Optimal Network Parameter Estimation: Single-Shot Exchange of Local Decisions

Saurabh Sihag and Ali Tajer

Abstract—**This letter considers a network of sensors that collectively sense a number of unknown parameters. Each sensor can possibly sense only a subset of the parameters, gather data only about these parameters, and has access to only the statistical model of the data that it collects locally. The objective is that each sensor forms optimal estimates for its designated parameters (i.e., the parameters that it can sense). This letter proposes an estimation cost function that strikes a balance between the sensors being autonomous in forming local estimates based on their locally available data and statistical models, and enforcing consistency among the local estimates formed for the parameters that are sensed by multiple sensors. Exact optimal estimators are characterized, and it is shown that the optimal estimators can be implemented in a distributed way, through a single-shot exchange of local decisions. Specifically, the distributed implementation consists of forming local estimates and exchanging certain sufficient statistics values in a single round of communication exchange among some of the sensors. Furthermore, the optimal performance under the proposed cost function is also compared analytically with the performance of the widely used mean squared error estimator.**

***Index Terms*—Autonomous systems, distributed algorithm, estimation theory, networks.**

I. Introduction

CONSIDER a network parameter estimation problem in which m sensors $\{S_1, \ldots, S_m\}$, collectively, estimate n mutually independent scalar parameters $\mathbf{X} \triangleq [X_1, \ldots, X_n]^T$. Each parameter might be sensed by only a subset of the sensors (e.g., due to the network being large). We define the column vector $\mathbf{A}_i \in \{0,1\}^n$ to account for the set of parameters sensed by sensor S_i. Specifically,

$$[\mathbf{A}_i]_j \triangleq \begin{cases} 1 & \text{if } X_j \text{ is sensed by } S_i \\ 0 & \text{otherwise} \end{cases} . \tag{1}$$

Accordingly, we define $\mathbf{X}_i \triangleq \mathbf{A}_i \circ \mathbf{X}$, where $\circ$ denotes the Hadamard product. Furthermore, we define the $\mathbf{Y}_i \in \mathbb{R}^{p\times 1}$ as the vector of p measurements generated by sensor S_i. Measurements $\mathbf{Y}_i$ can be any non-linear and arbitrary function of the parameters $\mathbf{X}_i$, and they are related via

$$\mathbf{Y}_i \sim f_i(\cdot \mid \mathbf{X}_i), \quad \text{for} \quad i \in \{1, \ldots, m\} , \tag{2}$$

where f_i denotes the conditional joint probability density function (pdf) of the measurements $\mathbf{Y}_i$. Furthermore, we denote the prior pdf of X_i by π_i. We assume that the pdfs f_i satisfy the regularity conditions [1].

Assumption: Each sensor S_i has access only to its local statistical model $f_i(\cdot \mid \mathbf{X}_i)$ and its local data $\mathbf{Y}_i$. This assumption is especially critical in large-scale networks in which it may be infeasible or undesirable for a sensor agent to know the complete statistical model of the network. For instance, in a power system with multiple control areas, it is infeasible to perform real-time exchange of the dynamically-changing system among the physically distant control areas.

Objective: The objective is that sensor S_i forms an estimate for its relevant parameters $\mathbf{X}_i$ such that the estimates optimize a joint global estimation cost function. The cost function will be specified in Section III.

Autonomy versus Consistency: We propose a natural estimation cost function that strikes a balance between:

1) sensors being autonomous in forming local estimates based on their local data and statistical models; and
2) enforcing consistency among the local estimates for the parameters that are being estimated by multiple sensors.

Main Results: The objective is that each sensor forms optimal estimates for the its respective unknown parameters while remaining oblivious to the statistical models of other sensors. Driven by the specific cost function adopted, we have the following main observations:

1) We start by allowing full coordination among the sensors in order to establish the optimal estimation performance. We provide closed-form characterizations of the optimal estimates when the sensors are allowed to share all their data and local statistical model.
2) We characterize the *minimal sufficient statistics* that the sensors need to share with each other in order to compute the optimal estimates found in the fully coordinated scenario.
3) The form of the sufficient statistics provided establish a distributed algorithm for finding the optimal estimates at the sensors, which we refer to as the single-shot exchange of local decisions (SELD) algorithm. This algorithm, specifically, consists in *only one* round of exchange of the sufficient statistics among some of the sensors, and that is sufficient for all the sensors to obtain the optimal estimates that they could form when having full coordination.
4) We also analytically compare the efficiency of the estimates under the proposed cost function with the performance of a widely-used mean-squared error (MSE) cost function where the MSE estimator has full access to all the data and the joint statistical models of the data at all sensors.

Manuscript received April 3, 2019; revised June 20, 2019; accepted June 21, 2019. Date of publication June 27, 2019; date of current version July 24, 2019. This work was supported by the U.S. National Science Foundation under the CAREER Award ECCS-1554482 and the Grant DMS-1737976. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Steeve Zozor. *(Corresponding author: Ali Tajer.)*

The authors are with the Electrical, Computer, and Systems Engineering Department, Rensselaer Polytechnic Institute, Troy, NY 12180 USA (e-mail: sihags@rpi.edu; tajer@ecse.rpi.edu).

Digital Object Identifier 10.1109/LSP.2019.2925539

II. Related Studies

The problem of network parameter estimation has a rich literature [1] and [2], and arises in a wide range of domains such as sensor networks [3]–[5], power systems [6]–[9], and social networks [10]–[12]. In many practical applications, it is desirable to distribute the inference among multiple autonomous entities in the network to gain scalability, resilience to failure, and alleviate privacy/security concerns [2]. For instance, the division of power systems into multiple control centers allows the design of reliable and computationally efficient inference rules [6]. However, in general, the inference decisions at different control centers in a power system may not be mutually independent due to the measurements and parameters associated with the tie lines connecting them [7]–[9]. Therefore, performing independent inference at individual control centers may not lead to optimal decisions. This motivates the design of inference frameworks with locally autonomous decisions while not compromising on the scalability and quality of decision rules.

In this letter, we focus on a network parameter estimation problem, in which the parameters are sensed, collectively, by m autonmous sensors. Similar problems have been studied in the context of Bayesian learning and multitask learning problems in multi-agent networks in [10]–[20]. In the context of Bayesian learning, each agent modifies its belief (posterior distribution of parameters) based on private information (i.e., information only accessible to agents that receive it) and multiple rounds of communication with neighboring agents with an aim to converge to a common decision that optimizes a cost or a reward (for e.g. MSE). In [13], a Bayesian filter is applied to a tracking problem where the agents reach consensus on the probability distribution of the states of a moving target. In [10], [14] and [16], the strategies for distributed estimation are characterized where the agents collaborate among themselves to reach a consensus on the belief of a fictitious fusion center that has access to the data at all agents.

The studies in [18]–[20] consider the settings where each agent in a multi-agent network is interested in learning only a subset of the complete set of unknowns. In these studies, data processing and inter-agent communication strategies are characterized by multiple rounds of communication among the agents that lead to convergence to the solution that optimizes a distributed objective function. In [18] and [19], sequential distributed estimation strategies based on diffusion least mean squares and consensus, respectively, are studied, whereas, a distributed optimization method based on alternating direction method of multipliers is adopted in [20]. The studies in [19] and [20] also apply their respective frameworks to a network flow problem.

The aforementioned works in the context of Bayesian learning in [10]–[17] focus primarily on achieving the performance of an estimator that optimizes MSE. In contrast, we propose an estimation framework based on a cost function that incorporates the quality of estimates based on autonomous decisions at all sensors and the consistency in the estimates of commonly observed parameters by different sensors. We provide closed form estimation rules that optimize the proposed cost. Furthermore, in contrast to the distributed estimation strategies in [18] and [19], we characterize the minimal sufficient statistics that the sensors must exchange among themselves in one round to form their respective optimal estimates. We also analytically and empirically compare the estimation performance of the proposed estimation rules with that of an estimator that optimizes MSE.

III. Estimation Framework

We aim to design an estimation framework for the network estimation problem specific in Section I. For this purpose, we define $\mathbf{Y} \triangleq [\mathbf{Y}_1, \ldots, \mathbf{Y}_m]$ as the collection of all measurements, and define $\hat{\mathbf{X}}_i(\mathbf{Y})$ as the estimate of $\mathbf{X}_i$ formed by sensor S_i. In general, there is a set of parameters that are estimated only by one sensor, and a set of parameters that are estimated by more than one sensor. We adopt a cost function that (i) provides the sensor with autonomy for forming estimates for their parameters that they sense exclusively, and (ii) forces the sensors sharing the same parameter form a consistent estimate for the shared parameter. The proposed cost function, as a result, consists of two separate costs specified next.

Throughout the rest of the letter for capturing the fidelity of a generic estimator U for any specific parameter X_i we define the quadratic cost function

$$C(X_i\ ,\ U) \triangleq \|X_i - U\|^2. \tag{3}$$

Autonomous Estimates: The parameters that S_i senses and aims to estimate are the non-zero elements of $\mathbf{X}_i = \mathbf{A}_i \circ \mathbf{X}$. We capture the fidelity of estimates generated by sensor S_i by defining the *local* average posterior cost function

$$\mathscr{L}_i(\hat{\mathbf{X}}_i(\mathbf{Y}) \,|\, \mathbf{Y}_i) \triangleq \mathbb{E}_i\left[C(\mathbf{X}_i\ ,\ \hat{\mathbf{X}}_i(\mathbf{Y})) \mid \mathbf{Y}\right]\ , \tag{4}$$

where expectation is with respect to f_i. This emphasizes the fact that each sensor used only its local data $\mathbf{Y}_i$ and its local statistical information $f_i(\cdot \mid \mathbf{X}_i)$.

Consistent Estimates: While each sensor can autonomously optimize (4) to ensure optimal estimator design based on its respective local data, it does not enforce consistency among the estimates of the parameters formed by multiple sensors. To ensure consistency, corresponding to each two arbitrary sensors S_i and S_j we define the following measure of *discrepancy* between the estimates of their shared parameters. The indexes of these shared parameters are specified by the non-zero elements of $\mathbf{A}_i \circ \mathbf{A}_j$.

$$\mathscr{D}_{ij}(\mathbf{Y}) \triangleq C(\mathbf{A}_i \circ \mathbf{A}_j \circ \hat{\mathbf{X}}_i(\mathbf{Y})\ ,\ \mathbf{A}_i \circ \mathbf{A}_j \circ \hat{\mathbf{X}}_j(\mathbf{Y})). \tag{5}$$

Aggregate Cost Function: We adopt a cost function that aggregates the local and discrepancy cost functions specified in (4) and (5), respectively. By setting $\mathbf{V} \triangleq [\mathbf{X}_1, \ldots, \mathbf{X}_m]$ and denoting its estimate by $\hat{\mathbf{V}}(\mathbf{Y}) \in \mathbb{R}^{n\times m}$ we define

$$\mathscr{J}(\hat{\mathbf{V}}(\mathbf{Y}) \mid \mathbf{Y}) \triangleq \sum_{i=1}^{m} \alpha_i \mathscr{L}_i(\hat{\mathbf{V}}_i(\mathbf{Y}) \,|\, \mathbf{Y}_i) + \beta \sum_{i\neq j} \mathscr{D}_{ij}(\mathbf{Y})\ , \tag{6}$$

where $\{\alpha_i\}_{i=1}^m$ are positive constants that can be tuned to place the appropriate emphases on the local cost functions and the inconsistency of the estimates for shared parameters. For instance, sensors can experience different levels of signal-to-noise ratio (SNR), and can be adjusted to contribute differently to the aggregate cost function $\mathscr{J}$. Similarly, β is a positive constant that controls the degree of consistency among the estimates of the same parameter formed by different sensors. Increasing β places more emphasis on consistency while penalizing the estimates at individual sensors.

This estimation cost allows us to set up an estimation framework that ensures that each sensor forms estimates with high fidelity while maintaining consistency with the estimates of its shared parameters formed by other sensors. We define the optimal estimates as the solutions to

$$\hat{\mathbf{V}}(\mathbf{Y}) \triangleq \arg\min_{\boldsymbol{U}(\mathbf{Y})} \mathscr{J}(\boldsymbol{U}(\mathbf{Y}) \mid \mathbf{Y}). \tag{7}$$

IV. Estimator: Optimal Design and Implementation

In this section, we provide the closed form characterization of the optimal estimator $\hat{\mathbf{V}}(\mathbf{Y})$ when we allow the sensors to have full coordination, i.e., they can exchange as much information as they wish. Once the optimal designs are characterized, in the next step we show that the designs are amenable to distributed implementation with limited and only round of information exchange among the sensors with shared parameters.

A. Optimal Estimators

We define matrix $\mathbf{A} \triangleq [\mathbf{A}_1, \ldots, \mathbf{A}_m]$, the unit vector $\boldsymbol{e}_j \in \mathbb{R}^{1\times n}$ that contains all zeros except at location j, and the unit vector $\boldsymbol{g}_i \in \mathbb{R}^{1\times m}$ that contains all zeros except at location i. The following theorem characterizes the optimal estimators.

Theorem 1 (Estimators): For all $j \in \{1, \ldots, n\}$, the $j-$th row of the optimal estimator $\hat{\mathbf{V}}(\mathbf{Y})$, i.e., $\boldsymbol{e}_j \cdot \hat{\mathbf{V}}$ satisfies the linear system of equations given by

$$\mathbf{D}_j \cdot [\boldsymbol{e}_j \cdot \hat{\mathbf{V}}]^T = [\boldsymbol{e}_j \cdot \mathbf{A}]^T \circ \mathbf{b}_j \ , \tag{8}$$

where $\mathbf{D}_j \in \mathbb{R}^{m\times m}$ and $\mathbf{b}_j \in \mathbb{R}^{m\times 1}$ are defined as

$$[\mathbf{D}_j]_{u,v} \triangleq \begin{cases} \boldsymbol{e}_j \cdot \mathbf{A}_u \cdot \dfrac{(\beta \|\boldsymbol{e}_j \cdot \mathbf{A}\|_1 - \beta + \alpha_u)}{\alpha_u} + (1 - \boldsymbol{e}_j \cdot \mathbf{A}_u) & \text{if } u = v \\ -\boldsymbol{e}_j \cdot \mathbf{A}_v \cdot \frac{\beta}{\alpha_u} \quad \text{if } u \neq v \end{cases} \tag{9}$$

$$\text{and} \quad [\mathbf{b}_j]_i \triangleq \mathbb{E}_i[X_j \mid \mathbf{Y}_i]. \tag{10}$$

Furthermore, the optimal estimates at sensor S_i are given by

$$\hat{\mathbf{V}}_i(\mathbf{Y}) = [\boldsymbol{g}_i \cdot \mathbf{D}_1^{-1} \cdot [\boldsymbol{e}_1 \cdot \mathbf{A}]^T \circ \mathbf{b}_1, \ldots, \boldsymbol{g}_i \cdot \mathbf{D}_n^{-1} \cdot [\boldsymbol{e}_n \cdot \mathbf{A}]^T \circ \mathbf{b}_n]^T, \tag{11}$$

where $\hat{\mathbf{V}}_i(\mathbf{Y}) \in \mathbb{R}^{n\times 1}$ is a vector formed by the $i-$th elements of the solutions to linear system of equations of the form (8) .

The matrix $\mathbf{D}_j$ defined in (9) establishes the linear relationships between the optimal estimates of X_j formed at different sensors and their corresponding locally formed posterior means of X_j. An important observation from Theorem 1 is that the optimal estimate of X_j at sensor S_i is given by the scalar $\boldsymbol{g}_i \cdot \mathbf{D}_j^{-1} \cdot [\boldsymbol{e}_j \cdot \mathbf{A}]^T \circ \mathbf{b}_j$, which is the weighted average of the elements of the vector of local posterior means of X_j, i.e., $\mathbf{b}_j$, and the weights are given by the elements of the $i-$th row of $\mathbf{D}_j^{-1}$. This important observation guides computing the optimal estimates $\hat{\mathbf{V}}$ in a distributed way, based on primarily forming local posterior means at each sensor based on only the locally available data and statistical models, and one round of communication for coordinating the decisions, as described in the next subsection.

B. Distributed Implementation: SELD Algorithm

The structure of $\hat{\mathbf{V}}_i(\mathbf{Y})$ given in (11) indicates that determining the globally optimal estimate of X_j is a weighted average of posterior means of this parameter at the sensors that sense X_j, as formalized in the following corollary.

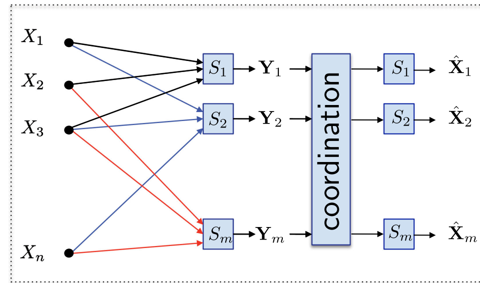

Fig. 1. Distributed network parameter estimation. An arrow from X_j to S_i indicates that sensor S_i senses parameter X_j.

Corollary 1 (Minimimal Sufficient Statistics): The local posterior means of the parameters formed by the sensors are the minimal sufficient statistics to be shared across the network in order to compute the optimal estimates $\hat{\mathbf{V}}(\mathbf{Y})$.

Hence, finding the globally optimal estimates can be carried out in the following three steps that form the SELD algorithm:

1) *Local Estimates:* Each sensor S_i forms the posterior means for all the parameters that it senses based on its locally available information and statistical models. Hence, sensor S_i computes $\mathbb{E}_i[\mathbf{X}_i \mid \mathbf{Y}_i]$.
2) *Coordination:* All sensors that are commonly observing a non-empty subset of unknown parameters exchange their corresponding posterior means among themselves. For instance, in Fig. 1, Sensors S_1 and S_2 commonly sense parameters X_1 and X_3. Hence, S_1 and S_2 exchange the following two scalar values for $j \in \{1, 3\}$
$$\boldsymbol{e}_j \cdot \mathbb{E}_1[\mathbf{X}_1 \mid \mathbf{Y}_1] \quad \text{and} \quad \boldsymbol{e}_j \cdot \mathbb{E}_2[\mathbf{X}_2 \mid \mathbf{Y}_2]. \tag{12}$$
3) *Aggregation:* Sensor S_i aggregates all the posterior means shared with it according to (11) in order to find the globally optimal estimates $\hat{\mathbf{X}}_i(\mathbf{Y}_i)$.

Therefore, the sensors can form the globally optimal estimates with only one round of communication for sharing their required sufficient statistics (posterior means). We comment that that the structure of the shared parameters is known to the sensors a-priori.

C. Efficiency and Connection to MSE

In this section we assess the efficiency of the optimal estimators for the proposed cost function $\mathscr{J}(\hat{\mathbf{V}}(\mathbf{Y}) \mid \mathbf{Y})$. We first characterize its asymptotic performance. Next, we compare the performance of the optimal estimators for the cost function $\mathscr{J}(\hat{\mathbf{V}}(\mathbf{Y}) \mid \mathbf{Y})$ with that of the estimators for the MSE cost function.

In Lemma 1, we provide the efficiency of the estimation framework established in Theorem 1. To describe the results in Lemma 1, we define $diag(\mathbf{a})$ for a one dimensional vector $\mathbf{a}$ as a diagonal matrix of size $|\mathbf{a}| \times |\mathbf{a}|$ with its diagonal elements given by the elements of $\mathbf{a}$. We also define

$$\tilde{\mathcal{I}}_j(\mathbf{X}) \triangleq \left[\boldsymbol{e}_j \cdot \mathbf{A}_1 [\mathcal{I}_1(\mathbf{X}_1)^{-1}]_{jj}, \ldots, \boldsymbol{e}_j \cdot \mathbf{A}_m [\mathcal{I}_m(\mathbf{X}_m)^{-1}]_{jj}\right]^T , \tag{13}$$

where $\mathcal{I}_i(\mathbf{X}_i)$ is the Fisher information matrix corresponding to the unknown parameters observed by sensor S_i, i.e.,

$$\mathcal{I}_i(\mathbf{X}_i) \triangleq -\mathbb{E}\left[\nabla^2_{\mathbf{X}_i} \log f_i(\mathbf{Y}_i \mid \mathbf{X}_i)\right]. \tag{14}$$

Lemma 1: The estimate $\hat{\mathbf{V}}_i(\mathbf{Y})$ formed by sensor S_i, in the asymptote of large number of samples p converges to a normal distribution according to

$$\sqrt{p}[\hat{\mathbf{V}}_i(\mathbf{Y}) - \mathbf{X}] \xrightarrow{d} \mathcal{N}(\mathbf{0}, \Sigma_i(\mathbf{X})) \ , \tag{15}$$

where the diagonal elements of $\Sigma_i(\mathbf{X}) \in \mathbb{R}^{n\times n}$ are given by

$$[\Sigma_i(\mathbf{X})]_{jj} = \|(diag(\boldsymbol{g}_i \cdot \mathbf{D}_j^{-1}))^2 \tilde{\mathcal{I}}_j(\mathbf{X})\|_1 \ . \tag{16}$$

Proof: The proof follows directly from the asymptotic efficiency of Bayesian estimators [21]. ■

The relationship in (15) indicates that the variance of the estimate of X_j by sensor S_i scales according to $\frac{1}{p}[\Sigma_i(\mathbf{X})]_{jj}$. We use the result of Lemma 1 to assess the asymptotic performance with respect to the optimal MSE-based estimator that estimates $\mathbf{X}$ directly by using all the data $\mathbf{Y}$. We denote such MSE estimator by

$$\hat{\mathbf{X}}_{\rm MSE}(\mathbf{Y}) = \arg\min_{\boldsymbol{U}(\mathbf{Y})} \|\boldsymbol{U}(\mathbf{Y}) - \mathbf{X}\|^2. \tag{17}$$

For this purpose, for each sensor S_i and parameters X_j that it senses we define

$$\eta_i^j(\mathbf{X}) \triangleq \frac{var(\boldsymbol{e}_j \cdot \hat{\mathbf{X}}_i)}{var(\boldsymbol{e}_j \cdot \hat{\mathbf{X}}_{\rm MSE})} \ , \tag{18}$$

which captures the relative estimation error variances.

Corollary 2: The ratio of the estimation error variances $\eta_i^j(\mathbf{X})$ in the asymptote of large p is given by

$$\eta_i^j(\mathbf{X}) = \frac{[\Sigma_i(\mathbf{X})]_{jj}}{[\mathcal{I}(\mathbf{X})^{-1}]_{jj}} \ , \tag{19}$$

where $\mathcal{I}(\mathbf{X})$ is the Fisher information matrix corresponding to the true set of parameters $\mathbf{X}$.

Remark 1 (Restricted collaboration): Note that the restriction on communication between any two arbitrary sensors S_i and S_j can be accommodated in the SELD algorithm by setting $\mathscr{D}_{ij}(\mathbf{Y}) = 0$ in the estimation cost (6). In this scenario, the optimal estimates can be obtained and their performance analyzed by the relative estimation error variances in Corollary 2 using similar technical arguments used to recover the results for full collaboration.

V. Numerical Evaluations

Consider a 4-sensor network collecting measurement about parameters $\{X_i : i \in \{1, \ldots, 5\}\}$, where the sensors and their designated parameters are specified in Fig. 2. The parameters are independent and distributed according to

$$\mathbf{X} \sim \mathcal{N}(\mathbf{0}, \Sigma) \ , \quad \text{where} \quad \Sigma = \text{diag}(1, 2, 4, 6, 6). \tag{20}$$

The p measurements at sensor S_i are given by

$$\mathbf{Y}_i = \mathbf{A}_i \circ \mathbf{X} + \boldsymbol{N}_i \ , \tag{21}$$

where $\boldsymbol{N}_i \sim \mathcal{N}(\mathbf{0}, \sigma_i^2 \ \mathbf{I})$, where $(\sigma_1^2, \sigma_2^2, \sigma_3^2, \sigma_4^2) = (2, 1, 4, 8)$.

We compare the empirical performance of proposed estimation framework based on single-shot exchange of local decisions (SELD) with that of an MSE estimator, an autonomous-only estimation framework (where each sensor forms final estimates locally and independently), a uniform aggregation framework where the sensors combine the minimal sufficient statistics with equal weights to form consistent estimates), and a consensus + innovations estimation strategy (CIRFE) proposed in [18]. We estimate the empirical root mean squared error (RMSE) over 10^5 Monte-Carlo simulations for different estimation strategies and use it as the criterion for comparison among them. For the results in Fig. 3, we set $\beta = 1$ and $\alpha_i = \sigma_i^{-2}$, for $i \in \{1, 2, 3, 4\}$ such that the asymptotic estimation error variances for the estimates formed using the SELD algorithm are within 5% of an MSE estimator.

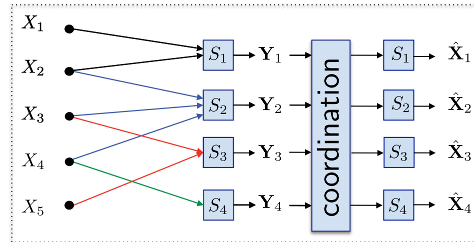

Fig. 2. 4-sensor distributed network for estimating 5 parameters.

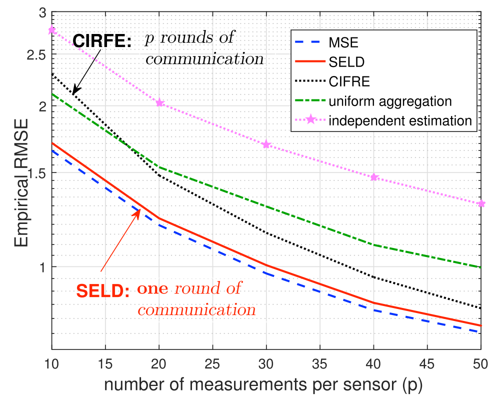

Fig. 3. Empirical RMSE versus the number of samples per sensor.

In Fig. 3, we observe that the RMSE of the estimates yielded by the SELD algorithm is very close to that of the MSE estimator. One the other hand, the CIRFE algorithm exhibits a considerable gap with SELD and MSE. This is despite the fact that in contrast to SELD, CIRFE is substantially more communication-intensive. Specifically, for p number of measurements per sensor, the final estimates using CIRFE are formed after p rounds of communication. The performance gap between CIRFE and SELD, nevertheless, diminishes asymptotically as the number of measurements per sensor increases. Furthermore, SELD outperforms the independent estimation and the uniform aggregation estimation strategies significantly, which underlines the significance of strategically combining the minimal sufficient statistics at all sensors.

VI. Conclusion

We have considered the problem of parameter estimation in a multi-sensor network. Each sensor can observe a non-exclusive subset of parameters. We have proposed an estimation framework based on a proposed cost function that consists of estimation qualities at all sensors and discrepancies among the estimates of same parameters formed by different sensors. Closed form optimal solution to the proposed cost is provided which characterizes the minimal sufficient statistics for each sensor to form optimal estimators. The estimation framework is amenable to distributed implementation with one round of exchange of minimal sufficient statistics among the sensors.

REFERENCES

[1] H. V. Poor, *An Introduction to Signal Detection and Estimation*, 2nd ed. New York, NY, USA: Springer-Verlag, 1998.

[2] A. H. Sayed, “Adaptation, learning, and optimization over networks,” *Found. Trends Mach. Learn.*, vol. 7, no. 4-5, pp. 311–801, 2014.

[3] S. Kar, J. M. Moura, and K. Ramanan, “Distributed parameter estimation in sensor networks: Nonlinear observation models and imperfect communication,” *IEEE Trans. Inf. Theory*, vol. 58, no. 6, pp. 3575–3605, Jun. 2012.

[4] A. Ribeiro and G. B. Giannakis, “Bandwidth-constrained distributed estimation for wireless sensor networks-part I: Gaussian case,” *IEEE Trans. Signal Process.*, vol. 54, no. 3, pp. 1131–1143, Mar. 2006.

[5] V. V. Veeravalli and P. K. Varshney, “Distributed inference in wireless sensor networks.” *Philos. Trans. Roy. Soc. A, Math., Phys. Eng. Sci.*, vol. 370, no. 1958, pp. 100–117, 2012.

[6] A. Gómez-Expósito, A. de la Villa Jaén, C. Gómez-Quiles, P. Rousseaux, and T. V. Cutsem, “A taxonomy of multi-area state estimation methods,” *Elect. Power Syst. Res.*, vol. 81, no. 4, pp. 1060–1069, 2011.

[7] G. N. Korres, “A distributed multiarea state estimation,” *IEEE Trans. Power Syst.*, vol. 26, no. 1, pp. 73–84, Apr. 2011.

[8] V. Kekatos and G. B. Giannakis., “Distributed robust power system state estimation,” *IEEE Trans. Power Syst.*, vol. 28, no. 2, pp. 1617–1626, Oct. 2013.

[9] L. Zhao and A. Abur, “Multi area state estimation using synchronized phasor measurements,” *IEEE Trans. Power Syst.*, vol. 20, no. 2, pp. 611–617, May 2005.

[10] E. Mossel, N. Olsman, and O. Tamuz., “Efficient Bayesian learning in social networks with Gaussian estimators,” in *Proc. Allerton Conf. Commun., Control, Comput.*, Monticello, IL, USA, Sep. 2016, pp. 425–432.

[11] D. Acemoglu, M. A. Dahleh, I. Lobel, and A. Ozdaglar, “Bayesian Learning in Social Networks,” *Rev. Econ. Stud.*, vol. 78, no. 4, pp. 1201–1236, Mar. 2011.

[12] R. G. Everitt, “Bayesian parameter estimation for latent Markov random fields and social networks,” *J. Comput. Graph. Statist.*, vol. 21, no. 4, pp. 940–960, Oct. 2012.

[13] S. Bandyopadhyay and S.-J. Chung, “Distributed estimation using Bayesian consensus filtering,” in *Proc. Amer. Control Conf.*, Portland, OR, USA, Jun. 2014, pp. 634–641.

[14] Y. Wang and P. M. Djurić, “Distributed Bayesian estimation of linear models with unknown observation covariances,” *IEEE Trans. Signal Process.*, vol. 64, no. 8, pp. 1962–1971, Apr. 2016.

[15] K. R. Rad and A. Tahbaz-Salehi, “Distributed parameter estimation in networks,” in *Proc. IEEE Conf. Decis. Control*, Atlanta, GA, USA, Dec. 2010, pp. 5050–5055.

[16] M. A. Rahimian and A. Jadbabaie, “Bayesian learning without recall,” *IEEE Trans. Signal Inf. Process. Over Netw.*, vol. 3, no. 3, pp. 592–606, Sep. 2017.

[17] P. M. Djurić and Y. Wang, “Distributed Bayesian learning in multiagent systems: Improving our understanding of its capabilities and limitations,” *IEEE Signal Process. Mag.*, vol. 29, no. 2, pp. 65–76, Mar. 2012.

[18] A. K. Sahu, D. Jakovetić, and S. Kar, “$\mathcal{CIRFE}$: A distributed random fields estimator,” *IEEE Trans. Signal Process.*, vol. 66, no. 18, pp. 4980–4995, Sep. 2018.

[19] R. Nassif, C. Richard, A. Ferrari, and A. H. Sayed, “Diffusion LMS for multitask problems with local linear equality constraints,” *IEEE Trans. Signal Process.*, vol. 65, no. 19, pp. 4979–4993, Oct. 2017.

[20] J. F. Mota, J. M. Xavier, P. M. Aguiar, and M. Püschel, “Distributed optimization with local domains: Applications in MPC and network flows,” *IEEE Trans. Autom. Control*, vol. 60, no. 7, pp. 2004–2009, Jul. 2015.

[21] E. L. Lehmann and G. Casella, *Theory of Point Estimation*. Berlin, Germany: Springer Science+Business Media, 2006.