Power Syst.*, vol. 28, no. 4, pp. 4162–4172, Nov. 2013.

[19] S. Huang and B. C. Levy, "Blind Calibration of Timing Offsets for Four-Channel Time-Interleaved ADCs," *IEEE Trans. Circuits Syst. I, Reg. Papers*, vol. 54, no. 4, pp. 863–876, Apr. 2007.

[20] J. Goodman, B. Miller, M. Herman, G. Raz, and J. Jackson, "Polyphase Nonlinear Equalization of Time-Interleaved Analog-to-Digital Converters," *IEEE J. Sel. Topics Signal Process.*, vol. 3, no. 3, pp. 362–373, Jun. 2009.

[21] R. Venkataramani and Y. Bresler, "Optimal sub-Nyquist nonuniform sampling and reconstruction for multiband signals," *IEEE Trans. Signal Process.*, vol. 49, no. 10, pp. 2301–2313, 2001.

[22] M. Mishali and Y. Eldar, "From Theory to Practice: Sub-Nyquist Sampling of Sparse Wideband Analog Signals," *IEEE J. Sel. Topics Signal Process.*, vol. 4, no. 2, pp. 375–391, Apr. 2010.

[23] Y. Eldar and A. Oppenheim, "Filterbank reconstruction of bandlimited signals from nonuniform and generalized samples," *IEEE Trans. Signal Process.*, vol. 48, no. 10, pp. 2864–2875, 2000.

[24] M. Mishali and Y. Eldar, "Sub-Nyquist Sampling," *IEEE Signal Process. Mag.*, vol. 28, no. 6, pp. 98–124, Nov. 2011.

[25] A. Nedic, A. Ozdaglar, and P. Parrilo, "Constrained Consensus and Optimization in Multi-Agent Networks," *IEEE Trans. Autom. Control*, vol. 55, no. 4, pp. 922–938, Apr. 2010.

[26] S. S. Ram, A. Nedic, and V. V. Veeravalli, "A new class of distributed optimization algorithms : application to regression of distributed data," *Optimization Methods and Software*, no. 1, pp. 37–41, Feb. 2012.

[27] S. Mou and A. Morse, "A fixed-neighbor, distributed algorithm for solving a linear algebraic equation," in *Proc' ECC' 2013*, Jul. 2013, pp. 2269–2273.

[28] X. Li and A. Scaglione, "Convergence and Applications of a Gossip-Based Gauss-Newton Algorithm," *IEEE Trans. Signal Process.*, vol. 61, no. 21, pp. 5231–5246, Nov. 2013.

[29] A. H. Sayed, *Fundamentals of Adaptive Filtering*, 1st ed. Wiley-IEEE Press, Jun. 2003.

[30] L. Tong and S. Perreau, "Multichannel blind identification: from subspace to maximum likelihood methods," *Proc. IEEE*, vol. 86, no. 10, pp. 1951–1968, 1998.

[31] G. Xu, H. Liu, L. Tong, and T. Kailath, "A least-squares approach to blind channel identification," *IEEE Trans. Signal Process.*, vol. 43, no. 12, pp. 2982–2993, 1995.

[32] R. A. Horn and C. R. Johnson, Eds., *Matrix Analysis*. Cambridge University Press, 1986.

[33] D. Brillinger, *Time Series: Data Analysis and Theory*. Society for Industrial and Applied Mathematics, 2001.

[34] M. Razaviyayn, M. Hong, and Z.-Q. Luo, "A Unified Convergence Analysis of Block Successive Minimization Methods for Nonsmooth Optimization," *SIAM Journal on Optimization*, vol. 23, no. 2, pp. 1126–1153, Jun. 2013.

[35] S. Boyd, A. Ghosh, B. Prabhakar, and D. Shah, "Randomized gossip algorithms," *IEEE Trans. Inf. Theory*, vol. 52, no. 6, pp. 2508–2530, Jun. 2006.

[36] D. P. Bertsekas, *Nonlinear Programming*. Athena Scientific, Sep. 1999.

[37] "IEEE Standard for Synchrophasors for Power Systems," *IEEE Std C37.118-2005 (Revision of IEEE Std 1344-1995)*, pp. 1–57, 2006.

[38] L. Grigsby, Ed., *Power System Stability and Control*, ser. Electrical Engineering Handbook. CRC Press, May 2007, vol. 20073061.

[39] A. V. Oppenheim, R. W. Schafer, and J. R. Buck, *Discrete-time Signal Processing (2nd Ed.)*. Upper Saddle River, NJ, USA: Prentice-Hall, Inc., 1999.

[40] L. Xiao and S. Boyd, "Fast linear iterations for distributed averaging," *Systems & Control Letters*, vol. 53, no. 1, pp. 65–78, Sep. 2004.

[41] R. D. Zimmerman, C. E. Murillo-Sanchez, and R. J. Thomas, "MATPOWER: Steady-State Operations, Planning, and Analysis Tools for Power Systems Research and Education," *IEEE Trans. Power Syst.*, vol. 26, no. 1, pp. 12–19, Feb. 2011.

[42] NASPI, "PMUs and Synchrophasor Data Flows in North America," https://www.naspi.org/File.aspx?fileID=1275, Mar. 2014.

[43] P. P. Vaidyanathan, *Multirate Systems and Filter Banks*. Prentice-Hall, Inc., 1993.

Consensus on State and Time: Decentralized Regression with Asynchronous Sampling

Hoi-To Wai and Anna Scaglione

Abstract

An implicit assumption made in several studies on sensor systems is that the time and frequency at which sensor measurements are taken is consistent across all the distributed sensing sites. In reality, the times of measurement often lack consistency and integrity, and this is an intrinsic vulnerability of wide area sensor system. Data logs coming from different Analog to Digital Converters (ADCs) are not in phase and may differ also in the sampling rate, in some cases because heterogeneity in the sensors and in others because the data are simply not refreshed in the data historians with the same frequency. Lack of good synchronization in sensing may be the result of a malfunction or also due to intentional delay attacks.

This premise motivates our work, where we advance the area of decentralized signal processing and consider explicitly timing errors and non-homogenous sampling rates in least square estimation problems with distributed sensing. For linear observations models, we provide a necessary and sufficient condition for identifiability of the time offsets. We propose an algorithm for the joint regression on the state vector and time offsets. The algorithm also exploits the asynchrony and redundancy in the spatial sampling to attain sub-Nyquist sampling resolution of the slow sensor feeds. Importantly, this also leads to the development of a novel decentralized algorithm. The efficacies of the proposed decentralized algorithm are shown by both convergence analysis and numerical simulations.

Index terms— decentralized state estimation, sampling offsets, sub-Nyquist sampling, smart grid.

EDICS: SPE-DETC (Detection and estimation in power grid), SPE-DP (Distributed processing and protocols in energy systems), SAM-APPL (Applications of sensor and array multichannel processing)

This material is based upon work supported by the Department of Energy under Award Number DE-OE0000097. Preliminary versions of this work were presented at IEEE SAM 2014 [1] and Asilomar 2014 [2].

H.-T. Wai and A. Scaglione are with Department of Electrical and Computer Engineering, University of California, Davis, CA 95616, USA. E-mails: {htwai,ascaglione}@ucdavis.edu.

I. Introduction

Today there is significant interest in developing *decentralized* signal processing techniques for solving regression problems that arises in array processing and control applications (e.g., see [3]–[12] and the references therein). These decentralized algorithms overcome the lack of *observability* in individual sensors by merging communication with computations in a resilient fashion, relying on (possibly randomized) near-neighbors communications.

One of the implicit assumptions made in the vast majority of related literatures is that the measurements are sampled in a synchronous manner. Such an assumption is valid only when: i) the system state evolution is sufficiently slow that the lack of synchrony in sampling is negligible; or ii) the timing information is sufficiently accurate (with the aid of for instance of a GPS receiver) to calibrate the ADCs; and iii) the sensors employed follow the same sampling rate. These assumption are quite limiting. For example, in power grid, local clocks in the measurement devices are prone to malicious attacks [13]; in sensor array processing, the desire of processing signals over high frequency carriers and wide bandwidths has made the design of hardwares for synchronization more challenging.

This paper attacks the problem of non-ideal sampling by addressing these issues in a unified manner, while considering the case of power grid as an immediate application example. Our formulation imposes very mild restrictions. Specifically, assuming that the relationship between the measurements and state is *memoryless*, we model the system state variables as band-limited continuous-time signals and the measurements as samples taken at different sampling frequencies and with an unknown time offset. The analysis and algorithms we propose rely only on the sampling expansion (i.e. *smoothness* in the state signal). On the practical side, by representing the down-sampled and time-shifted signals in the frequency domain, we tackle the new regression problem on state and time offsets using a novel decentralized algorithm. The decentralized algorithm is proven to converge both analytically and empirically.

The remainder of this paper is organized as follows. In Section II, we introduce the system model with linear measurement and non-ideal sampling. Specifically, we derive an equivalent frequency-domain representation where our analysis and algorithms are based upon. In Section III, we derive conditions under which the accurate state and times offsets are recoverable via solving the proposed regression problem. A decentralized algorithm for tackling the proposed regression problem will be discussed and analyzed in Section IV. In Section V, we discuss the extension to non-linear measurement models and its applications on the power grids. Finally, the paper is concluded by the simulation results that show the efficiencies of the proposed method in Section VI.

Notations: We follow the standard notations used in signal processing literature. The operator $\otimes$ denotes the Kronecker product, $(\cdot)^*$ denotes complex conjugate, $\mathrm{DTFT}\{\cdot\}$ denotes the standard DTFT transformation $\sum_{n=-\infty}^{\infty} \mathbf{x}[n]e^{-j\omega n}$.

A. Related works

Techniques for mitigating timing errors have long been considered in the control theory literatures, e.g., [14]–[17]. A common feature among these works is that they adopt a Kalman filtering approach and are often combined with a Taylor approximation to the system dynamics. As a result, applying these techniques requires an a-priori knowledge of the dynamic equations that represent the evolution of system state, and that the timing offset is sufficiently small such that the Taylor approximation is accurate. A good example of prior works that is relevant to ours is [18]. In this paper, Yang et. al. considered a static system where the effects of non-ideal sampling is resulted from down-converting signals with high-frequency carriers (e.g., in power system state estimation (PSSE) with linear measurements) and tackled the joint time offsets and state regression using a Taylor approximation.

What differentiates our work is the direct manner in which we model the effects of non-ideal sampling. In fact, we do not require any a-priori knowledge on the state evolution dynamics nor assumptions on the magnitude of time offsets. Our work is in the same spirit with sampling using time-interleaved ADCs [19], [20]. Furthermore, it overlaps with the recent works on sub-Nyquist recovery in [21]–[23], reviewed in [24]. In these works, it is usually required that the signal being sampled satisfies certain properties known a-priori to the system, while ours focuses on the blind calibration problem and fusion of measurements.

Decentralized algorithms over sensor network have been considered in [11], [12], [25]–[27]. For example, [11], [12] combines gradient (or quasi-Newton) descent and consensus protocols to develop the diffusion-based LMS (or RLS) algorithm; [28] proposes a Gossip-based Gauss-Newton algorithm. While most of these algorithms applies to general optimization problems, they are guaranteed to converge only when the optimization is convex. However, our regression problem with non-ideal sampling is non-convex and the aforementioned algorithms cannot be directly applied. We have proposed a new decentralized algorithm that exploits structures in the said regression problem. More importantly, the algorithm is proven to converge under certain conditions.

II. System model

Consider a sensor network equipped with P sensors. The sensor network monitors the system state that can be modelled as a continuous-time signal $\mathbf{x}_c(t) \in \mathbb{C}^N$. Specifically, we study the case when the

measurement is linear in the system state such that the pth sensor observes the following at time t:

$$\boldsymbol{\zeta}_p(t) = \mathbf{H}_p \mathbf{x}_c(t) + \boldsymbol{v}_p(t), \tag{1}$$

where $\mathbf{H}_p \in \mathbb{C}^{M_p \times N}$ is the measurement matrix with $M_p \leq N$ and $\boldsymbol{v}_p(t) \sim \mathcal{CN}(0, \sigma_w^2 \mathbf{I})$ is an additive white noise. For the applications of (1) on sensor array processing, we refer our readers to [6]–[8]; in addition, some recent applications on power system state estimation can be found in [3]–[5]. Our model can also be extended to the general case with non-linear measurement, i.e., when $\boldsymbol{\zeta}_p(t)$ is non-linear in $\mathbf{x}_c(t)$; see Section V.

The objective is to estimate $\mathbf{x}_c(t)$ using $\boldsymbol{\zeta}_p(t)$. In the control theory literatures [15]–[18], [29], the system state $\mathbf{x}_c(t)$ is often modelled by a linear/non-linear dynamical system. We consider the scenario when such a knowledge on the underlying dynamical system is not available. Instead, we study the state estimation problem under the model implied by the following assumptions:

Assumption 1. *The system state $\mathbf{x}_c(t)$ is band-limited by $W/2$ Hz, i.e., $\mathbf{x}_c(t)$ is a smooth signal.*

Assumption 2. *The measurements are collected from the continuous-time signal $\boldsymbol{\zeta}_p(t)$ under a non-ideal sampling model — the pth sensor samples $\boldsymbol{\zeta}_p(t)$ at time:*

$$t_p^n = (nA_p - b_p)T_s, \ n = 0, 1, 2, ..., \tag{2}$$

where $T_s = 1/W$ second is the Nyquist sampling period, $b_p \in \mathbb{R}$ is the normalized time offset at sensor p and the sampling factor $A_p \geq 1$ is an integer. In addition, we set $b_1 = 0$ to avoid ambiguity.

Assumption 3. *The down-sampling factor $A_p \in \mathbb{Z}_+$ is known while the time offset $b_p \in \mathcal{B}_p$ is unknown. The interval $\mathcal{B}_p \subseteq \mathbb{R}$ is convex and known.*

The consequences of Assumption 1 to 3 are discussed as follows. Assumption 1 is the key enabling assumption for estimating $\mathbf{x}_c(t)$ from samples of measurements in (1). Under the assumption, it suffices to obtain $\mathbf{x}[n] = \mathbf{x}_c(nT_s)$ in order to estimate $\mathbf{x}_c(t)$. In fact, by observing that

$$\boldsymbol{\zeta}_p(nT_s) = \mathbf{H}_p \mathbf{x}[n] + \boldsymbol{v}_p(nT_s), \tag{3}$$

the system state $\mathbf{x}[n]$ can be readily estimated by solving a least square optimization with the data $\{\boldsymbol{\zeta}_p(nT_s)\}_p$. In fact, Assumption 1 with the synchronous sampling model is one of the implicit assumptions made in most literatures on sensor array processing [3]–[9].

In reality, obtaining the set of data $\{\boldsymbol{\zeta}_p(nT_s)\}_p$ is impossible since it requires the physical system to sample at time $t_p^n = nT_s$ for *all* sensors. This requires synchronization, and it is hard to enforce

for sensors placed over a physical system that occupies a wide area. In this case, often measurement samples are collected at *heterogenous* sampling rate and with sampling offsets. This is the reason why Assumption 2 was imposed. Under Assumption 2, the *sampled* version of (1) can be expressed as:

$$\boldsymbol{\zeta}_p[n] \triangleq \boldsymbol{\zeta}_p(t_p^n) = \mathbf{H}_p \mathbf{x}_c((nA_p - b_p)T_s) + \boldsymbol{v}_p[n]. \tag{4}$$

Lastly, Assumption 3 is justified by the fact that the sampling rate of a sensor is usually known a-priori, while the time offsets are unpredictable.

The aim of this paper is to study the estimation problem of $\mathbf{x}[n]$ and b_p. We first observe that estimating $\mathbf{x}[n]$ and b_p on a sample-by-sample basis from $\{\boldsymbol{\zeta}_p[n]\}_{p=1}^K$ alone is impossible as the former terms do not appear in the right hand side of (4). As a remedy, we consider a frequency domain representation for (4) and leverage on the following observation:

Observation 1. *Let $\mathbf{x}_c(t)$ be a bandlimited signal with bandwidth $W/2$ Hz. We denote $\mathbf{x}[n] = \mathbf{x}_c(nT_s)$ with $T_s = 1/W$ as its discrete time equivalent and $\mathbf{X}(e^{j\omega}) = \mathrm{DTFT}\{\mathbf{x}[n]\}$ is the discrete time Fourier transform (DTFT) spectrum. Then:*

$$\mathrm{DTFT}\{\mathbf{x}_c((nA - b_p)T_s)\} = \frac{1}{A}\sum_{a=0}^{A-1} e^{-jb_p\Omega_A^a(\omega)} \mathbf{X}\left(e^{j\Omega_A^a(\omega)}\right), \tag{5}$$

where

$$\Omega_A^a(\omega) \triangleq \left(\frac{\omega}{A} - \frac{a}{A}2\pi\right) \bmod (-\pi, \pi]. \tag{6}$$

The proof of Observation 1 is relegated to Appendix A. In fact, the DTFT spectrum of $\mathbf{x}((nA - b_p)T_s)$ is a weighted combination of the stretched and shifted versions of $\mathbf{X}(e^{j\omega})$.

For ease of exploration, it will be useful to consider that the K sensors are sampling at the same (sub-Nyquist) rate. This can be done by creating further decimated samples from $\boldsymbol{\zeta}_p[n]$. In particular, we define the constants:

$$A \triangleq \mathsf{LCM}\{A_1, ..., A_P\} \text{ and } Q_p \triangleq A/A_p \tag{7}$$

and decompose $\boldsymbol{\zeta}_p[n]$ into Q_p streams of samples:

$$\boldsymbol{\zeta}_p^q[n] \triangleq \boldsymbol{\zeta}_p[Q_p n - q],\ q = 0, 1, ..., Q_p - 1. \tag{8}$$

Each of $\boldsymbol{\zeta}_p^q[n]$ is a sequence of samples of $\boldsymbol{\zeta}_p(t)$ downsampled by A and offsetted by $qA_p + b_p$ unit of time. Applying Observation 1 to (4) and (8) gives:

$$\begin{aligned}\mathbf{Z}_p^q(e^{j\omega}) &= \frac{1}{A}\times \\ &\sum_{a=0}^{A-1} e^{-j(b_p + qA_p)\Omega_A^a(\omega)} \mathbf{H}_p \mathbf{X}\left(e^{j\Omega_A^a(\omega)}\right) + \mathbf{V}_p(e^{j\omega})\end{aligned} \tag{9}$$

for $\omega \in (-\pi, \pi]$. Here, $\mathbf{X}(e^{j\omega}) = \sum_n \mathbf{x}[n]e^{-j\omega n}$ and $\mathbf{V}_p^q(e^{j\omega}) = \sum_n \boldsymbol{v}_p^q[n]e^{-j\omega n}$ are the DTFT spectrum of $\mathbf{x}[n]$ and $\boldsymbol{v}_p^q[n]$, respectively.

To simplify (9), we define the following extended state spectrum:

$$\mathbf{X}_A(\omega) \triangleq \left[\mathbf{X}\left(e^{j\Omega_A^0(\omega)}\right)^T \mathbf{X}\left(e^{j\Omega_A^1(\omega)}\right)^T \ \dots \ \mathbf{X}\left(e^{j\Omega_A^{A-1}(\omega)}\right)^T\right]^T \tag{10}$$

and the extended measurement matrix, i.e.,

$$\boldsymbol{\mathcal{H}}_p(b_p, \omega) \triangleq \boldsymbol{\Theta}_p(b_p, \omega) \otimes \mathbf{H}_p \tag{11}$$

where

$$\boldsymbol{\Theta}_p(b_p, \omega) \triangleq \frac{1}{A} \times \begin{bmatrix} e^{-j(b_p)\Omega_A^0(\omega)} & \dots & e^{-j(b_p)\Omega_A^{A-1}(\omega)} \\ e^{-j(b_p+A_p)\Omega_A^0(\omega)} & \dots & e^{-j(b_p+A_p)\Omega_A^{A-1}(\omega)} \\ \dots & \dots & \dots \\ e^{-j(b_p+(Q_p-1)A_p)\Omega_A^0(\omega)} & \dots & e^{-j(b_p+(Q_p-1)A_p)\Omega_A^{A-1}(\omega)} \end{bmatrix}.$$

Then, (9) can be conveniently expressed as:

$$\boldsymbol{\mathcal{Z}}_p(e^{j\omega}) = \boldsymbol{\mathcal{H}}_p(b_p, \omega)\mathbf{X}_A(\omega) + \boldsymbol{\mathcal{V}}_p(e^{j\omega}), \tag{12}$$

for $\omega \in (-\pi, \pi]$, where $\boldsymbol{\mathcal{Z}}_p(e^{j\omega})$ and $\boldsymbol{\mathcal{V}}_p(e^{j\omega})$ can be formed by vertically concatenating the vectors $\{\mathbf{Z}_p^q(e^{j\omega})\}_q$ and $\{\mathbf{V}_p^q(e^{j\omega})\}_q$, respectively.

Observe that there is a one-to-one correspondence between $\mathbf{X}_A(\omega)$ and $\mathbf{X}(e^{j\omega})$ since each entry in the extended spectrum $\mathbf{X}_A(\omega)$ is non-repeating as the intervals $\Omega_A^a((-\pi, \pi])$ and $\Omega_A^b((-\pi, \pi])$ are disjoint whenever $a \neq b$. Consequently, estimating $\mathbf{X}_A(\omega)$ is equivalent to estimating the time domain sequence $\{\mathbf{x}[n]\}_n$. The latter can be obtained by first converting $\mathbf{X}_A(\omega)$ to $\mathbf{X}(e^{j\omega})$, and then performing an inverse DTFT.

To conclude, we observe that the measured spectrum $\boldsymbol{\mathcal{Z}}_p(e^{j\omega})$ can be expressed as a linear transformation of $\mathbf{X}_A(\omega)$. In the sequel, we will study the model (12) from two different aspects — i) to derive a set of identifiability conditions such that we can uniquely identify b_p and $\mathbf{X}_A(\omega)$ from $\boldsymbol{\mathcal{Z}}_p(e^{j\omega})$; ii) to propose a tractable, decentralized algorithm for retrieving b_p and $\mathbf{X}_A(\omega)$.

III. Identifiability Condition

This section derives an identifiability condition for (12). Recall that our intention is to estimate *jointly* the time offsets and the state spectrum, i.e., the tuple $(\mathbf{X}_A(\omega), \mathbf{b})$ with $\mathbf{b} = (b_1, b_2, ..., b_K)$, from the linear system (12). Under such context, we define:

Definition 1. *The sensing system* $\{\mathbf{H}_p\}_{p=1}^{P}$ *is said to be identifiable*[1] *under non-ideal sampling if and only if for any* $\{\boldsymbol{\mathcal{Z}}_p(e^{j\omega})\}_p$ *that is generated by* $(\mathbf{X}_A(\omega), \mathbf{b})$*, the tuple* $(\mathbf{X}_A(\omega), \mathbf{b})$ *is the only one satisfying* (12).

In other words, if the system is *identifiable*, then one can recover the tuple $(\mathbf{X}_A(\omega), \mathbf{b})$ unambiguously from $\{\boldsymbol{\mathcal{Z}}_p(e^{j\omega})\}_p$. As a comment, blind identification conditions are explored in the absence of noise, see e.g., [30], [31]. Our main result is summarized by the following proposition:

Proposition 1. *Consider the following matrix:*

$$\hat{\boldsymbol{\mathcal{H}}}(\hat{\mathbf{b}}, \mathbf{b}, \omega) \triangleq \left[\overline{\boldsymbol{\mathcal{H}}}(\hat{\mathbf{b}}, \omega) \quad -\overline{\boldsymbol{\mathcal{H}}}(\mathbf{b}, \omega)\right], \tag{13}$$

where $\mathbf{b} = [b_1 \ b_2 \ \dots \ b_P]$ *and*

$$\overline{\boldsymbol{\mathcal{H}}}(\mathbf{b}, \omega) \triangleq \begin{bmatrix} \boldsymbol{\mathcal{H}}_1(b_1, \omega) \\ \dots \\ \boldsymbol{\mathcal{H}}_P(b_P, \omega) \end{bmatrix}. \tag{14}$$

Assuming that the measurement is noiseless, i.e., $\boldsymbol{\mathcal{V}}_p(e^{j\omega}) = \mathbf{0}$*,* $\mathbf{X}_A(\omega) \neq \mathbf{0}$ *and* $\mathrm{rank}(\overline{\boldsymbol{\mathcal{H}}}(\mathbf{b}, \omega)) = AN$ *for all* ω*. Then, the sensing system* $\{\mathbf{H}_p\}_p$ *is identifiable if and only if*

$$\mathrm{rank}(\hat{\boldsymbol{\mathcal{H}}}(\hat{\mathbf{b}}, \mathbf{b}, \omega)) = 2AN. \tag{15}$$

for all ω *whenever* $\hat{\mathbf{b}} \neq \mathbf{b}$ *(with* $b_1 = \hat{b}_1 = 0$*).*

Proof. Notice that any tuple $(\hat{\mathbf{X}}_A(\omega), \hat{\mathbf{b}})$ satisfying (12) with $\boldsymbol{\mathcal{V}}_p(e^{j\omega}) = \mathbf{0}$ must fulfill the homogeneous equation:

$$\hat{\boldsymbol{\mathcal{H}}}(\hat{\mathbf{b}}, \mathbf{b}, \omega) \begin{bmatrix} \hat{\mathbf{X}}_A(\omega) \\ \mathbf{X}_A(\omega) \end{bmatrix} = \mathbf{0}, \tag{16}$$

since $\boldsymbol{\mathcal{Z}}_p(e^{j\omega}) = \boldsymbol{\mathcal{H}}_p(b_p, \omega)\mathbf{X}_A(e^{j\omega})$ for all p and ω.

We first prove the sufficient condition. If $\hat{\boldsymbol{\mathcal{H}}}(\hat{\mathbf{b}}, \mathbf{b}, \omega)$ is full column rank whenever $\hat{\mathbf{b}} \neq \mathbf{b}$, then (16) implies that whenever $\mathbf{b} \neq \hat{\mathbf{b}}$, we have $\mathbf{X}_A(\omega) = \hat{\mathbf{X}}_A(\omega) = \mathbf{0}$, leading to a contradiction. As such, $\mathbf{b} = \hat{\mathbf{b}}$ and the assumption $\mathrm{rank}(\overline{\boldsymbol{\mathcal{H}}}(\mathbf{b}, \omega)) = AN$ guarantees that $\hat{\mathbf{X}}_A(\omega) = \mathbf{X}_A(\omega)$.

For the necessary condition, let us assume that $\hat{\boldsymbol{\mathcal{H}}}(\hat{\mathbf{b}}, \mathbf{b}, \omega)$ is not full rank for some $\hat{\mathbf{b}} \neq \mathbf{b}$. In this case, for some instances of $\mathbf{X}_A(\omega)$ there exists a tuple $(\hat{\mathbf{X}}_A(\omega), \hat{\mathbf{b}}) \neq (\mathbf{X}_A(\omega), \mathbf{b})$ such that (16) is satisfied. This contradicts the uniqueness of $(\mathbf{X}_A(\omega), \mathbf{b})$. **Q.E.D.**

[1] Identifiability is interchangeable here with the notion of observability that is prevalent in the control theory terminology, e.g., [3].

Proposition 1 provides the identifiability condition for which the joint recovery of both sampling offsets and state spectrum are possible. However, verifying the condition that $\text{rank}(\hat{\boldsymbol{\mathcal{H}}}(\hat{\mathbf{b}}, \mathbf{b}, \omega)) = 2AN$ for all ω is a non-trivial problem. In the following, we derive several insightful conditions for identifiability which are easy to verify.

We begin by examining an assumption made in Proposition 1 that $\text{rank}(\overline{\boldsymbol{\mathcal{H}}}(\mathbf{b}, \omega)) = AN$. This assumption is equivalent to enforcing identifiability on the system with *known* time offsets, which is necessitated by:

$$\sum_{p=1}^{P} \min\{Q_p \text{rank}(\mathbf{H}_p), AN_c(\mathbf{H}_p)\} \geq AN, \tag{17}$$

where $N_c(\mathbf{H}_p) \leq N$ is the number of non-zero columns in $\mathbf{H}_p$. The above condition is obtained by upper bounding $\overline{\boldsymbol{\mathcal{H}}}(\mathbf{b}, \omega)$ by the sum of rank of its sub-matrices, $\sum_p \text{rank}(\boldsymbol{\mathcal{H}}_p(b_p, \omega))$. Furthermore, we have:

$$\begin{aligned} \text{rank}(\boldsymbol{\mathcal{H}}_p(b_p, \omega)) &= \text{rank}(\boldsymbol{\Theta}_p(b_p, \omega) \otimes \mathbf{H}_p) \\ &= \min\{Q_p \text{rank}(\mathbf{H}_p), AN_c(\mathbf{H}_p)\}, \end{aligned} \tag{18}$$

where the second equality is due to the Vandermonde structure in $\boldsymbol{\Theta}_p(b_p, \omega)$ [32].

Another interesting observation is that if b_p's are not distinct, then the upper bound (17) is loose. To see why, we suppose that $b_p = b_q$, then the rank of the submatrix $[\boldsymbol{\mathcal{H}}_p(b_p, \omega)^T \ \boldsymbol{\mathcal{H}}_q(b_q, \omega)^T]^T = \boldsymbol{\Theta}_p(b_p, \omega) \otimes [\mathbf{H}_p^T \ \mathbf{H}_q^T]^T$ is upper bounded by

$$\text{rank}(\boldsymbol{\Theta}_p(b_p, \omega) \otimes [\mathbf{H}_p^T \ \mathbf{H}_q^T]^T) \leq Q_p N.$$

This is because the rank of $[\mathbf{H}_p^T \ \mathbf{H}_q^T]^T$ is upper bounded by N. Such observation suggests that the existence of time offsets may be beneficial, especially when the sensors are down-sampling.

Using the similar reasoning as in (17), we conclude that the identifiability condition in Proposition 1 is necessitated by:

Corollary 1. *The condition* $\text{rank}(\hat{\boldsymbol{\mathcal{H}}}(\hat{\mathbf{b}}, \mathbf{b}, \omega)) = 2AN$ *is fulfilled only if*

$$\sum_{p=1}^{P} \min\{Q_p \text{rank}(\mathbf{H}_p), AN_c(\mathbf{H}_p)\} \geq 2AN \tag{19}$$

provided that the sampling offsets b_p *are distinct.*

Eq. (19) provides a guideline for the deployment of sensors in a robust sensing system. In particular, it suggests that

$$\sum_{p=1}^{P} \min\{Q_p M_p, AN_c(\mathbf{H}_p)\} \geq 2AN. \tag{20}$$

We remark that (19) is only a necessary condition. However, our numerical experiments suggest that when $\mathbf{H}_p$ is a Gaussian random matrix and (20) is satisfied, then the identifiability condition in Proposition 1 holds with high probability.

IV. Consensus on State and Time

This section studies a method to recover the state spectrum and time offset in a decentralized fashion. We treat each sensor, indexed by p as in the previous discussions, as a computing node that processes the knowledge of $\mathbf{H}_p$ and $\{\boldsymbol{\zeta}_p[n]\}_{n=0}^{Q_pL-1}$, where L is a designated frame size. The sensors are connected through a communication network described by $G=(\mathcal{V},\mathcal{E})$. The goal is to estimate the unknowns (the state spectrum and time offset) while the sensors perform local computations. The sensors communicate with their neighbors only occasionally.

Our first step is to formulate the regression problem for estimating $\{b_p\}_p$ and $\mathbf{X}_A(\omega)$ in (12). As the sensors possesses only finite-length data $\{\boldsymbol{\zeta}_p[n]\}_{n=0}^{Q_pL-1}$, we resort to taking approximation and discretizing the both sides of Eq. (12) by a K-point discrete Fourier transform (DFT), i.e.,

$$\boldsymbol{\mathcal{Z}}_p[k] \approx \boldsymbol{\mathcal{H}}_p(b_p,\omega_k)\mathbf{X}_A(\omega_k)+\boldsymbol{\mathcal{V}}_p[k], \tag{21}$$

where $\omega_k \triangleq 2\pi(k-K+1)/(K)$ is the frequency that the kth DFT point is related to. The approximation above is exact when $\boldsymbol{z}_p[n]$ is a periodic sequence with length L. For general signals, the approximation error decays as $\mathcal{O}(L^{-1})$ [33][2]. The K-point DFT spectrum of $\boldsymbol{\mathcal{Z}}_p[k]$ is obtained from L samples of $\boldsymbol{z}_p[n]$ as:

$$\boldsymbol{\mathcal{Z}}_p[k] = \textstyle\sum_{n=0}^{L-1} \boldsymbol{z}_p[n]e^{-j\omega_k n}, \tag{22}$$

where $\boldsymbol{z}_p[n] \triangleq [\boldsymbol{\zeta}_p^0[n]^T \ \ldots \ \boldsymbol{\zeta}_p^{Q_p-1}[n]^T]^T$ (cf. (8)) is a concatenation of the Q_p streams of data and K satisfies $L \le K$. Notice that $\{\mathbf{X}_A(\omega_k)\}_{k=0}^{K-1}$ is a re-ordering of $\{\mathbf{X}(e^{j\tilde{\omega}_k})\}_k$ where $\tilde{\omega}_k \triangleq 2\pi(k-AK+1)/(AK)$. The latter corresponds to the state sequence $\{\mathbf{x}[n]\}_{n=0}^{AL-1}$ over the finite length's frame of interest.

The discretized noise spectrum $\boldsymbol{\mathcal{V}}_p[k]$ is Gaussian and white. Furthermore, if $K=L$, then $\boldsymbol{\mathcal{V}}_p[k]$ are independent. As such, the maximum likelihood estimation problem of $\mathbf{X}_A(\omega_k)$ and b_p can be given as:

$$\begin{aligned} \min_{\substack{\hat{\mathbf{X}}_A(\omega_k),\forall k,\\ \hat{b}_p,\forall p}} & \sum_{p=1}^{P}\sum_{k=0}^{K-1} \left\| \boldsymbol{\mathcal{Z}}_p[k]-\boldsymbol{\mathcal{H}}_p(\hat{b}_p,\omega_k)\hat{\mathbf{X}}_A(\omega_k)\right\|_2^2 \\ \text{s.t. } & \hat{b}_p \in \mathcal{B}_p, \ \forall\ p. \end{aligned} \tag{23}$$

[2]In [33, Lemma 3.4.1], it was shown that $\|\boldsymbol{\mathcal{Z}}_p[k]-\boldsymbol{\mathcal{H}}_p(b_p,\omega_k)\mathbf{X}_A(e^{j\omega_k})-\boldsymbol{\mathcal{V}}_p[k]\|$ is finitely bounded, while the squared magnitude of $\boldsymbol{\mathcal{Z}}_p[k]$ grows as $\mathcal{O}(L)$. The error is thus $\mathcal{O}(L^{-1})$.

The next step is to develop a decentralized algorithm for (23).

A. Decentralized optimization with nuisance parameter

Note that (23) is a non-convex optimization problem. In particular, it can be regarded as a regression problem with local, nuisance parameter $\{\hat{b}_p\}_p$. To this end, a natural approach for tackling the problem is to apply an alternating optimization (AO) strategy, which works by alternating between the updates of $\{\hat{\mathbf{X}}_A(\omega_k)\}_k$ and $\hat{\mathbf{b}}$. To fix ideas, we let $t \in \mathbb{N}$ be the iteration index and define the objective function in (23) as $f(\{\hat{\mathbf{X}}_A(\omega_k)\}_k, \hat{\mathbf{b}})$. A practical AO strategy can be described by the following recursion:

$$\begin{aligned} \{\mathbf{X}_A^{(t+1)}(\omega_k)\}_k &\leftarrow \arg\min_{\hat{\mathbf{X}}_A(\cdot)} f(\{\hat{\mathbf{X}}_A(\omega_k)\}_k, \mathbf{b}^{(t)}), \\ \mathbf{b}^{(t+1)} &\leftarrow \mathcal{P}_{\mathcal{B}}\big(\mathbf{b}^{(t)} - \beta \nabla_{\mathbf{b}} f(\{\mathbf{X}_A^{(t)}(\omega_k)\}_k, \mathbf{b}^{(t)})\big), \end{aligned} \tag{24}$$

where $\mathcal{P}_{\mathcal{B}}(\cdot)$ is the projection onto the convex set $\mathcal{B}_1 \times \cdots \times \mathcal{B}_P$ and $\beta > 0$ is a step size. Notice that the optimization of $\mathbf{b}$ is taken care by a projected gradient descent update instead of an exact minimization.

There are several motivations for us to apply AO to (23). First, we observe that update of the time offsets $\mathbf{b}^{(t+1)}$ can be computed locally, which is due to the fact that $f(\{\mathbf{X}_A(\omega_k)\}_k, \mathbf{b})$ can be written as a summation of P functions, each of them depending only on b_p and is known to the pth sensor. Second, the optimization of state spectrum $\{\mathbf{X}_A(\omega_k)\}_k$ is equivalent to solving a standard linear least square problem. The latter admits a closed form solution. Lastly, the recursion (24) can be analyzed as a special case of the Block Successive Minimization Method in [34]. In particular, the recursion is shown to converge to a stationary point of (23).

In the interest of tackling (23) distributively, we see that the first update in (24) involves data from all sensors. In particular,

$$\begin{aligned} \mathbf{X}_A^{(t+1)}(\omega_k) &= \mathbf{X}_{A,k}^{\star}(\mathbf{b}^{(t)}) \\ &= \Big(\overline{\boldsymbol{\mathcal{H}}\boldsymbol{\mathcal{H}}}(\mathbf{b}^{(t)}, \omega_k)\Big)^{-1} \overline{\boldsymbol{\mathcal{H}}\boldsymbol{\mathcal{Z}}}(\mathbf{b}^{(t)}, \omega_k), \end{aligned} \tag{25}$$

with

$$\overline{\boldsymbol{\mathcal{H}}\boldsymbol{\mathcal{H}}}(\mathbf{b}^{(t)}, \omega_k) \triangleq (1/P) \textstyle\sum_{p=1}^{P} \boldsymbol{\mathcal{H}}_p(b_p^{(t)}, \omega_k)^H \boldsymbol{\mathcal{H}}_p(b_p^{(t)}, \omega_k), \tag{26}$$

$$\overline{\boldsymbol{\mathcal{H}}\boldsymbol{\mathcal{Z}}}(\mathbf{b}^{(t)}, \omega_k) \triangleq (1/P) \textstyle\sum_{p=1}^{P} \boldsymbol{\mathcal{H}}_p(b_p^{(t)}, \omega_k)^H \boldsymbol{\mathcal{Z}}_p[k], \tag{27}$$

where $(\cdot)^H$ denotes Hermitian transpose. As seen, (25) requires the knowledge of $\overline{\boldsymbol{\mathcal{H}}\boldsymbol{\mathcal{H}}}(\cdot)$, $\overline{\boldsymbol{\mathcal{H}}\boldsymbol{\mathcal{Z}}}(\cdot)$, which is not available to the individual sensors.

There are a number of decentralized algorithm that can be applied to solve the linear least square problem in (24). To list a few options, the diffusion-based LMS and RLS algorithm are proposed in [11], [12], respectively; the ADMM method is applied in [4], etc.

It is preferred to apply a decentralized algorithm with fast convergence. In this regard, we propose to directly compute (25) in a decentralized fashion. The idea is to leverage on the fact that $\overline{\boldsymbol{\mathcal{H}}\boldsymbol{\mathcal{H}}}(\cdot)$ and $\overline{\boldsymbol{\mathcal{H}}\boldsymbol{\mathcal{Z}}}(\cdot)$ can be replaced by the averages of the collection $\{\boldsymbol{\mathcal{H}}_p(\cdot)^H\boldsymbol{\mathcal{H}}_p(\cdot)\}_p$ and $\{\boldsymbol{\mathcal{H}}_p(\cdot)^H\boldsymbol{\mathcal{Z}}_p[k]\}_p$, respectively. As a result, the computation of (25) can be treated as a decentralized averaging problem. To compute the averages, we apply the Gossip-based average consensus (G-AC) protocol in [35], which is described in the following.

Let us take

$$\boldsymbol{y}_{p,k}^0 \triangleq \begin{bmatrix} \boldsymbol{\mathcal{H}}\boldsymbol{\mathcal{Z}}_p^0[k] \\ \boldsymbol{\mathcal{H}}\boldsymbol{\mathcal{H}}_p^0[k] \end{bmatrix} \triangleq \begin{bmatrix} \mathbf{H}_p(b_p^{(t)},\omega_k)^H\boldsymbol{\mathcal{Z}}_p[k] \\ \mathrm{vec}(\mathbf{H}_p(b_p^{(t)},\omega_k)^H\mathbf{H}_p(b_p^{(t)},\omega_k)) \end{bmatrix},$$

which is a $((AN)^2+AN)$-dimensional complex vector. It suffices to compute (25) by obtaining the average of $\{\boldsymbol{y}_{p,k}^0\}_p$, i.e.,

$$\overline{\boldsymbol{y}}_k = (1/P)\sum_{p=1}^P \boldsymbol{y}_{p,k}^0. \tag{28}$$

The G-AC protocol achieves $\bar{\boldsymbol{y}}_k$ by performing the following recursions:

$$\boldsymbol{y}_{p,k}^\ell = \textstyle\sum_{q\in\mathcal{N}_p} W_{pq}^\ell \boldsymbol{y}_{q,k}^{\ell-1}, \tag{29}$$

where $\mathcal{N}_p \subseteq \mathcal{V}$ denotes the set of neighbors of sensor p. To guarantee convergence, the mixing matrix $\mathbf{W}^\ell = [W_{pq}^\ell]_{p,q}$ satisfies a certain set of mild conditions, e.g., it is required to be *doubly stochastic*, i.e., $\mathbf{W}^\ell\mathbf{1} = \mathbf{1}$ and $\mathbf{1}^T\mathbf{W}^\ell = \mathbf{1}^T$. For a more detailed discussion, see [35].

As seen in (29), at each G-AC step $\ell \in \mathbb{N}$, the sensor p only obtains information from its immediate neighbors, i.e., $q \in \mathcal{N}_p$. Moreover, as $\mathbf{W}^\ell$ can be time-varying, only a subset of links $\mathcal{E}^\ell \subseteq \mathcal{E}$ are required to be active at each G-AC step. The G-AC method requires only *local computation* and it allows *random* communication between the agents. Finally, the variable $\mathbf{X}_{p,A,k}^{\ell_t}(\mathbf{b}^{(t)})$ is computed using the *approximate* averages stored at the pth agent after ℓ_t G-AC steps, i.e.,

$$\mathbf{X}_{p,A,k}^{\ell_t}(\mathbf{b}^{(t)}) = \big(\mathrm{vec}^{-1}(\boldsymbol{\mathcal{H}}\boldsymbol{\mathcal{H}}_p^{\ell_t}[k])\big)^{-1}\big(\boldsymbol{\mathcal{H}}\boldsymbol{\mathcal{Z}}_p^{\ell_t}[k]\big). \tag{30}$$

Notice that we apply the G-AC protocol to compute approximates to all the K points in DFT spectrum $\mathbf{X}_{A,k}^\star(\mathbf{b}^{(t)})$.

Combining AO and G-AC results in a decentralized algorithm for (23). We call this algorithm the Gossip-based alternating optimization (G-AO) algorithm, as summarized in Algorithm 1.

B. Convergence analysis

To study the convergence of the G-AO algorithm, we first need to study the convergence rate of G-AC. As shown in [35], the recursion (29) converges to the true average vector $\bar{\boldsymbol{y}}_k$ under several

Algorithm 1 The G-AO algorithm for (23).

1: **Initialize:** $\{\{\mathbf{X}_{p,A}^{(0)}(\omega_k)\}_k\}_{p=1}^P$, $\{b_p^{(0)}\}_{p=1}^P$;

2: **for** $t = 0, 1, ...$ **do**

3: The network computes $\mathbf{X}_{p,A,k}^{\ell_t}(\mathbf{b}^{(t)})$ for each (p,k) using ℓ_t G-AC steps (cf. (30)).

4: **for** $p = 1, 2, ..., P$ **do**

5: Agent p updates its copies of $\mathbf{X}_{p,A}(\cdot)$ and b_p as:

$$\mathbf{X}_{p,A}^{(t+1)}(\omega_k) \leftarrow \mathbf{X}_{p,A,k}^{\ell_t}(\mathbf{b}^{(t)}),\ k = 0, ..., K-1,$$

$$b_p^{(t+1)} \leftarrow \mathcal{P}_{\mathcal{B}_p}\big(b_p^{(t)} - \beta \nabla_{b_p} f(\{\mathbf{X}_{p,A}^{(t+1)}(\omega_k)\}_k, \mathbf{b}^{(t)})\big)$$

As mentioned, the update of b_p can be performed distributively since f is separable.

6: **end for**

7: **end for**

8: **Return:** $\{\{\mathbf{X}_{p,A}^{(t+1)}(\omega_k)\}_k\}_{p=1}^P$, $\{b_p^{(t+1)}\}_{p=1}^P$.

assumptions on $\{\mathbf{W}^\ell\}_\ell$. In fact, the rate of convergence is exponential, i.e., $\|\boldsymbol{y}_{p,k}^\ell - \overline{\boldsymbol{y}}_k\|_2 = \mathcal{O}(\lambda_{\bar{W}}^\ell)$ where $0 < \lambda_{\bar{W}} < 1$ is the second largest eigenvalue of the matrix $\bar{\mathbf{W}} = \mathbb{E}_\ell\{\mathbf{W}^\ell\}$. Consequently, the accuracy on the approximation $\mathbf{X}_{p,A,k}^{\ell_t}(\mathbf{b}^{(t)})$ improves exponentially with ℓ_t:

Proposition 2. *Suppose that*

$$C_0 \cdot C_1 \cdot \lambda_W^{\ell_t} < 1, \tag{31}$$

where $C_0 = \max_{\mathbf{b},k} \|\sum_p \mathbf{H}_p(b_p, \omega_k)^H \mathbf{H}_p(b_p, \omega_k)\|$ *and* $C_1 = \max_{\mathbf{b},k} \|(\sum_p \mathbf{H}_p(b_p, \omega_k)^H \mathbf{H}_p(b_p, \omega_k))^{-1}\|$ *are finite constants, then the spectrum* $\mathbf{X}_{p,A,k}^{\ell_t}(\mathbf{b}^{(t)})$ *computed in* (30) *using* ℓ_t *G-AC steps satisfies:*

$$\big\|\mathbf{X}_{A,k}^\star(\mathbf{b}^{(t)}) - \mathbf{X}_{p,A,k}^{\ell_t}(\mathbf{b}^{(t)})\big\|_2 = \mathcal{O}(\lambda_{\bar{W}}^{\ell_t}),\ \forall\ p, k, \tag{32}$$

where ℓ_t *is the number of G-AC steps at the* t*th iteration.*

The proof is relegated to Appendix B. An important implication is that the approximation improves exponentially with ℓ_t, i.e., the number of G-AC steps per iteration.

We are ready to state the following theorem regarding the convergence of the G-AO algorithm.

Theorem 1. *Let* $(\{\mathbf{X}_{A,k}^\star\}_k, \mathbf{b}^\star)$ *be a local minimum to* (23)*. Suppose that* f *is* (m_o, M_o)*-strongly convex in the neighborhood* $\mathcal{N}_{R^\star}(\{\mathbf{X}_{A,k}^\star\}_k, \mathbf{b}^\star)$ *and it is Lipschitz continuous with constant* L_o*. Suppose that*

(31) holds, $\beta \leq 1/M_o$, $B = \max_t \|(\{\hat{\mathbf{X}}_{A,k}^{(t)}\}_k, \mathbf{b}^{(t)}) - (\{\mathbf{X}_{A,k}^{\star}\}_k, \mathbf{b}^{\star})\|_2 < \infty$ and $B \leq R^{\star}$, then we have:

$$\lim_{t\to\infty} \left\|(\{\hat{\mathbf{X}}_{A,k}^{(t)}\}_k, \mathbf{b}^{(t)}) - (\{\mathbf{X}_{A,k}^{\star}\}_k, \mathbf{b}^{\star})\right\|_2^2 = \sqrt{\mathcal{O}(\lambda_W^{\ell_{min}})}, \tag{33}$$

where $\ell_{min} = \min_t \ell_t$ is the minimum number of G-AC steps taken and $\hat{\mathbf{X}}_{A,k}^{(t)} \triangleq (1/P)\sum_{p=1}^{P} \mathbf{X}_{p,A}^{(t)}(e^{j\omega_k})$.

The proof is provided in Appendix C, which is based on studying the error dynamics of the G-AO algorithm as a second order dynamical system. In fact, Theorem 1 implies that if the G-AO algorithm stays close enough to a local minimum, then the algorithm converges to an approximate of that local minimum, where the approximation accuracy improves exponentially with ℓ_{min}.

The strong convexity assumption on f around a local minimum may appear restrictive at first. However, our numerical results indicate that Theorem 1 can accurately predict the performance of G-AO algorithm applied on (23).

V. Extensions & Applications

In this section, we extend the previous formulations/algorithms to the case with general measurement (e.g., non-linear). This model pertains to the case when the measured signal at the sensor is non-linearly related to the underlying system state. As an application example, we demonstrate that the formulation can be employed to robustify power system state estimation (PSSE) systems.

Considering Assumption 1 to 3 in Section II and the sampling architecture in Fig. 1, we can express the measured sample at sensor p as:

$$\begin{aligned} \boldsymbol{\zeta}_p[n] &= \overline{(\boldsymbol{h}_p \circ \mathbf{x}_c)}(T_s(nA_p - b_p)) + \boldsymbol{v}_p[n], \\ \overline{(\boldsymbol{h}_p \circ \mathbf{x}_c)}(t) &= \mathrm{LPF}\{(\boldsymbol{h}_p(\mathbf{x}_c(t))\}, \end{aligned} \tag{34}$$

where $\boldsymbol{h}_p : \mathbb{C}^N \to \mathbb{C}^{M_p}$ is a general measurement function and $\overline{(\boldsymbol{h}_p \circ \mathbf{x}_c)}(t)$ is the low-pass filtered version of $\boldsymbol{h}_p(\mathbf{x}_c(t))$ with a cutoff frequency at $W/2$ Hz. The reason for introducing the low-pass-filtered is explained in Remark 1.

Under the same spirit as in our previous developments, the next step is to consider the frequency domain equivalent to (34) via Observation 1. To simplify notations, we consider $A_p = A$ for all p in the following. The K-point DFT of $\boldsymbol{\zeta}_p[n]$ is given by (cf. (22)):

$$\begin{aligned} &\mathbf{Z}_p[k] \approx \frac{1}{A} \times \\ &\sum_{a=0}^{A-1} e^{-jb_p\Omega_A^a(\omega_k)} \sum_{n=0}^{AL-1} \boldsymbol{h}_p(\mathbf{x}[n]) e^{-j\Omega_A^a(\omega_k)n} + \mathbf{V}_p[k], \end{aligned} \tag{35}$$

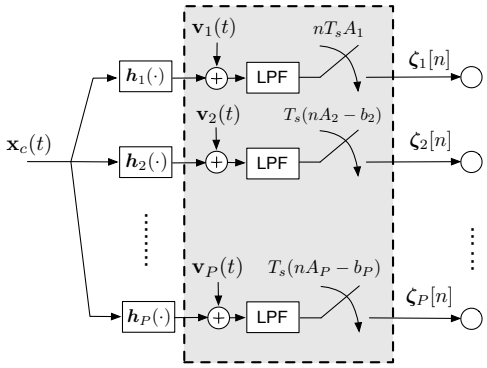

Fig. 1. The architecture of the sample & estimation system (cf. (4)). The low pass filters (LPF) have a cutoff frequency at $W/2$ Hz.

where we recall that $\omega_k = 2\pi(k-K+1)/K$. Analogous to the previous derivations, it follows that we can formulate the nonlinear regression problem as:

$$\min_{\{\hat{\mathbf{x}}[n]\}_{n=0}^{AL-1},\hat{b}_p,\forall p} \sum_{p=1}^{P}\sum_{k=0}^{K-1} \left\| \boldsymbol{g}_{p,k}\big(\{\hat{\mathbf{x}}[n]\}_{n=0}^{AL-1},\hat{b}_p\big)\right\|_2^2, \tag{36}$$

where

$$\begin{aligned}\boldsymbol{g}_{p,k}\big(\{\hat{\mathbf{x}}[n]\}_{n=0}^{AL-1},\hat{b}_p\big) \triangleq \mathbf{Z}_p[k] - \\ \frac{1}{A}\sum_{a=0}^{A-1} e^{-j\hat{b}_p\Omega_A^a(\omega_k)} \sum_{n=0}^{AL-1} \boldsymbol{h}_p(\hat{\mathbf{x}}[n]) e^{-jn\Omega_A^a(\omega_k)}.\end{aligned} \tag{37}$$

Notice that the above formulation allows us to consider hybrid measurement, i.e., $\boldsymbol{h}_p$ can be linear for some p.

Remark 1. We observe that in general $\boldsymbol{h}_p(\mathbf{x}_c(t))$ has a bandwidth higher than $W/2$ Hz due to the nonlinearity of $\boldsymbol{h}_p$. In particular, the measurement function cannot be analyzed separately in $\boldsymbol{h}_p(\mathbf{x}_c(t))$ as in the linear case. The low pass filter is introduced to remedy this such that Observation 1 can be applied. Moreover, it can be verified that

$$\overline{(\boldsymbol{h}_p \circ \mathbf{x}_c)}(nT_s) = \boldsymbol{h}_p(\mathbf{x}[n]),\ \forall\ n, \tag{38}$$

where the equality can be established by studying the spectrum of the both sides.

A. Extending the G-AO algorithm

Similar to (23), Problem (36) is also a non-convex problem. To develop a decentralized algorithm for (36), we apply a modified version of the G-AO algorithm. The modified algorithm is based on the Gossip-based Gauss Newton (GGN) method in [28] and the AO strategy.

We observe that the G-AO algorithm cannot be applied directly to (36). In fact, even when $\mathbf{b}$ is fixed in (36), the optimization problem remains non-convex and does not admit a closed form solution. We develop our algorithm by borrowing insights from the GGN method.

Let $t \in \mathbb{N}$ be the iteration index and assume that $\mathbf{b}^{(t)}$ is fixed. The damped GN direction at $\{\mathbf{x}^{(t)}[m]\}_m$ for $\mathbf{x}[n]$ is given as:

$$\begin{aligned}&\mathbf{d}_n(\{\mathbf{x}_p^{(t)}[m]\}_m, \mathbf{b}^{(t)}) = \\ &\left(\lambda_{GN}\mathbf{I} + \overline{\boldsymbol{\mathcal{G}}\boldsymbol{\mathcal{G}}}(\{\mathbf{x}_p^{(t)}[m]\}_m, \mathbf{b}^{(t)})\right)^{-1}\overline{\boldsymbol{\mathcal{G}}\boldsymbol{\mathcal{Z}}}(\{\mathbf{x}_p^{(t)}[m]\}_m, \mathbf{b}^{(t)}),\end{aligned} \tag{39}$$

where $\lambda_{GN} \geq 0$ is the damped GN parameter, $\overline{\boldsymbol{\mathcal{G}}\boldsymbol{\mathcal{G}}}$ and $\overline{\boldsymbol{\mathcal{G}}\boldsymbol{\mathcal{Z}}}$ are both defined in terms of the Jacobian matrix of $\boldsymbol{g}_{p,k}(\cdot)$:

$$\overline{\boldsymbol{\mathcal{G}}\boldsymbol{\mathcal{G}}}(\{\mathbf{x}_p[m]\}_m, \mathbf{b}) = \sum_{p=1}^{P}\sum_{k=0}^{K-1} \boldsymbol{G}_{p,k}^n(\cdot,\cdot)^H \boldsymbol{G}_{p,k}^n(\{\mathbf{x}[m]\}_m, b_p)$$

$$\overline{\boldsymbol{\mathcal{G}}\boldsymbol{\mathcal{Z}}}(\{\mathbf{x}_p[m]\}_m, \mathbf{b}) = \sum_{p=1}^{P}\sum_{k=0}^{K-1} \boldsymbol{G}_{p,k}^n(\cdot,\cdot)^H \boldsymbol{g}_{p,k}(\{\mathbf{x}[m]\}_m, b_p)$$

such that $\boldsymbol{G}_{p,k}^n(\{\mathbf{x}[m]\}_m, b_p)$ is the $N \times N$ Jacobian matrix of $\boldsymbol{g}_{p,k}(\cdot,\cdot)$ taken at $(\{\mathbf{x}[m]\}_m, b_p)$ with respect to the vector $\mathbf{x}[n]$. Notice that the state vector $\mathbf{x}[m]$ is updated in a Gauss-Seidal like manner [36].

Importantly, we observe that the GN direction can be calculated distributively using the G-AC protocol. This suggests us to combine the AO algorithm and the Gossip-GN algorithm to develop a decentralized algorithm for (36). In particular, the pseudo code for the GGN-AO algorithm for (36) is now summarized in Algorithm 2, where we have denoted the objective function in (36) as $f_{NL}(\{\mathbf{x}_p[n]\}_n, \mathbf{b})$.

In contrast to the G-AO algorithm studied in Section IV, the GGN-AO algorithm applied to (36) entails a higher complexity in general. For instance, the update of state variable $\mathbf{x}[n]$ in the GGN-AO algorithm is based on the GN method, which is an iterative method by nature. In contrast, the G-AO's counterpart of the update relies on a closed form solution. Intuitively, the G-AO algorithm will exhibit a faster convergence rate.

B. Application: Robustifying the PSSE systems

This subsection applies the models developed in this paper to robustify the power system state estimation (PSSE) systems. In particular, we consider the problem of PSSE in wide area measurement systems, where data from Phasor Measurement Units (PMUs) and the legacy Supervisory Control And Data Acquisition (SCADA) systems are combined to provide the state estimates.

Algorithm 2 The GGN-AO algorithm for (36).

Initialize: $\{\{\mathbf{x}_p^{(0)}[n]\}_n\}_{p=1}^P$, $\{b_p^{(0)}\}_{p=1}^P$;
for $t = 0, 1, ...$ **do**
  Update: $\mathbf{x}_p^{(t+1)}[n] \leftarrow \mathbf{x}_p^{(t)}[n]$ for all $n = 0, ..., AL-1$.
  for $n = 0, 1, ..., AL-1$ **do**
    The network computes the (approximate) GN direction $\mathbf{d}_{p,n}^{\ell_t}(\{\mathbf{x}_p^{(t+1)}[m]\}_m, \mathbf{b}^{(t)})$ for each p using ℓ_t G-AC steps (cf. (39)).
    for $p = 1, 2, ..., P$ **do**
      Agent p updates its copies of $\mathbf{x}_p[\cdot]$ and b_p as:

$$\mathbf{x}_p^{(t+1)}[n] \leftarrow \mathbf{x}_p^{(t)}[n] - \mathbf{d}_{p,n}^{\ell_t}(\{\mathbf{x}_p^{(t)}[m]\}_m, \mathbf{b}^{(t)}).$$

    end for
  end for
  for $p = 1, 2, ..., P$ **do**
    Agent p updates b_p by:

$$b_p^{(t+1)} \leftarrow \mathcal{P}_{\mathcal{B}_p}\big(b_p^{(t)} - \beta \nabla_{b_p} f_{NL}(\{\mathbf{x}_p^{(t+1)}[n]\}_n, \mathbf{b}^{(t)})\big)$$

  end for
end for
Return: $\{\{\mathbf{x}_p^{(t+1)}[n]\}_n\}_{p=1}^P$, $\{b_p^{(t+1)}\}_{p=1}^P$.

The system's state of interest is the complex envelope $\mathbf{x}_c^e(t)$ of the voltage on each bus, which is related to the actual voltage on bus i by $[\mathbf{x}_c(t)]_i = \Re\left\{[\mathbf{x}_c^e(t)]_i e^{j\omega_0 t}\right\}$, where ω_0 is the operating frequency (typically 60/50 Hz) of the power grid. Typically $\mathbf{x}_c^e(t)$ has a smaller bandwidth than 60 Hz.

There are two types of sensors/systems that are used in the power grid:

PMU — The PMU installed on bus i takes samples of the voltage phasor $[\mathbf{x}_c^e(t)]_i$ and the current flow phasors $I_{ij}^e(t)$ on branches that are connected to bus i. Typically, the PMU relies on a local oscillator, synchronized using GPS clocks, to take samples of $[\mathbf{x}_c^e(t)]_i$ and $I_{ij}^e(t)$ [37]. The GPS clocks may be tampered in the case of an attack. As such, we model the voltage measurements obtained at the PMU on bus i as:

$$V_i[n] = [\mathbf{x}_c^e[n]]_i = [\mathbf{x}_c^e(nT_s - b_p)]_i e^{-j\omega_0 b_p} + v_i[n], \tag{40}$$

where $v_i[n]$ is the measurement noise, b_p represents the sampling offset, and the current measurements:

$$\begin{aligned} I_{ij}^e[n] &= e^{-j\omega_0 b_p} Y_{ij} \times \\ &\big([\mathbf{x}_c^e(nT_s - b_p)]_i - [\mathbf{x}_c^e(nT_s - b_p)]_j\big) + v_{ij}[n], \end{aligned} \tag{41}$$

where Y_{ij} is the admittance of branch (i, j). Notice that we have assumed Nyquist sampling, i.e., $A_p = 1$, as the PMU has a sampling rate of 10-30 Hz [37].

Let $\mathcal{I}_p \subseteq \{1, ..., N\}$ be the buses included in the pth sensing site. By stacking $\{V_i[n]\}_{i\in\mathcal{I}_p}$ and $\{I_{ij}[n]\}_{j\in\mathcal{N}_i, i\in\mathcal{I}_p}$ vertically as $\boldsymbol{\zeta}_p[n]$, where $\mathcal{N}_i$ denotes the buses connected to bus i, the samples obtained at the pth PMUs' site is modelled as:

$$\boldsymbol{\zeta}_p[n] = \mathbf{H}_p \mathbf{x}_c^e(nT_s - b_p) e^{-j\omega_0 b_p} + \boldsymbol{v}_p[n]. \tag{42}$$

SCADAs — The SCADA installed on bus i samples on the injected complex power $S_i(t)$ as well as the complex power $S_{ij}(t)$ that flows to/from bus i. For example, the power flow phasor through branch (i, j) is given by

$$S_{ij}(t) = [\mathbf{x}_c^e(t)]_i \big([\mathbf{x}_c^e(t)]_i - [\mathbf{x}_c^e(t)]_j\big)^* Y_{ij}^*. \tag{43}$$

Similarly, the injected power is

$$S_i(t) = |[\mathbf{x}_c^e(t)]_i|^2 Y_i^S + \textstyle\sum_{j\in\mathcal{N}_i} S_{ij}(t) \tag{44}$$

where Y_i^S is the shunt admittance from bus i to the ground. In fact, these quantities are obtained by measuring the active power and reactive power. Besides the measurement models, another factor that differentiates SCADA systems from PMUs is the sample rate used. In fact, the sampling rate is only 0.2–0.5 Hz [38] for SCADA. In light of this, we stack the relevant $\{S_i(t)\}_{i\in\mathcal{I}_p}$ and $\{S_{ij}(t)\}_{j\in\mathcal{N}_i, i\in\mathcal{I}_p}$ to form the $\boldsymbol{h}_p(\mathbf{x}_c(t))$ in (34). Under the assumption of an LPF and using an appropriate choice of A_p, we see that the model described in (35) applies to the sampled SCADA data.

We observe that both of the models for PMUs and SCADA fit into the descriptions in (34). Subsequently, the *robust* PSSE problem can be formulated in a similar manner as (36).

Remark 2. Assume the case with only asynchronous PMUs. If $b_p \ll T_s$, we can assume that the power system state is stable relative to the time offsets in PMUs, i.e., we have $\mathbf{x}_c^e(nT_s - b_p) \approx \mathbf{x}_c^e(nT_s)$, the regression problem for $\mathbf{x}_c^e(nT_s)$ and b_p can be formulated as a special case of (23):

$$\min_{\hat{\mathbf{x}}_c^e(nT_s), \{\hat{b}_p\}} \sum_{p=1}^{K} \big\|\boldsymbol{\zeta}_p[n] - \mathbf{H}_p \hat{\mathbf{x}}_c^e(nT_s) e^{-j\omega_0 \hat{b}_p}\big\|_2^2. \tag{45}$$

The G-AO algorithm can then be applied to tackle (45).

VI. Numerical results

To highlight different aspects affecting the performance of the techniques for joint state and sampling offset estimation, this section will be divided into two parts — i) the first part includes simulations pertaining the linear model in Section II with sensing matrices that are complex Gaussian zero mean i.i.d. coefficients (e.g. Rayleigh fading); ii) the second part is focused on the application to PSSE problem with a non-linear measurement model.

We first explain how the non-ideally sampled measurements are generated. Notice that in (4), the discrete-time measurement $\boldsymbol{\zeta}_p[n]$ is equivalent to the infinite sum $\sum_{m=-\infty}^{\infty} \boldsymbol{\zeta}_p((n-m)T_s)\sin(\pi(A_p m - b_p))/(\pi(A_p m - b_p))$. It is impossible to evaluate the infinite sum, we thus truncate the latter by a finite sum from $m = 0$ to $m = AL - 1$. In this way, note that $\{\boldsymbol{\zeta}_p[n]\}_{n=0}^{Q_p L-1}$ retains the information from $\{\mathbf{x}[n]\}_{n=0}^{AL-1}$.

Moreover, our numerical experience shows that it is necessary to apply a pre-processing window to obtain the measured spectrum in (22), so as to reduce the modelling error introduced by discrete approximation. We instead take $\boldsymbol{Z}_p[k] = \sum_{n=0}^{L-1} w[n]\boldsymbol{z}_p[n]e^{-j\omega_k n}$, where $w[n]$ is the Blackman window [39]. Due to the windowing operation, the estimated state at the boundaries can be unreliable. As such, unless otherwise specified, we evaluate only the mean squared error (MSE) for the state $\mathbf{x}[n]$ in the middle of the frame. Specifically, the MSE is computed as the per sample error $\mathbb{E}[(1/(2G+1))\sum_{n=AL/2-G}^{AL/2+G} \|\mathbf{x}[n] - \hat{\mathbf{x}}[n]\|_2^2/N]$, where $\hat{\mathbf{x}}[n]$ is the estimated state and $G = \lfloor 0.15AL \rfloor$. The squared error for time offsets is calculated as $\mathbb{E}[\sum_{p=1}^{P}(b_p - \hat{b}_p)^2]$.

For the other simulation parameters, we fix $K = 192$ as the DFT size and $\sigma_w^2 = 10^{-2}$ as the noise variance. We perform 100 Monte-Carlo simulation trials to get the averages. The G-AO/GGN-AO algorithm is terminated when the relative decrease in objective value is less than 0.1%. The G-AO algorithm is initialized with $b_p = 0$ for all p. The communication network is generated as an Erdos-Renyi graph with parameter $p = 0.5$. We assume that the mixing matrix $\mathbf{W}$ is static with Metropolis-Hastings weight [40]. The error is evaluated as the maximum MSEs evaluated for each sensor.

A. Example: Rayleigh fading

In the following examples, we focus on the performance of proposeds method under the linear model specified in Section II. In particular, the states $\mathbf{x}[n]$ and measurement matrices $\mathbf{H}_p$ are generated as random vectors/matrices with unit variance i.i.d. complex Gaussian random entries.

Our first example considers a system with sub-Nyquist sampling, i.e., we set $A_p = 2$ for all p. The system dimensions are set as $M = 4, N = 8, P = 12$. The time offsets are uniformly drawn from

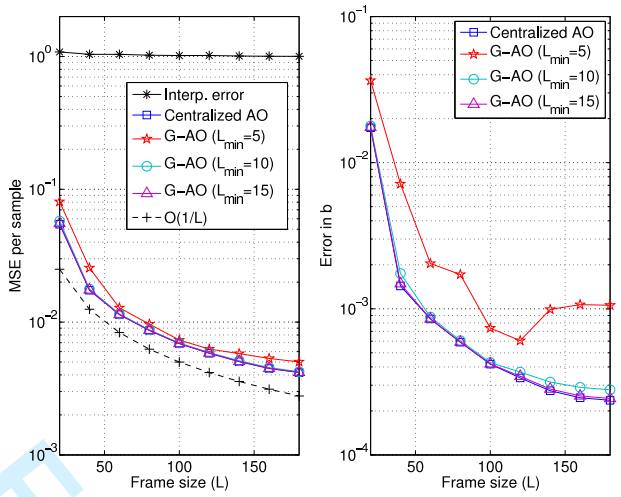

Fig. 2. Comparing the MSE performance against the frame size L. (Left) On estimating $\{\boldsymbol{x}[n]\}_n$. (Right) On estimating $\{b_p\}_p$.

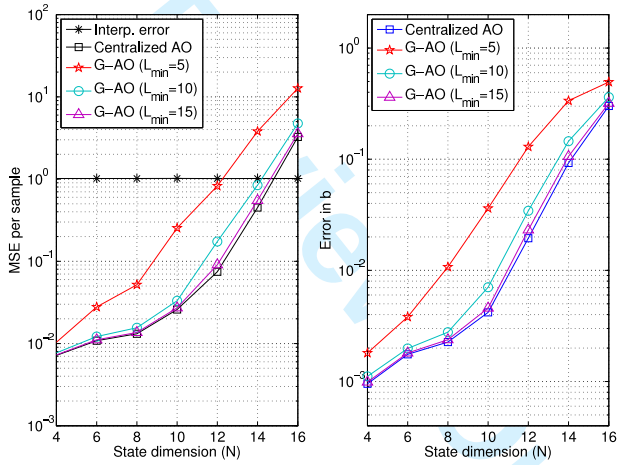

Fig. 3. Comparing the MSE performance against the state dimension N. (Left) On estimating $\{\boldsymbol{x}[n]\}_n$. (Right) On estimating $\{b_p\}_p$.

$\mathcal{B} = [-0.5, 0.5]$. Notice that under sub-Nyquist sampling, without exploiting the time offsets between the sensors, it is impossible to estimate the state vector $\mathbf{x}[n]$ for all n. Therefore, as a benchmark, we provide the MSE evaluated by comparing $\{\mathbf{x}[n]\}_n$ with an interpolated state sequence estimated from the sub-Nyquist measurements.

The simulation result from this example is depicted in Fig. 2, where we compare the MSE in state and in $\{b_p\}_p$ against the frame size L. From the figure, we see that the error metrics of the proposed algorithm decrease as L increases. It is due to the improved approximation to the true DTFT spectrum. In

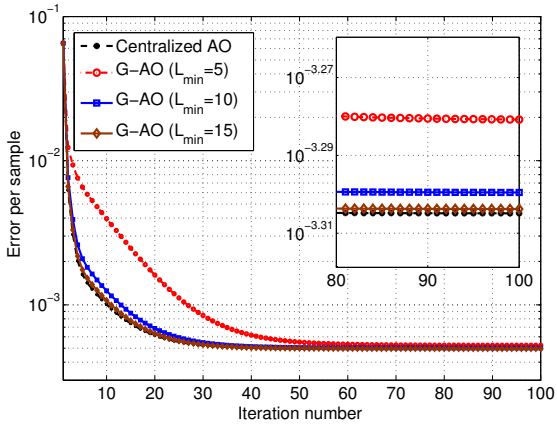

Fig. 4. Comparing the state estimation error against iteration number of the proposed algorithms. $M = 4, P = 12, N = 8, L = 120$.

fact, the MSE in state decays as $\mathcal{O}(L^{-1})$, coinciding with the discussions that follows (21). On the other hand, the G-AO algorithm achieves a similar performance with its centralized counterpart. Especially, as ℓ_{min} increases, the performance of the former approaches that of the latter. This observation is in line with the analysis results on G-AO from Theorem 1.

In the second example, we examine the identifiability condition in Section III. Specifically, the system parameters are $M = 3, P = 12, L = 120$ and the MSEs are compared with different state dimension N. Fig. 3 shows the result from this example. Recall that from Corollary 1, the identifiability condition is likely to be satisfied if $N \leq 9$. This is evident from the figure that the MSE increases significantly when $N \geq 10$. The discrepancy is due to the fact that Corollary 1 is derived based on a noiseless model and the G-AO algorithm may have been initialized close enough to the true optimum.

The next example, shown in Fig. 4, examines the convergence speed of the G-AO algorithm, for which we track the state estimation error as G-AO algorithm proceeds. In this example, we set $M = 4, P = 12, N = 8, L = 120$ and consider solving a randomly generated instance of (23). We observe that the error is gradually decreasing as the algorithm progresses and converges in about 30-40 iterations. Importantly, we see that the G-AO algorithm follows closely with its centralized counterpart, suggesting that they both achieve a similar performance.

Estimating the state via adaptive filtering — As an extension, we study a practical scheme for incorporating the proposed regression method into standard state estimators.

Assuming that the time offsets $\{\hat{b}_p\}_p$ are estimated with the G-AO algorithm in an earlier stage, our idea is to consider a *sub-optimal* adaptive filter, which reverts only the effects of non-ideal sampling by

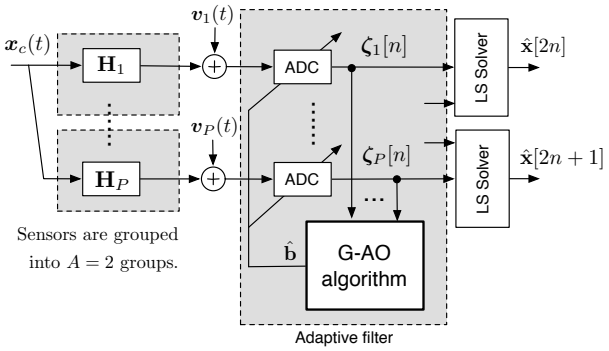

Fig. 5. Incorporating the proposed regression method into standard least squares state estimators. (Consider the special case when $A = A_p = 2$.)

adding delays to the ADCs. For example, if $A = A_p = 2$, the sensors can be divided into two groups, one with a delay tap of $\tau_p = 1 - \hat{b}_p$ and the other with $\tau_p = 2 - \hat{b}_p$, etc. Notice that we have created two groups of sensors, one sampling for $\mathbf{x}[2n]$, and the other one sampling for $\mathbf{x}[2n+1]$. The main advantage for applying the adaptive filter is to reduce complexity. In fact, the regression problem on the post-filtered data becomes a standard least-square optimization that admits a closed form solution; see Fig. 5.

Fig. 6 shows the MSE in state estimate against L using the inverse filter designed from the time offset estimate $\hat{b}_p$. The simulation settings are the same as in Fig. 2. The sensors grouping is done by grouping the first $P/2$ sensors into one group; and the rest of the sensors into another. As a benchmark, the 'offset-free (interleave)' refers to the system with *perfectly aligned*, time-interleaved sensors; and the 'CRLB' is the Cramer-Rao's lower bound in the error performance by assuming that the signal is periodic, i.e., the DFT approximation is exact. The 'offset-free (interleave)' benchmark requires perfect knowledge of $\{b_p\}_p$. From the simulation results, the MSE performance using the proposed algorithm is comparable to the benchmark when L is large.

B. Example: Power System State Estimation

In the following examples, we evaluate the performance of the proposed algorithms applied to PSSE. We consider the IEEE-30 bus test case in Fig. 7. To simulate the power system dynamics, the state vector is generated as $\mathbf{x}^e[n] = \mathbf{x}_c^e(nT_s) = \mathbf{x}_t + \mathbf{d}[n]$, where $\mathbf{x}_t$ is the voltage vector in the IEEE-30 test case in MATPOWER [41] and $\mathbf{d}[n] \sim \mathcal{CN}(0, \sigma_d^2\mathbf{I})$ models the fluctuation of voltages. Bus 1 is assumed to be the slack bus with zero phase angle.

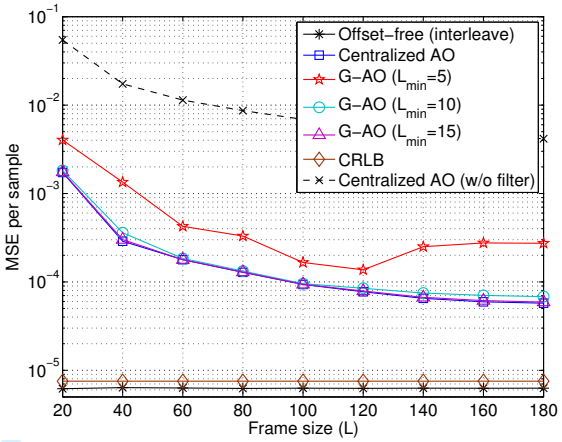

Fig. 6. Comparing the MSE in state estimate against L upon applying the adaptive filter. System settings are $M = 4, P = 12, N = 8$.

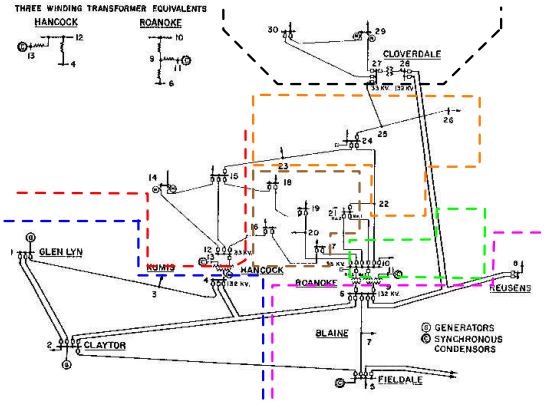

Fig. 7. The IEEE-30 bus system. We divide the PMUs into $P = 7$ sensing sites, as indicated by the colored dotted lines. The IEEE-30 bus system is partitioned as $\mathcal{A}_1 = \{1, 2, 3, 4\}, \mathcal{A}_2 = \{5, 6, 7, 8\}, \mathcal{A}_3 = \{9, 10, 11\}, \mathcal{A}_4 = \{12, 13, 14, 15\}, \mathcal{A}_5 = \{16, 17, 18, 19, 20, 21\}, \mathcal{A}_6 = \{22, 23, 24, 25, 26\}, \mathcal{A}_7 = \{27, 28, 29, 30\}$.

We first consider the case where only PMUs are employed. We assume that $b_p \ll T_s$ such that $\mathbf{x}_c^e(nT_s - b_p) \approx \mathbf{x}_c^e(nT_s)$. As noted by Remark 2, the corresponding joint regression problem (45) can be tackled using the G-AO algorithm. To ensure identifiability for both state and time offsets, the PMUs collect measurements from the buses $\{1, 2, 4, 5, 6, 9, 13, 15, 17, 19, 20, 21, 23, 25, 27, 29\}$.

The simulation result, as shown in Fig. 8, is performed with 1000 Monte-Carlo trials. In particular, we compare the MSE in state estimation against different range of sampling offset σ_b, where the sampling offsets b_p is generated as uniformly distributed over $[-\sigma_b, \sigma_b]$. As observed, the proposed method achieves

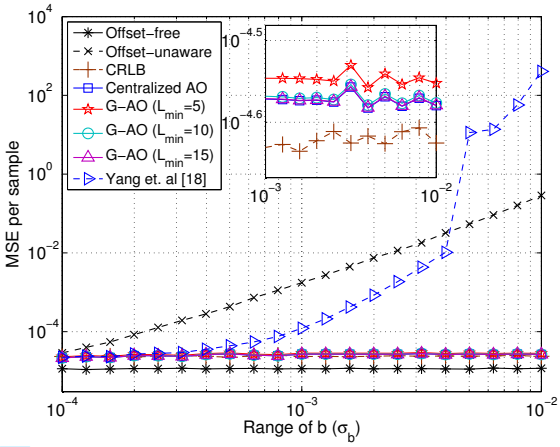

Fig. 8. Comparing the MSE per sample of an IEEE-30 bus system with only PMUs against the variance in sampling offset in b_p and we fix $\sigma_w = 10^{-2}$. We also have $\sigma_d = 10^{-2}$ and $\sum_p M_p = 62, N = 30$.

a performance on par with the method (Yang et. al.) from [18] when σ_b is small. The latter is outperformed by ours when $\sigma_b \geq 10^{-3}$. Moreover, our proposed algorithm achieves a MSE that is close to the Cramer Rao's lower bound (CRLB). This demonstrates the benefit of modelling non-ideal sampling directly.

The next simulation example considers the case where we combine both PMU and SCADA data. In particular, we consider the setting when only 8 PMUs are installed on bus $\{1, 4, 5, 9, 11, 15, 20, 24, 27\}$ in an IEEE-30 system. The number of installed PMUs is insufficient to provide identifiability if we only rely on PMU data. To remedy, we deploy 30 SCADA systems (each installed on a bus) on the power grid to monitor every bus and branch in the grid. We consider $L = 40$ and apply the centralized version of the GGN-AO algorithm, which is initialized by $\mathbf{x}_p^{(0)}[n] = \mathbf{x}_t$ for all n and $b_p = 0$ for all p.

Our aim is to demonstrate that the full power grid's state in *transient* can be revealed using the sub-Nyquist SCADA data and insufficient amount of PMU data. We assume that the complex envelope $\mathbf{x}_c^e(t)$ has a bandwidth of 0.5 Hz such that the Nyquist sampling period is $T_s = 1$ second. For simplicity, the SCADA systems has a sampling period of $A_p T_s = 2$ second, while the PMUs are able to capture samples with period shorter than $A_p T_s = 1$ second. A snapshot of the estimation result is shown in Fig. 9, in which we compare the state/time offset estimation error against iteration number. We have only shown the errors in estimating the voltage on buses $\{8, 13, 16, 17, 21, 26\}$. Notice that these buses are unobservable using the PMU data alone. That said, from the figure, we observe that the state estimation error from these buses is fairly low. More importantly, the error in estimating $\{b_p\}_p$ is at the order of 10^{-4} to 10^{-3}. This allows us to design the adaptive filter and treat the SCADA systems as the *time-interleaved* sensors

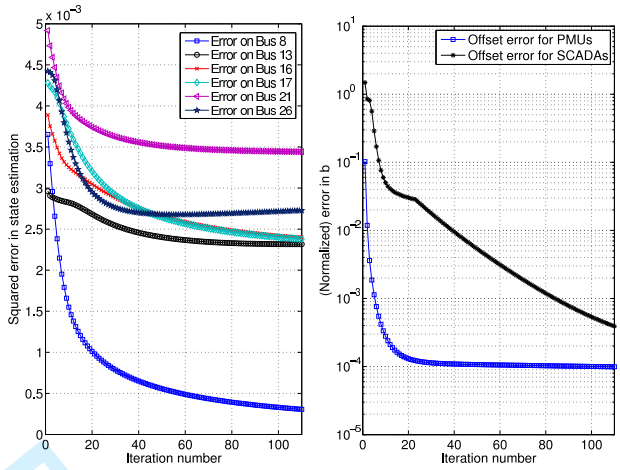

Fig. 9. The state estimation error performance against the iteration number over an IEEE-30 bus system. We have set $\lambda_{GN} = 10^2$ for the GGN-AO algorithm. (Left) On estimating the power system state at different bus. (Right) On estimating $\{b_p\}_p$ for PMUs and SCADAs. The latter error is normalized according to the range of time offsets b_p.

in a similar fashion as the last subsection.

C. Discussions

There are only ∼1,000 installed PMUs in total in the North America Power Grid as of 2014, while the number of buses exceeds 10,000 [42]. Although more PMUs are being installed, the number of PMUs employed is still insufficient to provide full identifiability of the entire power grid in the near future.

As demonstrated in the last simulation example, we see that deploying PMUs with SCADAs on the power grid is beneficial in the sense that the power system state in *transient* can be captured. In fact, using the result from Corollary 1, the fundamental limit on the maximum bandwidth for the power system state can be estimated. We assume that the PMUs are sampling at 10 Hz while the SCADAs are sampling at 0.5 Hz, each sensor takes $M_p = 3$ measurements which depends on the state of $N_c(\mathbf{H}_p) = 5$ buses. Now, set $Q_p = 1$ for the SCADAs, $Q_p = 20$ for the PMUs and using Eq. (20) yield:

$$\#\text{ PMUs} \cdot \min\{60, 5A\} + \#\text{ SCADAs} \cdot 3 \geq 20000A \tag{46}$$

In particular, when the number of PMUs and SCADAs is 1,000 and 10,000, respectively, the maximum possible A is 2. That is, the maximum possible bandwidth for system transient signal is 0.5 Hz, which is a two-fold improvement over the case with SCADAs alone. Naturally, the individual sensors renders transients variations at 10 Hz visible, but due to their relative scarcity the gain in resolving transients of

nearby locations is degraded. To what extent aliasing corrupts the reconstruction of faster transients is a topic of future research.

VII. Conclusions

In this paper, we have addressed the issue of non-ideal sampling in decentralized regression. Our contributions are multi-fold: i) we propose a new joint regression problem that estimates the system state and unknown time offsets under sub-Nyquist sampling; ii) we derive a set of identifiability conditions that guarantee perfect state and time offset estimation; iii) we develop a decentralized algorithm that solves the joint regression problem and prove its convergence. The efficacy of the proposed algorithm is demonstrated through numerical examples on both synthetic data and realistic power systems.

Appendix A
Proof of Observation 1

The observation is the consequence of a several properties from the Fourier transform. We first derive the DTFT spectrum of $\mathbf{x}_c((n-b_p)T_s)$. Under the Assumption 1 and by the Shannon's interpolation formula, we observe that:

$$\mathbf{x}_c((n-b_p)T_s) = \sum_{m=-\infty}^{\infty} \mathbf{x}[n-m]\frac{\sin(\pi(m-b_p))}{\pi(m-b_p)}, \tag{47}$$

which implies

$$\mathrm{DTFT}\{\mathbf{x}_c((n-b_p)T_s)\} = \mathbf{X}(e^{j\omega})e^{-jb_p\left(\omega \bmod (-\pi,\pi]\right)}. \tag{48}$$

Now, by decomposing the DTFT spectrum of $\mathbf{x}((nA-b_p)T_s)$ into its polyphase components and using results from Chapter 4 in [43], we can obtain (5).

Appendix B
Proof of Proposition 2

In the following proof, we shall abbreviate $\boldsymbol{\mathcal{H}}\boldsymbol{\mathcal{H}}_p^{\ell_t}$ as $\boldsymbol{\mathcal{H}}_p^{\ell_t}$, $\boldsymbol{\mathcal{H}}\boldsymbol{\mathcal{Z}}_p^{\ell_t}$ as $\boldsymbol{\mathcal{Z}}_p^{\ell_t}$, $\overline{\boldsymbol{\mathcal{H}}\boldsymbol{\mathcal{H}}}$ as $\overline{\boldsymbol{\mathcal{H}}}$ and $\overline{\boldsymbol{\mathcal{H}}\boldsymbol{\mathcal{Z}}}$ as $\overline{\boldsymbol{\mathcal{Z}}}$. We also drop the dependence on $\mathbf{b}$ and k of the vectors/matrices as they are irrelevant in the proof. Our goal is to show the following:

$$\|(\boldsymbol{\mathcal{H}}_p^{\ell_t})^{-1}\boldsymbol{\mathcal{Z}}_p^{\ell_t} - (\overline{\boldsymbol{\mathcal{H}}})^{-1}\overline{\boldsymbol{\mathcal{Z}}}\| = \mathcal{O}(\lambda_{\bar{W}}^{\ell_t}). \tag{49}$$

We now observe that:

$$\boldsymbol{\mathcal{H}}_p^{\ell_t} = \overline{\boldsymbol{\mathcal{H}}} + \delta\boldsymbol{\mathcal{H}}_p^{\ell_t},\ \boldsymbol{\mathcal{Z}}_p^{\ell_t} = \overline{\boldsymbol{\mathcal{Z}}} + \delta\boldsymbol{\mathcal{Z}}_p^{\ell_t} \tag{50}$$

where the error terms satisfy

$$\|\delta\boldsymbol{\mathcal{H}}_p^{\ell_t}\| \leq C_0\lambda_{\bar{W}}^{\ell_t},\ \|\delta\boldsymbol{\mathcal{Z}}_p^{\ell_t}\| \leq C_z\lambda_{\bar{W}}^{\ell_t}, \tag{51}$$

for some $C_z < \infty$. The error bounds in (51) are due to exponential convergence of the G-AC protocol [35]. Under assumption (31), the matrix inverse admits a series expansion [32]:

$$(\overline{\boldsymbol{\mathcal{H}}} + \delta\boldsymbol{\mathcal{H}}_p^{\ell_t})^{-1} = \sum_{q=0}^{\infty}(-1)^q(\overline{\boldsymbol{\mathcal{H}}}^{-1}\delta\boldsymbol{\mathcal{H}}_p^{\ell_t})^q(\overline{\boldsymbol{\mathcal{H}}})^{-1}. \tag{52}$$

Consequently, the left hand side in (49) can be upper bounded by:

$$\begin{aligned}
&\|(\boldsymbol{\mathcal{H}}_p^{\ell_t})^{-1}\boldsymbol{\mathcal{Z}}_p^{\ell_t} - (\overline{\boldsymbol{\mathcal{H}}})^{-1}\overline{\boldsymbol{\mathcal{Z}}}\| \\
&= \left\|\overline{\boldsymbol{\mathcal{H}}}^{-1}\delta\boldsymbol{\mathcal{Z}}_p^{\ell_t} + \sum_{q=1}^{\infty}(-1)^q(\overline{\boldsymbol{\mathcal{H}}}^{-1}\delta\boldsymbol{\mathcal{H}}_p^{\ell_t})^q\overline{\boldsymbol{\mathcal{H}}}^{-1}(\overline{\boldsymbol{\mathcal{Z}}} + \delta\boldsymbol{\mathcal{Z}}_p^{\ell_t})\right\| \\
&\leq \|\overline{\boldsymbol{\mathcal{H}}}^{-1}\|\Big(C_z\lambda_{\bar{W}}^{\ell_t} + \sum_{q=1}^{\infty}\|(\delta\boldsymbol{\mathcal{H}}_p^{\ell_t}\overline{\boldsymbol{\mathcal{H}}}^{-1})^q\|\|\overline{\boldsymbol{\mathcal{Z}}} + \delta\boldsymbol{\mathcal{Z}}_p^{\ell_t}\|\Big).
\end{aligned} \tag{53}$$

Assuming that $\|\overline{\boldsymbol{\mathcal{Z}}} + \delta\boldsymbol{\mathcal{Z}}_p^{\ell_t}\| \leq C_Z$, we have

$$\begin{aligned}
&\|(\boldsymbol{\mathcal{H}}_p^{\ell_t})^{-1}\boldsymbol{\mathcal{Z}}_p^{\ell_t} - (\overline{\boldsymbol{\mathcal{H}}})^{-1}\overline{\boldsymbol{\mathcal{Z}}}\| \\
&\quad\leq C_1\Big(C_z\lambda_{\bar{W}}^{\ell_t} + C_Z\sum_{q=1}^{\infty}\|(\delta\boldsymbol{\mathcal{H}}_p^{\ell_t}\overline{\boldsymbol{\mathcal{H}}}^{-1})^q\|\Big) \\
&\quad\leq C_1\Big(C_z\lambda_{\bar{W}}^{\ell_t} + C_Z\sum_{q=1}^{\infty}(C_0C_1\lambda_{\bar{W}}^{\ell_t})^q\Big)
\end{aligned} \tag{54}$$

Since $C_0C_1\lambda_{\bar{W}}^{\ell_t} < 1$, we have

$$\begin{aligned}
&\|(\boldsymbol{\mathcal{H}}_p^{\ell_t})^{-1}\boldsymbol{\mathcal{Z}}_p^{\ell_t} - (\overline{\boldsymbol{\mathcal{H}}})^{-1}\overline{\boldsymbol{\mathcal{Z}}}\| \\
&\quad\leq C_1\Big(C_z\lambda_{\bar{W}}^{\ell_t} + \frac{C_ZC_0C_1}{1 - C_0C_1\lambda_{\bar{W}}^{\ell_t}}\lambda_{\bar{W}}^{\ell_t}\Big) = \mathcal{O}(\lambda_{\bar{W}}^{\ell_t}).
\end{aligned} \tag{55}$$

The proof is completed.

APPENDIX C
PROOF OF THEOREM 1

To facilitate our discussions, we introduce the shorthand notations: $\boldsymbol{x}_p^{(t)} \triangleq \{\mathbf{X}_{p,A}^{(t)}(\omega_k)\}_k$, $\boldsymbol{x}_p^{\ell_t}(\mathbf{b}^{(t)}) \triangleq \{\mathbf{X}_{p,A,k}(\mathbf{b}^{(t)})\}_k$, $\boldsymbol{x}_t^{\star} \triangleq \mathbf{X}_{A,k}(\mathbf{b}^{(t-1)})$, $\boldsymbol{x}^{(t)} \triangleq (\boldsymbol{x}_1^{(t)}, ..., \boldsymbol{x}_P^{(t)})$ and $\hat{\boldsymbol{x}}^{(t)} \triangleq (1/P)\sum_{p=1}^{P}\boldsymbol{x}_p^{(t)}$. With a slight abuse of the notations, we let $f(\boldsymbol{x}^{(t)}, \mathbf{b}) \triangleq \sum_{p=1}^{P} f_p(\boldsymbol{x}_p^{(t)}, b_p)$. We also set $\beta = 1/M_o$.

We can assign upper bounds to the following norms of the differences:

$$\textstyle\sum_{p=1}^{P}\|\boldsymbol{x}_p^{(t)} - \boldsymbol{x}_t^{\star}\|_2 \leq \theta(\lambda_{\bar{W}}^{\ell_t}) \triangleq C_\theta\lambda_{\bar{W}}^{\ell_t}, \tag{56}$$

$$\sum_{p=1}^{P} \|\hat{\boldsymbol{x}}^{(t)} - \boldsymbol{x}_t^\star\|_2 \leq \phi(\lambda_{\bar{W}}^{\ell_t}) \triangleq C_\phi \lambda_{\bar{W}}^{\ell_t}, \tag{57}$$

$$\sum_{p=1}^{P} \|\boldsymbol{x}_p^{(t)} - \hat{\boldsymbol{x}}^{(t)}\|_2 \leq \psi(\lambda_{\bar{W}}^{\ell_t}) \triangleq C_\psi \lambda_{\bar{W}}^{\ell_t}. \tag{58}$$

The first inequality is due to Proposition 2. The latter two inequalities can be derived using triangular inequalities on (56).

Under the assumption in Theorem 1, for all t, $(\hat{\boldsymbol{x}}^{(t)}, \mathbf{b}^{(t)})$ stays in the neighborhood $\mathcal{N}_{R^\star}(\boldsymbol{x}^\star, \mathbf{b}^\star)$ such that f is strongly convex. Our idea is to study the dynamics of the following:

$$\Delta^{(t)} = f(\hat{\boldsymbol{x}}^{(t)}, \mathbf{b}^{(t)}) - f(\boldsymbol{x}^\star, \mathbf{b}^\star) \tag{59}$$

Observe that:

$$\begin{aligned}
&\Delta^{(t)} - \Delta^{(t-1)} \\
&= f(\hat{\boldsymbol{x}}^{(t)}, \mathbf{b}^{(t)}) - f(\hat{\boldsymbol{x}}^{(t-1)}, \mathbf{b}^{(t-1)}) \\
&= f(\boldsymbol{x}^{(t)}, \mathbf{b}^{(t)}) - f(\boldsymbol{x}^{(t)}, \mathbf{b}^{(t)}) + f(\hat{\boldsymbol{x}}^{(t)}, \mathbf{b}^{(t)}) \\
&\quad - f(\hat{\boldsymbol{x}}^{(t-1)}, \mathbf{b}^{(t-1)}) \\
&\leq L_o \psi(\lambda_{\bar{W}}^{\ell_t}) + f(\boldsymbol{x}^{(t)}, \mathbf{b}^{(t)}) - f(\hat{\boldsymbol{x}}^{(t-1)}, \mathbf{b}^{(t-1)}) \\
&\leq L_o \psi(\lambda_{\bar{W}}^{\ell_t}) + f(\boldsymbol{x}^{(t)}, \mathbf{b}^{(t-1)}) - \\
&\quad f(\hat{\boldsymbol{x}}^{(t-1)}, \mathbf{b}^{(t-1)}) - \frac{M_o}{2} \|\mathbf{b}^{(t)} - \mathbf{b}^{(t-1)}\|_2^2
\end{aligned} \tag{60}$$

where the first inequality is due to Lipschitz continuity of f and (58); the second inequality is due to the descent lemma [36]. Moreover, we have:

$$\begin{aligned}
f(\boldsymbol{x}^{(t)}, \mathbf{b}^{(t-1)}) &\leq f(\boldsymbol{x}_t^\star, \mathbf{b}^{(t-1)}) + L_o \theta(\lambda_{\bar{W}}^{\ell_t}) \\
&\leq f(\hat{\boldsymbol{x}}^{(t-1)}, \mathbf{b}^{(t-1)}) + L_o \theta(\lambda_{\bar{W}}^{\ell_t}),
\end{aligned} \tag{61}$$

where the first inequality is due to (56) and the second inequality is due to the optimality of $\boldsymbol{x}_t^\star$ with $\mathbf{b}^{(t-1)}$ fixed. Therefore,

$$\Delta^{(t)} - \Delta^{(t-1)} \leq L_o(\theta(\lambda_{\bar{W}}^{\ell_t}) + \psi(\lambda_{\bar{W}}^{\ell_t})) - \frac{M_o}{2} \|\mathbf{b}^{(t)} - \mathbf{b}^{(t-1)}\|_2^2. \tag{62}$$

Our next task is to lower bound $\|\mathbf{b}^{(t)} - \mathbf{b}^{(t-1)}\|_2^2$. To this end, we proceed by:

$$\begin{aligned}
\Delta^{(t)} &= f(\hat{\boldsymbol{x}}^{(t)}, \mathbf{b}^{(t)}) - f(\boldsymbol{x}^\star, \mathbf{b}^\star) \\
&= f(\boldsymbol{x}_t^\star, \mathbf{b}^{(t)}) - f(\boldsymbol{x}_t^\star, \mathbf{b}^{(t)}) + f(\hat{\boldsymbol{x}}^{(t)}, \mathbf{b}^{(t)}) - f(\boldsymbol{x}^\star, \mathbf{b}^\star) \\
&\leq L_o \phi(\lambda_{\bar{W}}^{\ell_t}) + \langle \nabla_b f(\boldsymbol{x}_t^\star, \mathbf{b}^{(t)}), \mathbf{b}^{(t)} - \mathbf{b}^\star \rangle + \\
&\quad B M_o \|\mathbf{b}^{(t)} - \mathbf{b}^{(t-1)}\|_2
\end{aligned} \tag{63}$$

where in the last inequality, we have used i) f is Lipschitz continuous and (57), ii) f is locally convex and iii)

$$\begin{aligned}&\langle \nabla_x f(\boldsymbol{x}_t^\star, \mathbf{b}^{(t)}), \boldsymbol{x}_t^\star - \boldsymbol{x}^\star \rangle \\ &= \langle \nabla_x f(\boldsymbol{x}_t^\star, \mathbf{b}^{(t)}) - \nabla_x f(\boldsymbol{x}_t^\star, \mathbf{b}^{(t-1)}), \boldsymbol{x}_t^\star - \boldsymbol{x}^\star \rangle \\ &\leq BM_o \|\mathbf{b}^{(t)} - \mathbf{b}^{(t-1)}\|_2.\end{aligned} \tag{64}$$

The equality is due to $\nabla_x f(\boldsymbol{x}_t^\star, \mathbf{b}^{(t-1)}) = \mathbf{0}$.

Our next endeavor is to upper bound $\langle \nabla_b f(\boldsymbol{x}_t^\star, \mathbf{b}^{(t)}), \mathbf{b}^{(t)} - \mathbf{b}^\star \rangle$. To this end, we observe

$$\begin{aligned}&\nabla_b f(\boldsymbol{x}_t^\star, \mathbf{b}^{(t)}) = \nabla_b f(\boldsymbol{x}_t^\star, \mathbf{b}^{(t)}) - \nabla_b f(\boldsymbol{x}^{(t)}, \mathbf{b}^{(t-1)}) + \\ &\tfrac{1}{\beta}\big((\mathbf{b}^{(t-1)} - \mathbf{b}^{(t)}) + \mathbf{b}^{(t)} - (\mathbf{b}^{(t-1)} - \beta \nabla_b f(\boldsymbol{x}^{(t)}, \mathbf{b}^{(t-1)})\big),\end{aligned} \tag{65}$$

since the latter terms cancel each other. Together with the following inequalities:

$$\begin{aligned}&\langle \nabla_b f(\boldsymbol{x}_t^\star, \mathbf{b}^{(t)}) - \nabla_b f(\boldsymbol{x}^{(t)}, \mathbf{b}^{(t-1)}), \mathbf{b}^{(t)} - \mathbf{b}^\star \rangle \leq \\ &BM_o\big(\theta(\lambda_{\bar{W}}^{\ell_t}) + \|\mathbf{b}^{(t)} - \mathbf{b}^{(t-1)}\|_2\big),\end{aligned} \tag{66}$$

which is a consequence of Cauchy-Schwarz and (56). Moreover, we have

$$\langle \mathbf{b}^{(t)} - (\mathbf{b}^{(t-1)} - \beta \nabla_b f(\boldsymbol{x}^{(t)}, \mathbf{b}^{(t-1)})), \mathbf{b}^{(t)} - \mathbf{b}^\star \rangle \leq 0, \tag{67}$$

since $\mathbf{b}^{(t)}$ is the projection of $\mathbf{b}^{(t-1)} - \beta \nabla_b f(\boldsymbol{x}^{(t)}, \mathbf{b}^{(t-1)})$ onto $\mathcal{B}$ and $\mathbf{b}^\star \in \mathcal{B}$.

Consequently,

$$\begin{aligned}&\langle \nabla_b f(\boldsymbol{x}_t^\star, \mathbf{b}^{(t)}), \mathbf{b}^{(t)} - \mathbf{b}^\star \rangle \leq BM_o\big(\theta(\lambda_{\bar{W}}^{\ell_t}) + \\ &\|\mathbf{b}^{(t)} - \mathbf{b}^{(t-1)}\|_2\big) + (B/\beta)\|\mathbf{b}^{(t)} - \mathbf{b}^{(t-1)}\|_2\end{aligned} \tag{68}$$

and $\Delta^{(t)}$ is upper bounded by:

$$\Delta^{(t)} \leq L_o \phi(\lambda_{\bar{W}}^{\ell_t}) + BM_o \theta(\lambda_{\bar{W}}^{\ell_t}) + 3BM_o \|\mathbf{b}^{(t)} - \mathbf{b}^{(t-1)}\|_2 \tag{69}$$

Plugging the above results back to (62) yields the following:

$$\begin{aligned}&\Delta^{(t)} - \Delta^{(t-1)} \leq L_o(\theta(\lambda_{\bar{W}}^{\ell_t}) + \psi(\lambda_{\bar{W}}^{\ell_t})) - \\ &\gamma_o \max\{\Delta^{(t)} - (L_o \phi(\lambda_{\bar{W}}^{\ell_t}) + BM_o \theta(\lambda_{\bar{W}}^{\ell_t})), 0\}^2\end{aligned} \tag{70}$$

where $\gamma_o = 1/(18B^2 M_o)$.

Since $\Delta^{(t)}$ is non-negative, we can simplify (70) by considering the upper bound $\xi^{(t)}$ such that $\Delta^{(t)} \leq \xi^{(t)}$ and $\mathcal{O}(\lambda_{\bar{W}}^{\ell_t}) \leq \mathcal{O}(\lambda_{\bar{W}}^{\ell_{min}})$ for all t:

$$\begin{aligned}&\xi^{(t)} - \xi^{(t-1)} = L_o(\theta(\lambda_{\bar{W}}^{\ell_{min}}) + \psi(\lambda_{\bar{W}}^{\ell_{min}})) - \\ &\gamma_o \max\{\xi^{(t)} - (L_o \phi(\lambda_{\bar{W}}^{\ell_{min}}) + BM_o \theta(\lambda_{\bar{W}}^{\ell_{min}})), 0\}^2\end{aligned} \tag{71}$$

A fixed point $\bar{\xi}$ to the above system must satisfy:

$$\begin{aligned} &L_o(\theta(\lambda_{\bar{W}}^{\ell_{min}}) + \psi(\lambda_{\bar{W}}^{\ell_{min}})) = \\ &\gamma_o \max\{\bar{\xi} - (L_o\phi(\lambda_{\bar{W}}^{\ell_{min}}) + BM_o\theta(\lambda_{\bar{W}}^{\ell_{min}})), 0\}^2 \end{aligned} \tag{72}$$

which implies

$$\begin{aligned} \bar{\xi} = &(L_o\phi(\lambda_{\bar{W}}^{\ell_{min}}) + BM_o\theta(\lambda_{\bar{W}}^{\ell_{min}})) + \\ &\sqrt{L_o(\theta(\lambda_{\bar{W}}^{\ell_{min}}) + \psi(\lambda_{\bar{W}}^{\ell_{min}}))/\gamma_o} = \sqrt{\mathcal{O}(\lambda_{\bar{W}}^{\ell_{min}})}. \end{aligned} \tag{73}$$

It can be verified that the above fixed point is stable. In fact, it is the only fixed point for the upper bound system (71).

Finally, from (73) and the local strong convexity of f, we have the following chain

$$\begin{aligned} \lim_{t\to\infty} \|(\hat{\boldsymbol{x}}^{(t)}, \mathbf{b}^{(t)}) - (\boldsymbol{x}^\star, \mathbf{b}^\star)\|^2 &\le \frac{2}{m_o} \lim_{t\to\infty} \Delta^{(t)} \\ &\le \sqrt{\mathcal{O}(\lambda_{\bar{W}}^{\ell_{min}})}, \end{aligned} \tag{74}$$

which completes the proof.

ACKNOWLEDGMENT

The authors would like to thank the anonymous reviewers for their valuable comments to improve the manuscript.

REFERENCES

[1] H.-T. Wai and A. Scaglione, "State Estimation with Sampling Offsets in Wide Area Measurement Systems," in *Proc' IEEE SAM 2014*, Jun. 2014, pp. 49–52.

[2] ——, "Decentralized Regression with Asynchronous Sub-Nyquist Sampling," to appear in *Proc' Asilomar 2014*.

[3] L. Xie, D.-h. Choi, S. Kar, and H. V. Poor, "Fully Distributed State Estimation for Wide-Area Monitoring Systems," *IEEE Trans. Smart Grid*, vol. 3, no. 3, pp. 1154–1169, Sep. 2012.

[4] V. Kekatos and G. B. Giannakis, "Distributed Robust Power System State Estimation," *IEEE Trans. Power Syst.*, vol. 28, no. 2, pp. 1617–1626, May 2013.

[5] X. Li and A. Scaglione, "Robust Decentralized State Estimation and Tracking for Power Systems via Network Gossiping," *IEEE J. Sel. Areas Commun.*, vol. 31, no. 7, pp. 1184–1194, Jul. 2013.

[6] A. G. Dimakis, S. Kar, J. M. F. Moura, M. G. Rabbat, and A. Scaglione, "Gossip Algorithms for Distributed Signal Processing," *Proc. IEEE*, vol. 98, no. 11, pp. 1847–1864, Nov. 2010.

[7] S. Kar, J. M. F. Moura, and K. Ramanan, "Distributed Parameter Estimation in Sensor Networks: Nonlinear Observation Models and Imperfect Communication," *IEEE Trans. Inform. Theory*, vol. 58, no. 6, pp. 3575–3605, Jun. 2012.

[8] I. D. Schizas, A. Ribeiro, and G. B. Giannakis, "Consensus in Ad Hoc WSNs With Noisy Links Part I: Distributed Estimation of Deterministic Signals," *IEEE Trans. Signal Process.*, vol. 56, no. 1, pp. 350–364, Jan. 2008.

[9] M. Rabbat and R. Nowak, "Distributed optimization in sensor networks," in *Proc' IPSN'04*, 2004.

[10] R. Olfati-Saber, J. A. Fax, and R. M. Murray, "Consensus and Cooperation in Networked Multi-Agent Systems," *Proc. IEEE*, vol. 95, no. 1, pp. 215–233, Jan. 2007.

[11] A. H. Sayed, S.-y. Tu, J. Chen, X. Zhao, and Z. J. Towfic, "Diffusion strategies for adaptation and learning over networks: an examination of distributed strategies and network behavior," *IEEE Signal Process. Mag.*, vol. 30, no. 3, pp. 155–171, May 2013.

[12] 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.

[13] D. P. Shepard, T. E. Humphreys, and A. Fansler, "Evaluation of the vulnerability of phasor measurement units to GPS spoofing attacks," *International Journal of Critical Infrastructure Protection*, vol. 5, no. 3-4, pp. 146–153, Dec. 2012.

[14] A. Ferreira and J. Fernandes, "A survey on time delay system estimation," *Proc' ECC 1997*.

[15] S. Julier and J. Uhlmann, "Fusion of time delayed measurements with uncertain time delays," in *Proc' ACC 2005*, pp. 4028–4033.

[16] J.-O. Nilsson, I. Skog, and P. Handel, "Joint state and measurement time-delay estimation of nonlinear state space systems," in *Proc' ISSPA 2010*, May 2010, pp. 324–328.

[17] J. Nilsson and P. Händel, "Timing estimation in distributed sensor and control systems with central processing," *arXiv preprint arXiv:1309.1864*, vol. 46, no. 0, 2013. [Online]. Available: http://arxiv.org/abs/1309.1864

[18] P. Yang, Z. Tan, A. Wiesel, and A. Nehorai, "Power System State Estimation Using PMUs With Imperfect Synchronization," *IEEE Trans. Power Syst.*, vol. 28, no. 4, pp. 4162–4172, Nov. 2013.

[19] S. Huang and B. C. Levy, "Blind Calibration of Timing Offsets for Four-Channel Time-Interleaved ADCs," *IEEE Trans. Circuits Syst. I, Reg. Papers*, vol. 54, no. 4, pp. 863–876, Apr. 2007.

[20] J. Goodman, B. Miller, M. Herman, G. Raz, and J. Jackson, "Polyphase Nonlinear Equalization of Time-Interleaved Analog-to-Digital Converters," *IEEE J. Sel. Topics Signal Process.*, vol. 3, no. 3, pp. 362–373, Jun. 2009.

[21] R. Venkataramani and Y. Bresler, "Optimal sub-Nyquist nonuniform sampling and reconstruction for multiband signals," *IEEE Trans. Signal Process.*, vol. 49, no. 10, pp. 2301–2313, 2001.

[22] M. Mishali and Y. Eldar, "From Theory to Practice: Sub-Nyquist Sampling of Sparse Wideband Analog Signals," *IEEE J. Sel. Topics Signal Process.*, vol. 4, no. 2, pp. 375–391, Apr. 2010.

[23] Y. Eldar and A. Oppenheim, "Filterbank reconstruction of bandlimited signals from nonuniform and generalized samples," *IEEE Trans. Signal Process.*, vol. 48, no. 10, pp. 2864–2875, 2000.

[24] M. Mishali and Y. Eldar, "Sub-Nyquist Sampling," *IEEE Signal Process. Mag.*, vol. 28, no. 6, pp. 98–124, Nov. 2011.

[25] A. Nedic, A. Ozdaglar, and P. Parrilo, "Constrained Consensus and Optimization in Multi-Agent Networks," *IEEE Trans. Autom. Control*, vol. 55, no. 4, pp. 922–938, Apr. 2010.

[26] S. S. Ram, A. Nedic, and V. V. Veeravalli, "A new class of distributed optimization algorithms : application to regression of distributed data," *Optimization Methods and Software*, no. 1, pp. 37–41, Feb. 2012.

[27] S. Mou and A. Morse, "A fixed-neighbor, distributed algorithm for solving a linear algebraic equation," in *Proc' ECC' 2013*, Jul. 2013, pp. 2269–2273.

[28] X. Li and A. Scaglione, "Convergence and Applications of a Gossip-Based Gauss-Newton Algorithm," *IEEE Trans. Signal Process.*, vol. 61, no. 21, pp. 5231–5246, Nov. 2013.

[29] A. H. Sayed, *Fundamentals of Adaptive Filtering*, 1st ed. Wiley-IEEE Press, Jun. 2003.

[30] L. Tong and S. Perreau, "Multichannel blind identification: from subspace to maximum likelihood methods," *Proc. IEEE*, vol. 86, no. 10, pp. 1951–1968, 1998.

[31] G. Xu, H. Liu, L. Tong, and T. Kailath, "A least-squares approach to blind channel identification," *IEEE Trans. Signal Process.*, vol. 43, no. 12, pp. 2982–2993, 1995.

[32] R. A. Horn and C. R. Johnson, Eds., *Matrix Analysis*. Cambridge University Press, 1986.

[33] D. Brillinger, *Time Series: Data Analysis and Theory*. Society for Industrial and Applied Mathematics, 2001.

[34] M. Razaviyayn, M. Hong, and Z.-Q. Luo, "A Unified Convergence Analysis of Block Successive Minimization Methods for Nonsmooth Optimization," *SIAM Journal on Optimization*, vol. 23, no. 2, pp. 1126–1153, Jun. 2013.

[35] S. Boyd, A. Ghosh, B. Prabhakar, and D. Shah, "Randomized gossip algorithms," *IEEE Trans. Inf. Theory*, vol. 52, no. 6, pp. 2508–2530, Jun. 2006.

[36] D. P. Bertsekas, *Nonlinear Programming*. Athena Scientific, Sep. 1999.

[37] "IEEE Standard for Synchrophasors for Power Systems," *IEEE Std C37.118-2005 (Revision of IEEE Std 1344-1995)*, pp. 1–57, 2006.

[38] L. Grigsby, Ed., *Power System Stability and Control*, ser. Electrical Engineering Handbook. CRC Press, May 2007, vol. 20073061.

[39] A. V. Oppenheim, R. W. Schafer, and J. R. Buck, *Discrete-time Signal Processing (2nd Ed.)*. Upper Saddle River, NJ, USA: Prentice-Hall, Inc., 1999.

[40] L. Xiao and S. Boyd, "Fast linear iterations for distributed averaging," *Systems & Control Letters*, vol. 53, no. 1, pp. 65–78, Sep. 2004.

[41] R. D. Zimmerman, C. E. Murillo-Sanchez, and R. J. Thomas, "MATPOWER: Steady-State Operations, Planning, and Analysis Tools for Power Systems Research and Education," *IEEE Trans. Power Syst.*, vol. 26, no. 1, pp. 12–19, Feb. 2011.

[42] NASPI, "PMUs and Synchrophasor Data Flows in North America," https://www.naspi.org/File.aspx?fileID=1275, Mar. 2014.

[43] P. P. Vaidyanathan, *Multirate Systems and Filter Banks*. Prentice-Hall, Inc., 1993.