2011 50th IEEE Conference on Decision and Control and
European Control Conference (CDC-ECC)
Orlando, FL, USA, December 12-15, 2011

Input driven consensus algorithm for distributed estimation and classification in sensor networks

Fabio Fagnani, Sophie M. Fosson, and Chiara Ravazzi

***Abstract*— This paper deals with the problem of simultaneously classifying sensors and estimating hidden parameters in a network with communication constraints. In particular, we consider a network where sensors measure a common parameter with different precision rank. The goal of each unit is to estimate the unknown parameter and its own specific type through local communication and computation. Here, we present a decentralized version of the centralized maximum likelihood (ML) estimator. Each sensor computes local sufficient statistics by using its own observations and transmits its local information to its neighborhood. By using an Input Driven Consensus Algorithm (IDCA), the local information can be gradually propagated through the entire network, allowing to estimate the global parameter. We prove the convergence of the proposed algorithm and we show that the relative classification error converges to that of the centralized ML as the network dimension goes to infinity. We also compare this strategy with implementation of expectation-maximization (EM) algorithm via numerical simulations.**

I. Introduction

The recent enhancements in wireless technology have favored the employment of sensor networks in widespread engineering and industrial applications. Consider the problem of estimating a parameter that each sensor can measure with a certain degree of precision. When no supervision is available, each unit is required to locally cooperate with its neighbors according to rules that allow to achieve a global consensus on the parameter's value (see, e.g., [1] and the references therein).

In real applications, sensors are subjected to failures owing to which their measurements may become strongly inaccurate. This motivates the recent research on fault-tolerant networks, which must (a) self-detect the corrupted sensors and (b) perform a reliable estimation on the unknown parameter despite the partial damage. The first task is a classification problem: the sensors cluster into two sets ("working" and "faulty" sensors), and each of them must evaluate its own states through local information sharing; the second one, instead, requires the derivation of distributed estimation algorithms robust to failures. Such a dual problem has been studied, e.g., in [1].

The aim of this work is the development of a decentralized algorithm that performs classification and estimation in the same iterative procedure. The model we will consider is the following: $y_i = \theta + T_i n_i$ is the measurement performed by sensor i, θ being the unknown parameter, n_i's independent gaussian noises $\mathsf{N}(0,1)$, and T_i's sensor standard deviations. If i is functioning regularly, $T_i = \alpha$, while if i is faulty, $T_i = \beta \gg \alpha$.

Analogously to the issue studied in [1], the problem of estimating T_i is envisaged in the general context of unsupervised clustering and of gaussian mixtures estimation [2], [3]. As a difference from [1], in which T_i is an additive term, in our model T_i is a multiplicative factor, which depicts a different scenario. Based on maximum likelihood (ML) estimation, the algorithm we propose in this work consists of an iterative, decentralized procedure and has been inspired by the consensus propagation protocol introduced in [4], [5].

A. Relation to prior literature and our contribution

1) Standard EM: The most popular technique in statistical estimation problems, such as mixture problems, is the Expectation Maximization (EM) algorithm [6]. In its centralized form, it is an iterative procedure which seeks to compute the ML estimates of the model parameters for which the observed data are the most likely. EM alternates an expectation step (E-step), which estimates the unobserved variables based on the observed data, with a maximization step (M-step), which maximizes the likelihood function using estimates obtained in the E-step. The convergence is guaranteed since each iteration increases the likelihood. As drawback, the implementation of EM on a sensor network requires a complete communication graph since the computations need the information of all nodes in the network. We refer the reader to [6] for a derivation of EM equations which can be easily adapted to our specific problem.

2) Distributed EM algorithms: In [3], [7] distributed versions of the EM algorithm are proposed for estimation of gaussian mixtures: a network is given where each node independently performs the E-step through local observations. In particular, in [7] a consensus filter is used to propagate the local information. The tricky point of such techniques is the choice of the number of averaging iterations between two consecutive M-steps, which must be sufficient to reach consensus.

3) Belief propagation in Bayesian networks: It is worth mentioning a recent line of research on the use of belief propagation (BP) as an asynchronous protocol to solve the classification problem [5], [4]. However, the convergence of these algorithms, providing each node with the most probable class based on all observations, is guaranteed only for particular topologies of the communication graph (when the network is a tree or a regular graph).

The authors are with Dipartimento di Matematica (DIMAT), Politecnico di Torino, Italy name.surname@polito.it

S.M. Fosson's research is supported by Regione Autonoma Valle d'Aosta and Fondo Sociale Europeo. C. Ravazzi's research is supported by MIUR under PRIN project no. 20087W5P2K.

B. Our contribution and outline of the paper

Here, we explore a different strategy for the ML estimation. More specifically, we propose a distributed algorithm, which we call Input Driven Consensus Algorithm (IDCA). What is new with respect to the previous literature is that the algorithm does not perform repetitive average calculations: it consists of computation of two averages and an hard decision. Moreover, it is possible to obtain for IDCA an accurate analysis of the convergence. Another advantage is that the algorithm can be generalized to scenarios where the parameters θ and T belong to multidimensional spaces.

The paper is organized as follows. In Section II, we formally present our model and we approach it by classical ML estimation; in Section III, we introduce IDCA and the theoretical results. Afterwards, we show the outcomes of some simulations (Section IV), and we collect some concluding remarks in Section V. An Appendix, containing the sketch of some theoretical proofs, completes the paper.

II. Problem statement

A. The model

In our model, we consider a network, represented by a directed graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$, where $\mathcal{E} \subseteq \mathcal{V} \times \mathcal{V}$. $\mathcal{G}$ represents the system communication architecture and $|\mathcal{V}| = N$ is the number of agents. We assume that each node $i \in \mathcal{V}$ measures the observable

$$y_i = \theta^\star + T_i^\star n_i \tag{1}$$

where, we recall, $\theta^\star \in \mathbb{R}$, n_i's are independent gaussian noises $\mathsf{N}(0,1)$ and each $T_i^\star \in \{\alpha, \beta\}$. In summary, data $y = (y_1, \ldots, y_N)$ are distributed according to the probability density function

$$f(y|\theta^\star, T^\star) = \prod_{i \in \mathcal{V}} f(y_i|\theta^\star, T_i^\star)$$

$$f(y_i|\theta^\star, T_i^\star) = \frac{1}{\sqrt{2\pi} T_i^\star} \mathrm{e}^{-\frac{(y_i - \theta^\star)^2}{2 T_i^{\star 2}}}.$$

We assume an a-priori distribution on the state of the sensors $T_i^\star$'s: they are assumed to be independent and we let p to be the probability that each $T_i^\star$ is equal β.

The goal of each unit is to estimate the parameter $\theta^\star$ and its own specific configuration $T_i^\star \in \{\alpha, \beta\}$ that is more likely to have generated the observation y_i. Notice that the presence of the common parameter $\theta^\star$ imposes a coupling between the different nodes.

B. ML-estimator

Given two vectors $v, w \in \{\alpha, \beta\}^N$ let $\mathrm{d_H}(v, w) = |\{i \in \{1, \ldots, N\} : v_i \neq \lambda\}|$ where $\mathbb{1} = (1, 1, \ldots, 1)^\mathsf{T} \in \mathbb{R}^N$. It would be natural to take as estimators of vector $T^\star = (T_1^\star, \ldots, T_N^\star)^\mathsf{T}$ and of $\theta^\star$ the ones which minimize $\mathbb{E}[\mathrm{d_H}(T^\star, \widehat{T})]$ and $\mathbb{E}[|\theta^\star - \widehat{\theta}|^2]$ over all $\widehat{T} \in \{\alpha, \beta\}^N$ and over all $\widehat{\theta} \in \mathbb{R}$. However, these optimal estimators are computationally intractable and the computation is difficult to decentralize. We consider instead the ML-estimator of the posterior probability, defined as follows:

$$(\widehat{\theta}_{ML}, \widehat{T}_{ML}) = \operatorname*{argmax}_{\theta \in \mathbb{R}, T \in \{\alpha,\beta\}^N} f(y|\theta, T) p(T) = \operatorname*{argmax}_{\theta \in \mathbb{R}, T \in \{\alpha,\beta\}^N} L(\theta, T)$$

where

$$L(\theta, T) = -\sum_{k=1}^N \frac{(y_k - \theta)^2}{2 T_k^2} - \eta \sum_{i=1}^N T_i \tag{2}$$

and $\eta = \frac{1}{\beta - \alpha} \ln \left[\frac{(1-p)}{p} \frac{\beta}{\alpha} \right]$.

The ML-estimator is typically used for finite mixtures classification problems [3] and does not have a closed form solution. Our goal is to find a distributed algorithm solving this optimization problem.

The max-problem can be easily solved for a fixed T:

$$\widehat{\theta}(T) = \operatorname*{argmax}_{\theta \in \mathbb{R}} L(\theta, T) = \frac{\sum_{k=1}^N \frac{y_k}{T_k^2}}{\sum_{k=1}^N \frac{1}{T_k^2}} \tag{3}$$

Clearly, $\widehat{T}_{ML} = \operatorname{argmax} L(\widehat{\theta}(T), T)$ and $\widehat{\theta}_{ML} = \widehat{\theta}(\widehat{T}_{ML})$.

Equivalently, the problem can be solved by maximizing $L(\theta, T)$ for a fixed θ. The classification rule becomes

$$\widehat{T}_i(\theta) = \begin{cases} \alpha & \text{if } |y_i - \theta| < \delta \\ \beta & \text{otherwise} \end{cases} \qquad \delta = \sqrt{2 \frac{\ln\left(\frac{1-p}{p} \frac{\beta}{\alpha}\right)}{\frac{1}{\alpha^2} - \frac{1}{\beta^2}}} \tag{4}$$

and $\widehat{\theta}_{ML} = \operatorname{argmax} L(\theta, \widehat{T}(\theta))$, $\widehat{T}_{ML} = \widehat{T}(\widehat{\theta}_{ML})$.

C. Performance metrics

As performance metrics to evaluate the goodness of an estimate $\widehat{T}$ of $T^\star$, we take the relative classification error over the network:

$$P_N(\widehat{T}) := \frac{1}{N} \mathbb{E}\left[\mathrm{d_H}(T^\star, \widehat{T})\right].$$

We denote by $P_N(\widehat{T}_{\mathrm{ML}})$ the classification error for the ML estimation.

Theorem 1:

$$\liminf_{N \to +\infty} P_N(\widehat{T}_{\mathrm{ML}}) \geq q(p, \alpha, \beta)$$

where

$$q(p, \alpha, \beta) = (1-p)\mathrm{erfc}^1\left(\frac{\delta}{\alpha\sqrt{2}}\right) + p\left[1 - \mathrm{erfc}\left(\frac{\delta}{\beta\sqrt{2}}\right)\right].$$

The proof, which is omitted for reasons of space, is based on the fact that $q = P_N(\widehat{T}_{LB})$ with

$$\widehat{T}_{LB} = \operatorname*{argmax}_{T \in \{\alpha,\beta\}^N} L\left(\operatorname*{argmax}_{\theta \in \mathbb{R}} L(\theta, T^\star), T\right).$$

The details can be found in [8].

For a fixed $\alpha, \beta \in \mathbb{R}$ the error probability $q(p, \alpha, \beta)$ vanishes when p goes to zero, as to be expected: $\lim_{p \to 0} q(p, \alpha, \beta) = 0$. Moreover, the dependence of function q on the parameters α and β is exclusively through their ratio β/α. In particular, we have

$$\lim_{\beta/\alpha \to +\infty} q(p, \alpha, \beta) = 0 \qquad \lim_{\beta/\alpha \to 1} q(p, \alpha, \beta) = 1.$$

[1] $\mathrm{erfc}(x)$ is the complementary error function defined as $\frac{2}{\sqrt{\pi}} \int_x^{+\infty} \mathrm{e}^{-t^2} \mathrm{d}t$

D. A naïve iterative algorithm: Hard-EM

Expressions (3) and (4) suggest a possible iterative algorithm. Set the initial estimate $\widehat{\theta}^{(0)} = \frac{1}{N}\sum_{i\in\mathcal{V}} y_i$ (which is a usual assumption that the estimation process begins from the average of initial conditions). Given the current estimate $\widehat{\theta}^{(t)}$:

$$\widehat{T}_i^{(t+1)} = \begin{cases} \alpha & \text{if } |y_i - \widehat{\theta}^{(t)}| < \delta \\ \beta & \text{otherwise} \end{cases}$$

$$\widehat{\theta}^{(t+1)} = \sum_{j=1}^N y_j \left[\widehat{T}_j^{(t+1)}\right]^{-2} \Big/ \sum_{j=1}^N \left[\widehat{T}_j^{(t+1)}\right]^{-2}.$$

The algorithm halts when $|\widehat{\theta}^{(t+1)} - \widehat{\theta}^{(t)}| < \epsilon$, where ϵ is a prescribed tolerance. Notice that this iterative algorithm can be seen as an hard version of the EM-algorithm. As drawback, the implementation of this iterative algorithm on a sensor network requires a complete communication graph since the computations need the information of all units in the network. Moreover, the convergence is not guaranteed in general cases.

III. Input driven consensus algorithm

A. Description of the algorithm

We propose an Input Driven Consensus Algorithm (IDCA) to make Hard-EM algorithm distributed. It seeks to estimate the average quantities $\frac{1}{N}\sum_{i\in\mathcal{V}} y_i T_i^{-2}$ and $\frac{1}{N}\sum_{i\in\mathcal{V}} T_i^{-2}$ and simultaneously uses them to classify T in an iterative and distributed way.

Formally, IDCA is parametrized by:

- a non-negative stochastic matrix P, adapted to the communication graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$, namely $P_{ij} > 0$ if $(j,i) \in \mathcal{E}$ and 0 otherwise;
- a real sequence $\gamma^{(t)} \in (0,1)$ $\forall t \in \mathbb{N}$, $\gamma^{(t)} \searrow 0$.

Every node i has two messages stored in its memory at time t, denoted with $\mu_i^{(t)}, \nu_i^{(t)}$. Given the initial conditions $\mu_i^{(0)} = 0, \nu_i^{(0)} = 0$ and the initial estimate $\widehat{T}_i^{(0)} = \alpha$, the dynamics consists of a convex combination, weighted by $\gamma^{(t)}$, of two contributions:

- the first term is the consensus part ($\sum_j P_{ij}\mu_j$ and $\sum_j P_{ij}\nu_j$, respectively)
- the second one is an input which takes into account the observations y_i and the current estimate on type $\widehat{T}_i^{(t)}$ ($y_i/[\widehat{T}_i^{(t)}]^2$ and $1/[\widehat{T}_i^{(t)}]^2$).

The update of parameters $\widehat{\theta}_i^{(t+1)}$ and of $\widehat{T}_i^{(t+1)}$ is then based on $\mu_i^{(t+1)}$ and $\nu_i^{(t+1)}$.

Input driven consensus algorithm (IDCA)

1) Initialize $\mu^{(0)} = 0, \nu^{(0)} = 0, \widehat{T}^{(0)} = \alpha\mathbb{1}$
2) For time $t \in \mathbb{N}$ do: for all $i \in \mathcal{V}$ do

$$\mu_i^{(t+1)} = (1-\gamma^{(t)}) \underbrace{\sum_j P_{ij}\mu_j^{(t)}}_{\text{consensus part}} + \gamma^{(t)} \underbrace{\frac{y_i}{\left[\widehat{T}_i^{(t)}\right]^2}}_{\text{input}} \quad (5a)$$

$$\nu_i^{(t+1)} = (1-\gamma^{(t)}) \underbrace{\sum_j P_{ij}\nu_j^{(t)}}_{\text{consensus part}} + \gamma^{(t)} \underbrace{\frac{1}{\left[\widehat{T}_i^{(t)}\right]^2}}_{\text{input}} \quad (5b)$$

$$\widehat{\theta}_i^{(t+1)} = \mu_i^{(t+1)}/\nu_i^{(t+1)} \quad (6a)$$

$$\widehat{T}_i^{(t+1)} = \begin{cases} \alpha & \text{if } |y_i - \widehat{\theta}_i^{(t+1)}| < \delta \\ \beta & \text{otherwise} \end{cases} \quad (6b)$$

3) end for

Since $\gamma^{(t)} \overset{t\to+\infty}{\longrightarrow} 0$ the injection of inputs vanishes when the number of iterations increases.

This algorithm provides a distributed protocol since the computations at each node require only information that is locally available.

B. Convergence and performance

The following theorem ensures the convergence of IDCA.

Theorem 2: Let $\gamma^{(t)} \in (0,1)$ $\forall t \in \mathbb{N}$, $\gamma^{(t)} \searrow 0$ and $\sum_t \gamma^{(t)} = +\infty$. Let $P \in \mathbb{R}_+^{N\times N}$ be doubly-stochastic, symmetric, primitive and such that all its eigenvalues are non-negative. Then,

1) there exist $t_0 \in \mathbb{N}$ and $\widehat{T}^{(\infty)} \in \{\alpha,\beta\}^N$ such that

$$\widehat{T}^{(t)} \overset{\text{a.s.}}{=} \widehat{T}^{(\infty)} \quad \forall t \geq t_0$$

$$\lim_{t\to+\infty} \widehat{\theta}^{(t)} \overset{\text{a.s.}}{=} \widehat{\theta}^{(\infty)} = \frac{\sum_{k=1}^N y_k \left[\widehat{T}_k^{(\infty)}\right]^{-2}}{\sum_{k=1}^N \left[\widehat{T}_k^{(\infty)}\right]^{-2}} \mathbb{1};$$

2) $(\widehat{\theta}^{(\infty)}, \widehat{T}^{(\infty)})$ is a local maximum of log-likelihood function $L(\theta, T)$ defined in (2).

This result guarantees that the estimates $\widehat{T}^{(t)}$ converge in a finite time almost surely. Moreover, it will be clear from the proof (see Appendix) that $||\widehat{\theta}^{(t)} - \widehat{\theta}^{(\infty)}|| = O(\gamma^{(t)})$ for $t \to \infty$.

Notice that the requirement of positive eigenvalues does not affect the generality of the problem. P in fact must match the network's topology (i.e., the zero entries are fixed), but its non-zero entries can always be chosen so that the eigenvalues are positive (it is sufficient to assign a sufficiently large weight on the diagonal).

It is easy to verify that $\liminf_{N\to+\infty} P_N(\widehat{T}^{(\infty)}) \geq q(p,\alpha,\beta)$ where q is defined in Theorem 1. We will show via simulations that for our algorithm the classification error is exactly given by q, for sufficiently large N.

IV. Simulations

In this section, we test IDCA and we compare it with the standard centralized EM and Hard-EM in terms of fraction of nodes that are not correctly identified. The sensors measure data according to the model in (1) with $\theta = 0$, $\alpha = 0.3$ and $\beta = 10$. The prior probability of having $T_i = \beta$ is set to $p = 0.25$. We consider different communication architectures: the complete graph, a circulant graph, a 2-dimensional grid and a random geometric graph with confidence radius $r = 0.3$. IDCA is parametrized by the matrix P adapted to the topology of the graph and by a sequence $\{\gamma^{(t)}\}_{t\in\mathbb{N}}$. More precisely, letting $\deg(i)$ be the number of agents communicating with i (i itself included), we have chosen $P_{ij} = 1/\deg(i)$ if $(j,i) \in \mathcal{E}$ and $P_{ij} = 0$

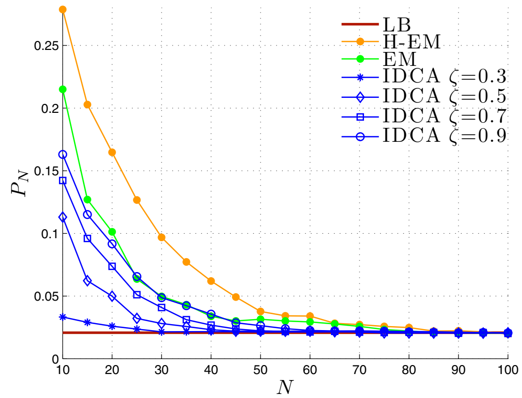

Fig. 1. Complete graph.

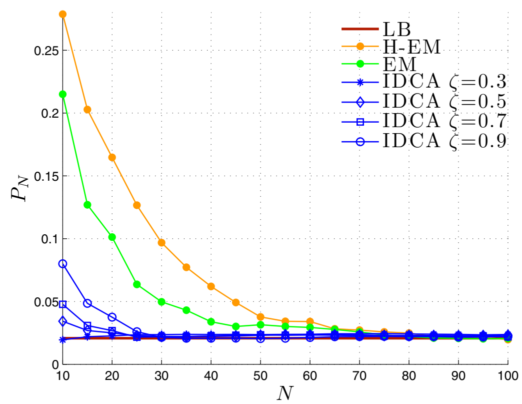

Fig. 2. Circulant graph.

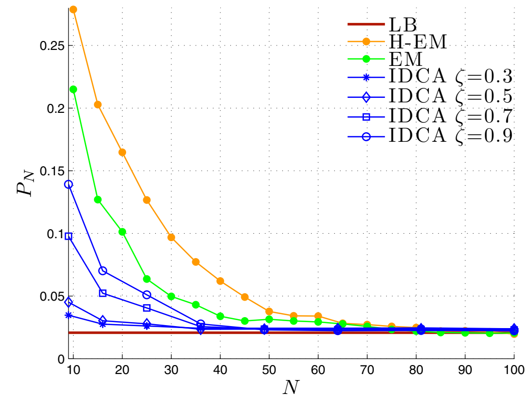

Fig. 3. 2d-grid.

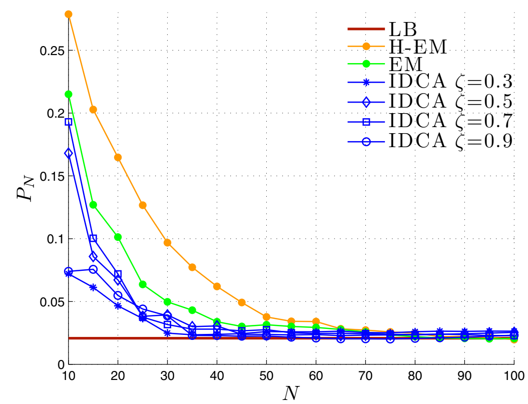

Fig. 4. Random geometric graph.

otherwise. Although the classification procedure requires a finite number of iterations, the choice of sequence $\{\gamma^{(t)}\}_{t\in\mathbb{N}}$ is critical, since it determines both the final accuracy and the convergence time of the continuous parameter θ. This raises the question of how to choose the sequence parameters $\{\gamma^{(t)}\}_{t\in\mathbb{N}}$ for a particular graph. Figs. 1-4 show the average (over 1000 Monte Carlo runs) of the relative classification error as a function of the number of nodes in the network. Results are obtained with $\gamma^{(t)} = 1/(t+1)^{\zeta}$ for different choices of $\zeta \in \{0.3, 0.5, 0.7, 0.9\}$. These figures put into evidence that IDCA has better performance than EM and Hard-EM, and the asymptotically behavior is very close to the lower bound predicted in Theorem 1.

V. Concluding Remarks

In this paper, the problem of distributively estimating hidden parameters in sensor networks with limited communication capability is studied. In particular, a distributed protocol (IDCA) is proposed. The main contribution includes the convergence of the algorithm to a local maximum of the ML-estimator. Different variants are possible, such as the generalization to multiple classes with unknown prior probabilities and to multidimensional estimation-classification problems. The choice of sequence $\{\gamma^{(t)}\}_{t\in\mathbb{N}}$ affects both convergence time and the final accuracy. The determination of a protocol for the adaptive search of sequence $\{\gamma^{(t)}\}_{t\in\mathbb{N}}$ and the study of the tightness of lower bound in Theorem 1 for both ML and IDCA are left for a future work.

VI. Acknowledgment

The authors wish to thank Sandro Zampieri and Luca Schenato for interesting discussions and suggestions.

References

[1] 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 (Accepted)*, 2011.

[2] D. M. Titterington, A. F. M. Smith, and U. E. Makov, *Statistical Analysis of Finite Mixture Distributions*. New York: John Wiley, 1985.

[3] R. Nowak, "Distributed EM algorithms for density estimation and clustering in sensor networks," *IEEE Transactions on Signal Processing*, vol. 51, no. 8, pp. 2245–2253, 2003.

[4] C. Moallemi and B. V. Roy, "Consensus propagation," *IEEE Transactions on Information Theory*, vol. 52, no. 11, pp. 4753 – 4766, 2006.

[5] V. Saligrama, M. Alanyali, and O. Savas, "Asynchronous distribution detection in sensor networks," *Preprint*, 2005.

[6] A. P. Dempster, N. M. Laird, and D. B. Rubin, "Maximum likelihood from incomplete data via the EM algorithm," *J. Roy. Statist. Soc. Ser. B*, vol. 39, no. 1, pp. 1–38, 1977.

[7] D. Gu, "Distributed EM algorithms for gaussian mixtures in sensor networks," *IEEE Transactions on Neural Networks*, vol. 19, no. 7, pp. 1154–1166, 2008.

[8] F. Fagnani, S. M. Fosson, and C. Ravazzi. (2011, August) Input driven consensus algorithm for distributed estimation and classification in sensor networks (extended version). [Online]. Available: http://calvino.polito.it/ ravazzi/research.html

[9] F. R. Gantmacher, *The theory of matrices*. New York: Chelsea publ., 1959.

APPENDIX
PROOF OF THEOREM 2

Our aim is to prove the convergence of (6) through the analysis of system (5). The proof is organized as follows.

1) First, we show that for sufficiently large t $\mu^{(t)}, \nu^{(t)}$, and $\widehat{\theta}^{(t)}$ are close to consensus vectors and we prove their convergence, assuming $\widehat{T}^{(t)}$ has already stabilized.
2) Second, we prove the stabilization of $\widehat{T}^{(t)}$ in finite time, by modeling (5)- (6) as a switching dynamical system.
3) Finally, combining these facts together we conclude the proof of Theorem 2.

Some proofs are omitted for reasons of space. The interested reader is referred to [8] for further details.

1) Towards consensus: Let $\Omega = I - \frac{1}{N}\mathbb{1}\mathbb{1}^\top$. Given $x \in \mathbb{R}^N$, let $\overline{x} = \frac{1}{N}\mathbb{1}^\top x$ be its average. Then $x = \overline{x}\mathbb{1} + \Omega x$. Given a bounded sequence $u^{(t)} \in \mathbb{R}^N$, consider the dynamics

$$x^{(t+1)} = \left(1 - \gamma^{(t)}\right) P x^{(t)} + \gamma^{(t)} u^{(t)} \quad t \in \mathbb{N} \qquad (7)$$

where, we recall the standing assumptions, $\gamma^{(t)} \in (0,1)$ and $\gamma^{(t)} \searrow 0$, and $P \in \mathbb{R}_+^{N\times N}$ is a doubly-stochastic, symmetric, primitive (irreducible and aperiodic) matrix. $x^{(0)}$ is any fixed vector.

Lemma 3:

$$\lim_{t\to+\infty} \Omega x^{(t)} = 0.$$

Proof: Let $z^{(t)} = \Omega x^{(t)}$. From (7) and the fact that $\Omega P = P\Omega$ we get

$$z^{(t+1)} = (1-\gamma^{(t)}) P z^{(t)} + \gamma^{(t)} \left(u^{(t)} - \overline{u}^{(t)}\mathbb{1}\right)$$

from which

$$||z^{(t+1)}||_2 \leq \prod_{s=0}^{t} \left(1-\gamma^{(s)}\right) ||z^{(0)}||_2 + \sum_{s=0}^{t} \gamma^{(s)} \left|\left|P^{t-s}u^{(s)} - \overline{u}^{(s)}\mathbb{1}\right|\right|_2 .$$

Perron-Frobenius Theorem [9] and boundness of $u^{(t)}$ imply that there exists $|\rho| < 1$ and $c \in \mathbb{R}$ such that $||P^{t-s}u^{(s)} - \overline{u}^{(s)}\mathbb{1}||_2 \leq c|\rho|^t$. We finally get

$$||z^{(t+1)}||_2 \leq \mathrm{e}^{-\sum_{s=0}^t \gamma^{(s)}} ||z^{(0)}||_2 + c\sum_{s=0}^t \gamma^{(t-s)} |\rho|^s.$$

Since $\sum_t \gamma^{(t)} = +\infty$ from hypothesis, the first term vanishes when $t \to +\infty$. The second term can be splitted as follows

$$\sum_{s=0}^{t} \gamma^{(t-s)}|\rho|^s \leq \sum_{s=0}^{\lfloor t/3 \rfloor} \gamma^{(t-s)}|\rho|^s + \sum_{s=\lfloor t/3\rfloor+1}^{t} \gamma^{(t-s)}|\rho|^s$$
$$\leq \frac{\gamma^{(\lfloor \frac{2}{3}t\rfloor)} + |\rho|^{\lfloor t/3\rfloor}}{1-|\rho|} \overset{t\to\infty}{\longrightarrow} 0.$$

■

We now go back to $\widehat{\theta}^{(t)}$.

Lemma 4:

$$\widehat{\theta}^{(t)} = \frac{\bar{\mu}^{(t)}}{\bar{\nu}^{(t)}}\mathbb{1} + \frac{1}{\bar{\nu}^{(t)}}\Omega\left(\mu^{(t)} - \frac{\bar{\mu}^{(t)}}{\bar{\nu}^{(t)}}\nu^{(t)}\right) + o\left(\gamma^{(t)}\right) \qquad (8)$$

Proof: For any $i \in \mathcal{V}$,

$$\frac{\mu_i^{(t)}}{\nu_i^{(t)}} - \frac{\bar{\mu}^{(t)}}{\bar{\nu}^{(t)}} = \frac{1}{\bar{\nu}^{(t)}}\left(\Omega\mu^{(t)}\right)_i - \frac{\mu_i^{(t)}}{\nu_i^{(t)}\bar{\nu}^{(t)}}\left(\Omega\nu^{(t)}\right)_i .$$

It follows from Lemma 3 that $\mu^{(t)} = \bar{\mu}^{(t)}\mathbb{1} + o(1)$ and $\nu^{(t)} = \bar{\nu}^{(t)}\mathbb{1} + o(1)$ for $t \to +\infty$. This yields

$$\frac{\mu_i^{(t)}}{\nu_i^{(t)}\bar{\nu}^{(t)}}\left(\Omega\nu^{(t)}\right)_i = \frac{\bar{\mu}^{(t)}}{\bar{\nu}^{(t)}}\frac{\left(\Omega\nu^{(t)}\right)_i}{\bar{\nu}^{(t)} + o\left(1\right)} + o\left(\gamma^{(t)}\right).$$

Finally, since $\bar{\nu}_{(t)}$ is positive and bounded away from 0, we can write $\left[\bar{\nu}^{(t)} + o\left(1\right)\right]^{-1} = 1/\bar{\nu}^{(t)}\left(1 + o\left(1\right)\right)$ from which the thesis follows. ■

The following corollary is a straightforward consequence of Lemmas 3 and 4.

Corollary 5:

$$\lim_{t\to+\infty} \Omega\widehat{\theta}^{(t)} = 0.$$

Corol. 5 says that the estimate $\widehat{\theta}^{(t)}$ is close to a consensus for sufficiently large t. The following proposition guarantees the convergence under the assumption of input stabilization.

Proposition 6: If $\exists\ t_0 \in \mathbb{N}$ s.t. $u^{(t)} = u\ \forall t \geq t_0$ then

$$\lim_{t\to+\infty} x^{(t)} = \overline{u}\mathbb{1}.$$

Proof: The vector $x^{(t)}$ can be written as $x^{(t)} = \overline{x}^{(t)}\mathbb{1} + \Omega x^{(t)}$. From Lemma 3 it is sufficient to prove that $\lim_{t\to+\infty} \overline{x}^{(t)}\mathbb{1} = \overline{u}\mathbb{1}$. Since P is a doubly-stochastic matrix, we have

$$\overline{x}^{(t+1)} - \overline{u} = \textstyle\prod_{t=t_0}^{t}(1-\gamma^{(t)})(\overline{x}^{(t)} - \overline{u})$$

which goes to zero from the non-summability of $\gamma^{(t)}$. ■

If $\widehat{T}^{(t)}$ stabilizes at finite time in $\widehat{T}^{(\infty)}$, Prop. 6 guarantees that $\mu^{(t)}$ and $\nu^{(t)}$ converge to $\frac{1}{N}\sum_{i\in\mathcal{V}} y_i[\widehat{T}_i^{(\infty)}]^{-2}\mathbb{1}$ and $\frac{1}{N}\sum_{i\in\mathcal{V}}[\widehat{T}_i^{(\infty)}]^{-2}\mathbb{1}$, respectively.

Corollary 7: If $\exists\ t_0 \in \mathbb{N}$ s.t. $\widehat{T}^{(t)} = \widehat{T}^{(\infty)}\ \forall t \geq t_0$ then

$$\lim_{t\to+\infty} \widehat{\theta}^{(t)} = \widehat{\theta}^{(\infty)} = \frac{\sum_{k=1}^N y_k\left[\widehat{T}_k^{(\infty)}\right]^{-2}}{\sum_{k=1}^N \left[\widehat{T}_k^{(\infty)}\right]^{-2}}\mathbb{1}.$$

2) Stabilization (a.s) of $\widehat{T}^{(t)}$: Since $\widehat{T}^{(t)}$ can only assume values in a finite set, equations in (5) and (6) can be conveniently modeled by a switching system. Indeed, given observations y and $\omega \in \{\alpha,\beta\}^N$, define

$$\Theta_\omega := \{x \in \mathbb{R}^N : |x_i - y_i| < \delta, \text{if } \omega_i = \alpha, |x_i - y_i| \geq \delta, \text{if } \omega_i = \beta\}.$$

When $\widehat{\theta}^{(t)} \in \Theta_\omega$ we, abstractly, rewrite (5a) and (5b) as

$$\mu^{(t+1)} = f_\omega(t,\mu^{(t)}) \qquad \nu^{(t+1)} = g_\omega(t,\nu^{(t)}).$$

Together with $\widehat{\theta}_i^{(t)} = \mu_i^{(t)}/\nu_i^{(t)}$, it describes a closed-loop switching system, the switching policy being determined by $\widehat{\theta}^{(t)}$. It is clear that the stabilization of $\widehat{T}^{(t)}$ is equivalent to the fact that there exist an $\omega^\star \in \{\alpha,\beta\}^N$ and a time t_0 such that $\widehat{\theta}^{(t)} \in \Theta_{\omega^\star}$ for all $t \geq t_0$. From Corol. 7 candidate limit points for $\widehat{\theta}^{(t)}$ are

$$\bar{y}_\omega = \frac{\sum_{i\in\mathcal{V}} y_i\omega_i^{-2}}{\sum_{i\in\mathcal{V}}\omega_i^{-2}}\mathbb{1} \qquad \omega \in \{\alpha,\beta\}^N.$$

From Corol. 5, the dynamics can be analyzed by studying it in a neighborhood of the line $\Lambda = \{\lambda\mathbb{1} | \lambda \in \mathbb{R}\}$. Notice that the line Λ crosses facets of regions almost surely. Given $\epsilon > 0$, define the following sets

$$\Theta^\epsilon := \{x \in \mathbb{R}^N : ||\Omega x||_2 < \epsilon\}, \quad \Theta^\epsilon_\omega := \Theta^\epsilon \cap \Theta_\omega$$
$$\Gamma := \{\omega \in \{\alpha, \beta\}^N : \ \Theta_\omega \cap \Lambda \neq \emptyset\}.$$

In the sequel, we will use the natural ordering on Γ. Two elements $\omega,\ \omega' \in \Gamma$ are called consequent if

C1) $\omega_i = \omega'_i$ for all $i \neq i_0$ and $\omega_{i_0} \neq \omega'_{i_0}$;
C2) $\Theta_\omega \cap \Lambda < \Theta_{\omega'} \cap \Lambda$.

Given two consequent $\omega, \omega' \in \Gamma$, consider the following subsets of $\mathbb{R}^N$:

$$\mathcal{M}^\epsilon_\omega := \{x \in \Theta^\epsilon_\omega : \overline{x}\mathbb{1} + \Omega z \in \Theta^\epsilon_\omega,\ \forall z : ||z||_2 < \epsilon\}$$
$$\mathcal{L}^\epsilon_{\omega,\omega'} := \{x \in \Theta^\epsilon : \mathcal{M}^\epsilon_\omega \cap \Lambda < \bar{x} < \mathcal{M}^\epsilon_{\omega'} \cap \Lambda\}.$$

We clearly have $\Theta^\epsilon = \bigcup_{\omega,\omega' \in \Gamma} \mathcal{M}^\epsilon_\omega \cup \mathcal{L}^\epsilon_{\omega,\omega'}$ (see Fig. 5).

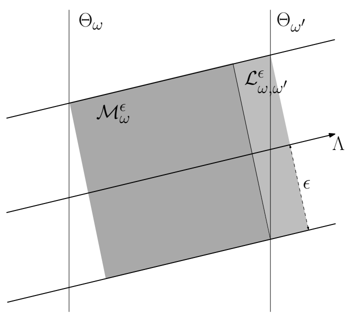

Fig. 5. Given the couple (ω, ω') the sets $\mathcal{L}^\epsilon_{\omega,\omega'}$ and $\mathcal{M}^\epsilon_\omega$ are visualized.

Notice that we can always choose $\epsilon_0 > 0$ such that

$$\bar{y}_\omega \mathbb{1}, y_i \mathbb{1} \in \bigcup_{\omega' \in \Gamma} \mathcal{M}^{\epsilon_0}_{\omega'} \qquad \forall \omega \in \Gamma, \forall i \in \mathcal{V}$$

and, moreover, there exists $\bar{c} > 0$ such that

$$d\left(\bigcup_{\omega' \in \Gamma} \partial\left(\mathcal{M}^\epsilon_{\omega'} \cap \Lambda\right), \{\bar{y}_\omega, y_i\}\right) \geq \bar{c}, \quad \forall \epsilon \leq \epsilon_0 \qquad (9)$$

where $\partial(s)$ indicates the endpoints of segment s and d is the Euclidean distance between sets.

Fix now $\epsilon \leq \epsilon_0$ and choose t_ϵ such that $\widehat{\theta}^{(t)} \in \Theta^\epsilon$ for all $t \geq t_\epsilon$ (it exists by Corol. 5). From now on we consider times $t \geq t_\epsilon$.

Lemma 8: If $\widehat{\theta}^{(t)} \in \Theta_\omega$ then there exist constants $0 < c_1 < c^{(t)} < c_2$ such that

$$\frac{\overline{\mu}^{(t+1)}}{\overline{\nu}^{(t+1)}} - \frac{\overline{\mu}^{(t)}}{\overline{\nu}^{(t)}} = c^{(t)}\gamma^{(t)}\left(\bar{y}_\omega - \frac{\overline{\mu}^{(t)}}{\overline{\nu}^{(t)}}\right)$$

Proposition 9: If $\bar{y}_\omega \mathbb{1} \in \Theta_\omega$, there exists $t_0 \geq t_\epsilon$ such that

$$\widehat{\theta}^{(t)} \in \mathcal{M}^\epsilon_\omega \ \Rightarrow\ \widehat{\theta}^{(t+1)} \in \mathcal{M}^\epsilon_\omega \quad \forall t \geq t_0\,.$$

Proof: A straightforward application of Lemmas 4 and 8 allows to write

$$\overline{\widehat{\theta}}^{(t+1)} = \overline{\widehat{\theta}}^{(t)} + c^{(t)}\gamma^{(t)}\left(\bar{y}_\omega - \overline{\widehat{\theta}}^{(t)}\right) + r \qquad (10)$$

where $r = o(\gamma^{(t)})$. If $\widehat{\theta}^{(t)} \in \mathcal{M}^\epsilon_\omega$, we have, by convexity, that

$$z := \overline{\widehat{\theta}}^{(t)} + c^{(t)}\gamma^{(t)}\left(\bar{y}_\omega - \overline{\widehat{\theta}}^{(t)}\right) \in \mathcal{M}^\epsilon_\omega.$$

Moreover, because of (9) and the fact that $c^{(t)}$ is bounded away from 0, there exists $c' > 0$ such that $d(z, \partial(\mathcal{M}^\epsilon_\omega \cap \Lambda)) \geq c'\gamma^{(t)}$. Proof is then completed by selecting t_0 such that $|r| < c'\gamma^{(t)}/2$. ■

Our next goal is to prove that if $\overline{y}_\omega \mathbb{1} \notin \mathcal{M}^\epsilon_\omega$, then, at a certain time t, $\widehat{\theta}^{(t)}$ will definitely be outside $\mathcal{M}^\epsilon_\omega$. A technical lemma based on convexity arguments is required [8].

Lemma 10: Suppose $\omega, \omega' \in \Gamma$ are consequent. Then

$$\bar{y}_\omega > \Theta_\omega \cap \Lambda \quad \Rightarrow \quad \bar{y}_{\omega'} > \Theta_\omega \cap \Lambda$$
$$\bar{y}_{\omega'} < \Theta_{\omega'} \cap \Lambda \quad \Rightarrow \quad \bar{y}_\omega < \Theta_{\omega'} \cap \Lambda.$$

Proposition 11: If $\overline{y}_\omega \mathbb{1} \notin \Theta_\omega$, then there exists $t_0 \in \mathbb{N}$ such that $\widehat{\theta}^{(t)} \notin \Theta^\epsilon_\omega \ \forall t > t_0$.

Proof: Suppose $\overline{y}_\omega > \Theta_\omega \cap \Lambda$ (the case when is $<$ can be treated analogously). Let $\omega' \in \Gamma$ such that ω, ω' are consequent. Lemma 10 implies that $\overline{y}_{\omega'} > \Theta_\omega \cap \Lambda$. Put

$$A := \{x \in \Theta^\epsilon_\omega \cup \Theta^\epsilon_{\omega'} \mid \overline{x} \leq \alpha := \min\{\overline{y}_\omega, \overline{y}_{\omega'}\} - \overline{c}/2\}.$$

Choose t_1 in such a way that $c_2(\max\{y_i\} - \min\{y_i\})\gamma^{(t)} + r < \overline{c}/2$ and $|r| < c_1\overline{c}\gamma^{(t)}/4$ for all $t \geq t_1$. It follows from (10) that, if for some $t \geq t_1$ $\widehat{\theta}^{(t)} \in A$, then,

$$\overline{\widehat{\theta}}^{(t+1)} \geq \overline{\widehat{\theta}}^{(t)} + c_1\overline{c}\ \gamma^{(t)}/4.$$

Owing to the non-summability of $\gamma^{(t)}$ it follows that if $\widehat{\theta}^{(t)}$ enters in Θ^ϵ_ω for some $t \geq t_1$, then, in finite time it will enter into $A \setminus \Theta^\epsilon_\omega$ and then it will finally exit A. In particular there must exist $t_2 \geq t_1$ such that $\overline{\widehat{\theta}}^{(t_2)} > \alpha$. Now, by the way t_1 has been chosen, if $\overline{\widehat{\theta}}^{(t)} > \alpha$ for $t \geq t_1$, necessarily $\overline{\widehat{\theta}}^{(t+1)} > \min\{\overline{y}_\omega, \overline{y}_{\omega'}\} - \overline{c}$. This implies that after time t_2, $\widehat{\theta}^{(t)}$ will never enter Θ^ϵ_ω again. ■

Vector $\widehat{\theta}^{(t)}$ in principle might belong definitively to a boundary region $\mathcal{L}^\epsilon_{\omega,\omega'}$, keeping on switching from Θ_ω and $\Theta_{\omega'}$. This is not the case and the following proposition holds [8].

Proposition 12: If $P \in \mathbb{R}^{N\times N}_+$ and admits positive eigenvalues, then there exists $s \in \mathbb{N}$ such that $\widehat{\theta}^{(s)} \notin \bigcup_{\omega,\omega' \in \{\alpha,\beta\}^N} \mathcal{L}_{\omega,\omega'}$ for all $t > s$.

3) Proof of Theorem 1: From Prop. 11 we have that if $\overline{y}_\omega \mathbb{1} \notin \Theta_\omega$, then $\widehat{\theta}^{(t)}$ will be definitely outside Θ_ω. Recalling that the (finite) union of the regions Θ_ω, $\omega \in \{\alpha, \beta\}^N$, is a partition of $\mathbb{R}^N$, $\widehat{\theta}^{(t)}$ definitely must belong to a region Θ_ω containing $\overline{y}_\omega \mathbb{1}$ and Prop. 9 and 12 guarantees that $\widehat{T}^{(t)}$ stabilizes at ω. Finally, from Corol. 7 we get that $\widehat{\theta}^{(\infty)} = \lim_{t\to+\infty} \widehat{\theta}^{(t)} = \overline{y}_\omega$. This proves point 1).

For the second part, let $\epsilon > 0$ and θ be such that $||\widehat{\theta}^{(\infty)} - \theta|| < \epsilon$ and $\widehat{T}(\theta)$ defined in (4). Then we have

$$L(\theta, T) \leq \max_{T \in \{\alpha,\beta\}^N} L(\theta, T) = L(\theta, \widehat{T}(\theta)).$$

If ϵ is sufficiently small then $L(\theta, \widehat{T}(\theta)) = L(\theta, \widehat{T}^{(\infty)})$ and

$$L(\theta, T) \leq L(\theta, \widehat{T}^{(\infty)}) \leq \max_{\theta \in \mathbb{R}} L(\theta, \widehat{T}^{(\infty)}) = L(\widehat{\theta}^{(\infty)}, \widehat{T}^{(\infty)})$$

where the last equality follows from (3).