Chalmers Publication Library

Compressed Sensing in Wireless Sensor Networks without Explicit Position Information

(article starts on next page)

Compressed Sensing in Wireless Sensor Networks without Explicit Position Information

Christopher Lindberg, Alexandre Graell i Amat, *Senior Member, IEEE*,
and Henk Wymeersch, *Member, IEEE*

Abstract—Reconstruction in compressed sensing relies on knowledge of a sparsifying transform. In a setting where a sink reconstructs a field based on measurements from a wireless sensor network, this transform is tied to the locations of the individual sensors, which may not be available to the sink during reconstruction. In contrast to previous works, we do not assume that the sink knows the position of each sensor to build up the sparsifying basis. Instead, we propose the use of spatial interpolation based on a predetermined sparsifying transform, followed by random linear projections and ratio consensus using local communication between sensors. For this proposed architecture, we upper bound the reconstruction error induced by spatial interpolation, as well as the reconstruction error induced by distributed compression. These upper bounds are then utilized to analyze the communication cost tradeoff between communication to the sink and sensor-to-sensor communication.

I. INTRODUCTION

WIRELESS sensor networks (WSNs) have received a lot of attention due to the possibility of dense deployment of sensors with communication capabilities to cooperatively sense and monitor a physical phenomenon of interest. Most works in the technical literature consider the WSN to be energy-constrained, and focus on developing energy-efficient sensing strategies. However, smart phones and similar devices carrying powerful batteries have become ubiquitous and may serve as a platform for WSNs [1], in which energy is not the limiting resource. The number of sensors in large WSNs is typically in the order of hundreds to several thousands. Consequently, this makes sending the sensed data from the network to the sink expensive in terms of bandwidth. Hence, there is a need for in-network compression to decrease the bandwidth cost of conveying the information to the sink. There exist several techniques for in-network compression. Examples include Slepian-Wolf coding [2], [3], which uses the fact that sensor measurements are spatially and/or temporally correlated; distributed transform coding using the Karhunen-Loève transform [4]; or wavelet transforms [5]. Approaches relying on the theory of compressed sensing (CS) [6]–[8] have recently gained attention due to their scalability in the number of sensors, and the fact that they do not require precise models of the monitored phenomenon, only that the phenomenon can

be described with a limited number of components in a suitable basis.

Prior work on CS-based in-network compression and reconstruction at a sink can be divided into three categories. First of all, distributed compressed sensing (DCS) [9], [10] utilizes correlation present in sensor measurements by representing the measured field as a sum of a sparse common component and a sparse innovation component unique to each sensor. The second category is compressive wireless sensing (CWS) [11]–[13]. This approach lets every sensor first compute its contribution to the complete CS compression. The sensors then send their contributions to the sink simultaneously, usually by analog transmission. The contributions are directly added at the sink by superposition of the waveforms. Finally, the third category involves DCS with pre-distribution using consensus or gossip algorithms [14], [15]. Similar to CWS, the sensors first calculate their own contribution to the CS compression. Then, these contributions are spread in the network by a consensus or gossip algorithm.

Comparing the three categories in terms of communication cost, we observe the following. DCS is not well suited for scenarios where the goal is to decrease the amount of communication between sensors and the sink, since all sensors need to communicate with the sink. CWS is performed either by analog transmissions between sensors and the sink [11], [12], which requires tight synchronization and channel state information, or by forming a spanning tree of the sensors [13] to route the information towards the sink, which does not scale well in the number of sensors. Finally, DCS with pre-distribution (distribution of compression by linear projections before transmission to the sink) relies largely on transmissions between sensors, but since the measurements are spread in the network, only few sensors need to be queried by the sink. This allows for a tradeoff between communication to the sink and communication among sensors. Common to all three categories is that, in order to find the sparsifying basis, position information from all sensors is required at the sink, which defeats the purpose of decreasing the transmission cost between sensors and the sink, as this information scales linearly with the number of sensors.

In this paper, we propose a novel variation of DCS with pre-distribution, where the sensor locations are unknown to the sink. The sensors measure the intensity of a spatial field in their respective locations, and then locally interpolate the field at certain predetermined locations, known a priori by the sink and the network. The sensors then compress and distribute the interpolated data, after which a limited number

of sensors transmit it to the sink, which reconstructs the data using the basis pursuit denoising (BPDN) algorithm. Our contributions are (i) a framework to use DCS with pre-distribution for in-network compression of measurement data, without the requirement of complete position knowledge at the sink; (ii) an analytical description of the reconstruction error induced by spatial interpolation, based on nearest neighbor sensing, in in-network compression; (iii) a study of the tradeoff between local and global communication to achieve a target reconstruction error.

Notation: We use boldface lowercase letters $\boldsymbol{x}$ for column vectors, and boldface uppercase letters $\boldsymbol{X}$ for matrices. In particular, $\boldsymbol{I}_M$ denotes an $M \times M$ identity matrix, $\boldsymbol{1}_{N \times 1}$ is the all-one vector of size $N \times 1$, and $\boldsymbol{0}_{N \times 1}$ is the all-zero vector of size $N \times 1$. The (i,j)th element of a matrix $\boldsymbol{X}$ is expressed as $[\boldsymbol{X}]_{i,j}$ or X_{ij}. Sets are described by calligraphic letters $\mathcal{X}$. The cardinality of a set is denoted by $|\mathcal{X}|$. The transpose of a vector or matrix is denoted by $[\cdot]^\mathsf{T}$. Expectation and variance (or covariance matrix) of a random variable with respect to the probability measure P is denoted by $\mathbb{E}_P \{\cdot\}$ and $\mathrm{Var}_P (\cdot)$, respectively.

II. System Model

We consider a two-dimensional WSN consisting of N sensors spread uniformly in a square area with sensor density ρ. The sensors measure the intensity of a random spatial field $x(\boldsymbol{s})$, where $x : \mathbb{R}^2 \to \mathbb{R}$ and $\boldsymbol{s}$ denotes a generic location. The spatial field $x(\boldsymbol{s})$ is considered to be low-pass, wavenumber-limited (equivalent of band-limited for multidimensional signals), and to have an isotropic spectrum. Hence, letting $\boldsymbol{f} = [f_X, f_Y]^\mathsf{T}$ denote the spatial frequency, then the power spectral density (PSD) $W(\boldsymbol{f})$ of $x(\boldsymbol{s})$ is a function of $\|\boldsymbol{f}\|_2$ and

$$W(\boldsymbol{f}) = 0, \ \ \|\boldsymbol{f}\|_2 \geq f_u, \tag{1}$$

where f_u is the maximum spatial frequency. Without loss of generality, we normalize the area in which we want to reconstruct the field to be $\Omega_{\mathrm{field}} = [0,1] \times [0,1]$, while the sensors are deployed in $\Omega_{\mathrm{network}} = [-\delta, 1+\delta] \times [-\delta, 1+\delta]$, $\delta > 0$. The coordinate of sensor k, $k = 0, \ldots, N-1$, is denoted $\boldsymbol{s}_k \in \Omega_{\mathrm{network}}$. The measurement of sensor k is given by

$$z_k = x(\boldsymbol{s}_k) + n_k, \tag{2}$$

where $x(\boldsymbol{s}_k)$ is the intensity of the field in location $\boldsymbol{s}_k$, and n_k is a spatially white Gaussian noise sample with variance σ_n^2. The quantities $x(\boldsymbol{s}_k)$, z_k, and n_k are stacked in vectors $\boldsymbol{x} = [x(\boldsymbol{s}_0), \ldots, x(\boldsymbol{s}_{N-1})]^\mathsf{T}$, $\boldsymbol{z} = [z_0, \ldots, z_{N-1}]^\mathsf{T}$, and $\boldsymbol{n} = [n_0, \ldots, n_{N-1}]^\mathsf{T}$, respectively.

Sensors can communicate with neighbors within a distance R_{comm} and with the sink (or base station). Communication to the sink incurs a fixed communication cost C_{glo}, while inter-sensor communication incurs a communication cost C_{loc}. For later use, we denote by $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ the undirected graph describing the network, where $\mathcal{V}$ is the set of sensor nodes and $\mathcal{E}$ the set of edges connecting the sensor nodes. Additionally, the set of neighboring sensors to sensor i is defined as $\mathcal{M}_i = \{j : (i,j) \in \mathcal{E}\}$. Associated with the graph $\mathcal{G}$

we have the adjacency matrix $\boldsymbol{E}$, where $E_{ij} = 1$ if $(i,j) \in \mathcal{E}$ and zero otherwise, and the degree matrix $\boldsymbol{D} = \mathrm{diag}(\boldsymbol{E1})$. In order for the graph $\mathcal{G}$ to be connected with high probability, i.e., there exists a path from every node to every other node, we assume $R_{\mathrm{comm}} \geq \sqrt{4.52/\rho\pi}$ [16]. In order to achieve good reconstruction, we assume $\rho \geq \pi f_u^2$ [17].

The sink is assumed to know the statistical properties of the field, the number of nodes N, and the measurement variance σ_n^2. The sink does not know the precise locations of all of the nodes. Furthermore, the sink can broadcast information to all sensors without any communication cost.

Goal: Given the observations $\boldsymbol{z}$ and the system model, the goal is to reconstruct the sampled field at the sink such that a certain reconstruction error can be guaranteed. In addition, $\boldsymbol{z}$ should be compressed in such a way that the compression scheme scales well in the number of sensors and that the sink does not need to know each individual $\boldsymbol{s}_k$.

III. Compressed Sensing

In this section, we provide the basics of CS and highlight challenges in its use in wireless networks.

A. Compressed Sensing Basics

Let $\boldsymbol{A} \in \mathbb{R}^{M \times N}$ be an $M \times N$ matrix with $M \ll N$, and let $\boldsymbol{x} \in \mathbb{R}^{N \times 1}$ be a vector of samples of a field $x(\boldsymbol{s})$, i.e., $\boldsymbol{x} = [x(\boldsymbol{s}_0), \ldots, x(\boldsymbol{s}_{N-1})]^\mathsf{T}$. We define the noisy compression of $\boldsymbol{x}$

$$\boldsymbol{y} = \boldsymbol{Ax} + \boldsymbol{e}, \tag{3}$$

where $\boldsymbol{e} = \boldsymbol{Am}$, in which $\boldsymbol{m}$ is a vector whose entries are spatially white noise samples with variance σ_i^2, $i = 0, \ldots, N-1$. Since $M \ll N$, recovering $\boldsymbol{x}$ from $\boldsymbol{y}$ is an ill-posed problem, as there are infinitely many vectors $\boldsymbol{x}^\star$ that satisfy $\boldsymbol{y} = \boldsymbol{Ax}^\star$. However, if $\boldsymbol{x}$ is sparse, or has a sparse representation in some transform basis, the theory of CS enables the recovery of $\boldsymbol{x}$. Assume that $\boldsymbol{x}$ is such that there exists a transformation $\boldsymbol{T}$ such that

$$\boldsymbol{\theta} = \boldsymbol{Tx}, \tag{4}$$

where $\boldsymbol{\theta}$ is K-sparse, i.e., contains no more than K nonzero entries. Assume further that $\boldsymbol{AT}^{-1}$ satisfies the restricted isometry property (RIP) [6] of order $2K$, i.e., there exists a constant $\delta_{2K} \geq 0$ such that the following inequality holds for all $2K$-sparse vectors $\boldsymbol{u}$

$$(1 - \delta_{2K})\|\boldsymbol{u}\|_2^2 \leq \|\boldsymbol{AT}^{-1}\boldsymbol{u}\|_2^2 \leq (1 + \delta_{2K})\|\boldsymbol{u}\|_2^2. \tag{5}$$

In general, it is hard to determine if a given matrix satisfies the RIP. However, in [6] and [18] it was shown that if $\boldsymbol{AT}^{-1}$ is a Gaussian random matrix with independent and identically distributed (i.i.d.) entries $[\boldsymbol{AT}^{-1}]_{i,j} \sim \mathcal{N}(0, 1/M)$, then $\boldsymbol{AT}^{-1}$ satisfies the RIP with high probability (w.h.p.). If $\boldsymbol{T}^{-1}$ is a unitary matrix, we can let the entries of $\boldsymbol{A}$ be i.i.d. and distributed as $A_{ij} \sim \mathcal{N}(0, 1/M)$ in order for $\boldsymbol{AT}^{-1}$ to satisfy the RIP w.h.p. In the remainder of the paper we assume that the sparsifying transform $\boldsymbol{T}$ is such that $\boldsymbol{T}^{-1}$ is unitary. Additionally, the event that $\boldsymbol{AT}^{-1}$ does not satisfy

RIP is neglected since the probability of this happening is exponentially decaying in M [6].

We now consider the following convex ℓ_1-minimization problem, called BPDN [7],

$$\text{minimize } \|\boldsymbol{\theta}\|_1 \tag{6a}$$

$$\text{subject to } \|\boldsymbol{AT}^{-1}\boldsymbol{\theta} - \boldsymbol{y}\|_2^2 \leq \varepsilon^2, \tag{6b}$$

where ε is the denoising parameter. When $\delta_{2K} < \sqrt{2} - 1$ and $\varepsilon^2 \geq \|\boldsymbol{e}\|_2^2$, we obtain so-called stable recovery, and the ℓ_2-norm of the reconstruction error of BPDN can be shown to be upper bounded as [7], [19]

$$\|\boldsymbol{x} - \boldsymbol{x}^\star\|_2 \leq \frac{C_0}{\sqrt{K}}\|\boldsymbol{T}^{-1}(\boldsymbol{\theta} - \boldsymbol{\theta}_K)\|_1 + C_1\varepsilon, \tag{7}$$

where $\boldsymbol{x}^\star = \boldsymbol{T}^{-1}\boldsymbol{\theta}^\star$, in which $\boldsymbol{\theta}^\star$ is the solution to (6), $\boldsymbol{\theta}_K$ is the best K-sparse approximation of $\boldsymbol{\theta}$, and $C_0, C_1 \geq 0$ are constants [7].

We observe that $\|\boldsymbol{e}\|_2^2$ is unknown to the sink, but provided that N and M are sufficiently large, $\boldsymbol{e}$ can be approximated by a zero-mean Gaussian random variable with covariance matrix $\bar{\sigma}^2 N/M \boldsymbol{I}_M$, where $\bar{\sigma}^2 = 1/N \sum_{i=1}^{N} \sigma_i^2$. Hence, $\|\boldsymbol{e}\|_2^2$ has a $\chi^2(M)$-distribution with $\mathbb{E}\left\{\|\boldsymbol{e}\|_2^2\right\} = N\bar{\sigma}^2$ and $\text{Var}\left(\|\boldsymbol{e}\|_2^2\right) = 2N^2\bar{\sigma}^4/M$. Thus, we can set

$$\varepsilon^2 = N\bar{\sigma}^2\left(1 + \mu\sqrt{2/M}\right). \tag{8}$$

where μ is a confidence level [7, eq. (17)], which serves as a tuning to the denoising parameter in the BPDN reconstruction algorithm, where a higher μ means that we hedge against higher perturbation levels of $\boldsymbol{y}$ at the expense of reconstruction performance (when the perturbation does not exceed ε^2).

B. Compressed Sensing in WSNs

CS in WSNs faces two fundamental challenges: first of all, the sparsifying transform $\boldsymbol{T}$ depends on the positions of the individual sensors, with typical examples including the wavelet or Fourier transforms [12], [20]. Secondly, the sensors must compute $\boldsymbol{y}$ and send it to the sink, in such a way that it is less costly than sending the individual sensor measurements to the sink. The latter challenge was addressed in [11]–[13] through analog transmission and in [14], [15] using pre-distribution using consensus. However, the former challenge remains and will be the focus of this paper.

IV. PROPOSED FRAMEWORK

In this section, we give a high-level description of the framework that we propose for DCS with pre-distribution *without knowledge of the locations of all sensors at the sink*. The framework is shown in Fig. 1, and comprises the following steps. First, the sink decides on a sparsifying transform $\boldsymbol{T}$, defines N_G locations $\tilde{\boldsymbol{s}} = [\tilde{\boldsymbol{s}}_0, \ldots, \tilde{\boldsymbol{s}}_{N_G-1}]^\mathsf{T}$, where $\tilde{\boldsymbol{s}}_i \in \Omega_{\text{field}}$, associated with the transform $\boldsymbol{T}$, and broadcasts a description of the desired locations to all sensors (step A in Fig. 1). These N_G locations should (i) have an associated sparsifying transform, related to the sparsity of the scenario (e.g., a grid with a Fourier transform would be suitable in our scenario); (ii) have a compact description (e.g., an origin, a horizontal
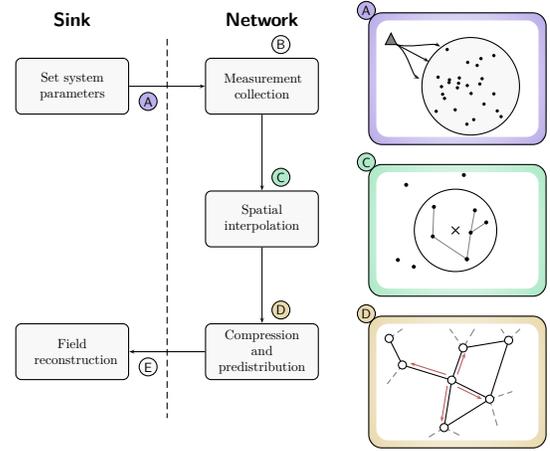
Figure 1. Overview of the steps performed by the network and the sink. Box (A) illustrates the sink distributing the system parameters to the WSN; box (B) represents the measurement collection phase, at locations unknown to the sink; box (C) shows the interpolation phase, and how a subset of the sensors form a spanning tree to interpolate the field at a given location; box (D) depicts the compression and pre-distribution phase, and the network over which the sensors exchange local information.

resolution, and a vertical resolution for a two-dimensional grid) so that the broadcast overhead is independent of N_G or N. Then, the sensors measure the field in their respective locations $\boldsymbol{s}_k$, which are unknown to the sink (step B in Fig. 1). In step C, the sensors collectively determine the measurements at the desired locations through spatial interpolation. This leads to a virtual measurement at coordinate $\tilde{\boldsymbol{s}}_k$

$$\hat{z}_k = x(\tilde{\boldsymbol{s}}_k) + m_k, \tag{9}$$

where m_k is the combined measurement and interpolation error when interpolating the field at $\tilde{\boldsymbol{s}}_k$. The values $\hat{z}_k, x(\tilde{\boldsymbol{s}}_k)$, and m_k are stacked in the vectors $\hat{\boldsymbol{z}} = [\hat{z}_0, \ldots, \hat{z}_{N_G-1}]^\mathsf{T}$, $\hat{\boldsymbol{x}} = [x(\tilde{\boldsymbol{s}}_0), \ldots, x(\tilde{\boldsymbol{s}}_{N_G-1})]^\mathsf{T}$, and $\boldsymbol{m} = [m_0, \ldots, m_{N_G-1}]^\mathsf{T}$, respectively. In step D in Fig. 1, the sensors compress $\hat{\boldsymbol{x}}$ to $\boldsymbol{y}$ with the measurement matrix $\boldsymbol{A} \in \mathbb{R}^{M \times N_G}$ in a distributed manner using a specialized consensus algorithm. Finally, the sink queries a subset of the sensors for their compressed estimates, which the sink combines and uses to reconstruct the sampled field using the theory described in Section III. The querying of the sensors is done by sampling uniformly at random from the N sensors, known to sink, and then broadcasting this request to the sensors.

The framework above relies on spatial interpolation and consensus, both of which affect the compression perturbation $\boldsymbol{e}$ in (3), and thus also the choice of the denoising parameter ε in (6). In the subsequent sections, we will detail how ε should be selected to enable stable field reconstruction.

V. SPATIAL INTERPOLATION

In order to estimate the values at the predetermined locations $\tilde{\boldsymbol{s}}$, we rely on spatial interpolation [21]. The design and properties of the interpolator depends on the statistics of the underlying field. In this section, we will describe the statistics using the *semivariogram*, which is subsequently used to design spatial interpolators. Finally, we will quantify the interpolation error.

A. Semivariogram

Most techniques for spatial interpolation rely on the assumption of stationarity of the random process. We consider a weak type of stationarity for isotropic processes: *intrinsic stationarity* [21], whereby $x(\boldsymbol{s}_1) - x(\boldsymbol{s}_2)$ is second-order stationary, which implies that it has a zero mean and a covariance that depends only on $h = \|\boldsymbol{s}_1 - \boldsymbol{s}_2\|_2$. The field is defined by the variogram $2\gamma(\|\boldsymbol{s}_1 - \boldsymbol{s}_2\|_2) = \mathrm{Var}(x(\boldsymbol{s}_1) - x(\boldsymbol{s}_2))$. The semivariogram $\gamma(h)$ can be related to the covariance $C(h)$ of the field by $\gamma(h) = C(0) - C(h)$, whenever the covariance function exists. Semivariograms provide a convenient way to model, describe, and predict spatial processes. There exist different models for the semivariogram, among which the *exponential*, *Gaussian*, *power,* and *spherical* are the most common. Due to the smooth nature of the fields we are concerned with, the semivariogram model we consider is the Gaussian semivariogram

$$\gamma(h) = \begin{cases} (\gamma_\infty - \gamma_0)\left(1 - e^{-h^2/h_r^2}\right) + \gamma_0, & h > 0 \\ 0 & h = 0, \end{cases} \quad (10)$$

where $\gamma_\infty = \lim_{h \to \infty} \gamma(h)$, γ_0 is used to account for the measurement noise, and h_r is a correlation distance. Note that $\gamma(h)$ has a discontinuity at $h = 0$. In practice, the parameters γ_0 and h_r of the semivariogram can be learned by the sensors in a distributed manner [22]. Here, we assume that each sensor has knowledge of $\gamma(h)$. In particular, we approximate $W(\boldsymbol{f})$, the PSD of the field, with a Gaussian, so that we can set $h_r \propto 1/f_u$ and $\gamma_0 = \sigma_n^2$.

B. Spatial Interpolation

Given a semivariogram, spatial interpolation operates as follows: an interpolated value $\hat{z}_k$ at a desired location $\tilde{\boldsymbol{s}}_k$ is obtained from a set $\mathcal{K}_k$ of measurements z_i with associated sensor locations $\boldsymbol{s}_i$ as a linear combination

$$\hat{z}_k = \sum_{i \in \mathcal{K}_k} v_i z_i, \quad (11)$$

where v_i's are the non-negative interpolation weights, which satisfy $\sum_{i \in \mathcal{K}_k} v_i = 1$. The weights are set to minimize the interpolation error variance, and exploit the spatial dependency of the field through the semivariogram. It can be shown that the prediction error variance is given by [21, eq. (3.2.17)]

$$\sigma_K^2 = 2 \sum_{i \in \mathcal{K}_k} v_i \bar{\gamma}_{ik} - \sum_{i,j \in \mathcal{K}_k} v_i v_j \gamma_{ij}, \quad (12)$$

where $\gamma_{ij} = \gamma(\|\boldsymbol{s}_i - \boldsymbol{s}_j\|_2)$, and $\bar{\gamma}_{ik} = \gamma(\|\boldsymbol{s}_i - \tilde{\boldsymbol{s}}_k\|_2)$. Hence, the sensors in the set $\mathcal{K}_k$ determine $\boldsymbol{v}$ by solving the following optimization problem

$$\text{minimize } \sigma_K^2 \quad (13a)$$
$$\text{s.t. } \mathbf{1}^\mathsf{T} \boldsymbol{v} = 1. \quad (13b)$$

Practically, this problem is solved with the method of Lagrangian multipliers. In the sensors, this boils down to solving a linear system of equations in a distributed fashion [21]. The set $\mathcal{K}_k$ will affect the performance of the interpolation, with more sensors in $\mathcal{K}_k$ leading to lower prediction variance but

also higher complexity and more communication overhead. In general, sensors at a distance greater than h_r away from $\tilde{\boldsymbol{s}}_k$ will not improve the quality of the interpolation. We will consider two choices for $\mathcal{K}_k$: $\mathcal{K}_k = \{i : \|\boldsymbol{s}_i - \tilde{\boldsymbol{s}}_k\|_2 \le R_{\text{int}}\}$ for some interpolation radius R_{int}; and nearest neighbor sensing (NNS) wherein $\mathcal{K}_k = \{\arg\min_i \|\boldsymbol{s}_i - \tilde{\boldsymbol{s}}_k\|_2\}$. Note that for NNS, $v_i = 1$ and $\sigma_{\text{NNS}}^2(\|\tilde{\boldsymbol{s}}_k - \boldsymbol{s}_i\|_2) = 2\gamma(\|\tilde{\boldsymbol{s}}_k - \boldsymbol{s}_i\|_2)$. To simplify the analysis, we will assume that $\mathcal{K}_k$ is not empty for any k and that the sensors in $\mathcal{K}_k$ form a connected network. How removing these assumptions affects the framework is revisited at the end of Appendix D. The communication overhead associated with spatial interpolation scales as $\mathcal{O}(R_{\text{int}}^2 \rho)$ for each location $\tilde{\boldsymbol{s}}_k$.

C. Interpolation Error

We have seen in Section III-A that the sink does not need to know the interpolation error σ_K^2 for each individual point $\tilde{\boldsymbol{s}}_k$, but only the average. This average can be computed with a standard consensus scheme. For NNS, the average error can be obtained in closed-form.

Lemma 1. *Assume a random distance to the nearest sensor $d_0 \in [0, \infty)$. Using the NNS to interpolate the given grid points with the semivariogram given by (10), the average interpolation error is*

$$\mathbb{E}_{d_0}\left\{\sigma_{\text{NNS}}^2\right\} = \frac{2(\gamma_0 - \gamma_\infty)\pi\rho}{1/h_r^2 + \pi\rho} + 2\gamma_\infty \triangleq \phi. \quad (14)$$

Proof: The proof is given in Appendix A. ∎

We observe that when $\rho \gg 1/(h_r^2 \pi)$, then $\phi \approx 2\gamma_0$, so for a sufficiently large density of sensors, the spatial interpolation error will be dominated by the measurement error.

VI. Ratio Consensus

After the interpolation phase, each sensor i has knowledge of $\hat{z}_k$ and $|\mathcal{K}_k|$, for all $\{k : i \in \mathcal{K}_k\}$. Hence a sensor i may have access to multiple $\hat{z}_k$ or no z_k at all (when $i \notin \mathcal{K}_k, \forall k$). A value $\hat{z}_k$ can also be known to multiple sensors. Due to the assumption that $\mathcal{K}_k$ is never empty, each $\hat{z}_k$ is known to at least one sensor. We recall that $\hat{\boldsymbol{z}} = [\hat{z}_0, \hat{z}_1, \ldots, \hat{z}_{N_G-1}]^\mathsf{T}$ is the vector of stacked interpolated measurements.

At this stage, any local measurement made by sensor i is discarded. The compression by linear projections of the interpolated measurements $\hat{z}_k$'s is now

$$\hat{\boldsymbol{y}} = \boldsymbol{A}\hat{\boldsymbol{z}} = \boldsymbol{A}\hat{\boldsymbol{x}} + \boldsymbol{e}_{\text{int}}, \quad (15)$$

where $\hat{\boldsymbol{x}}$ is the true value of the interpolated field and $\boldsymbol{e}_{\text{int}} = \boldsymbol{A}\boldsymbol{m}$ is the perturbation due to interpolation error (cf. (3)). The expression in (15) can be rewritten as

$$\hat{\boldsymbol{y}} = N_G \bar{\boldsymbol{w}}, \quad (16)$$

where

$$\bar{\boldsymbol{w}} = \frac{1}{N_G} \sum_{k=0}^{N_G-1} \boldsymbol{w}_k,$$

and where $\boldsymbol{w}_k = \boldsymbol{a}_k \hat{z}_k$, in which $\boldsymbol{a}_k$ is the kth column of $\boldsymbol{A}$. The expression in (16) includes the average $\bar{\boldsymbol{w}}$ of the N_G vectors $\boldsymbol{w}_i$, which allows us to compute $\hat{\boldsymbol{y}}$ using an

average consensus algorithm. A standard choice of consensus algorithm to solve this kind of problem is, e.g., using a Perron matrix as in [23], which assumes that all sensors have one value each, and that this value is unique. However, since a sensor may have access to a single $\hat{z}_k$, multiple $\hat{z}_k$'s, or no $\hat{z}_k$ at all, standard average consensus is not applicable. Here, we compute $\bar{w}$ by *ratio consensus*[1] [24]–[26].

A. Ratio Consensus Operation

Ratio consensus can be applied to scenarios where sensors compute the average of a set of values, and where each value may be associated with more than one sensor and one sensor may be associated with more than one value or even no value. Ratio consensus comprises two parallel consensus algorithms, whereby the average is achieved by taking the ratio of the two resulting values. The two initial values of sensor i are

$$\boldsymbol{\beta}_i(0) = \sum_{k:i \in \mathcal{K}_k} \frac{\boldsymbol{w}_k}{|\mathcal{K}_k|} \qquad (17)$$

$$\alpha_i(0) = \sum_{k:i \in \mathcal{K}_k} \frac{1}{|\mathcal{K}_k|}, \qquad (18)$$

with the tacit assumption that summations over empty sets are set to zero. The initialization ensures that $\sum_{i=0}^{N} \boldsymbol{\beta}_i(0) = \sum_{i=0}^{N_G-1} \boldsymbol{w}_i$ and $\sum_{i=0}^{N} \alpha_i(0) = N_G$. At each iteration l, every node updates $\boldsymbol{\beta}_i$ and α_i as follows

$$\boldsymbol{\beta}_i(l) = \frac{\boldsymbol{\beta}_i(l-1)}{|\mathcal{M}_i|+1} + \sum_{j \in \mathcal{M}_i} \frac{\boldsymbol{\beta}_j(l-1)}{|\mathcal{M}_j|+1}, \qquad (19)$$

and

$$\alpha_i(l) = \frac{\alpha_i(l-1)}{|\mathcal{M}_i|+1} + \sum_{j \in \mathcal{M}_i} \frac{\alpha_j(l-1)}{|\mathcal{M}_j|+1}, \qquad (20)$$

where $\mathcal{M}_i$ denotes the set of neighboring sensors to sensor i. We can express (19)–(20) for the whole network in matrix form as $\boldsymbol{B}(l) = \boldsymbol{P}\boldsymbol{B}(l-1) = \boldsymbol{P}^l \boldsymbol{B}(0)$ and $\boldsymbol{\alpha}(l) = \boldsymbol{P}\boldsymbol{\alpha}(l-1) = \boldsymbol{P}^l \boldsymbol{\alpha}(0)$, where $\boldsymbol{P} = (\boldsymbol{I}_N + \boldsymbol{E})(\boldsymbol{I}_N + \boldsymbol{D})^{-1} \in \mathbb{R}^{N \times N}$ is the consensus matrix defined by the updating equations, $\boldsymbol{B}(l) = [\boldsymbol{\beta}_0(l), \ldots, \boldsymbol{\beta}_{N-1}(l)]^\mathsf{T} \in \mathbb{R}^{N \times M}$, and $\boldsymbol{\alpha}(l) = [\alpha_0(l), \ldots, \alpha_{N-1}(l)]^\mathsf{T} \in \mathbb{R}^{N \times 1}$. The ratio consensus estimate at iteration l is defined as

$$\hat{\boldsymbol{W}}(l) = \boldsymbol{P}^l \boldsymbol{B}(0) \oslash \boldsymbol{P}^l \boldsymbol{\alpha}(0) \mathbf{1}_{1 \times M}, \qquad (21)$$

where $\oslash$ denotes the Hadamard division, i.e., element-wise division of two matrices, and $\hat{\boldsymbol{W}}(l) \in \mathbb{R}^{N \times M}$. It should be noted that while (19)–(20) are defined for all $l > 0$, (21) is only defined for $l \geq \kappa$, where κ is the first iteration for which $\alpha_i(l) > 0$ for all i. The value of κ is related to the maximum number of hops separating a sensor with no initial information to its nearest sensor with initial information. The ith row of $\hat{\boldsymbol{W}}(l)$ serves as sensor i's approximation of $\bar{w}$ at iteration l. The convergence and tightness of this approximation is discussed below.

[1]Sometimes also referred to as sum-weight consensus, or push-sum consensus.

B. Convergence Properties

1) Convergence: By following the updating procedure described in the previous subsection, $\boldsymbol{P}$ is guaranteed to be a column stochastic matrix, i.e., $\sum_{i=1}^{N} P_{ij} = 1$ for all $j = 1, \ldots, N$, where $P_{ij} \triangleq [\boldsymbol{P}]_{i,j}$. This means that all columns converge to a stationary distribution $\boldsymbol{\pi}$, i.e., $\lim_{l \to +\infty} \boldsymbol{P}^l = \boldsymbol{\pi} \mathbf{1}_{1 \times N}$. Since the consensus algorithm operates independently on every compression dimension $m = 1, \ldots, M$, we show that the algorithm converges on one dimension. Let $\boldsymbol{b}(0)$ be the mth column of $\boldsymbol{B}(0)$. Then, for the consensus update of the mth element of $\hat{\boldsymbol{w}}_i(l)$ in every node i

$$\lim_{l \to +\infty} \hat{w}_{i,m}(l) = \lim_{l \to +\infty} \frac{\left[\boldsymbol{P}^l \boldsymbol{b}(0)\right]_i}{\left[\boldsymbol{P}^l \boldsymbol{\alpha}(0)\right]_i} \qquad (22)$$

$$= \frac{\lim_{l \to +\infty} \sum_{j=1}^{N} P_{ij}^l b_j(0)}{\lim_{l \to +\infty} \sum_{j=1}^{N} P_{ij}^l \alpha_j(0)} \qquad (23)$$

$$\overset{\text{(a)}}{=} \frac{\pi_i \sum_{j=1}^{N} b_j(0)}{\pi_i \sum_{j=1}^{N} \alpha_j(0)} = \frac{\sum_{j=0}^{N_G-1} w_j}{N_G}, \qquad (24)$$

where π_i is the ith element of the stationary distribution $\boldsymbol{\pi}$. In (a) above, we used the fact that all columns of $\boldsymbol{P}$ converge to the stationary distribution $\boldsymbol{\pi}$. Since (24) holds for all dimensions m, the ratio consensus algorithm converges to the average of the projected virtual measurements, i.e.,

$$\lim_{l \to \infty} \boldsymbol{B}(l) \oslash \boldsymbol{\alpha}(l) \mathbf{1}_{1 \times M} = \frac{\mathbf{1}_{N \times 1}}{N_G} \sum_{k=0}^{N_G-1} \boldsymbol{w}_k^\mathsf{T} = \mathbf{1}_{N \times 1} \bar{\boldsymbol{w}}^\mathsf{T}. \qquad (25)$$

Thus, in the limit $l \to +\infty$, all sensors have access to $\bar{w}$.

2) Convergence Rate: Now that we know that the algorithm converges to the desired value, we upper bound the speed at which it converges. The following lemma upper bounds the convergence rate of the ratio consensus algorithm as a function of the number of iterations l.

Lemma 2. *Let $\boldsymbol{P}$ be the column stochastic ratio consensus matrix of a fixed, connected, and undirected graph, and let λ_2 be the second largest eigenvalue of $\boldsymbol{P}$. Then, the ℓ_2-norm of the error of the ratio consensus algorithm after $l \geq \kappa$ iterations is upper bounded by*

$$\|\boldsymbol{b}(l) \oslash \boldsymbol{\alpha}(l) - (\bar{b}/\bar{\alpha})\mathbf{1}\|_2 \leq \left|\lambda_2^l\right| \|\mathbf{1} \oslash \boldsymbol{P}^l \boldsymbol{\alpha}(0)\|_2$$
$$\times \left(\|\boldsymbol{b}(0) - \bar{b}\boldsymbol{\pi}\|_2 + \frac{|\bar{b}|}{|\bar{\alpha}|}\|\boldsymbol{\alpha}(0) - N_G\boldsymbol{\pi}\|_2 \right)$$

where $\boldsymbol{\pi}$ is the limiting column distribution of $\lim_{l \to \infty} \boldsymbol{P}^l$, $\bar{b} = \sum_{k=0}^{N_G-1} w_{m,k}$ for the mth dimension of $\boldsymbol{w}_i$, and $\bar{\alpha} = \sum_{j=1}^{N} \alpha_j(0)$.

Proof: The proof is given in Appendix B. ∎

We note that after $l \geq \kappa$ iterations of ratio consensus (see Appendix C)

$$\left\|\mathbf{1} \oslash \boldsymbol{P}^l \boldsymbol{\alpha}(0)\right\|_2 \approx \frac{N^{3/2}}{N_G}. \qquad (26)$$

Remark regarding tightness of the bound: The bound in Lemma 2 describes how the error evolves as a function of the number of iterations l. The error is exponentially decaying

as a function of l, and the rate is determined by the second largest eigenvalue λ_2 of $\boldsymbol{P}$. The approximation in (26) turns out to be a good approximation for reasonable values of N and N_G. This is partly due to the fact that the consensus converges much faster than the eigenvalue bound states. After only a few iterations, most values are close to consensus, i.e., $1/\alpha_i(l) \approx N_G/N$. This is a consequence of the bound in Lemma 2 being loose in general, and the actual convergence rate depends on the structure of the initial $\boldsymbol{b}(0)$ and $\boldsymbol{\alpha}(0)$ within the eigenspace of $\boldsymbol{P}$.

VII. TRANSMISSION TO SINK AND RECONSTRUCTION

After a certain number of iterations $I \geq \kappa$ of ratio consensus, a set $\mathcal{L} \subseteq \mathcal{V}$ of $|\mathcal{L}| = L$ randomly chosen sensors communicate their estimates $\hat{\boldsymbol{w}}_i(I)$ of $\bar{\boldsymbol{w}}$. The sink estimates $\hat{\boldsymbol{y}}$ from the set of received packets as

$$\hat{\boldsymbol{y}} = N_G \left(\frac{1}{L} \sum_{i \in \mathcal{L}} \hat{\boldsymbol{w}}_i(I) \right), \tag{27}$$

which is assumed to be a good approximation of $\boldsymbol{A}\hat{\boldsymbol{x}} + \boldsymbol{e}_{\text{int}}$. Finally, $\hat{\boldsymbol{y}}$ is used in (6) to reconstruct $\hat{\boldsymbol{x}}$. The perturbed compression $\hat{\boldsymbol{y}}$ can be decomposed as $\hat{\boldsymbol{y}} = \boldsymbol{A}\hat{\boldsymbol{x}} + \boldsymbol{e}$, where $\boldsymbol{e} = \boldsymbol{e}_{\text{int}} + \boldsymbol{e}_{\text{cons}}$ is due to measurement noise, interpolation error, and consensus error.

A. Choice of Denoising Parameter

The following theorem states how the denoising parameter ε should be chosen in order to ensure that $\varepsilon^2 \geq \|\boldsymbol{e}\|_2^2$ with high probability while still providing a sparse solution to BPDN, and in turn that we can upper bound the reconstruction error as in (7).

Theorem 3. *Let $\boldsymbol{P}$ be the column stochastic ratio consensus matrix of a fixed, connected, and undirected graph, and let λ_2 be the second largest eigenvalue of $\boldsymbol{P}$. Let also all $\mathcal{K}_k$ be nonempty. Given the system model in Section II, $\boldsymbol{A}$ as described in Section III-A, the NNS, the compression in (27), and $1 \leq L \leq N$, the denoising parameter ε for stable field reconstruction using BPDN should be chosen such that*

$$\varepsilon^2 = \left(\frac{N_G}{M}\phi + \sigma_\Delta^2 \right) \left(M + \mu\sqrt{2M} \right), \tag{28}$$

where

$$\sigma_\Delta^2 = \frac{4}{L} \left(\frac{N-L}{N-1} \right) \frac{E_X + N_G\phi}{M} \lambda_2^{2I} N^3, \tag{29}$$

ϕ was defined in Lemma 1, $E_X = \|\hat{\boldsymbol{x}}\|_2^2$, and μ is the desired confidence level.

Proof: The proof is given in Appendix D. ∎

We observe that ε^2 comprises a contribution due to (i) measurement error and spatial interpolation error (related to ϕ) and (ii) ratio consensus and compression estimation (related to σ_Δ^2). The first term $N_G\phi/M$ can be made arbitrarily close to the measurement error (i.e., $2N_G\sigma_n^2/M$) by using a higher density of nodes, while the second term can be made arbitrarily small by employing more consensus iterations.

In the next section, we present an analysis of how this term behaves with respect to the different system parameters.
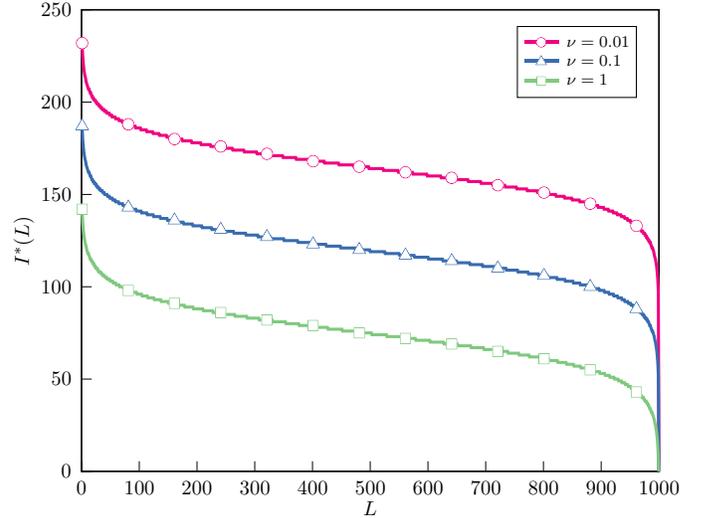
Figure 2. Minimal number of consensus iteration $I^*(L)$ as a function of the number of sensor-to-sink transmissions L, for different levels of $\nu \in \{0.01, 0.1, 1\}$.

B. Communication Cost Trade-off Analysis

1) Trade-off between ratio consensus iterations and number of queried nodes: In the error term ε, we have limited control over the interpolation error ϕ, other than increasing the node density. However, by tuning how many sensors transmit to the sink L, and choosing the number of consensus iterations I, we can reduce σ_Δ to a level so that the consensus error is negligible, i.e., such that $\sigma_\Delta \leq \nu\sqrt{N_G\phi/M}$, for some small value of $\nu \in (0, 1]$. For a fixed value of L, there is a minimal number of iterations

$$I^*(L) = \frac{\left\lceil \log\nu + \frac{1}{2}\log\left(\frac{M}{E_X + N_G\phi}\frac{L}{4N^3}\left(\frac{N-1}{N-L} \right) \right) \right\rceil}{\log\lambda_2} \tag{30}$$

needed to satisfy $\sigma_\Delta \leq \nu\sqrt{N_G\phi/M}$, with the understanding that $I^*(L) \geq 0$. Fig. 2 shows $I^*(L)$ for $\nu \in \{0.01, 0.1, 1\}$, fixing $E_X = 1000$, $N = 1000$, $N_G = 1000$, $\phi = 1$, and $\lambda_2 = 0.7$. We observe that the curves are rather flat, indicating that it is preferred to have a low value of L. Nevertheless, a too low value of L leads to a need to run many more consensus iterations. To reduce σ_Δ with an order of magnitude, it can be seen that we can employ two strategies: increasing the number of iterations with about 50, or drastically increasing the number of sensor-to-sink transmissions. This indicates that it is generally preferred to run more consensus iterations to reduce the impact of disagreement among nodes. For example, increasing the number of consensus iterations from 100 to 140 when $\nu = 0.1$, enables us to lower L from around 900 to around 100.

2) Trade-off between local and global communication: To obtain further insight into the communication trade-offs, we must include the communication cost. The (local) communication cost for performing the spatial interpolation is small compared to the cost of consensus, it is therefore ignored. Given a certain number of queried nodes L and a
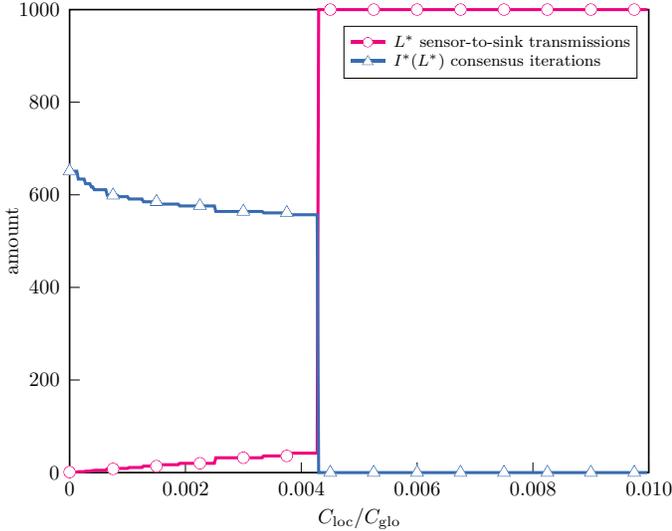
Figure 3. This plot shows the values of $I^*(L^*)$ and L^* for different values of the global-local communication cost ratio $C_{\text{loc}}/C_{\text{glo}}$.

Figure 4. Expected value of the ℓ_2-norm of the interpolation error variance as a function of ρh_r^2 of the NNS bound ϕ, simulations of NNS σ_{NNS}^2, and simulations of spatial interpolation using all nodes σ_K^2.

corresponding number of consensus iterations $I^*(L)$, the total communication cost is given by

$$C_{\text{tot}}(L) = C_{\text{glo}}L + C_{\text{loc}}\, I^*(L)\, N. \qquad (31)$$

For a given local-global cost ratio $C_{\text{loc}}/C_{\text{glo}}$, we can thus determine the number of nodes to query to minimize the total communication cost:

$$L^* = \arg\min_L \left[L + (C_{\text{loc}}/C_{\text{glo}})I^*(L)N \right]. \qquad (32)$$

In Fig. 3, we show the number of sensors transmitting to the sink L^* as a function of $C_{\text{loc}}/C_{\text{glo}}$, and the corresponding number of consensus iterations $I^*(L^*)$, of the cost-optimal communication strategy. We see a strong threshold effect: for low values of $C_{\text{loc}}/C_{\text{glo}}$, the cost-optimal communication strategy should be almost all sensor-to-sensor communication, the sink querying only a small fraction of the sensors, while after a threshold value of $C_{\text{loc}}/C_{\text{glo}}$, the cost-optimal communication strategy is to let all sensors transmit directly to the sink without any sensor-to-sensor communication. This threshold effect can be explained as follows. Assume $C_{\text{glo}} = 1$ is fixed and we vary C_{loc}. For very small C_{loc}, L^* will be small and monotonically increasing in C_{loc} (since $I^*(L)$ is monotonically decreasing in L). Hence, there exists a value of C_{loc} (say $C_{\text{loc}}^{\text{tip}}$) and thus of $L^*(C_{\text{loc}}^{\text{tip}}) < N$ for which $C_{\text{tot}}(L^*(C_{\text{loc}}^{\text{tip}}))$ will be slightly below N. When C_{loc} is increased slightly beyond this point, say to $C_{\text{loc}}^{\text{tip}} + \delta$, $L^*(C_{\text{loc}}^{\text{tip}} + \delta)$ will become N, since $I^*(N) = 1$, and suddenly it will become cheaper to have no local communication and instead let all sensors send directly to the sink.

VIII. RESULTS AND DISCUSSION

A. Evaluation of Bounds

In this section, we evaluate the tightness of the bounds derived in Lemma 1 and Lemma 2.

1) Spatial Interpolation: Fig. 4 shows the mean of the prediction error variance for NNS (i.e., σ_{NNS}^2), and the analytical expression ϕ from Lemma 1 as a function of ρh_r^2, for $h_r^2 = 0.1$, $2\gamma_0 = \sigma_n^2 = 0.01$, and $\gamma_\infty = 1$. Results were averaged over 400 Monte Carlo runs. While the bound is tight for NNS, there is a gap to more sophisticated interpolation strategies, which lead to much lower interpolation errors. As an extreme case, Fig. 4 also shows the mean of the prediction error variance for an interpolator that uses all sensor nodes. The performance was obtained by solving (13) and computing (12). As expected, the average prediction error variance (denoted by σ_K^2) is much lower than that of the NNS, which is due to the former method using all available information, while NNS only uses information from the closest sensor. For very low densities, σ_{NNS}^2 and σ_K^2 will tend to $2\gamma_\infty$, while for very high densities, σ_{NNS}^2 and σ_K^2 will tend to $2\gamma_0$, though they do so at different rates.

2) Ratio Consensus: We evaluate the actual performance of ratio consensus and compare with the bounds from Lemma 2. We consider a scenario with $N = 100$, and $N_G = 50$, within an area of $1\,\text{m} \times 1\,\text{m}$ (so that $\rho \approx 100$). We set $R_{\text{comm}} = 0.25\,\text{m}$, so that the network is connected with high probability. To have non-unique data association, we emulate the spatial interpolation procedure described in Section V-B. Data to sensor association is based on an interpolation radius of $R_{\text{int}} = 0.2\,\text{m}$. However, if an interpolation point has no sensor in its vicinity, the nearest sensor is associated with that data point. In Fig. 5, we plot the average ℓ_2-norm of the ratio-consensus error for different number of iterations. The ℓ_2-norm of the error is averaged over instances of a randomly generated sensor network and randomly generated data. We compare the performance to two versions of Lemma 2: one for which we use the simulations to evaluate $\|\mathbf{1} \oslash \boldsymbol{\alpha}(l)\|_2$, and the other where we use the approximation (26). From Fig. 5, it is clear that only after a few iterations, the error
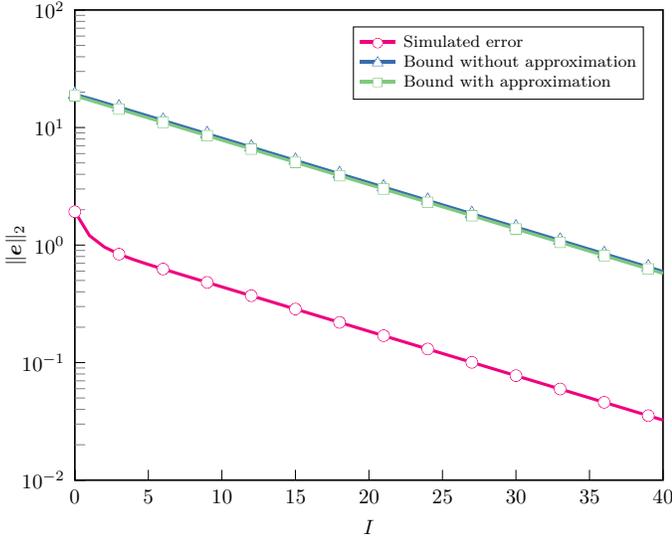
Figure 5. The average of the ℓ_2-norm of the ratio consensus error with respect to the number of iterations, for the bound in Lemma 2 with the approximation (26), without the approximation (26), and for simulations of ratio consensus.

Figure 6. The average ℓ_2-norm of the reconstruction error for the two approaches with respect to the position error variance.

in the simulation is dominated by λ_2, whereas in the first iterations, the error decreases much faster (see the discussion after (26)). Moreover, there is a gap between the simulated error and the upper bounds. The gap is due to two reasons: (i) we assume a worst case scenario for the error, i.e., it is aligned only in the part of the eigenspace corresponding to λ_2; (ii) in the derivation of the upper bound, we use the triangle inequality and Cauchy-Schwarz inequality on multiple occasions to bound ℓ_2-norms. Inspecting the figure, we deduce that the gap induced by bounding the ℓ_2-norms using Cauchy-Schwarz is dominant over the error from the eigenspace alignment, since the error drop in the first iterations of the simulated consensus is the smaller part of the gap. This implies that a practical system can use far fewer iterations than predicted using Lemma 2.

B. System-level Impact of Position Uncertainty

We generate a sparse field $x(s)$ over an area of 30 m $\times$ 30 m (using a 2001×2001 grid). The field has a maximum spatial frequency of $f_u = 1/15\,\mathrm{m}^{-1}$. N sensors are dropped in the area, of which $N_G = 400$ sensors are assumed to lie on a 20 m $\times$ 20 m grid, while the remaining $N - N_G$ nodes are uniformly distributed over the entire 30 m $\times$ 30 m area. To model a mismatch between the sensor locations assumed by the sink and the actual locations, we distinguish between the true locations of each sensor (s_i) and the assumed locations at the sink ($\hat{s}_i$), related by $\hat{s}_i = s_i + n_p$, in which $n_p \sim \mathcal{N}(0, I_2\delta)$. The sink has perfect knowledge of the grid locations and the number of sensors, but not of the individual sensors' locations. Each sensor has perfect knowledge of its own location. When $\delta = 0$, the sink knows the locations of all sensors.

We compare two methods, each of which relies on N_G observations.

- *Method 1:* The sink utilizes the direct measurements from the N_G sensors that are assumed to be on the grid for
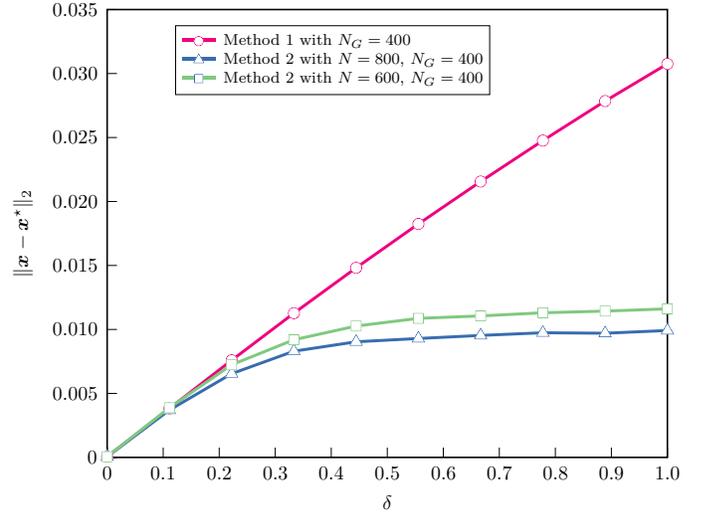
compression, under the assumption that $\delta = 0$. This method corresponds to the conventional approach, but with location errors.

- *Method 2:* The sink utilizes the interpolated measurement at the N_G grid locations for compression. These interpolated measurements are computed using the direct measurement from the nearest sensor of each grid location. This method corresponds to the framework proposed in this paper.

In both scenarios, we assume convergence of (ratio) consensus, so that the sink has perfect knowledge of the compressed vector. We set the compression ratio to $M/N_G = 0.3$, which we have observed to be sufficient for the scenario under consideration. The measurement data is reconstructed from y by solving (6) using CVX. The reconstruction error is calculated as $\|x(s_g) - x^\star\|_2$, where $x(s_g)$ is the generated field on the uniform sampling grid and $x^\star$ is the reconstructed measurement vector.

Fig. 6 shows the average reconstruction error versus the position uncertainty δ over 100 Monte Carlo simulations. We see that the error for all cases is monotonically increasing with δ, which is to be expected as the mismatch between the reconstructed spatial position of a measurement and its sampling position increases as δ increases. However, the reconstruction error in the NNS cases (method 2) saturates to a level much lower than for method 1. This effect is due to the possibility of using a closer sample of the field than the originally associated sensor. Hence, there is a maximum position error lower than δ for the NNS cases, and in turn this yields a maximum reconstruction error. The performance difference between the two methods can be traced to how they use position information. For method 1, explicit knowledge of all sensors' positions is needed centrally at the sink, while for method 2 position information is only needed locally to support the interpolation. Note also that one can expect even better performance for method 2 when considering more advanced interpolation than NNS, such as ordinary kriging.

IX. CONCLUSION

We studied the problem of distributed compressed sensing with in-network compression and pre-distribution, for field reconstruction at a sink in a wireless sensor network. To avoid the use of global position information at the sink during reconstruction, we proposed a novel framework comprising three phases: (i) distributed spatial interpolation, (ii) distributed compression using ratio consensus, (iii) transmission to the sink by selection of a subset of sensors. We derived an upper bound on the reconstruction error accounting for the error induced by the interpolation, distributed compression, selected transmission, and sparse reconstruction. Moreover, we established a tradeoff between local and global communication to meet a target upper bound on the reconstruction error. Based on numerical simulations, we quantified the impact of position uncertainty on the reconstruction error of the proposed framework and demonstrated improved performance over a conventional approach.

APPENDIX

A. Proof of Lemma 1

In this proof we approximate the uniform sensor placement with a Poisson point process (PPP). We note that $\sigma_{\mathrm{NNS}}^2(d_0)$ only depends on d_0 through $\exp\left(-d_0^2/h_r^2\right)$. For a network of nodes placed according to a PPP with intensity (approximately the sensor density) ρ, the probability density function of the distance d_0 from an arbitrary point to its nearest node is given by $f(d_0) = 2\pi\rho d_0 \exp(-\pi\rho d_0^2)$ [27]. Hence,

$$\mathbb{E}_{d_0}\{e^{-d_0^2/h_r^2}\} = \int_0^\infty 2\pi\rho d_0 e^{-\pi\rho d_0^2} e^{-d_0^2/h_r^2}\mathrm{d}d_0 \tag{33}$$

$$= \frac{\pi\rho}{1/h_r^2 + \pi\rho}, \tag{34}$$

from which the result follows immediately.

B. Proof of Lemma 2

Since $\bar{b}\boldsymbol{\pi} \oslash \bar{\alpha}\boldsymbol{\pi} = (\bar{b}/\bar{\alpha})\mathbf{1}$, the ℓ_2-norm of the error at iteration l, $l \geq \kappa$, is

$$\left\|\boldsymbol{b}(l) \oslash \boldsymbol{\alpha}(l) - (\bar{b}/\bar{\alpha})\mathbf{1}\right\|_2 \tag{35}$$

$$= \left\|\boldsymbol{P}^l \boldsymbol{b}(0) \oslash \boldsymbol{P}^l \boldsymbol{\alpha}(0) - \bar{b}\boldsymbol{\pi} \oslash \bar{\alpha}\boldsymbol{\pi}\right\|_2 \tag{36}$$

$$= \left\|\boldsymbol{P}^l \boldsymbol{b}(0) \oslash \boldsymbol{P}^l \boldsymbol{\alpha}(0) - \bar{b}\boldsymbol{\pi} \oslash \boldsymbol{P}^l \boldsymbol{\alpha}(0)\right.$$
$$\left. + \bar{b}\boldsymbol{\pi} \oslash \boldsymbol{P}^l \boldsymbol{\alpha}(0) - \bar{b}\boldsymbol{\pi} \oslash \bar{\alpha}\boldsymbol{\pi}\right\|_2 \tag{37}$$

$$= \left\|\left(\boldsymbol{P}^l \boldsymbol{b}(0) - \bar{b}\boldsymbol{\pi}\right) \oslash \boldsymbol{P}^l \boldsymbol{\alpha}(0)\right.$$
$$\left. + \left(\bar{b}\boldsymbol{\pi} \circ \bar{\alpha}\boldsymbol{\pi} - \bar{b}\boldsymbol{\pi} \circ \boldsymbol{P}^l \boldsymbol{\alpha}(0)\right) \oslash \left(\boldsymbol{P}^l \boldsymbol{\alpha}(0) \circ \bar{\alpha}\boldsymbol{\pi}\right)\right\|_2, \tag{38}$$

where $\circ$ denotes the Hadamard product. Using the triangle inequality, we upper bound

$$\left\|\boldsymbol{b}(l) \oslash \boldsymbol{\alpha}(l) - (\bar{b}/\bar{\alpha})\mathbf{1}\right\|_2 \tag{39}$$

$$\leq \left\|\left(\boldsymbol{P}^l \boldsymbol{b}(0) - \bar{b}\boldsymbol{\pi}\right) \oslash \boldsymbol{P}^l \boldsymbol{\alpha}(0)\right\|_2 \tag{40}$$

$$+ \left\|\bar{b}\boldsymbol{\pi} \circ \left(\bar{\alpha}\boldsymbol{\pi} - \boldsymbol{P}^l \boldsymbol{\alpha}(0)\right) \oslash \left(\boldsymbol{P}^l \boldsymbol{\alpha}(0) \circ \bar{\alpha}\boldsymbol{\pi}\right)\right\|_2 \tag{41}$$

$$= \left\|\mathbf{1} \oslash \boldsymbol{P}^l \boldsymbol{\alpha}(0)\right\|_2 \tag{42}$$

$$\times \left(\left\|\boldsymbol{P}^l \boldsymbol{b}(0) - \bar{b}\boldsymbol{\pi}\right\|_2 + \frac{|\bar{b}|}{|\bar{\alpha}|}\left\|\boldsymbol{P}^l \boldsymbol{\alpha}(0) - \bar{\alpha}\boldsymbol{\pi}\right\|_2\right)$$

$$= \left\|\mathbf{1} \oslash \boldsymbol{P}^l \boldsymbol{\alpha}(0)\right\|_2 \left(\left\|\left(\boldsymbol{P}^l - \boldsymbol{P}^\infty\right)\left(\boldsymbol{b}(0) - \bar{b}\boldsymbol{\pi}\right)\right\|_2\right. \tag{43}$$

$$\left. + \frac{|\bar{b}|}{|\bar{\alpha}|}\left\|\left(\boldsymbol{P}^l - \boldsymbol{P}^\infty\right)\left(\boldsymbol{\alpha}(0) - \bar{\alpha}\boldsymbol{\pi}\right)\right\|_2\right)$$

$$\leq |\lambda_2^l|\left\|\mathbf{1} \oslash \boldsymbol{P}^l \boldsymbol{\alpha}(0)\right\|_2 \tag{44}$$

$$\times \left(\left\|\left(\boldsymbol{b}(0) - \bar{b}\boldsymbol{\pi}\right)\right\|_2 + \frac{|\bar{b}|}{|\bar{\alpha}|}\left\|\left(\boldsymbol{\alpha}(0) - \bar{\alpha}\boldsymbol{\pi}\right)\right\|_2\right),$$

where we have used the convergence of ratio consensus and, in the last step, the eigen-decomposition of $\boldsymbol{P}$ with eigenvalues $1 = \lambda_1 > |\lambda_2| \geq \cdots \geq |\lambda_N|$.

C. Approximation

The definition of the ℓ_2-norm gives

$$\left\|\mathbf{1}_{N\times 1} \oslash \boldsymbol{P}^l \boldsymbol{\alpha}(0)\right\|_2 = \sqrt{\sum_{i=0}^{N-1}\left(\frac{1}{\alpha_i(l)}\right)^2}. \tag{45}$$

For large networks, and after a sufficient number of iterations l, $\alpha_i(l) \approx (N_G/N)$ for all i. Therefore,

$$\left\|\mathbf{1}_{N\times 1} \oslash \boldsymbol{P}^l \boldsymbol{\alpha}(0)\right\|_2 \approx \sqrt{\sum_{i=0}^{N-1}\left(\frac{N}{N_G}\right)^2} \tag{46}$$

$$= \frac{N^{3/2}}{N_G}. \tag{47}$$

D. Proof of Theorem 3

The observation at the sink can be decomposed as follows:

$$\hat{\boldsymbol{y}} = N_G/L \left(\sum_{i\in\mathcal{L}}\hat{\boldsymbol{w}}_i(I)\right)$$

$$= \frac{N_G}{L}\left(\sum_{i\in\mathcal{L}}\frac{\boldsymbol{b}_i(l)}{\alpha_i(l)}\right)$$

$$= \frac{N_G}{L}\left(\sum_{i\in\mathcal{L}}\left[\frac{\sum_{j=0}^{N_G}\boldsymbol{w}_j}{N_G} + \boldsymbol{\Delta}_i\right]\right)$$

$$= \frac{N_G}{L}\left(\sum_{i\in\mathcal{L}}\frac{\boldsymbol{A}\hat{\boldsymbol{x}} + \boldsymbol{e}_{\mathrm{int}}}{N_G}\right) + \frac{N_G}{L}\sum_{i\in\mathcal{L}}\boldsymbol{\Delta}_i$$

$$= \boldsymbol{A}\hat{\boldsymbol{x}} + \boldsymbol{e}_{\mathrm{int}} + \underbrace{\frac{N_G}{L}\sum_{i\in\mathcal{L}}\boldsymbol{\Delta}_i}_{=\boldsymbol{e}_{\mathrm{cons}}},$$

where $\boldsymbol{\Delta}_i$ is the consensus error vector at node i. The total error is $\boldsymbol{e} = \boldsymbol{e}_{\mathrm{int}} + \boldsymbol{e}_{\mathrm{cons}}$, with mean $\mathbb{E}\{\boldsymbol{e}\} = \mathbb{E}\{\boldsymbol{e}_{\mathrm{int}}\} +$

$\mathbb{E}\{\boldsymbol{e}_{\text{cons}}\}$, and variance $\text{Var}(\boldsymbol{e}) = \text{Var}(\boldsymbol{e}_{\text{int}}) + \text{Var}(\boldsymbol{e}_{\text{cons}})$ due to $\boldsymbol{e}_{\text{int}}$ and $\boldsymbol{e}_{\text{cons}}$ being uncorrelated which we show in the following lemma.

Lemma 4. *The two error vectors $\boldsymbol{e}_{\text{cons}}$ and $\boldsymbol{e}_{\text{int}}$ are uncorrelated, i.e., $\mathbb{E}\{\boldsymbol{e}_{\text{int}}\boldsymbol{e}_{\text{cons}}^{\mathsf{T}}\} = \boldsymbol{0}_{M \times M}$.*

Proof: First of all, $\mathbb{E}\{\boldsymbol{e}_{\text{int}}\boldsymbol{e}_{\text{cons}}^{\mathsf{T}}\}$ is a diagonal matrix, since the M compression dimensions are mutually independent. Hence, we focus on one of the dimensions $j \in \{1, \ldots, M\}$. We aim to show that $\mathbb{E}\{e_{\text{int},j}e_{\text{cons},j}\} = 0$. For some $l \geq \kappa$ we have

$$\mathbb{E}\{e_{\text{int},j}e_{\text{cons},j}\} \tag{48}$$

$$= \mathbb{E}\left\{e_{\text{int},j}\frac{N_G}{L}\sum_{i \in \mathcal{L}}(\hat{w}_{j,i}(l) - \bar{w}_j)\right\} \tag{49}$$

$$= \frac{N_G}{L}\sum_{i \in \mathcal{L}}\left(\mathbb{E}\{e_{\text{int},j}\hat{w}_{j,i}(l)\} - \mathbb{E}\{e_{\text{int},j}\bar{w}_j\}\right). \tag{50}$$

Since ratio consensus is mean preserving at every iteration $l \geq \kappa$ it follows that

$$\mathbb{E}\{e_{\text{int},j}e_{\text{cons},j}\}$$
$$= \frac{N_G}{L}\sum_{i \in \mathcal{L}}\left(\mathbb{E}\{e_{\text{int},j}\hat{w}_{j,i}(l)\} - \sum_{n=0}^{N-1}\frac{\mathbb{E}\{e_{\text{int},j}\hat{w}_{j,n}(l)\}}{N}\right).$$

Furthermore, we know that $\mathbb{E}\{e_{\text{int},j}\hat{w}_{j,n}(l)\} = \mathbb{E}\{e_{\text{int},j}\hat{w}_{j,i}(l)\}$ for any $i, n \in \{0, \ldots, N-1\}$, and thus

$$\mathbb{E}\{e_{\text{int},j}e_{\text{cons},j}\}$$
$$= \frac{N_G}{L}\sum_{i \in \mathcal{L}}\left(\mathbb{E}\{e_{\text{int},j}\hat{w}_{j,i}(l)\} - \mathbb{E}\{e_{\text{int},j}\hat{w}_{j,i}(l)\}\right)$$
$$= 0.$$

∎

Continuing with the proof of Theorem 3: when employing NNS, we have that $\boldsymbol{e}_{\text{int}}$ is zero-mean with covariance matrix $\phi(N_G/M)\boldsymbol{I}_M$ when averaging over d_0, using the same argument as in Section III-A. It follows immediately that

$$\mathbb{E}\{\|\boldsymbol{e}_{\text{int}}\|_2^2\} = N_G\phi,$$

and

$$\text{Var}\left(\|\boldsymbol{e}_{\text{int}}\|_2^2\right) = \frac{2N_G^2}{M}\phi^2.$$

We will now characterize $\boldsymbol{e}_{\text{cons}}$ in terms of its first two moments. We readily find that the mean of $\boldsymbol{e}_{\text{cons}}$ is zero as a consequence of the mean of the consensus error vector being zero. Next, we observe that the variance induced by the random querying of sensors, in a specific dimension m, can be expressed as [15]

$$\text{Var}_{\mathcal{L}}(e_{\text{cons},m}) = \frac{N_G^2}{L}\left(\frac{N-L}{N-1}\right)\sigma_w^2,$$

where $\sigma_w^2 = \|\boldsymbol{w}(l) - \bar{w}\boldsymbol{1}_{N \times 1}\|$ is the sample variance among the N sensors after l iterations of consensus, when $\boldsymbol{A}$, $\boldsymbol{n}$, and

d_0 are fixed. Now, by considering $\boldsymbol{A}$, $\boldsymbol{n}$, and d_0 to be random and applying expectation to both sides we get

$$\text{Var}_{\mathcal{L},\boldsymbol{A},\boldsymbol{n},d_0}(e_{\text{cons},m}) = \frac{N_G^2}{L}\left(\frac{N-L}{N-1}\right)\mathbb{E}_{\boldsymbol{A},\boldsymbol{n},d_0}\{\sigma_w^2\}.$$

Due to Lemma 2, it follows that

$$\mathbb{E}_{\boldsymbol{A},\boldsymbol{n},d_0}\{\sigma_w^2\}$$
$$= \mathbb{E}_{\boldsymbol{A},\boldsymbol{n},d_0}\left\{\left\|\boldsymbol{b}(l) \oslash \boldsymbol{\alpha}(l) - \bar{b}\boldsymbol{\pi} \oslash \bar{\alpha}\boldsymbol{\pi}\right\|_2^2\right\}$$
$$\approx \lambda_2^{2l}\frac{N^3}{N_G^2}\left(\mathbb{E}_{\boldsymbol{A},\boldsymbol{n},d_0}\left\{\left\|\boldsymbol{b}(0) - \bar{b}\boldsymbol{\pi}\right\|_2^2\right\}\right.$$
$$+ 2\frac{\mathbb{E}_{\boldsymbol{A},\boldsymbol{n},d_0}\{|\bar{b}|\}}{N_G}\|\boldsymbol{\alpha}(0) - \bar{\alpha}\boldsymbol{\pi}\|_2$$
$$\times \mathbb{E}_{\boldsymbol{A},\boldsymbol{n},d_0}\left\{\left\|\boldsymbol{b}(0) - \bar{b}\boldsymbol{\pi}\right\|_2\right\}$$
$$\left.+ \frac{\mathbb{E}_{\boldsymbol{A},\boldsymbol{n},d_0}\left\{|\bar{b}|^2\right\}}{N_G^2}\|\boldsymbol{\alpha}(0) - \bar{\alpha}\boldsymbol{\pi}\|_2^2\right).$$

We upper bound $\|\boldsymbol{\alpha}(0) - \bar{\alpha}\boldsymbol{\pi}\|_2 \leq N_G$ (the upper bound is achieved when one sensor holds all N_G interpolated measurements). Due to Jensen's inequality, $\mathbb{E}\{\sqrt{X}\} \leq \sqrt{\mathbb{E}\{X\}}$, we find that the cross-term can be expressed as

$$\mathbb{E}_{\boldsymbol{A},\boldsymbol{n},d_0}\left\{\left\|\boldsymbol{b}(0) - \bar{b}\boldsymbol{\pi}\right\|_2\right\}\mathbb{E}_{\boldsymbol{A},\boldsymbol{n},d_0}\{|\bar{b}|\}$$
$$\leq \sqrt{\mathbb{E}_{\boldsymbol{A},\boldsymbol{n},d_0}\left\{|\bar{b}|^2\right\}\mathbb{E}_{\boldsymbol{A},\boldsymbol{n},d_0}\left\{\left\|\boldsymbol{b}(0) - \bar{b}\boldsymbol{\pi}\right\|_2^2\right\}}.$$

Hence

$$\mathbb{E}_{\boldsymbol{A},\boldsymbol{n},d_0}\{\sigma_w^2\} \tag{51}$$
$$\leq \lambda_2^{2l}\frac{N^3}{N_G^2}\left(\sqrt{\mathbb{E}_{\boldsymbol{A},\boldsymbol{n},d_0}\left\{\left\|\boldsymbol{b}(0) - \bar{b}\boldsymbol{\pi}\right\|_2^2\right\}}\right.$$
$$\left.+ \sqrt{\mathbb{E}_{\boldsymbol{A},\boldsymbol{n},d_0}\left\{|\bar{b}|^2\right\}}\right)^2.$$

We easily find that

$$\mathbb{E}_{\boldsymbol{A},\boldsymbol{n},d_0}\left\{|\bar{b}|^2\right\} = \frac{E_X + N_G\phi}{M},$$

where $E_X = \|\hat{\boldsymbol{x}}\|_2^2$, due to the independence of the N_G variables $b_i(0) = a_i(x_i + m_i)$ that are non-zero (recall that $N - N_G$ variables $b_i(0)$ are equal to zero) and since $\mathbb{E}_{d_0}\{\sigma_{\text{NNI}}^2\} = \phi$. Secondly,

$$\mathbb{E}_{\boldsymbol{A},\boldsymbol{n},d_0}\left\{\left\|\boldsymbol{b}(0) - \bar{b}\boldsymbol{\pi}\right\|_2^2\right\}$$
$$= \mathbb{E}_{\boldsymbol{A},\boldsymbol{n},d_0}\left\{\|\boldsymbol{b}(0)\|^2 + \bar{b}^2\|\boldsymbol{\pi}\|^2 - 2\bar{b}\boldsymbol{\pi}^{\mathsf{T}}\boldsymbol{b}(0)\right\}$$
$$= \frac{E_X + N_G\phi}{M}(1 + \|\boldsymbol{\pi}\|^2) - 2\boldsymbol{\pi}^{\mathsf{T}}\mathbb{E}_{\boldsymbol{A},\boldsymbol{n},d_0}\{\bar{b}\boldsymbol{b}(0)\}$$
$$\leq \frac{E_X + N_G\phi}{M}(1 + \|\boldsymbol{\pi}\|^2),$$

since $\mathbb{E}_{\boldsymbol{A},\boldsymbol{n},d_0}\{\bar{b}b_i(0)\} = \mathbb{E}_{\boldsymbol{A},\boldsymbol{n},d_0}\{b_i^2(0)\}$. It is readily verified that $\|\boldsymbol{\pi}\|^2 \leq 2/N \ll 1$, so we can neglect this term. Substituting back into (51) leads to

$$\mathbb{E}_{\boldsymbol{A},\boldsymbol{n},d_0}\{\sigma_w^2\} \leq \lambda_2^{2l}\frac{4N^3}{N_G^2}\frac{E_X + N_G\phi}{M}$$

and hence

$$\mathrm{Var}(e_{\mathrm{cons},m}) \lesssim \frac{4}{L}\left(\frac{N-L}{N-1}\right)\lambda_2^{2l}N^3\frac{E_X+N_G\phi}{M} \triangleq \sigma_\Delta^2.$$

Finally, since the m dimensions of compression are i.i.d., it follows that [11], [16], [18]

$$\mathbb{E}\left\{\|\boldsymbol{e}_{\mathrm{cons}}\|_2^2\right\} = \sigma_\Delta^2 M,$$

and

$$\mathrm{Var}\left(\|\boldsymbol{e}_{\mathrm{cons}}\|_2^2\right) = 2M\sigma_\Delta^4.$$

Hence, we can set ε^2 as specified in the Theorem, using (8).

Remark 5. Throughout the proof, we assumed that $\mathcal{K}_k$ was not empty and that the sensors in $\mathcal{K}_k$ formed a connected network. In case $\mathcal{K}_k$ is empty for some k, this means that $\bar{\boldsymbol{w}}$ is missing a contribution $\boldsymbol{a}_k\hat{z}_k$. Considering $0 \le E < N_G$ empty sets $\mathcal{K}_k$ then (ignoring the consensus error):

$$\hat{\boldsymbol{y}} = \frac{N_G}{N_G-E}\left(\boldsymbol{A}-\bar{\boldsymbol{A}}\right)\hat{\boldsymbol{x}} + \frac{N_G}{N_G-E}(\boldsymbol{A}-\bar{\boldsymbol{A}})\boldsymbol{m}, \quad (52)$$
$$= c\boldsymbol{A}\hat{\boldsymbol{x}} + c\bar{\boldsymbol{A}}\hat{\boldsymbol{x}} + c(\boldsymbol{A}-\bar{\boldsymbol{A}})\boldsymbol{m}, \quad (53)$$

where $\bar{\boldsymbol{A}}$ is a matrix of all zeros and $c = N_G/(N_G-E) \le 1$, except with columns equal to $\boldsymbol{a}_k$ when $\mathcal{K}_k$ is empty. We see from (52) that (i) the solution to (6) must be scaled by c; (ii) the total interpolation error is reduced; (iii) there is an additional signal-dependent error term. The statistics of similar errors were investigated in [15], where information is lost due to packet erasures. A detailed analysis is beyond the scope of the current paper.

In case the sensors in $\mathcal{K}_k$ do not form a connected network, ratio consensus as proposed here will not work. Supposing $\mathcal{K}_k$ comprised N_k clusters, with interpolated measurements $\hat{z}_{k,n}$, $n = 1,\ldots,N_k$, then ratio consensus will converge to

$$\frac{\boldsymbol{A}}{\sum_{k=1}^{N_G}N_k}\left[\begin{array}{c}\sum_{n=1}^{N_1}\hat{z}_{1,n}\\ \vdots\\ \sum_{n=1}^{N_{N_G}}\hat{z}_{N_G,n}\end{array}\right],$$

which is not the desired result. Hence, clusters must be aggregated *before* ratio consensus. This can be achieved through an additional consensus (e.g., max-consensus based on a unique hash value per cluster), based on which the communication range can be increased.

References

[1] S. S. Kanhere, "Participatory sensing: Crowdsourcing data from mobile smartphones in urban spaces," in *Proc. 12th IEEE Int. Conf. on Mobile Data Management*, vol. 2, pp. 3–6, Jun. 2011.

[2] D. Slepian and J. K. Wolf, "Noiseless coding of correlated information sources," *IEEE Trans. Inf. Theory*, vol. 19, pp. 471–480, Jul. 1973.

[3] Z. Xiong, A. D. Liveris, and S. Cheng, "Distributed source coding for sensor networks," *IEEE Sig. Proc. Mag.*, vol. 21, pp. 80–94, Sep. 2004.

[4] M. Gastpar, P. L. Dragotti, and M. Vetterli, "The distributed Karhunen-Loève transform," *IEEE Trans. Inf. Theory*, vol. 52, pp. 5177–5196, Dec. 2006.

[5] R. S. Wagner, R. G. Baraniuk, S. Du, D. B. Johnson, and A. Cohen, "An architecture for distributed wavelet analysis and processing in sensor networks," in *Proc. 5th Int. Conf. on Information Processing in Sensor Networks*, pp. 243–250, Apr. 2006.

[6] E. J. Candès and T. Tao, "Decoding by linear programming," *IEEE Trans. Inf. Theory*, vol. 51, pp. 4203–4215, Dec. 2005.

[7] E. J. Candès, J. K. Romberg, and T. Tao, "Stable signal recovery from incomplete and inaccurate measurements," *Communications on Pure and Applied Mathematics*, vol. 59, pp. 1207–1223, Mar. 2006.

[8] D. L. Donoho, "Compressed sensing," *IEEE Trans. Inf. Theory*, vol. 52, pp. 1289–1306, Apr. 2006.

[9] M. F. Duarte, M. B. Wakin, D. Baron, and R. G. Baraniuk, "Universal distributed sensing via random projections," in *Proc. 5th Int. Conf. on Information Processing in Sensor Networks*, pp. 177–185, Apr. 2006.

[10] D. Baron, M. B. Wakin, M. F. Duarte, S. Sarvotham, and R. G. Baraniuk, "Distributed compressed sensing," *preprint*, 2006.

[11] W. Bajwa, J. Haupt, A. Sayeed, and R. Nowak, "Compressive wireless sensing," in *Proc. 5th Int. Conf. on Information Processing in Sensor Networks*, pp. 134–142, Apr. 2006.

[12] J. Haupt, W. Bajwa, M. Rabbat, and R. Nowak, "Compressed sensing for networked data," *IEEE Sig. Proc. Mag.*, vol. 25, pp. 92–101, Mar. 2008.

[13] C. Luo, F. Wu, J. Sun, and C. W. Chen, "Compressive data gathering for large-scale wireless sensor networks," in *Proc. 15th Annual Int. Conf. on Mobile Computing and Networking*, pp. 145–156, Sep. 2009.

[14] M. Rabbat, J. Haupt, A. Singh, and R. Nowak, "Decentralized compression and predistribution via randomized gossiping," in *Proc. 5th Int. Conf. on Information Processing in Sensor Networks*, pp. 51–59, Apr. 2006.

[15] C. Lindberg, A. Graell i Amat, and H. Wymeersch, "Distributed compressed sensing for sensor networks with packet erasures," in *Proc. IEEE Global Communications Conf.*, pp. 13–19, Dec. 2014.

[16] P. Balister, A. Sarkar, and B. Bollobás, "Percolation, connectivity, coverage and colouring of random geometric graphs," in *Handbook of large-scale random networks*, pp. 117–142, Springer, 2008.

[17] H. Landau, "Necessary density conditions for sampling and interpolation of certain entire functions," *Acta Mathematica*, vol. 117, pp. 37–52, Jul. 1967.

[18] R. Baraniuk, M. Davenport, R. Devore, and M. Wakin, "A simple proof of the restricted isometry property for random matrices," *Constructive Approximation*, vol. 28, pp. 253–263, Dec. 2008.

[19] E. J. Candès, "The restricted isometry property and its implications for compressed sensing," *Académie des Sciences*, vol. I, pp. 589–592, May 2008.

[20] M. Lustig, D. Donoho, and J. M. Pauly, "Sparse MRI: The application of compressed sensing for rapid MR imaging," *Magnetic Resonance in Medicine*, vol. 58, pp. 1182–1195, Dec. 2007.

[21] N. A. C. Cressie, *Statistics for spatial data.* Wiley New York, 1993.

[22] G. Hernandez-Penaloza and B. Beferull-Lozano, "Field estimation in wireless sensor networks using distributed kriging," in *Proc. IEEE Int. Conf. Communications*, pp. 724–729, Jun. 2012.

[23] R. Olfati-Saber, A. Fax, and R. M. Murray, "Consensus and cooperation in networked multi-agent systems," *Proceedings of the IEEE*, vol. 95, pp. 215–233, Jan. 2007.

[24] D. Kempe, A. Dobra, and J. Gehrke, "Gossip-based computation of aggregate information," in *Proc. 44th Ann. IEEE Symposium on Foundations of Computer Science*, pp. 482–491, Oct. 2003.

[25] F. Bénézit, V. Blondel, P. Thiran, J. Tsitsiklis, and M. Vetterli, "Weighted gossip: Distributed averaging using non-doubly stochastic matrices," in *Proc. IEEE Int. Symp. Information Theory*, pp. 1753–1757, Jun. 2010.

[26] A. D. Dominguez-Garcia and C. N. Hadjicostis, "Coordination and control of distributed energy resources for provision of ancillary services," in *Proc. 1st IEEE Int. Conf. Smart Grid Communications*, pp. 537–542, Oct. 2010.

[27] D. Moltchanov, "Distance distributions in random networks," *Ad Hoc Networks*, vol. 10, pp. 1146–1166, Aug. 2012.

Christopher Lindberg received the B.Sc. degree in electrical engineering, M.Sc. degree in engineering mathematics and computational science, and degree of Licentiate of Engineering in electrical engineering from Chalmers University of Technology in 2010, 2012, and 2015, respectively. Since 2012 he is working towards the Ph.D. degree in the Communication Systems group at the Department of Signals and Systems, Chalmers University of Technology, Gothenburg, Sweden. His research interests involve cooperative distributed processing in networks based on consensus algorithms and belief propagation.

Henk Wymeersch (S'99–M'05) received the Ph.D. degree in Electrical Engineering/Applied Sciences in 2005 from Ghent University, Belgium. He is a Professor in Communication Systems at the Department of Signals and Systems, Chalmers University of Technology, Sweden. Prior to joining Chalmers, he was a postdoctoral researcher from 2005 until 2009 with the Laboratory for Information and Decision Systems at the Massachusetts Institute of Technology. He served as Associate Editor for IEEE Communication Letters (2009-2013), IEEE Transactions on Wireless Communications (since 2013), and IEEE Transactions on Communications (since 2016). His current research interests include cooperative systems and intelligent transportation.

Alexandre Graell i Amat (S'01–M'05–SM'10) received the M.Sc. degree in telecommunications engineering from the Universitat Politècnica de Catalunya, Barcelona, Catalonia, Spain, in 2001, and the M.Sc. and Ph.D. degrees in electrical engineering from the Politecnico di Torino, Turin, Italy, in 2000 and 2004, respectively. From 2001 to 2002, he was a Visiting Scholar with the University of California at San Diego, La Jolla, CA, USA. From 2002 to 2003, he held a visiting appointment with the Universitat Pompeu Fabra, Barcelona, Spain, and the Telecommunications Technological Center of Catalonia, Barcelona. From 2001 to 2004, he also held a part-time appointment with the STMicroelectronics Data Storage Division, Milan, Italy, as a Consultant on coding for magnetic recording channels. From 2004 to 2005, he was a Visiting Professor with the Universitat Pompeu Fabra. From 2006 to 2010, he was with the Department of Electronics, Telecom Bretagne, Brest, France. In 2011, he joined the Department of Signals and Systems, Chalmers University of Technology, Gothenburg, Sweden, where he is currently a Professor. His research interests include the areas of (modern) coding theory, distributed storage, and optical communications. He was an Associate Editor of the IEEE TRANSACTIONS ON COMMUNICATIONS from 2011 to 2015, and the IEEE COMMUNICATIONS LETTERS from 2011 to 2013. He is currently an Editor-at-Large of the IEEE TRANSACTIONS ON COMMUNICATIONS. He was the General Co-Chair of the seventh International Symposium on Turbo Codes and Iterative Information Processing, Gothenburg, in 2012. He received the post-doctoral Juan de la Cierva Fellowship from the Spanish Ministry of Education and Science and the Marie Curie Intra-European Fellowship from the European Commission. He received the IEEE Communications Society 2010 Europe, Middle East, and Africa Region Outstanding Young Researcher Award.