DISTRIBUTED BAYESIAN ESTIMATION OF ARRIVAL RATES IN ASYNCHRONOUS MONITORING NETWORKS

Angelo Coluccia and Giuseppe Notarstefano

Department of Engineering, Università del Salento
{angelo.coluccia,giuseppe.notarstefano}@unisalento.it.

ABSTRACT

In this paper we consider a network of agents monitoring a spatially distributed traffic process. Each node measures the number of arrivals seen at its monitoring point in a given time-interval. We propose an asynchronous distributed approach based on a hierarchical Bayes model with unknown hyper-parameter, which allows each node to compute the minimum mean square error (MMSE) estimator of the local arrival rate by suitably fusing the information from the whole network. Simulation results show that the distributed scheme improves the estimation accuracy compared to a purely decentralized setup and is reliable even in presence of limited local data. An ad-hoc algorithm with reduced complexity is also proposed, which performs very closely to the optimal MMSE estimator.

Index Terms— distributed estimation, Empirical Bayes, consensus, traffic network, monitoring.

1. INTRODUCTION

Rate estimation in arrival processes, like those arising in traffic networks, is an important problem with several practical applications. Typical scenarios are the estimation of traffic flows in packet-switched networks [1], both wired and wireless, and more in general queue analysis [2]. Recently, increasing attention has been devoted to Intelligent Transportation Systems (ITS) for smart cities [3]. In several application scenarios, like e.g. traffic jam avoidance, safety at road intersections, etc., networks of sensor devices are used to collect information concerning the operating environment [4, 5].

Distributed estimation has received a widespread attention in the distributed computation literature, especially as a natural application of linear (average) consensus algorithms, see, e.g. [6] for a recent survey. Nodes typically perform local estimation based on local data, and then interact iteratively with their neighbors to carry out the estimation task [7–9]. In alternative (dynamic) methods, the nodes keep collecting new measurements while interacting with each other [10, 11].

Maximum-Likelihood (ML) approaches for distributed estimation have been considered for parameter estimation [12]. In [13, 14] consensus-based algorithms have been developed to compute the solution of a ML optimization problem. In [15] consensus-based algorithms for parameter estimation have been developed in a linear Bayesian framework. In [7] a distributed Alternating Direction Method of Multipliers (ADMM) has been developed for distributed ML estimation of vector parameters in a wireless sensor network. In [16] we have introduced a distributed estimator for binary event probabilities, based on a hierarchical Bayesian approach and a dual decomposition distributed optimization algorithm.

Following the idea in [16], we propose a distributed estimation scheme based on the Empirical Bayes approach, in which the arrival rates are treated as random variables whose prior distribution is parametrized by a suitable unknown hyperparameter. We show that the ML estimator of the hyperparameter is obtained by a separable optimization problem that can be solved in a distributed way over the network. Moreover, we propose an alternative ad-hoc distributed estimator that, although suboptimal, performs comparably to the optimal one. The resulting algorithm has a simple updated rule based on linear consensus protocols and exhibits the same appealing convergence properties as consensus protocols.

2. MONITORING NETWORK AND ARRIVAL-RATE ESTIMATION PROBLEM

We consider a *network of monitors* having *sensing*, *communication* and *computation* capabilities. Each monitor can measure the number of arrivals in a given fixed *timescale interval* (e.g., 1 second or 1 minute), share local data with neighboring agents, and perform local computations on its own and its neighbors' data. The objective is to fuse the data in order to improve the estimation of the arrival rates. Each node collects measurements *asynchronously* in a certain period of time, i.e. in an *observation window*, over which the underlying process can be assumed to be stationary. Accordingly, for each monitor $i \in \{1, \dots, N\}$ we introduce the following variables:

- λ_i unknown arrival rate
- $y_{i,\ell}$ the ℓ-th measurement of the number of arrivals in the observation window, i.e. $y_{i,\ell}|\lambda_i \sim \text{Poisson}(\lambda_i)$ i.i.d.
- $n_i \geq 1$ number of measurements $y_{i,\ell}$ collected in the observation window.

978-1-4799-2893-4/14/$31.00 ©2014 IEEE

We assume that the network evolution is triggered by a *universal slotted time*, $t \in \mathbb{Z}_{\geq 0}$, not necessarily known by the monitors. The monitors communicate according to a time-dependent directed communication graph $t \mapsto \mathcal{G}(t) = (\{1,\ldots,N\}, E(t))$, where $\{1,\ldots,N\}$ are the monitor *identifiers* and the edge set $E(t)$ describes the communication among monitors: $(i,j) \in E(t)$ if monitor i communicates to j at time $t \in \mathbb{Z}_{\geq 0}$. For each node i, the nodes sending information to i at time t, i.e., the set of $j \in N$ such that $(i,j) \in E(t)$ are called *in-neighbors* of i at time t. The set of in-neighbors of i at t is denoted by $N_i^I(t)$. We make the following minimal assumption on the graph connectivity:

Assumption 2.1 (Uniform joint strong connectivity)**.** *There exists an integer* $Q \geq 1$ *such that the graph* $\left(\{1,\ldots,N\}, \bigcup_{\tau=tQ}^{(t+1)Q-1} E(\tau)\right)$ *is strongly connected* $\forall\, t \geq 0$.

3. DISTRIBUTED ARRIVAL-RATE ESTIMATION VIA EMPIRICAL BAYES

In a *decentralized set-up*, in which nodes do not communicate, each node could estimate λ_i based on the sample $\{y_{i,\ell}, \ell = 1,\ldots,n_i\}$ by simply computing the empirical mean of the available measures. That is, the decentralized estimator is $\hat{\lambda}_i^{\text{dec}} = \frac{1}{n_i}\sum_{\ell=1}^{n_i} y_{i,\ell} = \frac{\sigma_i}{n_i}$, where $\sigma_i \stackrel{\text{def}}{=} \sum_{\ell=1}^{n_i} y_{i,\ell}$ is the cumulative sum of the measurements at node i.

Notice that, the decentralized estimator turns out to be the Maximum Likelihood estimator of λ_i when node i can use only its own data. However, decentralized estimation yields reliable estimates only when the number of samples n_i is large enough. In our heterogenous set-up, it may happen that some nodes satisfy such a condition, while other ones do not have enough data, thus resulting in a poor estimation. In this paper we propose a distributed estimation scheme in which every node, especially the ones with fewer measurements, can take advantage from communicating with neighboring nodes.

3.1. Empirical Bayes approach in monitoring networks

In applying a Bayesian estimation approach to a network context, the assumption that the prior is known to all monitors may be too strong. To avoid this drawback we adopt the Empirical Bayes approach in which only the class of the prior is known, while the parameters need to be estimated.

Formally, in our network set-up, we assume that each λ_i is also a random variable. In particular, the λ_is are independent identically distributed (i.i.d) Gamma variables, i.e. $\lambda_i \sim \text{Gamma}(a,b)$, where the shape parameter a is known, while the scale parameter b is unknown[1].

[1]The assumption that a is known, while only b is unknown, says that only the shape of the Gamma distribution (determined by the parameter a) is known, while the scaling is not. This assumption is reasonable in many applications, since it is a way to embed a rough information on the phenomenon, and is customary for the sake of mathematical tractability [17]. Notice though that our model is more general than the classical Gamma-Poisson hierarchy, since the n_i's are different (non-homogeneous sample).

The hyperparameter b can be estimated via a Maximum Likelihood (ML) procedure. To this aim, it is necessary to derive the joint distribution of all measurements $\{y_{i,\ell}\}_{\ell=1,\ldots,n_i}$ for each agent i. Since these are independent, the likelihood function for the estimate of the hyperparameter b is the product of the marginal distributions of all agents:

$$L(\boldsymbol{y}_1,\ldots,\boldsymbol{y}_N|b) = \prod_{i=1}^{N} p(\boldsymbol{y}_i|b) \tag{1}$$

where $\boldsymbol{y}_i = [y_{i,1} \ \cdots \ y_{i,n_i}]^T$. The marginal distribution of agent i is derived from the joint distribution of $\boldsymbol{y}_i$ and λ_i,

$$\begin{aligned} p(\boldsymbol{y}_i|b) &= \int_0^\infty \left(\prod_{\ell=1}^{n_i} f(y_{i,\ell}|\lambda_i,b)\right) p(\lambda_i|b)\mathrm{d}\lambda_i \\ &= \int_0^\infty \frac{\lambda_i^{\sigma_i} e^{-n_i\lambda_i}}{\prod_{\ell=1}^{n_i} y_{i,\ell}!} \frac{\lambda_i^{a-1} e^{-\lambda_i/b}}{\Gamma(a) b^a}\mathrm{d}\lambda_i \\ &= \frac{\Gamma(\sigma_i + a)}{\Gamma(a) b^a \prod_{\ell=1}^{n_i} y_{i,\ell}!}\left(\frac{b}{n_i b+1}\right)^{\sigma_i + a} \end{aligned} \tag{2}$$

By using eq. (2) into eq. (1) the likelihood is rewritten as:

$$L(\boldsymbol{y}_1,\ldots,\boldsymbol{y}_N|b) \propto \frac{1}{b^{aN}} \prod_{i=1}^{N}\left(\frac{b}{n_i b+1}\right)^{\sigma_i + a} \tag{3}$$

from which the ML estimator $\hat{b}$ of b can be found by solving the following optimization problem:

$$\hat{b} = \underset{b \in \mathbb{R}_+}{\arg\min}\left\{aN\log b - \sum_{i=1}^{N}(\sigma_i + a)\log\left(\frac{b}{n_i b+1}\right)\right\}, \tag{4}$$

which can be shown to be strictly convex.

The problem can be solved in closed-form only for the *homogeneous* case where all n_i's are equal. This means that $n_i = n/N$, being n the total number of measurements. In this case the ML estimator of b based on the entire set of measurements is given by

$$\hat{b}^{\text{hom}} = \frac{1}{an}\sum_{i=1}^{N}\sigma_i = \frac{\sigma}{an} \tag{5}$$

where $n \stackrel{\text{def}}{=} \sum_{i=1}^{N} n_i$, and $\sigma \stackrel{\text{def}}{=} \sum_{i=1}^{N}\sigma_i$.

After obtaining an estimate for b, the Empirical Bayes estimator of the arrival rate λ_i that minimizes the Mean Square Error (MMSE) can be obtained by computing the conditional mean of the posterior distribution $p(\lambda_i|\boldsymbol{y}_i,b)$. The latter is given by the ratio between the joint pdf $p(\boldsymbol{y}_i,\lambda_i|b)$ and the marginal pdf $p(\boldsymbol{y}_i|b)$ as from eq. (2), i.e.:

$$p(\lambda_i|\boldsymbol{y}_i,b) = \frac{\lambda_i^{\sigma_i+a-1} e^{-\lambda_i \frac{n_i b+1}{b}}}{\Gamma(\sigma_i + a)}\left(\frac{b}{n_i b+1}\right)^{-\sigma_i - a} \tag{6}$$

Eq. (6) is a Gamma pdf with parameters $(\sigma_i + a, \frac{b}{n_i b+1})$, hence the Empirical Bayes MMSE estimator of each λ_i is

$$\hat{\lambda}_i = \mathrm{E}[\lambda_i | \boldsymbol{y}_i, \hat{b}] = \frac{\hat{b}}{\hat{b} n_i + 1}(a + \sigma_i). \quad (7)$$

In the following we will also consider an *ad-hoc estimator* obtained by using $\hat{b}^{\text{hom}}$ instead of the optimal $\hat{b}$, i.e.:

$$\hat{\lambda}_i^{\text{ad-hoc}} \stackrel{\text{def}}{=} \frac{\hat{b}^{\text{hom}}}{\hat{b}^{\text{hom}} n_i + 1}(a + \sigma_i) = \frac{\sigma}{an + \sigma n_i}(a + \sigma_i).$$

3.2. Distributed Empirical Bayes estimator

From eq. (7) it is clear that each agent can compute the Empirical Bayes MMSE estimator provided it knows $\hat{b}$. The optimization problem (4) giving the ML estimator of b has a separable cost function (i.e., the total cost is the sum of N local costs), hence it can be solved by using available distributed optimization algorithms for unconstrained optimization. An example of algorithm working on general asynchronous networks (under Assumption 2.1 of uniform joint strong connectivity) is [18]. We assume that each node implements the local update rule of the chosen distributed optimization algorithm (e.g., equation (1) in [18]). We let each node i have a local state x_i and an estimate $\hat{b}_i$ of $\hat{b}$. At each $t \in \mathbb{N}$ the node runs the local update rule

$$\begin{aligned}&(\hat{b}_i(t+1), x_i(t+1)) = \\ &\qquad \texttt{update\_local}\left(\hat{b}_i(t), x_i(t), \{x_j(t)\}_{j \in N_i^I(t)}; \gamma(t)\right),\end{aligned} \quad (8)$$

where $\{x_j(t)\}_{j \in N_i^I(t)}$ is the collection of states of the in-neighbors of node i, `update_local` is the local update of the chosen distributed optimization algorithm, and γ is an algorithm parameter as, e.g., a step-size. The distributed optimization algorithm guarantees that all node reach consensus on the minimizer of (4). That is,

$$\lim_{t \to \infty} \hat{b}_i(t) = \hat{b}, \qquad \text{for all } i \in \{1, \dots, N\}.$$

The distributed estimation algorithm is as follows. At each $t \in \mathbb{N}$, each agent i stores a local state $x_i(t)$, an estimate $\hat{b}_i(t)$ of $\hat{b}$ and an estimate $\hat{\lambda}_i(t)$ of $\hat{\lambda}_i$. The node initializes its local state x_i according to the chosen distributed optimization algorithm, sets $\hat{b}_i(0) = \sigma_i/(an_i)$ (which would be the solution of (4) if i were the only agent) and $\hat{\lambda}_i(0) = \frac{\hat{b}_i(0)}{\hat{b}_i(0) n_i + 1}(a + \sigma_i)$. Then it updates its estimate of $\hat{b}$ by using (8) and updates the current estimate $\lambda_i(t)$ by using (7).

From the convergence properties of the chosen distributed optimization algorithm it follows immediately that the proposed distributed estimator asymptotically computes at each node i the Empirical Bayes MMSE estimator of λ_i. However, most of the available distributed optimization algorithms, as the ones in [18,19], need the tuning of a global parameter (we denoted it γ). Also, typically they exhibit a sub-exponential convergence even in static graphs. To overcome these drawbacks we propose an alternative *ad-hoc distributed estimator* that, although suboptimal, performs comparably to the optimal one. The algorithm is defined as follows.

For each $t \in \mathbb{N}$, each node $i \in \{1, \dots, N\}$ stores in memory two local states $s_i(t)$ and $\eta_i(t)$, an estimate $\hat{b}_i^{\text{hom}}(t)$ of $\hat{b}^{\text{hom}}$, and an estimate $\hat{\lambda}_i^{\text{ad-hoc}}(t)$ of $\hat{\lambda}_i^{\text{ad-hoc}}$. For $(i,j) \in E(t)$ let $w_{ij}(t) \in \mathbb{R}_+$ be a set of weights, satisfying $\sum_{i=1}^n w_{ij}(t) = 1$.

Initialization: $s_i(0) = \sigma_i$, $\eta_i(0) = n_i$, $\hat{b}_i^{\text{hom}}(0) = \sigma_i/(an_i)$, $\hat{\lambda}_i^{\text{ad-hoc}}(0) = \frac{\hat{b}_i^{\text{hom}}(0)}{\hat{b}_i^{\text{hom}}(0) n_i + 1}(a + \sigma_i)$.

Iterate:

$$\begin{aligned} s_i(t+1) &= \sum_{j \in N_i^I(t) \cup \{i\}} w_{ij}(t) s_j(t) \\ \eta_i(t+1) &= \sum_{j \in N_i^I(t) \cup \{i\}} w_{ij}(t) \eta_j(t) \\ \hat{b}_i^{\text{hom}}(t+1) &= \frac{s_i(t+1)}{a\, \eta_i(t+1)} \\ \hat{\lambda}_i^{\text{ad-hoc}}(t+1) &= \frac{\hat{b}_i^{\text{hom}}(t+1)}{\hat{b}_i^{\text{hom}}(t+1) n_i + 1}(a + \sigma_i). \end{aligned} \quad (9)$$

Proposition 3.1. *Assume that Assumption 2.1 holds. Then the distributed algorithm* (9) *computes the ad-hoc estimator, i.e.,*

$$\lim_{t \to \infty} \hat{b}_i^{hom}(t) = \hat{b}^{hom}$$
$$\lim_{t \to \infty} \hat{\lambda}_i^{ad\text{-}hoc}(t) = \hat{\lambda}_i^{ad\text{-}hoc}.$$

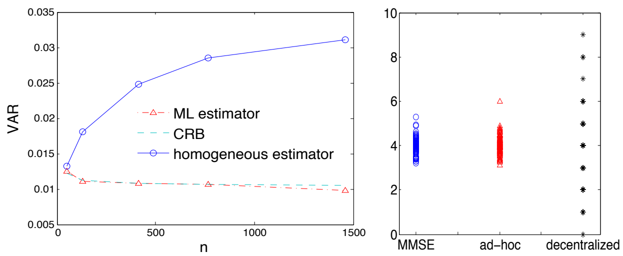
Following [20], the proof relies on: (i) bounding $\max_{i \in \{1,\dots,N\}} |\hat{b}_i^{\text{hom}}(t) - \hat{b}| \le \max_{i \in \{1,\dots,N\}} |\hat{b}_i^{\text{hom}}(0)| f(t)$, where $f(t)$ is a decreasing function of time, and (ii) showing the weak ergodicity of the weight matrices, by using Assumption 2.1. The proof is omitted due to lack of space.

Remark 3.2. *The distributed algorithm* (9) *exhibits the exponential convergence of linear consensus protocols, as opposed to, for example, the $O((\log t)/\sqrt{t})$ rate of [18].*

Remark 3.3. *We want to stress that t is a universal time that the monitors do not need to know to implement the estimation algorithm above. Thus, the algorithm is well-suited for a completely asynchronous implementation. Simply, it is enough to assume that when node i is not active, then there are no edges going out nor coming in agent i (i.e., $N_i^I(t) = \emptyset$).*

4. PERFORMANCE ANALYSIS

The Cramer-Rao (lower) Bound (CRB) for the estimation of b turns out to be $\mathrm{CRB} = \frac{b/a}{\sum_{i=1}^N \frac{n_i}{n_i b + 1}}$, which for $n_i = n/N$ becomes $\mathrm{CRB}^{\text{hom}} = \frac{b}{an}\left(\frac{n}{N} b + 1\right)$. The variance of the homo-



(a) estimation of b (b) estimation of λ_i

Fig. 1. inhomogeneous

geneous estimator is $\mathrm{VAR}\left[\hat{b}^{\mathrm{hom}}\right] = \frac{b}{an}\left(\frac{b}{n}\sum_{i=1}^{N} n_i^2 + 1\right)$, which is generally higher than $\mathrm{VAR}[\hat{b}]$.

For large sample size one can expect that the ML estimator $\hat{b}$ attains the CRB due to asymptotic efficiency. Lacking a closed-form, the variance of $\hat{b}$ can be however obtained via Monte Carlo simulation. For a network of $N = 20$ agents performing a different number of measurements, Fig. 1(a) reports the variance of $\hat{b}$ along with the CRB and the variance of the homogeneous estimator $\hat{b}^{\mathrm{hom}}$. The different points are obtained by increasing the number of measurements of each node at different rates (logarithmic, linear and quadratic). The total number $n = \sum_{i=1}^{N} n_i$ is reported in the abscissa. The plot reveals that the ML estimator $\hat{b}$ exhibits excellent performance and attains the CRB for the variance, even with moderate sample size[2]. The performance of $\hat{b}^{\mathrm{hom}}$ are generally worse, and depend on the "level of inhomogeneity" in the number of measurements across agents. In fact, as the n_i's distribute more homogeneously, i.e., their values do not change much across agents, it is easy to verify that $\mathrm{VAR}[\hat{b}^{\mathrm{hom}}] \to \mathrm{CRB}^{\mathrm{hom}}$, as one might expect due to the optimality of $\hat{b}^{\mathrm{hom}}$ for the homogeneous case. Nonetheless, the estimator of λ_i based on $\hat{b}^{\mathrm{hom}}$, i.e., $\hat{\lambda}_i^{\text{ad-hoc}}$, performs comparably to the optimal MMSE estimator. This is a merit of the hierarchical Bayesian formulation, which is robust to deviation from the exact prior distribution [17]. Fig. 1(b) illustrates this appealing property for a generic agent i with $n_i = 1$ and $\lambda_i \approx 4$ for 100 realizations: remarkably, $\hat{\lambda}_i^{\text{ad-hoc}}$ is very close to the optimal $\hat{\lambda}_i$, while the decentralized estimator has a very poor accuracy being it based on the sole $y_{i,1}$.

A more insightful performance assessment is reported in Fig. 2. The root mean square error (RMSE) of the proposed ad-hoc estimator stays always very close to the optimal MMSE estimator, irrespective of the local number of measurements n_i, and the accuracy improves as n grows. This means that the local estimation task benefits from the increasing number of measurements performed by the network (better estimate of b) even when the local n_i remains fixed.

[2]Notice that values (slightly) below the CRB can occur due to the possible (small) bias for finite sample size which may trade-off some variance, while the homogeneous estimator is always unbiased.

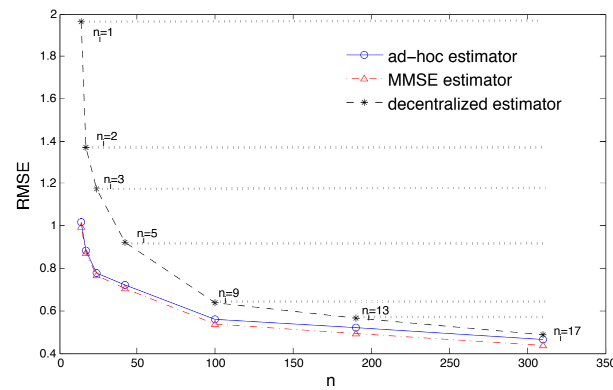


Fig. 2. Root mean square error for the estimation of λ_i as a function of the total number of measurements n.

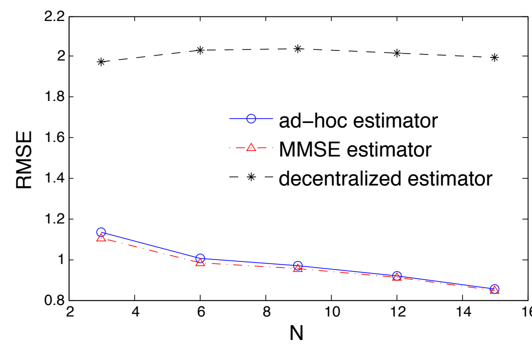


Fig. 3. Root mean square error for the estimation of λ_i as a function of the number of monitors N.

Conversely, the decentralized estimator provides reliable estimates only for sufficiently large n_i — several values are reported in the figure, with dotted horizontal lines highlighting the inability of $\hat{\lambda}_i^{\mathrm{dec}}$ to gain information from the network.

Finally, Fig. 3 reports the RMSE as function of the network size N. The figure shows that the estimation accuracy improves as more agents are available, since even a single measurements per-node will increase the aggregate n hence, in turn, the performance of the distributed estimators.

5. CONCLUSION

In this paper we have proposed a novel distributed scheme, based on a hierarchical Bayes approach, for the estimation of arrival rates in asynchronous monitoring networks. The proposed distributed approach gains information from the network and generally outperforms the decentralized estimator, especially when the node local information is insufficient. In particular, we have developed an ad-hoc distributed algorithm that performs closely to the optimal MMSE estimator, but is much simpler to implement and exhibits faster convergence. A formal theoretical analysis of the estimator performance is part of our ongoing work.

6. REFERENCES

[1] F. Ricciato, A. Coluccia, A. D'Alconzo, D. Veitch, P. Borgnat, and P. Abry, "On the role of flows and sessions in internet traffic modeling: an explorative toy-model," in *IEEE GLOBECOM*, 2009.

[2] L. Kleinrock, *Queueing Systems: Theory, Volume 1*. Wiley, 1976.

[3] T. Le, C. Cai, and T. Walsh, "Adaptive signal-vehicle cooperative controlling system," in *14th International IEEE Conference on Intelligent Transportation Systems (ITSC)*, 2011.

[4] S. Giri and R. Wall, "A safety critical network for distributed smart traffic signals," *IEEE Instrumentation Measurement Magazine*, vol. 11, no. 6, 2008.

[5] D. DeVoe and R. Wall, "A distributed smart signal architecture for traffic signal controls," in *IEEE International Symposium on Industrial Electronics*, 2008.

[6] F. Garin and L. Schenato, "A survey on distributed estimation and control applications using linear consensus algorithms," *Networked Control Systems*, pp. 75–107, 2011.

[7] I. D. Schizas, A. Ribeiro, and G. B. Giannakis, "Consensus in ad hoc WSNs with noisy links Part I: Distributed estimation of deterministic signals," *IEEE Transactions on Signal Processing*, vol. 56, no. 1, pp. 350–364, 2008.

[8] I. D. Schizas, G. B. Giannakis, S. I. Roumeliotis, and A. Ribeiro, "Consensus in ad hoc WSNs with noisy links Part II: Distributed estimation and smoothing of random signals," *IEEE Transactions on Signal Processing*, vol. 56, no. 4, pp. 1650–1666, 2008.

[9] S. Sardellitti, M. Giona, and S. Barbarossa, "Fast distributed average consensus algorithms based on advection-diffusion processes," *IEEE Transactions on Signal Processing*, vol. 58, no. 2, 2010.

[10] F. Cattivelli, C. G. Lopes, and A. H. Sayed, "Diffusion recursive least-squares for distributed estimation over adaptive networks," *IEEE Transactions on Signal Processing*, vol. 56, no. 5, 2008.

[11] P. Braca, S. Marano, and V. Matta, "Enforcing consensus while monitoring the environment in wireless sensor networks," *IEEE Transactions on Signal Processing*, vol. 56, no. 7, 2008.

[12] S. Barbarossa and G. Scutari, "Decentralized maximum-likelihood estimation for sensor networks composed of nonlinearly coupled dynamical systems," *IEEE Transactions on Signal Processing*, vol. 55, no. 7, 2007.

[13] F. Fagnani, S. M. Fosson, and C. Ravazzi, "Input driven consensus algorithm for distributed estimation and classification in sensor networks," in *50th IEEE Conference on Decision and Control and European Control Conference (CDC-ECC)*. IEEE, 2011, pp. 6654–6659.

[14] A. Chiuso, F. Fagnani, L. Schenato, and S. Zampieri, "Gossip algorithms for simultaneous distributed estimation and classification in sensor networks," *IEEE Journal of Selected Topics in Signal Processing*, vol. 5, no. 4, pp. 691–706, 2011.

[15] D. Varagnolo, G. Pillonetto, and L. Schenato, "Distributed consensus-based bayesian estimation: sufficient conditions for performance characterization," in *American Control Conference (ACC), 2010*. IEEE, 2010, pp. 3986–3991.

[16] A. Coluccia and G. Notarstefano, "Distributed estimation of binary event probabilities via hierarchical bayes and dual decomposition," in *52nd IEEE Conference on Decision and Control*, 2013.

[17] E. Lehmann and G. Casella, *Theory of Point Estimation*. Springer, 1998.

[18] A. Nedic and A. Olshevsky, "Distributed optimization over time-varying directed graphs," *arXiv preprint arXiv:1303.2289*, 2013.

[19] F. Zanella, D. Varagnolo, A. Cenedese, G. Pillonetto, and L. Schenato, "Asynchronous newton-raphson consensus for distributed convex optimization," in *3rd IFAC Workshop on Distributed Estimation and Control in Networked Systems (NecSys'12)*, 2012.

[20] F. Bénézit, V. Blondel, P. Thiran, J. Tsitsiklis, and M. Vetterli, "Weighted gossip: Distributed averaging using non-doubly stochastic matrices," in *IEEE International Symposium on Information Theory Proceedings (ISIT)*. IEEE, 2010, pp. 1753–1757.