Submitted for publication. Author copy -- do not redistribute.

Decentralized Regression with Asynchronous Sub-Nyquist Sampling

Hoi-To Wai and Anna Scaglione

Department of Electrical and Computer Engineering, University of California at Davis, Davis, CA 95616, USA
Email: {htwai,ascaglione}@ucdavis.edu

***Abstract*—When capturing data on a sensor field to uncover its latent structure, there are often nuisance parameters in the observation model that turn even linear regression problems into non-convex optimizations. One common case is the lack of common timing source in ADCs, therefore samplings are done with time offsets. Motivated by the desire of estimating jointly the sensor field and nuisance parameters in a wide area deployment, this paper derives a new decentralized algorithm that combines alternating optimization and gossip-based learning. The proposed algorithm is shown to converge to the neighborhood of a local minimum, both analytically and empirically.**

***Index Terms*—decentralized regression, asynchronous sampling, sub-Nyquist sampling, gossip algorithm**

I. Introduction

Recently, a number of papers sought numerical methods for decentralized optimization in sensor networks. A major thrust behind these developments is that a large amount of data is ubiquitously acquired by a large number of sensors. As these sensors are physically distributed over a wide area, their data are often acquired through a mesh network of data concentrators. Efficient decentralized numerical method could harness the computational power available in the network itself.

One of the earliest studied problems which calls for decentralized optimization is linear regression. The linear regression problem constitutes a convex optimization with a *common variable* and admits a closed form solution when solved at a central station. In the context of sensor networks, a decentralized method is designed to coordinate the sensors such that a common variable can be found to minimize the objective function. In fact, decentralized methods applicable to this type of problems have been widely studied [1]–[5]; and considered in practical systems [6]–[9].

In certain scenarios (e.g., cyber physical systems), it is more natural to assume that the underlying sensing model beneath the linear regression formulation includes local nuisance parameters. For example, when the individual agents are not synchronized in time, [10], [11] have shown that a local parameter has to be introduced for modeling the time offsets between sensors. Notice that the resulting regression problem will be non-convex and a different solution approach will be needed.

The rest of this paper is organized as follows. In Section II & III, we derive a decentralized algorithm using on the ideas of gossip-based average consensus and alternating optimization. The convergence of the algorithm will be analyzed. In Section IV, we demonstrate that the developed algorithm is applicable to the joint regression problem under asynchronous and sub-Nyquist sampling. Lastly, the paper will be concluded with some numerical results in Section V.

This material is based upon work supported by the Department of Energy under Award Number DE-OE0000097.

II. Problem Statement

Consider a regression problem of the following type:

$$\min_{\boldsymbol{x},\boldsymbol{b}} f(\boldsymbol{x},\boldsymbol{b}) \triangleq \sum_{p=1}^{N} \|\boldsymbol{g}_p(\boldsymbol{x},\boldsymbol{b}_p)\|_2^2 \text{ s.t. } \boldsymbol{x}\in\mathbb{R}^n,\ \boldsymbol{b}_p\in\mathcal{B}_p, \quad (1)$$

where $\mathcal{B}_p \subseteq \mathbb{R}^m$ is a convex set and $\boldsymbol{g}_p$ is linear in $\boldsymbol{x}$, i.e.,

$$\boldsymbol{g}_p(\boldsymbol{x},\boldsymbol{b}_p) = \boldsymbol{\zeta}_p(\boldsymbol{b}_p) - \mathbf{H}_p(\boldsymbol{b}_p)\boldsymbol{x}. \quad (2)$$

Problem (1) may be *non-convex* even though $\boldsymbol{g}_p$ is linear.

Our goal is to solve (1) using N networked agents, described by an undirected graph $\mathcal{G} = (\mathcal{V},\mathcal{E})$. Specifically, we assume that the partial objective function $\boldsymbol{g}_p(\boldsymbol{x},\boldsymbol{b}_p)$ is only known to the pth agent. In this way, we may regard $\boldsymbol{x}$ as the common variable that has to be agreed upon each of the agents; and $\boldsymbol{b}_p$ is a local variable to the pth agent.

We notice that (1) has the following properties.

1) When $\boldsymbol{b}_p$ are fixed, the optimization (1) is *convex* in $\boldsymbol{x}$. In particular, the problem can be solved by:

$$\boldsymbol{x}^\star(\boldsymbol{b}) = \Big(\sum_{p=1}^{N}\mathbf{H}_p(\boldsymbol{b}_p)^T\mathbf{H}_p(\boldsymbol{b}_p)\Big)^{-1}\Big(\sum_{p=1}^{N}\mathbf{H}_p(\boldsymbol{b}_p)^T\boldsymbol{\zeta}_p(\boldsymbol{b}_p)\Big) \quad (3)$$

2) When $\boldsymbol{x}$ is fixed, the optimization (1) is *separable* in $\boldsymbol{b}_p$. In other words, given $\boldsymbol{x}$, agent p is able to optimize $\boldsymbol{b}_p$ independently.

These observations suggest us to adopt the alternating optimization (AO) approach as the backbone for tackling (1).

Algorithm 1 The Gossip-based AO algorithm for (1).

Initialize: $\{\boldsymbol{x}_p^0\}_{p=1}^N$, $\{\boldsymbol{b}_p^0\}_{p=1}^N$;

for $k = 0, 1, ...$ **do**

  The network compute $\tilde{\boldsymbol{x}}_p^{\ell_k}(\boldsymbol{b}^k)$ for each p using ℓ_k G-AC steps (see Section III-A). The solution $\tilde{\boldsymbol{x}}_p^{\ell_k}(\boldsymbol{b}^k)$ is an approximate solution to $\boldsymbol{x}^\star(\boldsymbol{b}^k)$ in (3).

  for $p = 1, 2, ..., N$ **do**

    Agent p updates its copies of $\boldsymbol{x}$ and $\boldsymbol{b}_p$ as:

$$\boldsymbol{x}_p^{k+1} \leftarrow \tilde{\boldsymbol{x}}_p^{\ell_k}(\boldsymbol{b}^k), \tag{4}$$

$$\boldsymbol{b}_p^{k+1} \leftarrow \mathcal{P}_{\mathcal{B}_p}\big(\boldsymbol{b}_p^k - \beta \nabla_{\boldsymbol{b}_p} \|\boldsymbol{g}_p(\boldsymbol{x}_p^{k+1}, \boldsymbol{b}_p^k)\|_2^2\big) \tag{5}$$

    where $\mathcal{P}_{\mathcal{B}_p}(\cdot)$ is the projection operator onto $\mathcal{B}_p$ and $\beta > 0$ is a step size.

  end for

end for

Return: $\{\boldsymbol{x}_p^k\}_{p=1}^N$, $\{\boldsymbol{b}_p^k\}_{p=1}^N$.

III. Gossip-based AO algorithm

The AO algorithm for (1) includes two steps, in which only *one* optimization variable ($\boldsymbol{x}$ or $\boldsymbol{b}$) is updated at a time. Since the problem in (1) is separable in terms of $\boldsymbol{b}_p$ when $\boldsymbol{x}$ is fixed, this stage of the AO algorithm is inherently decentralized. Our remaining task is to tackle the optimization of $\boldsymbol{x}$ in a decentralized manner.

The optimization of $\boldsymbol{x}$ in (1) while fixing $\boldsymbol{b}$ is equivalent to solving a standard least square problem. To this end, a number of existing decentralized algorithm can be employed, e.g., the diffusion-based recursive least square (RLS) [1] and least mean square (LMS) [4].

In this paper, we adopt a different approach by studying a gossip-based average consensus (G-AC) method for distributively computing the closed form solution (3). This approach has been demonstrated in [10]–[12] to provide a good performance for linear regression. Another advantage for this method is that an error bound to the obtained solution can be easily obtained and exploited in analyzing the algorithm convergence properties, as we will discuss in the paper. We first summarize the proposed algorithm in Algorithm 1.

A. Gossip-based average consensus for computing (3)

To begin our endeavor, we observe that the computation of (3) merely involves the averages of $\{\mathbf{H}_p(\boldsymbol{b}_p)^T\mathbf{H}_p(\boldsymbol{b}_p)\}_p$ and $\{\mathbf{H}_p(\boldsymbol{b}_p)^T\boldsymbol{\zeta}_p(\boldsymbol{b}_p)\}_p$, i.e., we can replace (3) by

$$\boldsymbol{x}^\star(\boldsymbol{b}) = \Big(\frac{1}{N}\sum_{p=1}^N \mathbf{H}_p(\boldsymbol{b}_p)^T\mathbf{H}_p(\boldsymbol{b}_p)\Big)^{-1}\Big(\frac{1}{N}\sum_{p=1}^N \mathbf{H}_p(\boldsymbol{b}_p)^T\boldsymbol{\zeta}_p(\boldsymbol{b}_p)\Big). \tag{6}$$

This suggests us to treat the task of computing (3) as a decentralized averaging problem. Specifically, we apply the gossip-based average consensus (G-AC) method in [13]. Now let us take $\boldsymbol{z}_p^0 \triangleq [\mathbf{H}_p(\boldsymbol{b}_p)^T\boldsymbol{\zeta}_p(\boldsymbol{b}_p); \mathrm{vec}(\mathbf{H}_p(\boldsymbol{b}_p)^T\mathbf{H}_p(\boldsymbol{b}_p))]$. It suffices to compute (3) by obtaining the average of $\{\boldsymbol{z}_p^0\}_p$, i.e.,

$$\bar{\boldsymbol{z}} = (1/N)\textstyle\sum_{p=1}^N \boldsymbol{z}_p^0. \tag{7}$$

The G-AC method performs the following recursions:

$$\boldsymbol{z}_p^\ell = \textstyle\sum_{q\in\mathcal{N}_p} W_{pq}^\ell \boldsymbol{z}_q^{\ell-1}, \tag{8}$$

where $\mathcal{N}_p \subseteq \mathcal{V}$ denotes the set of neighbors of agent p. The mixing matrix $\mathbf{W}^\ell = [W_{pq}^\ell]_{p,q}$ satisfies a certain set of mild conditions (see [13]), e.g., it is required to be *doubly stochastic*, i.e., $\mathbf{W}^\ell\mathbf{1} = \mathbf{1}$ and $\mathbf{1}^T\mathbf{W}^\ell = \mathbf{1}^T$.

As indicated by (8), at each G-AC step ℓ, the agent 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 $\tilde{\boldsymbol{x}}_p^{\ell_k}(\boldsymbol{b}^k)$ is computed using the *approximate* averages of $\{\mathbf{H}_p(\boldsymbol{b}_p)^T\mathbf{H}_p(\boldsymbol{b}_p)\}_p$ and $\{\mathbf{H}_p(\boldsymbol{b}_p)^T\boldsymbol{\zeta}_p(\boldsymbol{b}_p)\}_p$ stored at the pth agent after ℓ_k G-AC steps.

We conclude this subsection by commenting on the convergence of G-AC. As shown in [13], the recursion (8) converges to the true average vector $\bar{\boldsymbol{z}}$. In fact, the rate of convergence is exponential, i.e., $\|\boldsymbol{z}_p^\ell - \bar{\boldsymbol{z}}\| = \mathcal{O}(\lambda_{\bar{W}}^\ell)$ where $0 < \lambda_{\bar{W}} < 1$ is the second largest eigenvalue of the expected mixing matrix $\bar{\mathbf{W}} = \mathbb{E}\{\mathbf{W}^\ell\}$. Consequently, the accuracy on $\tilde{\boldsymbol{x}}_p^{\ell_k}(\boldsymbol{b}^k)$ also improves exponentially with ℓ_k:

Lemma 1. *Suppose that*

$$C_0 \cdot C_1 \cdot \lambda_{\bar{W}}^{\ell_k} < 1, \tag{9}$$

where $C_0 = \max_{\boldsymbol{b}} \|\sum_p \mathbf{H}(\boldsymbol{b}_p)^T\mathbf{H}(\boldsymbol{b}_p)\|$ *and* $C_1 = \max_{\boldsymbol{b}} \|(\sum_p \mathbf{H}(\boldsymbol{b}_p)^T\mathbf{H}(\boldsymbol{b}_p))^{-1}\|$ *are finite constants, then* $\tilde{\boldsymbol{x}}_p^{\ell_k}(\boldsymbol{b}^k)$ *computed using* ℓ_k *G-AC steps satisfies:*

$$\textstyle\sum_{p=1}^N \|\boldsymbol{x}^\star(\boldsymbol{b}^k) - \tilde{\boldsymbol{x}}_p^{\ell_k}(\boldsymbol{b}^k)\| \le \theta(\lambda_{\bar{W}}^{\ell_{min}}), \tag{10}$$

$$\textstyle\sum_{p=1}^N \|\boldsymbol{x}^\star(\boldsymbol{b}^k) - (1/N)\sum_{q=1}^N \tilde{\boldsymbol{x}}_p^{\ell_k}(\boldsymbol{b}^k)\| \le \phi(\lambda_{\bar{W}}^{\ell_{min}}), \tag{11}$$

$$\textstyle\sum_{p=1}^N \|\tilde{\boldsymbol{x}}_p^{\ell_k}(\boldsymbol{b}^k) - (1/N)\sum_{q=1}^N \tilde{\boldsymbol{x}}_p^{\ell_k}(\boldsymbol{b}^k)\| \le \psi(\lambda_{\bar{W}}^{\ell_{min}}), \tag{12}$$

where $\ell_{min} = \min_k \ell_k$ *and* $\theta(x), \psi(x), \phi(x)$ *are functions that grows linearly with* x.

The proof is in [11] and it is closely related to that in [12], which relies on a series expansion of the matrix inverse.

B. Convergence of the gossip-based AO algorithm

We have shown that the G-AC step for obtaining $\tilde{\boldsymbol{x}}_p^{\ell_k}(\boldsymbol{b}^k)$ in a decentralized manner provides a good approximation for $\boldsymbol{x}^\star(\boldsymbol{b}^k)$. In this subsection, we study the convergence of the proposed gossip-based AO algorithm in Algorithm 1.

To provide some insights, let us consider the special case when the closed form solution (3) is obtained by performing a centralized computation. In this case, the *centralized* AO algorithm can be analyzed as a Block Successive Minimization Method (BSUM) in [14]. Specifically, [14] shows that the algorithm converges to a local minimum as $k \to \infty$.

On the other hand, from Lemma 1 we observe that the gossip-based AO algorithm can be analyzed as a *perturbed* BSUM. This suggests that the gossip-based AO algorithm should also converge to the neighborhood of a local minimum, as confirmed in the following:

Theorem 1. *Let* $(\boldsymbol{x}^\star, \boldsymbol{b}^\star)$ *be a local minimum to* (1) *and* f *be* (m_o, M_o)*-strongly convex in the neighborhood* $\mathcal{N}_{R^\star}(\boldsymbol{x}^\star, \boldsymbol{b}^\star)$ *with radius* $R^\star$ *and it has a Lipschitz constant of* L_o*. Suppose* $\beta < 1/M_o$*,* $B \triangleq \max_k \|(\boldsymbol{x}^k, \boldsymbol{b}^k) - (\boldsymbol{x}^\star, \boldsymbol{b}^\star)\| < \infty$ *and* $B \leq R^\star$*, then we have:*

$$\lim_{k\to\infty} \|(\hat{\boldsymbol{x}}^k, \boldsymbol{b}^k) - (\boldsymbol{x}^\star, \boldsymbol{b}^\star)\|^2 \leq \rho(\lambda_W^{\ell_{min}}), \tag{13}$$

where $\hat{\boldsymbol{x}}^k \triangleq (1/N) \sum_{p=1}^N \boldsymbol{x}_p^k$ *and*[1]

$$\rho(\lambda_W^{\ell_{min}}) \triangleq \mathcal{O}\Big(\sqrt{\lambda_W^{\ell_{min}}}\Big) = \frac{2}{m_o} \times \left(L_o \phi(\lambda_{\bar{W}}^{\ell_{min}}) + B M_o \theta(\lambda_{\bar{W}}^{\ell_{min}}) + \sqrt{\frac{\theta(\lambda_{\bar{W}}^{\ell_{min}}) + \psi(\lambda_{\bar{W}}^{\ell_{min}})}{1/(18 B^2 L_o M_o)}} \right)$$

The proof can be found in [11]. In particular, one of the key steps is to show that

$$\begin{aligned} &(f(\hat{\boldsymbol{x}}^k, \boldsymbol{b}^k) - f(\boldsymbol{x}^\star, \boldsymbol{b}^\star)) - (f(\hat{\boldsymbol{x}}^{k-1}, \boldsymbol{b}^k) - f(\boldsymbol{x}^\star, \boldsymbol{b}^\star)) \\ &\leq L_o(\theta(\lambda_{\bar{W}}^{\ell_{min}}) + \psi(\lambda_{\bar{W}}^{\ell_{min}})) - (M_o/2)\|\boldsymbol{b}^k - \boldsymbol{b}^{k-1}\|_2^2. \end{aligned} \tag{14}$$

The proof follows by i) lower bounding $\|\boldsymbol{b}^k - \boldsymbol{b}^{k-1}\|_2^2$ using the convexity of f in the neighborhood; ii) studying the error dynamics as a discrete dynamical system.

The theorem shows that the gossip-based AO algorithm converges to an approximate *local minimum* solution, where the approximation accuracy decays *exponentially* with the number of G-AC steps taken per iteration. Notice that the (locally) strongly convex assumption is observed to be true in our applications and the simulation results in Section V corroborate the trend predicted from Theorem 1.

IV. Regression under Asynchronous and sub-Nyquist Sampling

The aim of this section is to formulate the regression problem under asynchronous and sub-Nyquist sampling. We highlight that the required regression problem can be written in the form of (1), such that decentralized regression will be possible by applying the gossip-based AO algorithm.

Our scenario of interest is a sensor network that captures the continuous-time sensor field $\boldsymbol{x}_c(t) \in \mathbb{R}^n$ using N sensors. Assume that the signal is band-limited by $1/(2T_s)$ Hz. We focus on the case when the sensors are taking memory-less, asynchronous and sub-Nyquist measurements on $\boldsymbol{x}_c(t)$. Specifically, the ith sample recorded at the pth sensor is:

$$\boldsymbol{\zeta}_p[i] = \mathbf{H}_p \boldsymbol{x}_c((iA_p - b_p)T_s) + \boldsymbol{w}_p[i], \tag{15}$$

where $A_p \geq 1, A_p \in \mathbb{Z}$ is the down-sampling factor of the pth sensor and $b_p \in \mathbb{R}$ is the time offset in sampling, $\boldsymbol{w}_p[n] \sim \mathcal{N}(0, \sigma^2 \mathbf{I})$ is an additive noise and $\mathbf{H}_p \in \mathbb{R}^{m \times n}$ represents the measurement matrix for the pth sensor.

As $\boldsymbol{x}_c(t)$ is bandlimited, it suffices to estimate the Nyquist-rate samples of $\boldsymbol{x}_c(t)$, i.e., $\boldsymbol{x}[i] \triangleq \boldsymbol{x}_c(iT_s)$, for our task. We assume that the down-sampling factor A_p is known, while the time offsets b_p is the unknown nuisance parameter. The field $\boldsymbol{x}_c(t)$ is also shift invariant such that we do not distinguish between $\boldsymbol{x}_c(t)$ and $\boldsymbol{x}_c(t - t_0)$. In this case, we can set $b_1 = 0$ without loss of generality.

In fact, the time offsets are essential to the recovery of $\boldsymbol{x}[i]$ from the possibly sub-Nyquist samples in (15), when exploited correctly; see [11] for some discussions.

A. Problem formulation

Our next endeavor is to derive the regression problem under model (15) with asynchronous and sub-Nyquist sampling. Note that from (15), it is impossible to infer a single sample $\boldsymbol{x}[i]$ by taking one snapshot of the measured signal $\{\boldsymbol{\zeta}_p[i]\}_{p=1}^N$, as the latter also depends on the samples $\{\boldsymbol{x}[j]\}_{j \neq i}$.

We resort to an *offline processing* scheme that performs a batch regression of the Fourier series of a long data stream. In particular, we observe that the frequency-domain equivalent model to (15) admits the following representation:

Observation 1 *Let* $\boldsymbol{x}_c(t)$ *be bandlimited by* $1/(2T_s)$ *Hz and*

$$\Omega_{A_p}^a(\omega) \triangleq \big(\frac{\omega}{A_p} - \frac{a}{A_p} 2\pi\big) \bmod (-\pi, \pi]. \tag{16}$$

The measurement model (15) *is equivalent to the following:*

$$\mathbf{Z}_p(e^{j\omega}) = \frac{1}{A_p} \sum_{a=0}^{A_p - 1} e^{-j b_p \Omega_{A_p}^a(\omega)} \mathbf{H}_p \mathbf{X}\left(e^{j\Omega_{A_p}^a(\omega)}\right) + \mathbf{V}_p(e^{j\omega}), \tag{17}$$

where $\mathbf{Z}_p(e^{j\omega})$*,* $\mathbf{V}_p(e^{j\omega})$ *and* $\mathbf{X}(e^{j\omega})$ *are the discrete-time Fourier transform (DTFT) of* $\boldsymbol{\zeta}_p[i]$*,* $\boldsymbol{v}_p[i]$ *and* $\boldsymbol{x}[i]$*, respectively.*

A key to verifying the above observation is to decompose $\boldsymbol{x}[i]$ into its polyphase components and study the spectrum of the down-sampled signal; see [11], [15].

Observation 1 shows that the observation on $\mathbf{X}(e^{j\omega})$ is subjected to linear transformation, linear phase shift and aliasing. These effects, when combined, can be viewed as a linear transformation on $\mathbf{X}(e^{j\omega})$ with given time offset b_p. To see this, we first need to ensure that the measured samples

[1] The bounds obtained are not optimized and can be improved.

are obtained at the same sampling rate. It can be achieved by generating the following samples from $\boldsymbol{\zeta}_p[i]$:

$$\boldsymbol{\zeta}_p^q[i] = \boldsymbol{\zeta}_p[Q_p i - q],\ q = 0, 1, ..., Q_p - 1, \tag{18}$$

where $Q_p \triangleq A/A_p$ and $A \triangleq \mathsf{LCM}\{A_1, ..., A_N\}$. These samples are equivalent to those sampled at $1/A$ of the Nyquist rate. Next, by noting that $\Omega_A^a((-\pi,\pi])$ is disjoint with $\Omega_A^b((-\pi,\pi])$ for $a \neq b$, we can define the following extended spectrum:

$$\tilde{\mathbf{X}}(e^{j\omega}) = \left[\mathbf{X}\left(e^{j\Omega_A^0(\omega)}\right)^T \ \dots \ \mathbf{X}\left(e^{j\Omega_A^{A-1}(\omega)}\right)^T\right], \tag{19}$$

where $A \triangleq \mathsf{LCM}\{A_1, ..., A_N\}$; and the extended matrix:

$$\tilde{\mathbf{H}}_p(b_p, e^{j\omega}) = \frac{1}{A}\left[e^{-jb_p\Omega_A^0(\omega)} \ \dots \ e^{-jb_p\Omega_A^{A-1}(\omega)}\right] \otimes \mathbf{H}_p. \tag{20}$$

Using Eq. (19) & (20), the observed spectrum in (17) can be simplified as:

$$\mathbf{Z}_p^q(e^{j\omega}) = \tilde{\mathbf{H}}_p(b_p - qQ_p, e^{j\omega})\tilde{\mathbf{X}}(e^{j\omega}) + \mathbf{V}_p^q(e^{j\omega}), \tag{21}$$

which is a linear function in $\tilde{\mathbf{X}}(e^{j\omega})$.

We now discuss how the spectrum $\mathbf{Z}_p^q(e^{j\omega})$ can be obtained from a finite number of measurements. To this end, we approximate $\mathbf{Z}_p^q(e^{j\omega})$ by the following K-point discrete Fourier transform (DFT) spectrum:

$$\mathbf{Z}_p^q[k] = \textstyle\sum_{m=0}^{L-1} \boldsymbol{\zeta}_p^q[m] e^{-j\omega_k m},\ k = 0, ..., K-1, \tag{22}$$

where $\omega_k \triangleq 2\pi(k - K + 1)/(K)$ and $K \geq L$ is required. In this way, the sequence $\{\boldsymbol{x}[i]\}_{i=0}^{AL-1}$ can be inferred from the collection of spectrum $\{\mathbf{Z}_p^q[k]\}_{p,k}$.

Finally, from (21) we can derive the following regression problem for estimating $\{\boldsymbol{x}[i]\}_{i=0}^{AL-1}$:

$$\begin{aligned} \min_{\{\mathbf{x}[i]\}_{i=0}^{AL-1}, \{b_p\}_{p=2}^N} \ & \sum_{p=1}^N f_p(\{\boldsymbol{x}[i]\}_{i=0}^{AL-1}, b_p) \\ \text{s.t.} \quad & \text{Eq. (19)},\ b_p \in \mathcal{B}_p,\ \forall\ p, \end{aligned} \tag{23}$$

where, denoting by $\mathbf{X}(e^{j\omega_k}) = \sum_{m=0}^{AL-1} \boldsymbol{x}[m] e^{-j\omega_k m}$,

$$\begin{aligned} & f_p(\{\boldsymbol{x}[i]\}_{i=0}^{AL-1}, b_p) \triangleq \\ & \sum_{q=0}^{Q_p-1}\sum_{k=0}^{K-1} \left\| \mathbf{Z}_p^q[k] - \tilde{\mathbf{H}}_p(b_p - qQ_p, e^{j\omega_k})\tilde{\mathbf{X}}(e^{j\omega_k}) \right\|_2^2. \end{aligned} \tag{24}$$

Since $\tilde{\mathbf{X}}(e^{j\omega_k})$ is a linear function of $\{\boldsymbol{x}[i]\}_{i=0}^{AL-1}$ (cf. (19)), the regression problem (23) is a special case of (1). Hence, the gossip-based AO algorithm can be applied. Note, however, that the model is exact only in the limit, thus there is a trade-off between the complexity of solving (23) and the quality of the approximation, that improves when the number of DFT points K is large.

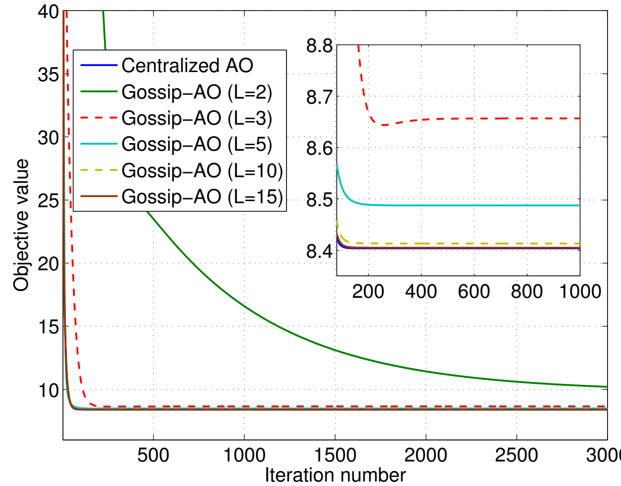

Fig. 1: The objective value against iteration number for solving an instance of (23). The data to the regression problem (23) is obtained with $L = 40$.

V. Numerical Results

This section presents numerical results that demonstrate the efficacy of the proposed algorithm. We test the decentralized regression under asynchronous sub-Nyquist sampling using synthetic data. The elements in the vectors $\boldsymbol{x}[i]$ and matrices $\mathbf{H}_p$ are generated as (complex) i.i.d. Gaussian random variable with zero-mean and unit variance.

Throughout this section, we assume the system parameter of $n = 6$, $m = 2$ and $N = 12$. The down-sampling factor is $A = 2$ for all sensors, while the time offset b_p is generated as a uniform random variable over $[-0.5, 0.5]$. We also set $\mathcal{B}_p = [-0.5, 0.5]$ to account for the said range of b_p. The noise variance is fixed at $\sigma^2 = 10^{-2}$. For the gossip-based AO algorithm, we assume that the underlying communication network is generated as an Erdos-Renyi graph with a connectivity of $p = 0.5$. The G-AC step is performed by a static mixing matrix $\mathbf{W}$, which is computed from the generated graph using the Metropolis-Hastings rule [16].

We set $K = 192$ for the DFT approximation (cf. (22)). To remedy the modelling error incurred in (22), we pre-process the data $\boldsymbol{\zeta}_p^q[i]$ by applying a length L Blackman window [17]. The estimation performance of $\boldsymbol{x}[n]$ is compared by calculating the mean square error (MSE) for the samples in the middle portion of the frame:

$$\text{MSE} = \frac{1}{2G+1}\sum_{m=AL/2-G}^{AL/2+G} \frac{\|\boldsymbol{x}[m] - \hat{\boldsymbol{x}}[m]\|_2^2}{n}, \tag{25}$$

with $G = \lfloor 0.3L \rfloor$ and $\hat{\boldsymbol{x}}[m]$ is the estimate produced by tackling (23). For the gossip-based AO algorithm, the MSE is evaluated as the maximum of the MSE errors of $\{\hat{\boldsymbol{x}}_p[n]\}_p$, i.e., $\text{MSE} = \max_p \text{MSE}_p$.

The first numerical example illustrates how the performance of the gossip-based AO algorithm corroborate Theorem 1. We

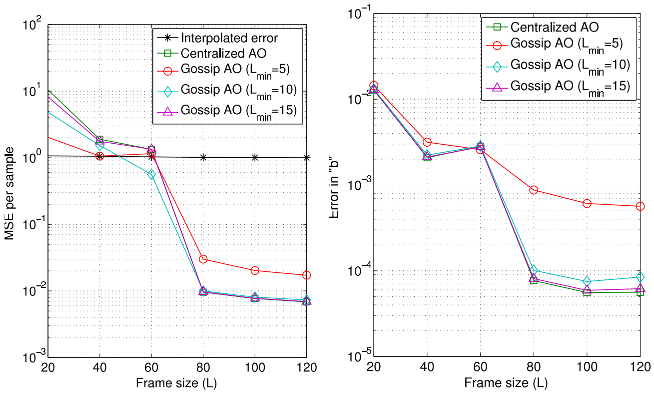

Fig. 2: The MSE performance against the frame size L with different algorithm for tackling (23). (Left) the MSE performance on estimating $\boldsymbol{x}[i]$. (Right) the MSE performance on estimating b_p.

consider an instance of (23) and tackle it using the centralized AO and gossip-based AO algorithms. Subsequently, in Fig. 1 we compare the objective values versus the iteration number of different cases. We observe that the objective value of the gossip-based AO algorithm converges to a value close to that achieved by the centralized AO algorithm. Moreover, the difference between the (converged) objective values decreases rapidly as the number of G-AC steps (ℓ_{min}) increases. This agrees with the conclusions in Theorem 1, where the difference between the converged solution in gossip-based AO and a local minimum shrinks exponentially with ℓ_{min}. We also observe that the convergence is generally fast, where a reasonable solution is found within the first 100 iterations.

The second numerical example aims at showing the regression problem (23) can be used to recover the sensor field $\boldsymbol{x}_c(t)$ from the asynchronous sub-Nyquist samples. Specifically, we compare the MSE performance (of both $\boldsymbol{x}[i]$ and b_p) of the proposed algorithms. The algorithms are terminated when the number of iterations reaches 200. We also benchmark the MSE evaluated with first performing a linear regression on the sub-Nyquist samples and then interpolating the obtained samples, labeled as 'interpolated error'. The MSE errors are obtained with 100 Monte-Carlo trials.

The numerical results are shown in Fig. 2. We observe that the MSE performance in the estimation of the time offsets b_p improves as the frame size L increases. It is also observed that the MSE hits an error floor with L exceeding 80, whose most plausible cause is the residual modeling error introduced by the DFT approximation. Lastly, we compare the MSE performance with centralized AO and gossip-based AO. We see that the decentralized algorithm achieves a similar performance as its centralized counterpart when the number of G-AC steps exceeds 10.

VI. Conclusion

In this paper, we have proposed a new decentralized algorithm for optimization problems with nuisance local parameters. The proposed algorithm combines the AO strategy with gossip-based average consensus. The conditions for convergence of the proposed algorithm is also provided. As an application, we have shown that the decentralized regression problem under asynchronous and sub-Nyquist sampling can be cast into a special case of the interested type of optimization problem. The simulation results verify the correctness of the theory and demonstrates that the regression problem can be solved under the impact of imperfect sampling devices.

References

[1] 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.

[2] 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.

[3] 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.

[4] 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.

[5] 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.

[6] 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.

[7] 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.

[8] 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.

[9] G. Mateos, J. A. Bazerque, and G. B. Giannakis, "Distributed Sparse Linear Regression," *IEEE Trans. Signal Process.*, vol. 58, no. 10, pp. 5262–5276, Oct. 2010.

[10] 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.

[11] ——, "Consensus on state and time: Decentralized regression with asynchronous sampling," *Submitted to IEEE Trans. Signal Process.*, Aug. 2014.

[12] 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.

[13] 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.

[14] 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.

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

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

[17] A. V. Oppenheim, R. W. Schafer, and J. R. Buck, *Discrete-time Signal Processing (2nd Ed.)*. Prentice-Hall, Inc., 1999.