View metadata, citation and similar papers at core.ac.uk

brought to you by CORE

provided by INRIA a CCSD electronic archive server

Distributed estimation from relative measurements of heterogeneous and uncertain quality

Chiara Ravazzi, Nelson Chan, Paolo Frasca

To cite this version:

Chiara Ravazzi, Nelson Chan, Paolo Frasca. Distributed estimation from relative measurements of heterogeneous and uncertain quality. IEEE transactions on Signal and Information Processing over Networks, IEEE, 2019, 5 (2), pp.203-217. 10.1109/TSIPN.2018.2869117 . hal-01877623

HAL Id: hal-01877623

https://hal.archives-ouvertes.fr/hal-01877623

Submitted on 20 Sep 2018

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Distributed estimation from relative measurements of heterogeneous and uncertain quality

Chiara Ravazzi, *Member* Nelson P. K. Chan, *Student Member* Paolo Frasca, *Senior Member, IEEE*

***Abstract*— This paper studies the problem of estimation from relative measurements in a graph, in which a vector indexed over the nodes has to be reconstructed from pairwise measurements of differences between its components associated to nodes connected by an edge. In order to model heterogeneity and uncertainty of the measurements, we assume them to be affected by additive noise distributed according to a Gaussian mixture. In this original setup, we formulate the problem of computing the Maximum-Likelihood (ML) estimates and we design two novel algorithms, based on Least Squares regression and Expectation-Maximization (EM). The first algorithm (LS-EM) is centralized and performs the estimation from relative measurements, the soft classification of the measurements, and the estimation of the noise parameters. The second algorithm (Distributed LS-EM) is distributed and performs estimation and soft classification of the measurements, but requires the knowledge of the noise parameters. We provide rigorous proofs of convergence for both algorithms and we present numerical experiments to evaluate their performance and compare it with solutions from the literature. The experiments show the robustness of the proposed methods against different kinds of noise and, for the Distributed LS-EM, against errors in the knowledge of noise parameters.**

I. Introduction

Whenever measurements are used to estimate a quantity of interest, measurement errors must be properly taken into account and the statistical properties of these errors should be identified to enable an efficient estimation. In this paper, we look at this broad issue in a specific problem within the context of network systems. Namely, we consider the problem of distributed estimation from relative measurements, defined as follows. We assume to have a real vector that is indexed over the nodes of a graph with a known topology: the nodes are allowed to take pairwise measurements of the differences between their vector entries and those of their neighbors in the graph. The estimation problem consists in reconstructing the original vector, up to an additive constant. This prototypical problem can be applied in a variety of contexts [1]. One example is relative localization of mobile automated vehicles, where the vehicles have to locate themselves by using only distance measurements [2]. Another example is statistical ranking, where a set of items needs to be sorted according to their quality, which can only be evaluated comparatively [3], [4]. In all these scenarios, the noise affecting the measurements can be drastically heterogeneous and, more importantly, its distribution may not be known a priori. For instance, in vehicle localization, distances between vehicles may be measured by more or less accurate sensors; in a ranking system, the items upon evaluation can be compared by more or less trustworthy entities. It is thus important to identify unreliable measurements and weight them differently in the estimation. In order to model this uncertainty, in this paper we assume that the measurement noise is sampled from a mixture of two Gaussian distributions with different variances, representing good and poor measurements, respectively. Our solution to this problem builds on the classical Expectation-Maximization (EM) approach [5], [6], where the likelihood is maximized by alternating operations of expectation and maximization.

We are particularly interested in finding efficient *distributed* algorithms to solve this problem. More precisely, we say that an algorithm is distributed if it requires each node to use information that is directly available at the node itself or from its immediate neighbors. Actually, many distributed algorithms for relative estimation are available [7], [1], [8], [9], [10], [11], [12], [13], [14], but they assume that the quality of the measurements is known beforehand. At the same time, there is a large literature on robust estimation that also covers estimation from relative measurements, but often provides algorithms that are not distributed; see for instance [15] and references therein. Perhaps the only work on robust distributed relative estimation is the recent [16]: their approach is very different from ours as it is based on ℓ_1 optimization. By proposing our distributed EM algorithm, we contribute to the growing body of research on distributed algorithms for network-related estimation problems with heterogeneous and unknown measurements [17], [18], [19], [16], [20], where distributed algorithms based on consensus and ranking procedures have been proposed to approximate Maximum-Likelihood (ML) estimates. Other authors have used EM to estimate Gaussian mixtures' parameters in other problems of distributed inference in sensor networks [21], [22], [23]. In these works, a network is given and each node independently performs the E-step from local observations and this information is suitably propagated to collaboratively perform the M-step. EM is also a key instrument to design

Chiara Ravazzi is with the National Research Council (CNR), Institute of Electronics, Computers and Telecommunication Engineering (IEIIT), c/o Politecnico di Torino, Italy. E-mail: chiara.ravazzi@ieiit.cnr.it.

Nelson P. K. Chan is with the University of Groningen, Groningen, the Netherlands. E-mail: n.p.k.chan@rug.nl.

Paolo Frasca is with Univ. Grenoble Alpes, CNRS, Inria, Grenoble INP, GIPSA-lab, F-38000 Grenoble, France and Research Associate at the IEIIT-CNR, Torino, Italy. E-mail: paolo.frasca@gipsa-lab.fr.

The research leading to this paper has been partly developed while C. Ravazzi was at Department of Electronics and Telecommunications (DET), Politecnico di Torino, Italy, and N. P. K. Chan and P. Frasca were with the Department of Applied Mathematics, University of Twente, Enschede, the Netherlands. The authors' research has been partially supported by the International Bilateral Joint CNR Lab COOPS and by IDEX Université Grenoble Alpes under C2S2 "Strategic Research Initiative" grant.

reliable learning systems based on unreliable information reported by users in the context of social sensing [24], [25].

Our contribution: In this paper, we define the problem of robust estimation from relative measurements when measurement noise is drawn from a Gaussian mixture and we design two iterative algorithms that solve it. Both algorithms are based on combining classical Weighted Least Squares (WLS) with Expectation-Maximization (EM), which is a popular tool in statistical estimation problems involving incomplete data [26], [27]. The first algorithm is centralized, whereas the second algorithm is distributed but requires to know the two variances of the Gaussian mixture. This knowledge is not necessary for the centralized version. Both algorithms are proved to converge and their performance is compared on synthetic data. We observe that the centralized algorithm has better performance, achieving smaller estimation errors. The centralized algorithm also requires less iterations to converge, but each iteration involves more computations. Both algorithms are fairly robust to uncertainties in the parameters.

Organization of the paper: We formally present the problem of relative estimation in Section II, where we also review some state-of-the-art algorithms. The centralised LS-EM algorithm is described in Section III and the Distributed LS-EM algorithm in Section IV. Section V contains some numerical examples and Section VI our conclusions. The details of our proofs are postponed to the Appendix.

Notation: Throughout this paper, we use the following notational conventions. Real and nonnegative integer numbers are denoted by $\mathbb{R}$ and $\mathbb{Z}_{\geq 0}$, respectively. Open intervals are denoted by parentheses and closed intervals by square brackets. Given a finite set $\mathcal{V}$, the Eucliean space of real vectors with components labelled by elements of $\mathcal{V}$ is denoted by $\mathbb{R}^{\mathcal{V}}$. We denote column vectors with small letters, and matrices with capital letters. Given $x \in \mathbb{R}^{\mathcal{V}}$, we denote its v-th element by x_v or $(x)_v$. Given $x \in \mathbb{R}^{\mathcal{V}}$ and $A \in \mathbb{R}^{\mathcal{V}\times\mathcal{V}}$, we denote the ℓ_p norm of vector x with the symbol $\|x\|_p$ (the ℓ_2 norm is taken when subscript p is omitted), and with $\|A\|$ the spectral norm of matrix A. The support set of a vector x is defined by $\mathrm{supp}(x) = \{i \in \mathcal{V} : x_i \neq 0\}$ and we define $\Sigma_k = \{x \in \mathbb{R}^{\mathcal{V}} : \|x\|_0 \leq k\}$ with $\|x\|_0 = |\mathrm{supp}(x)|$ denoting the ℓ_0-pseudonorm. Given $\mathcal{E}$ with finite cardinality $|\mathcal{E}|$, we define $\mathcal{P}_\ell : [0,1]^{\mathcal{E}} \to \Sigma_{|\mathcal{E}|-\ell}$ as the projection that zeroes the ℓ smallest components of the given vector. It should be noticed that in general the projection of a vector could be not unique: we assume that $\mathcal{P}_\ell(x)$ consistently selects one of the possible projections by a tie-breaking rule. Given matrix M, $M^\top$ denotes its transpose. Given vector x, we denote by $\mathrm{diag}(x)$ the diagonal matrix whose diagonal entries are the elements of x.

An (undirected) graph is a pair $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ where $\mathcal{V}$ is a finite set of vertices and $\mathcal{E} \subseteq \{\{v, w\} : v, w \in \mathcal{V}\}$ is the set of edges. Graph $\mathcal{G}$ is connected if, for all $i, j \in \mathcal{V}$, there exist vertices $i_1, \dots i_s$ such that $\{i, i_1\}, \{i_1, i_2\}, \dots, \{i_s, j\} \in \mathcal{E}$. We let $A \in \{0, \pm 1\}^{\mathcal{E}\times\mathcal{V}}$ be the edge incidence matrix of the graph $\mathcal{G}$, defined as follows. The rows and the columns of A are indexed by elements of $\mathcal{E}$ and $\mathcal{V}$, respectively. We assume to have an order on set $\mathcal{V}$, such as it would be for $\mathcal{V} = \{1, \dots, n\}$. By this order, the orientation of the edges is conventionally assumed to be such that, if $u < v$, then edge $\{v, u\}$ originates in u and terminates in v. The (e, w)-entry of A is 0 if vertex w and edge e are not incident, and otherwise it is 1 or -1 according as e originates or terminates at w:

$$A_{ew} = \begin{cases} +1 & \text{if } e = (v, w) \\ -1 & \text{if } e = (w, v) \\ 0 & \text{otherwise.} \end{cases}$$

II. Robust estimation from relative measurements

A set of nodes $\mathcal{V} = \{1, \dots, N\}$ is considered, each of them endowed with an unknown scalar quantity $\widetilde{x}_v \in \mathbb{R}$ with $v \in \mathcal{V}$. Starting from a set of noisy measurements, the nodes' goal is to estimate their own absolute position. More precisely, each node $u \in \mathcal{V}$ is interested in estimating the scalar value $\widetilde{x}_u$, based on noisy measurements of differences $\widetilde{x}_u - \widetilde{x}_v$ with v and u in $\mathcal{V}$. The set of available measurements can be conveniently represented by graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$, where each edge represents a measurement: $A \in \{0, \pm 1\}^{\mathcal{E}\times\mathcal{V}}$ is the edge incidence matrix of graph $\mathcal{G}$. We let $b \in \mathbb{R}^{\mathcal{E}}$ be the vector collecting the measurements

$$b = A\widetilde{x} + \eta,$$

where $\eta_e, e \in \mathcal{E}$ are mutually independent random variables distributed according to a Gaussian distribution $\mathcal{N}(0, \sigma_e^2)$, having

$$\sigma_e = (1 - z_e)\alpha + z_e\beta,$$

with $0 < \alpha < \beta$, with z_e distributed as a Bernoulli distribution $z_e \sim \mathcal{B}(p)$ and $p \in (0, 1/2)$. Provided $\alpha \ll \beta$, the value $z_e = 1$ is associated to a measurement that is unreliable. With this formulation the random variables $\{\eta_e\}_{e\in\mathcal{E}}$ are Gaussian mixtures, whose model is completely described by three parameters: p, α and β. For convenience, from now on we consider p fixed and known. This choice is done for simplicity and does not entail a significant restriction to our analysis: on the one hand, the algorithms we propose are fairly robust to small errors in the estimate of p; on the other hand, our framework can be easily extended to include the estimation of p as an unknown parameter.

Our main goal is to obtain a robust estimate of the state vector $\widetilde{x}$ by suitably taking into account the different quality of the measurements. We thus consider a joint Maximum Likelihood estimation

$$\widehat{x}^{\mathrm{ML}} = \underset{x\in\mathbb{R}^{\mathcal{V}}, \alpha>0, \beta>0}{\mathrm{argmax}} \; L(x, \alpha, \beta) \tag{1}$$

where $L(x, \alpha, \beta) := \log f(b|x, \alpha, \beta)$ and

$$f(b|x, \alpha, \beta) = \prod_{e\in\mathcal{E}} \left[\frac{1-p}{\sqrt{2\pi\alpha^2}} \exp\left(-\frac{(b - Ax)_e^2}{2\alpha^2}\right) + \frac{p}{\sqrt{2\pi\beta^2}} \exp\left(-\frac{(b - Ax)_e^2}{2\beta^2}\right) \right]. \tag{2}$$

The computational complexity of optimization problem (1) makes a brute force approach infeasible for large graphs.

A. Estimation via Weighted Least Squares

Problem (1) becomes much simpler if we assume to know the distribution that has produced the noise term for each measurement. Using the noise source information α, β, and $\widetilde{z}_e$ for all $e \in \mathcal{E}$, where $\widetilde{z}_e$ is the realization of z_e, the ML-estimation becomes

$$X_{\mathrm{ML}} = \operatorname*{argmax}_{x \in \mathbb{R}^{\mathcal{V}}} \log f(b|x, \widetilde{z}, \alpha, \beta) \tag{3}$$

where X_{ML} is the set of maximizing values of the log-likelihood

$$f(b|x, \widetilde{z}, \alpha, \beta) = \prod_{e \in \mathcal{E}} \left[\frac{1 - \widetilde{z}_e}{\sqrt{2\pi\alpha^2}} \exp\left(-\frac{(b - Ax)_e^2}{2\alpha^2} \right) + \frac{\widetilde{z}_e}{\sqrt{2\pi\beta^2}} \exp\left(-\frac{(b - Ax)_e^2}{2\beta^2} \right) \right].$$

Noticing that $\log(\prod_e x_e) = \sum_e \log(x_e)$ and that $\widetilde{z}$ is a binary vector, it is easy to see that ML is equivalent to solving the Weighted Least Square (WLS) problem

$$\operatorname*{argmin}_{x \in \mathbb{R}^{\mathcal{V}}} \frac{1}{2} \|b - Ax\|_W^2 = \operatorname*{argmin}_{x \in \mathbb{R}^{\mathcal{V}}} \frac{1}{2} (b - Ax)^\top W (b - Ax) \tag{4}$$

with $W = \mathrm{diag}((1 - \widetilde{z}_e)\, \alpha^{-2} + \widetilde{z}_e \beta^{-2})$.

The following lemma describes the solutions of (4).

Lemma 1 (WLS estimator): Let the graph $\mathcal{G}$ be connected and X be the set of solutions of (4), and let $L_W := A^\top W A$ denote the weighted Laplacian of the graph. The following facts hold:

1) $x \in X_{\mathrm{ML}}$ if and only if $A^\top W A x = A^\top W b$;
2) there exists a unique $\widehat{x}^{\mathrm{wls}} \in X_{\mathrm{ML}}$ such that $\|\widehat{x}^{\mathrm{wls}}\|_2 = \min_{x \in X_{\mathrm{ML}}} \|x\|_2$;
3) $$\widehat{x}^{\mathrm{wls}} = L_W^\dagger A^\top W b, \tag{5}$$
   where $L_W^\dagger$ denotes the Moore-Penrose pseudo-inverse of the weighted Laplacian L_W.

We recall that $\mathbf{1}^\top L_W = 0$ and $\mathbf{1}^\top L_W^\dagger = 0$.

Further useful properties are collected in the following result [28, Sect. 5.4].

Proposition 2 (Moments of WLS estimator): Provided $\mathcal{G}$ is connected, it holds that

$$\mathbb{E}[\widehat{x}^{\mathrm{wls}}] = \big(I - \frac{1}{N} \mathbb{1}\mathbb{1}^\top\big) \widetilde{x}$$
$$\mathbb{E}[(\widehat{x}^{\mathrm{wls}} - \mathbb{E}[\widehat{x}^{\mathrm{wls}}])(\widehat{x}^{\mathrm{wls}} - \mathbb{E}[\widehat{x}^{\mathrm{wls}}])^\top] = L_W^\dagger,$$

where $\mathbb{1}$ is a vector of length N whose entries are all 1.

It should be stressed that determining the state vector $\widetilde{x}$ from relative measurements is only possible up to an additive constant, being $A\mathbb{1} = 0$, and $X_{\mathrm{ML}} = \widehat{x}^{\mathrm{wls}} + \mathrm{span}(\mathbb{1})$. This ambiguity can be avoided by assuming the centroid of the nodes as the origin of the Cartesian coordinate system. n view of this comment and of the results above, we shall assume from now on that $\mathcal{G}$ is *connected* and $\mathbb{1}^\top \widetilde{x} = 0$.

As shown in Lemma 1, the WLS solution is explicitly known and can be easily computed solving a linear system. Furthermore, the following distributed computation is also possible, by using a gradient descent algorithm. Observe that the gradient of the cost function in (4) is given by $L_W x - A^\top W b$. Set an initial condition $x^{(0)} = 0$ and fix $\tau > 0$ and consider

$$x^{(t+1)} = (I - \tau L_W) x^{(t)} + \tau A^\top W b. \tag{6}$$

Provided $\tau < 2\|L_W\|_2^{-1}$, the gradient descent algorithm (6) converges to the WLS solution [29], [9].

B. Relations with literature and numerical example

The WLS estimation and the subsequent developments in this paper share some ideas with several approaches in literature. We recall two methods based on optimization that focus on *Sparse outliers detection* [15] and *Least absolute estimation* [30]. Then we will summarize the main advantages of WLS in the considered setting in contrast with these methods.

The problem of finding the smallest set that contains the outliers is considered in [15]. Using the same rationale of the *big M trick approach*, introduced in [15, Section III.C], and recalling that $|y_e - (A\widetilde{x})_e| \leq 3\sigma_e$ with a probability close to 1 (about 0.997), a reasonable adaptation of [15] can be formalized as an optimization problem in the ℓ_0-pseudonorm:

$$\begin{aligned} &\min_{x \in \mathbb{R}^{\mathcal{V}},\, z \in \{0,1\}^{\mathcal{E}}:\ \mathbb{1}^\top x = 0} \|z\|_0 \\ &\quad \text{s.t. } |y_e - (Ax)_e| \leq 3\alpha + 3z_e(\beta - \alpha) \quad \forall e \in \mathcal{E}. \end{aligned} \tag{7}$$

The decision variables $z_e \in \mathcal{E}$ play the role of indicator variables for each measurement $e \in \mathcal{E}$. The test to label the measurements is based on a confidence interval: if z_e is 0 then the measurement is trusted, if z_e is 1 then the measurement is not trusted. This problem is combinatorial and becomes intractable for large scale problems.

For this reason, a standard approach is resorting to least absolute estimation [30], also known as ℓ_1-minimization. Problem (7) is relaxed by replacing the ℓ_0-pseudonorm with the ℓ_1-norm, which is expected to promote sparsity [31]:

$$\begin{aligned} &\min_{x \in \mathbb{R}^{\mathcal{V}},\, z \in \mathbb{R}^{\mathcal{E}}:\ \mathbb{1}^\top x = 0} \|z\|_1 \\ &\quad \text{s.t. } |y_e - (Ax)_e| \leq 3\alpha + 3z_e(\beta - \alpha) \quad \forall e \in \mathcal{E} \end{aligned} \tag{8}$$

or

$$\min_{x \in \mathbb{R}^{\mathcal{V}}:\ \mathbb{1}^\top x = 0} \|y - Ax\|_1 \tag{9}$$

The problem in (9) has also a probabilistic characterization and can be interpreted as ML estimation assuming that the noise is distributed according to a Laplace distribution. The ℓ_1-norm is less sensitive to outliers [30] and performs better than LS-estimator in presence of different types of corrupted measurements. It should be noticed that the problem in (9) is not smooth but is still convex, indeed it is a linear program (LP) and can be solved efficiently by iterative algorithms, e.g. using subgradient methods [32] or iterative reweighted least squares (IRLS, [33]) that admit a distributed implementation. Observe that the subgradient of the cost function in (9) is

given by $A^\top \mathrm{sgn}(y - Ax)$. Set an initial condition $x^{(0)} = 0$ and fix $\tau > 0$ and consider

$$x^{(t+1)} = x^{(t)} + \tau A^\top \mathrm{sgn}(y - Ax). \tag{10}$$

Despite these interesting features, LAE has some drawbacks. First, there are no guarantees that the solution of (9) has the minimum cardinality property. Moreover, there are no theoretical conditions under which the problem in (7) is equivalent to (9). Extensive numerical results show that ℓ_1-norm encourage sparsity but in general the solution of (7) and (9) do not coincide [15]. Using the noise source information α, β, and $\widetilde{z}_e$ for all $e \in \mathcal{E}$, problems (7) and (8) reduce to a linear feasibility program. If $\widetilde{z}_e$ is 0, then the measurement is selected, and if z_e is 1, then the measurement is detected as outlier and not taken into account in the search of x satisfying the constraints. WLS instead uses all the measurements in the estimation by mitigating the effect of outliers: its covariance is given in Proposition 2. Furthermore, finding the optimal estimate using WLS approach is equivalent to solving a network of resistors [34]. This intuitive electrical interpretation highlights the role of the topology of the measurement graph and allows distinguishing between topologies that lead to small or large estimation errors [35], [9]. In particular, using Proposition 2 and the electrical interpretation, one can relate the measurement graph $\mathcal{G}$ to the error in the estimation. Such analysis of performance is not available for ℓ_0 or ℓ_1-minimization.

Finally, we provide a numerical example for illustration.

Example 1: Consider the connected network in Figure 1 with $N = 5$ nodes and set of edges $\mathcal{E} = \{(2,1),(5,1),(3,2),(5,2),(4,3),(5,4)\}$. Let $\widetilde{x} = [0.737, 0.088, 0.410, 0.125, -1.362]^\mathsf{T}$, $\widetilde{z} = [0,0,0,0,1,1]^\mathsf{T}$, $\alpha = 0.1$, and $\beta = 1$. Then, the incidence matrix A and the vector of measurements can be easily constructed as

$$A = \begin{bmatrix} 1 & -1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & -1 \\ 0 & 1 & -1 & 0 & 0 \\ 0 & 1 & 0 & 0 & -1 \\ 0 & 0 & 1 & -1 & 0 \\ 0 & 0 & 0 & 1 & -1 \end{bmatrix}$$

and $b = [0.658, 2.105, -0.322, 1.450, -0.094, 1.190]^\mathsf{T}$. The resulting estimates are $\widehat{x}^{\mathrm{wls}} = [0.737, 0.078, 0.397, 0.156, -1.368]^\mathsf{T}$ by weighted least squares, $\widehat{x}^{\mathrm{ls}} = [0.803, 0.084, 0.222, 0.132, -1.242]^\mathsf{T}$ by unweighted least squares and $\widehat{x}^{\mathrm{lae}} = [0.803, 0.144, 0.242, 0.112, -1.302]^\mathsf{T}$ by ℓ_1-minimization. We obtain that $\|\widehat{x}^{\mathrm{wls}} - \widetilde{x}\|^2/\|\widetilde{x}\|^2 = 4.89 \cdot 10^{-4}$ and $\|\widehat{x}^{\mathrm{ls}} - \widetilde{x}\|^2/\|\widetilde{x}\|^2 = 2.09 \cdot 10^{-2}$ and $\|\widehat{x}^{\mathrm{lae}} - \widetilde{x}\|^2/\|\widetilde{x}\|^2 = 1.52 \cdot 10^{-2}$. ■

III. Centralized algorithm

In this section, we tackle the likelihood maximization problem (1) in its full generality. Since (1) does not admit a closed form solution, we propose an iterative algorithm that provides a solution in an iterative fashion. Preliminarily to designing our algorithm, we convert the Maximum Likelihood problem into a minimization problem by the following result, whose proof is postponed to Appendix A.

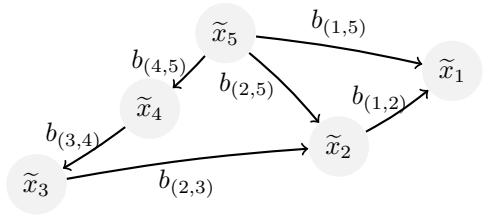

Fig. 1. The network of 5 nodes considered in Example 1.

Theorem 3: The following optimization problems have the same solutions

$$\max_{\alpha,\beta} \max_{x} L(x, \alpha, \beta) \tag{11}$$

$$-\min_{\alpha,\beta} \min_{x} \min_{\pi \in [0,1]^{\mathcal{E}}} V(x, \pi, \alpha, \beta) \tag{12}$$

where

$$\begin{aligned} V(x,\pi,\alpha,\beta) &= \frac{1}{2}\sum_{e\in\mathcal{E}} (b - Ax)_e^2 \left(\frac{1-\pi_e}{\alpha^2} + \frac{\pi_e}{\beta^2}\right) \\ &\quad + \sum_{e\in\mathcal{E}} \left[-\pi_e \log\frac{p}{\beta} - (1-\pi_e)\log\frac{1-p}{\alpha} - H(\pi_e)\right] \end{aligned} \tag{13}$$

and $H : [0,1] \to \mathbb{R}$ is the natural entropy function $H(\xi) = -\xi\log\xi - (1-\xi)\log(1-\xi)$.

Note that, with respect to the original problem (11), problem (12) explicitly introduces the variable $\pi \in [0,1]^{\mathcal{E}}$ which represents the estimated probabilities that the edges have large variances. Actually, instead of solving problem (12), we will solve a suitably modified problem, which we are going to define next. This modification marks a key difference with classical EM approaches. Namely, we shall solve

$$\min_{\alpha,\beta} \min_{x} \min_{\pi \in \Sigma_{|\mathcal{E}|-s}} \tilde{V}(x, \pi, \alpha, \beta, \epsilon), \tag{14}$$

where $\tilde{V} : \mathbb{R}^{\mathcal{V}} \times [0,1]^{\mathcal{E}} \times \mathbb{R}_{>0} \times \mathbb{R}_{>0} \times \mathbb{R}_{\geq 0} \to \mathbb{R}$ is

$$\begin{aligned} \tilde{V}(x,\pi,\alpha,\beta,\epsilon) &= \frac{1}{2}\sum_{e\in\mathcal{E}} \left((b-Ax)_e^2 + \frac{\epsilon}{|\mathcal{E}|}\right)\left(\frac{1-\pi_e}{\alpha^2} + \frac{\pi_e}{\beta^2}\right) \\ &\quad + \sum_{e\in\mathcal{E}} \left[-\pi_e \log\frac{p}{\beta} - (1-\pi_e)\log\frac{1-p}{\alpha} - H(\pi_e)\right]. \end{aligned} \tag{15}$$

Compared to (12), the optimization problem (14) introduces

- the positive variable ϵ, which has the goal to avoid possible singularities when one of the Gaussian components of the mixture collapses to one point;
- the constraint set $\Sigma_{|\mathcal{E}|-s}$, which implies that at least $s \geq 1$ measurements are classified as reliable.

As will become clear in the proofs, these modifications are instrumental to guarantee the convergence of the algorithms that we design. By defining function $\tilde{V}$, we do not intend

to pose any additional assumption in our original problem statement (1). However, problems (12) and (14) are not equivalent: instead, Problem (14) should be seen as a treatable approximation of (12). The mismatch between the two problems is meant to be small, since ε is bound to be small and it suffices to choose s as small as 1.

The following lemma summarizes the main properties of $\tilde{V}$ in minimization problems that only involve one variable at the time. Its proof can be obtained by differentiating $\tilde{V}$.

Proposition 4 (Partial minimizations): Let us define

$$\widehat{x} = \widehat{x}(\pi, \alpha, \beta, \epsilon) = \operatorname*{argmin}_{x \in \mathbb{R}^{\mathcal{V}}} \tilde{V}(x, \pi, \alpha, \beta, \epsilon)$$

$$\widehat{\pi} = \widehat{\pi}(x, \alpha, \beta, \epsilon) = \operatorname*{argmin}_{\pi \in \Sigma_{|\mathcal{E}|-s}} \tilde{V}(x, \pi, \alpha, \beta, \epsilon)$$

$$\widehat{\alpha} = \widehat{\alpha}(x, \pi, \beta, \epsilon) = \operatorname*{argmin}_{\alpha > 0} \tilde{V}(x, \pi, \alpha, \beta, \epsilon)$$

$$\widehat{\beta} = \widehat{\beta}(x, \pi, \alpha, \epsilon) = \operatorname*{argmin}_{\beta > 0} \tilde{V}(x, \pi, \alpha, \beta, \epsilon)$$

and denote $W = \operatorname{diag}\left(\frac{1-\pi}{\alpha^2} + \frac{\pi}{\beta^2}\right)$, $L_W = A^\top W A$, and $\xi_e = f(z_e = 1 | x, \alpha, \beta)$. Then, it holds true that

$$\widehat{x} = L_W^\dagger A^\top W b$$

$$\widehat{\pi} = \mathcal{P}_s(\xi)$$

$$\widehat{\alpha} = \sqrt{\frac{\sum_e (1-\pi_e)|b_e - (Ax)_e|^2 + \epsilon}{\|\mathbb{1} - \pi\|_1}}$$

$$\widehat{\beta} = \sqrt{\frac{\sum_e \pi_e |b_e - (Ax)_e|^2 + \epsilon}{\|\pi\|_1}}$$

where $\mathcal{P}_s : [0,1]^{\mathcal{E}} \to \Sigma_{|\mathcal{E}|-s}$ is the projection that zeroes the s smallest components of the given vector.
From the expressions of $\widehat{\alpha}$ and $\widehat{\beta}$, we can notice that the regularization term makes them greater than zero, and consequently also $\|\pi\|_1$.

Using the insights obtained by Proposition 4, we propose an alternating method for the minimization of (14). The resulting method is a combination of Iterative Reweighted Least Squares (IRLS) and Expectation Maximization. The algorithm, which is detailed in Algorithm 1, is based on the following four fundamental steps, which are iteratively repeated until convergence.

WLS solution: Given the relative measurements b and current parameters π_e, α, β, a new estimation of the variable x is obtained by solving the WLS problem with weights

$$w_e = (1-\pi_e)\alpha^{-2} + \pi_e \beta^{-2}, \quad \forall e \in \mathcal{E}.$$

Expectation: The posterior distribution ξ of the noise associated to the edges is evaluated, based on the current x, α, β.

Projection: The vector $\pi = \mathcal{P}_s(\xi)$ is the best $(|\mathcal{E}| - s)$-approximation of the posterior probability ξ. Therefore, the s smallest components of the posterior probabilities ξ are set to zero to make sure that at least s measurements are used in the WLS estimation problem.

Maximization: Given the projected posterior probability π, we use it to re-estimate the mixture parameters α and β.

The procedure is iterated until a suitable Stopping Criterion (SC) is satisfied, e.g. a maximum number of iterations $T_{\max}$ can be fixed (SC $=\{t \leq T_{\max}\}$) or the algorithm can be run until the estimate stops changing (SC $=\{\|x^{(t+1)} - x^{(t)}\| / \|x^{(t)}\| < \mathsf{tol}\}$ for some $\mathsf{tol} > 0$).

Algorithm 1 LS-EM

Require: Data: (b, A). Parameters: $c_1, c_2 > 0, p \in (0, \frac{1}{2})$, $\mathsf{tol} > 0$.

Initialization:
$t \leftarrow 0,\ \alpha^{(t)} \leftarrow \alpha_0,\ \beta^{(t)} \leftarrow \beta_0,\ \pi^{(t)} \leftarrow 0,\ \epsilon^{(t)} \leftarrow 1,$ SC $\leftarrow 1$.

while SC $\geq \mathsf{tol}$ **do**

Computation of weights: $\forall e \in \mathcal{E}$

$$w_e^{(t+1)} \leftarrow \frac{1-\pi_e^{(t)}}{(\alpha^{(t)})^2} + \frac{\pi_e^{(t)}}{(\beta^{(t)})^2}$$

WLS solution: $W^{(t+1)} \leftarrow \operatorname{diag}(w^{(t+1)})$

$$x^{(t+1)} \leftarrow L_{W^{(t+1)}}^\dagger A^\top W^{(t+1)} b,$$

Posterior distribution evaluation: $\forall e \in \mathcal{E}$

$$\xi_e^{(t+1)} \leftarrow f(z_e = 1 | x^{(t+1)}, \alpha^{(t)}, \beta^{(t)})$$

Best $(|\mathcal{E}| - s)$-approximation:

$$\pi^{(t+1)} \leftarrow \mathcal{P}_s(\xi^{(t+1)})$$

Regularization parameter:

$$\kappa^{(t+1)} \leftarrow \dim(\ker(L_{W^{(t+1)}}))$$

$$\theta^{(t+1)} \leftarrow \frac{1}{\log(t+1)} + c_1 \|x^{(t+1)} - x^{(t)}\| + c_2(\kappa^{(t+1)} - 1)$$

$$\epsilon^{(t+1)} \leftarrow \min\left(\epsilon^{(t)}, \theta^{(t+1)}\right)$$

Parameters estimation:

$$\alpha^{(t+1)} \leftarrow \sqrt{\frac{\epsilon^{(t)} + \sum_e (1-\pi_e^{(t+1)})|b_e - (Ax^{(t+1)})_e|^2}{\|\mathbb{1} - \pi^{(t+1)}\|_1}}$$

$$\beta^{(t+1)} \leftarrow \sqrt{\frac{\epsilon^{(t)} + \sum_e \pi_e^{(t+1)} |b_e - (Ax^{(t+1)})_e|^2}{\|\pi^{(t+1)}\|_1}}$$

Evaluate SC

$t \leftarrow t + 1$

end while

Although Algorithm 1 is a modified version of classical EM, this fact is not sufficient to guarantee the convergence of the proposed method. In fact, as observed in [26], a generic EM algorithm is not guaranteed to converge to a limit point but only to produce a sequence of points along which the log-likelihood function does not decrease. Hence, an explicit convergence proof is required in all specific cases. Algorithm 1 also includes a regularization sequence $\epsilon^{(t)}$, which appears in the "Maximization" step and is designed

to be monotonic and to go to zero upon convergence of the algorithm (see Step 7). The presence of such regularization is actually instrumental to prove the convergence to a local maximum of the log-likelihood function.

We underline that the proposed method (see Algorithm 1) can be interpreted also as an IRLS with a specific choice of the weights [36]. In contrast to classical IRLS, LS-EM allows to perform a classification of the measurements and the weights depend on the weighted energy based on this classification and this marks its difference with IRLS where the weights associated to edge e of the residual, chosen with the aim of approximating the ℓ_1-norm of residual, turn out to depend exclusively on the residual of edge $|b_e - (Ax)_e|$. The combination of EM with IRLS has been shown to outperform classical IRLS for in terms of speed and robustness in presence of noise in sparse estimation problems [37].

In order to state the convergence result, denote $\zeta^{(t)} = (x^{(t)}, \pi^{(t)}, \alpha^{(t)}, \beta^{(t)}, \epsilon^{(t)})$: then Algorithm 1 can be seen as a map from $\mathbb{R}^{\mathcal{V}} \times [0,1]^{\mathcal{E}} \times \mathbb{R}_{>0} \times \mathbb{R}_{>0} \times \mathbb{R}_{\geq 0}$ to itself that produces the sequence of iterates $\{\zeta^{(t)}\}_{t \in \mathbb{Z}_{\geq 0}}$.

Theorem 5 (LS-EM convergence): For any $b \in \mathbb{R}^{\mathcal{E}}$, the whole sequence $\zeta^{(t)}$ converges to $\zeta^{\infty} = (x^{\infty}, \pi^{\infty}, \alpha^{\infty}, \beta^{\infty}, \epsilon^{\infty})$ such that

$$x^{\infty} = L^{\dagger}_{W^{\infty}} A^{\top} W^{\infty} b, \quad W^{\infty} = \mathrm{diag}(w^{\infty}) \tag{16a}$$

$$w_e^{\infty} = \frac{1 - \pi_e^{\infty}}{|\alpha^{\infty}|^2} + \frac{\pi_e^{\infty}}{|\beta^{\infty}|^2} \tag{16b}$$

$$\pi_e^{\infty} = \mathcal{P}_s\Big(f(z_e = 1 | x^{\infty}, \alpha^{\infty}, \beta^{\infty})\Big) \tag{16c}$$

$$\alpha^{\infty} = \sqrt{\frac{\sum_e (1 - \pi_e^{\infty}) |b_e - (Ax^{\infty})_e|^2 + \epsilon^{\infty}}{\|\mathbb{1} - \pi^{\infty}\|_1}} \tag{16d}$$

$$\beta^{\infty} = \sqrt{\frac{\sum_e \pi_e^{\infty} |b_e - (Ax^{\infty})_e|^2 + \epsilon^{\infty}}{\|\pi^{\infty}\|_1}}. \tag{16e}$$

The converge point ζ^{∞} is a fixed point of the algorithm and a local minimum of $\tilde{V}(\cdot, \cdot, \cdot, \cdot, \epsilon^{\infty})$. If $\epsilon^{\infty} = 0$, then ζ^{∞} locally maximizes the log-likelihood.

The proof of Theorem 5 is based on observing that function $\tilde{V}$ in (14) is a Lyapunov function that is not increasing along the sequence of iterates. Details are postponed to Appendix B.

IV. Distributed algorithm

In this section, we design and study a distributed algorithm to solve problem (1), starting from the centralized one proposed in the previous section. Preliminary, let us examine steps 3–8 in Algorithm 1 in order to identify whether they are amenable to a distributed computation. Steps 3 and 5 only require information depending on edge e and are thus inherently decentralized. Furthermore, we already know that the least squares problem in Step 4 can be solved by a distributed procedure. Instead, steps 6–8 involve global information and can not easily be distributed.

Based on this discussion, we propose a simple but effective variation of LS-EM algorithm, detailed in Algorithm 2. Algorithm 2 is totally distributed and can be performed by the nodes: at each iteration, every node $v \in \mathcal{V}$ computes $w_e^{(t)}, \pi_e^{(t)}$ for every edge incident to it (see Step 3 and Step 5 in Algorithm 2) and the estimate of position $x_v^{(t)}$ (see Step 4 in Algorithm 2). The new algorithm is based on two design choices. The first choice is inspired by the distributed gradient dynamics (6): instead of fully solving a WLS problem at each iteration, we only perform *one step* of the corresponding gradient iteration. In the second choice, we assume α and β to be known, thus removing the need for their estimation. A further advantage is that keeping α and β fixed during the evolution of the algorithm avoids certain difficulties in the convergence analysis in Appendix B and namely removes the need for regularization and projection steps. Consequently, this distributed algorithm solves the exact estimation problem (12). Even though the knowledge of α and β can be a restrictive assumption, we have observed that the algorithm is fairly robust to uncertainties in these values: this quality is further discussed in Remark 1 below. The convergence of Algorithm 2 can be proved similarly to

Algorithm 2 Distributed LS-EM

Require: Data: (b, A). Parameters: $p \in (0, \frac{1}{2}), \tau > 0$, $0 < \alpha \ll \beta$, $\mathsf{tol} > 0$.

1: Initialization: $t \leftarrow 0,\ \pi^{(t)} \leftarrow 0,\ x^{(t)} \leftarrow 0,\ \mathrm{SC} \leftarrow 1$
2: **while** $\mathrm{SC} \geq \mathsf{tol}$ **do**
3: Computation of weights: $\forall e \in \mathcal{E}$
$$w_e^{(t+1)} \leftarrow \frac{1 - \pi_e^{(t)}}{\alpha^2} + \frac{\pi_e^{(t)}}{\beta^2}$$
4: Gradient step: $W \leftarrow \mathrm{diag}(w^{(t+1)})$
$$x^{(t+1)} \leftarrow (I - \tau L_W) x^{(t)} + \tau A^{\top} W b,$$
5: Posterior distribution evaluation: $\forall e \in \mathcal{E}$
$$\pi_e^{(t+1)} \leftarrow f(z_e = 1 | x^{(t+1)}, \alpha, \beta)$$
6: Evaluate SC
7: $t \leftarrow t + 1$
8: **end while**

Theorem 5, under the condition that the parameter τ belongs to a certain range: details are postponed to Appendix C.

Theorem 6 (Distributed LS-EM convergence): If $\tau < \alpha / \|A\|^2$, then for any $b \in \mathbb{R}^{\mathcal{E}}$ the sequence $(x^{(t)}, \pi^{(t)})$ generated by Algorithm 2 converges to $(x^{\infty}, \pi^{\infty})$ such that

$$x^{\infty} = L^{\dagger}_{W^{\infty}} A^{\top} W^{\infty} b, \quad W^{\infty} = \mathrm{diag}(w^{\infty}) \tag{17a}$$

$$w_e^{\infty} = \frac{1 - \pi_e^{\infty}}{\alpha^2} + \frac{\pi_e^{\infty}}{\beta^2} \tag{17b}$$

$$\pi_e^{\infty} = f(z_e = 1 | x^{\infty}, \alpha, \beta). \tag{17c}$$

The limit point $(x^{\infty}, \pi^{\infty})$ is a fixed point of the algorithm and a local minimum of V.

V. Numerical results

In this section, we provide simulations illustrating the performance of the proposed algorithms: we are mainly interested in comparing them in terms of their convergence times and final estimation errors.

A. Performance analysis of proposed algorithms

We examine how the performance of proposed techniques depends on the parameters of the problem, such as the parameters of the Gaussian mixture (α, β, p) and the topology of the measurement graph $\mathcal{G}$. As a measure of performance, we consider the normalized quadratic error (NQE), defined as $\text{NQE} = \|\widehat{x} - \widetilde{x}\|^2 / \|\widetilde{x}\|^2 * 100 \ [\%]$.

Let us begin by describing our baseline simulation setup in details. First, we generate synthetic data to define the estimation problem. The number of nodes is set to $N = 50$. The N components of the state vector are generated randomly according to a uniform distribution in the interval $(0, 1)$: then, the mean is subtracted yielding a state vector with mean 0. The topology is generated as Erdős-Rényi random graphs with edge probability p_{edge} ranging from 0.1 up to 1 (i.e., an edge is created between two arbitrary nodes with probability p_{edge}). In the extreme case of $p_{\text{edge}} = 1$, the graph generated is the complete graph where all nodes are connected to all others. We fix $\alpha = 0.05$ and β/α in the range from 2 to 10. Also the probability of getting a bad measurement p is taken between 0 and 1/2. The noise vector is then sampled from a normal distribution using a combination of the above parameters.

Next, we simulate the iterative algorithms. We initialize the state vector to be all zeros and also the vector π to be all zeros, meaning that all measurements are initially presumed to be good. For the LS-EM, an initial value for α and β is specified: α is randomly chosen from the set $\{0.1, 0.2, 0.3, 0.4, 0.5\}$ and $\beta = 2\alpha$ in order to meet the constraint $\beta > \alpha$. They are then held fixed for the different trials. For the Distributed LS-EM, the fixed values for α and β are taken to be the true values. The stopping criterion SC is chosen according to a tolerance $\text{tol} = 10^{-4}$, which has been verified to be small enough to represent numerical convergence. For the $(|\mathcal{E}| - s)$-approximation, we choose $s = N - 1$. This "optimistic" choice accounts to assume a number of valid measurements that could suffice to construct a spanning tree: this assumption is not imposed on our synthetic data. Similar $\mathcal{P}_s$ projections have already shown useful to accelerate convergence of iterative reweighted least square methods for estimation problems with sparsity prior [37].

Then, we simulate different trials whereby for each trial the vector $\widetilde{z}$ is regenerated. In order to illustrate the evolution of the algorithms, we plot the NQE against the iteration count for Algorithm 1 in Fig. 2 and Algorithm 2 in Fig. 3. We have chosen four trials out of a set of 250: the same trials (that is, the same random graphs and measurements) are chosen for both algorithms. We can observe that Algorithm 1 converges faster than Algorithm 2, but the two algorithms achieve similar final errors in a majority of trials.

The comparison between Algorithms 1 and Algorithm 2 is further explored in Fig. 4, where the distributions of the final NQEs for all the 250 trials mentioned above are summarized via the `boxplot` command in `MATLAB` with the default settings, showing the 25th (lower edge), 50th or median (central mark) and 75th (upper edge) percentiles. In order to make the comparison more complete and provide benchmarks, we also include the weighted least squares (WLS) as per (5) and the "naive" unweighted least squares estimator (LS) $\widehat{x}^{\text{ls}}$, in which we assume all measurements to be good. As expected, WLS outperforms all other estimators, thanks to using a-priori information on the noise parameters α, β and complete knowledge of the type of measurements. Instead, the naive LS has the worst performance.

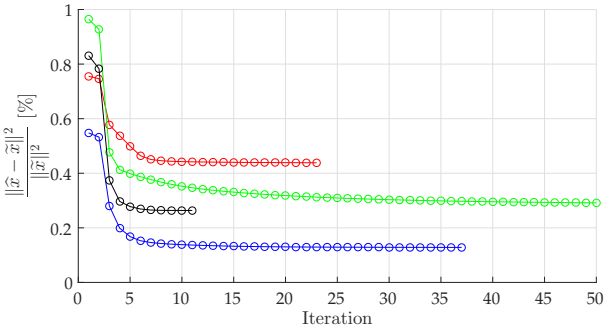

Fig. 2. NQE plotted against iteration count for 4 randomly chosen trials obtained using LS-EM Algorithm (each color represents the result of a trial); the parameter set is $N = 50, p_{\text{edge}} = 0.3, p = 0.1, \alpha = 0.05, \beta/\alpha = 5$. Note that the color of each chosen trial matches its counterpart in Fig. 3

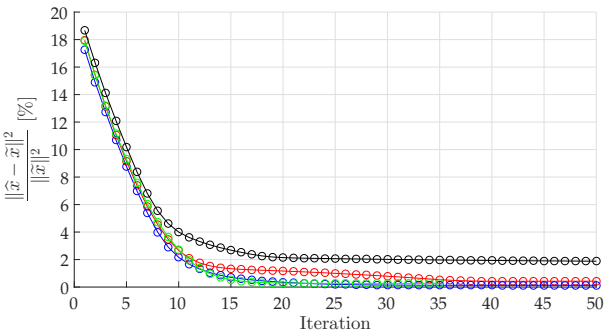

Fig. 3. NQE plotted against iteration count for 4 randomly chosen trials obtained using Distributed LS-EM Algorithm (each color represents the result of a trial); the parameter set is $N = 50, p_{\text{edge}} = 0.3, p = 0.1, \alpha = 0.05, \beta/\alpha = 5$. Note that the color of each chosen trial matches its counterpart in Fig. 2

We can observe that all our approaches have a median that is clearly lower than the median of the LS approach. Actually, the bulks of the error distributions are very similar to the WLS benchmark, except for few trials of the Distributed LS-EM that perform more poorly. A careful inspection of these few trials shows that these large errors are due to incorrect classification of the type of a small number of edges. This phenomenon is not observed in LS-EM, possibly thanks to the fact that in a centralized algorithm the information of all nodes is used at each iteration.

We also performed a parameter study in order to quantify the behavior of the mean NQE with respect to p_{edge}, p and β. In Fig. 5, we can observe that the mean NQE decreases with increasing p_{edge}: as the graph becomes more connected, there are more measurements available to estimate the state variables. From the same figure, we can also observe that starting from $p_{\text{edge}} = 0.4$, the performance of the Distributed

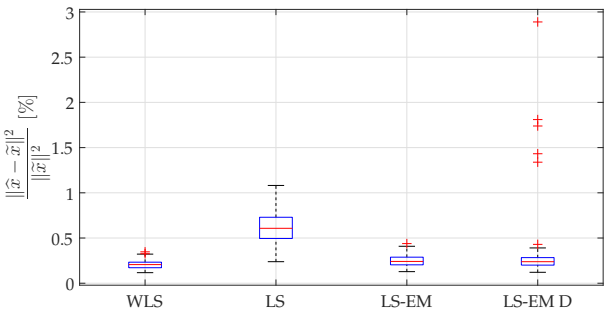

Fig. 4. Boxplot showing NQE for the different approaches; the parameter set is $N = 50, p_{\text{edge}} = 0.3, p = 0.1, \alpha = 0.05, \beta/\alpha = 5$. + are the NQE considered as outliers by the `boxplot` command;

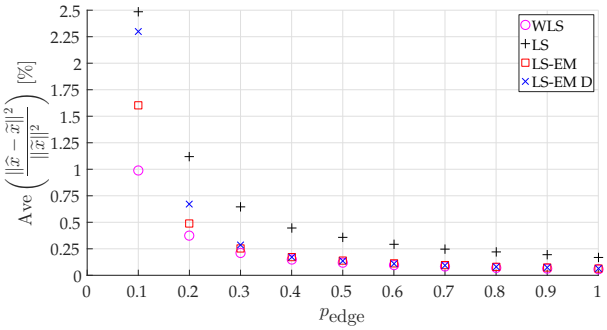

Fig. 5. Mean NQE with respect to p_{edge}; the parameter set is $N = 50, p = 0.1, \alpha = 0.05, \beta/\alpha = 5$ and the number of trials is 1000. ○ = WLS, + = LS, □ = LS-EM, × = Distributed LS-EM.

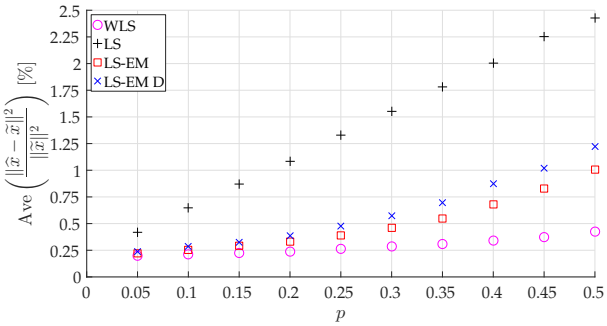

Fig. 6. Mean NQE with respect to p; the parameter set is $N = 50, p_{\text{edge}} = 0.3, \alpha = 0.05, \beta/\alpha = 5$ and the number of trials is 1000. ○ = WLS, + = LS, □ = LS-EM, × = Distributed LS-EM.

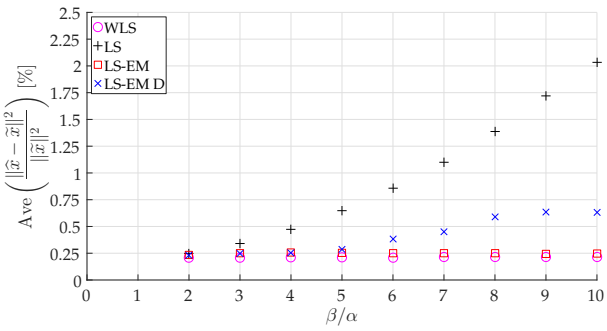

Fig. 7. Mean NQE with respect to β; the parameter set is $N = 50, p_{\text{edge}} = 0.3, p = 0.1, \alpha = 0.05$ and the number of trials is 1000. ○ = WLS, + = LS, □ = LS-EM, × = Distributed LS-EM.

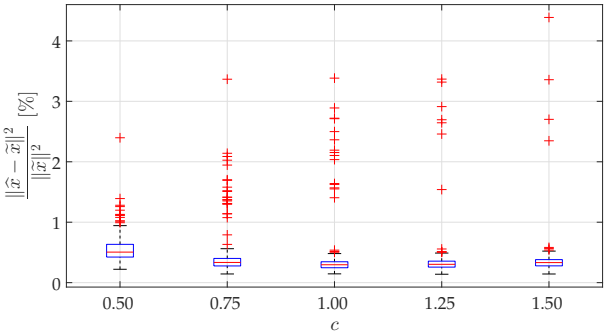

Fig. 8. Boxplot of NQE for different initialization of α and β obtained using Distributed LS-EM; we consider β/α to be known and the actual α and β values are set as c times their real values, with $c \in \{0.50, 0.75, 1.00, 1.25, 1.50\}$. The parameter set is $N = 50, p_{\text{edge}} = 0.3, p = 0.1, \alpha = 0.05, \beta/\alpha = 5$.

LS-EM is similar to that of LS-EM. In Fig. 6, the mean NQE increases with increasing p: this is due to the presence of more bad measurements. A similar reasoning explains the increase of NQE for increasing ratios β/α in Fig. 7. These dependencies on the parameters are consistent with intuition. From these three figures, it is clear that the Distributed LS-EM has larger average error than the centralised LS-EM (which has in turn a larger error than the WLS estimate). However, we should recall that the mean error of the Distributed LS-EM is driven up by the aforementioned sporadic errors: a comparison of median values would show a smaller gap from the centralized approach.

Remark 1 (Parameter uncertainty in Distributed LS-EM): Crucially, in Algorithm 2 the values for α and β are assumed to be known a priori. In practice one would usually not be able to have this information: hence we want to explore the sensitivity of the algorithm to incorrect choices of α and β. In Fig. 8, we assume not to know the actual values of α and β, but only their ratio β/α: we can observe that choosing α and β to be smaller than their actual value yields a higher median, while larger values than the real one yield similar results. In Fig. 9, we assume to know α but not β: we can observe that an incorrect and too large value of β increases the presence of large errors, even though the bulk of the error distribution remains similar. Overall, we conclude that the algorithm is fairly robust to moderate uncertainties in the knowledge of the parameters.

B. Distributed LS-EM versus Distributed LAE

In this section we compare Distributed LS-EM with a distributed version of LAE (see (10) for the update). In order to perform a fair comparison we introduce a mismatch in the measurements model. The measurements are not generated as a Gaussian mixtures and we consider an experiment coming from the geometric estimation problems in multi robot localization [15]. More precisely, we consider the following setting. Ground truth $\widetilde{x}$ of positions of $n = 30$ nodes are drawn from a uniform distribution over the interval $[-1, 1]$. Connections among nodes are generated according to the Erdős-Rényi random graph model, where each edge is

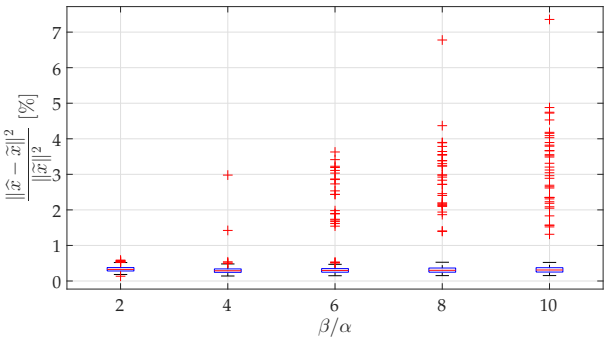

Fig. 9. Boxplot of NQE for different initialization of β obtained using Distributed LS-EM; we consider α to be known and let the actual ratio β/α to be $\{2, 4, 6, 8, 10\}$. The parameter set is $N = 50, p_{\text{edge}} = 0.3, p = 0.1, \alpha = 0.05, \beta/\alpha = 5$.

included in the graph with probability p_{edge} independently from every other edge. The outliers indicator vector $\widetilde{z} \sim \mathcal{B}(0.1)$ and the measurements are generated as follows

$$b = A\widetilde{x} + (1 - \widetilde{z})\alpha\eta + \widetilde{z}\gamma$$

where η is white Gaussian noise and γ_e is distributed according to a uniform distribution over the interval $[\Delta/4, \Delta/4]$ with Δ is the size of the environment. In Figure 11 we show a comparison between Distributed LS-EM and Distributed LAE in terms of speed of convergence for different values of $p_{\text{edge}} \in \{0.25, 0.5, 0, 75\}$. In order to perform a fair comparison we have fixed $p = 0.2$ (and not equal to 0.1) and $\alpha = 0.05$ and $\beta = 5\alpha$. The figure depicts the NQE averaged over 50 experiments as a function of number of iterations. The efficiency of the proposed algorithm allows to reduce the number of iterations required to achieve a satisfactory level of accuracy. As can be noticed, Distributed LS-EM need fewer updates (about 40 itarations) than Distributed LAE (more than 300 iterations) to achieve $\text{NQE} = 10^{-3}$ if $p_{\text{edge}} = 0.25$. This gain reduces when the graph originated by the measurement become denser and more connected: when $p_{\text{edge}} = 0.5$ few iterations (about 5) are needed to guarantee the convergence of Distributed LS-EM and about 90 for Distributed LAE.

VI. Concluding Remarks

In this paper, we have studied the problem of estimation from relative measurements with heterogeneous quality. We have introduced a novel formulation for the problem and we have proposed two original algorithms based on the method of Expectation-Maximization. One of the algorithms has the important feature of being distributed and thus amenable to applications where communication is limited or expensive. The other algorithm also distinguishes itself from standard EM approaches, due to the presence of regularization variables and of a projection step, which help dealing with the graph-dependent nature of the problem. Besides designing the algorithms, we have proved their convergence to a local maximum of the log-likelihood function (or to an approximation when regularization is employed). We note here that, as per the convergence properties, the projection step could be dispensed with at the price of a more involved proof: however, its role is not only in allowing for a proof but also in improving the performance in terms of speed, as we discussed in Section V-A. We have also presented a number of simulations that support the good performance of the algorithms and their robustness against uncertainties in the choice of the parameters. Despite a generally good performance, the algorithms (and particularly the distributed one) may perform poorly on some instances: this could be explained by the local nature of the optimality results. It is worth mentioning that the model considered in this paper considers only one measurement per node. Our choice derives from the need to make the theoretical analysis as simple as possible. The proposed algorithms can be adapted to the case when each node in the network has access to multiple measurements of relative distances. The arguments used to prove convergence can be adapted to this case, while repeated measurements would allow for smaller estimation errors.

Several other interesting problems remain open: we mention three of them here. First, one could further investigate the role of the topology of the measurement graphs in determining the performance of the algorithms: namely, some topologies could be more effective for the same number of measurements taken. Second, one could look for distributed algorithms that need not to assume the knowledge of the mixture parameters α and β. Third, one could propose algorithms that perform a "hard" classification of the measurements, as opposed to the "soft" classification that is done in this paper, where the measurements are assigned a probability of being of type α or β: some preliminary results and designs with hard classification are available in [38].

Appendix

A. Properties of the likelihood

Theorem 3 converts the ML problem (1) into a minimization problem. Before its proof, we recall expression (2) and introduce some useful notation:

$$f(b_e|x,\alpha,\beta) = \frac{1-p}{\sqrt{2\pi\alpha^2}} \exp\left(-\frac{(b-Ax)_e^2}{2\alpha^2}\right) + \frac{p}{\sqrt{2\pi\beta^2}} \exp\left(-\frac{(b-Ax)_e^2}{2\beta^2}\right), \tag{18}$$

$$f(b_e, z_e|x,\alpha,\beta) = \left[(1-z_e)\frac{1-p}{\sqrt{2\pi\alpha^2}} \exp\left(-\frac{(b-Ax)_e^2}{2\alpha^2}\right) + z_e \frac{p}{\sqrt{2\pi\beta^2}} \exp\left(-\frac{(b-Ax)_e^2}{2\beta^2}\right)\right], \tag{19}$$

$$f(b_e|z_e,x,\alpha,\beta) = \left[\frac{1-z_e}{\sqrt{2\pi\alpha^2}} \exp\left(-\frac{(b-Ax)_e^2}{2\alpha^2}\right) + \frac{z_e}{\sqrt{2\pi\beta^2}} \exp\left(-\frac{(b-Ax)_e^2}{2\beta^2}\right)\right] \tag{20}$$

and $f(z_e|b_e,x,\alpha,\beta) = \frac{f(b_e,z_e|x,\alpha,\beta)}{f(b_e|x,\alpha,\beta)}$.

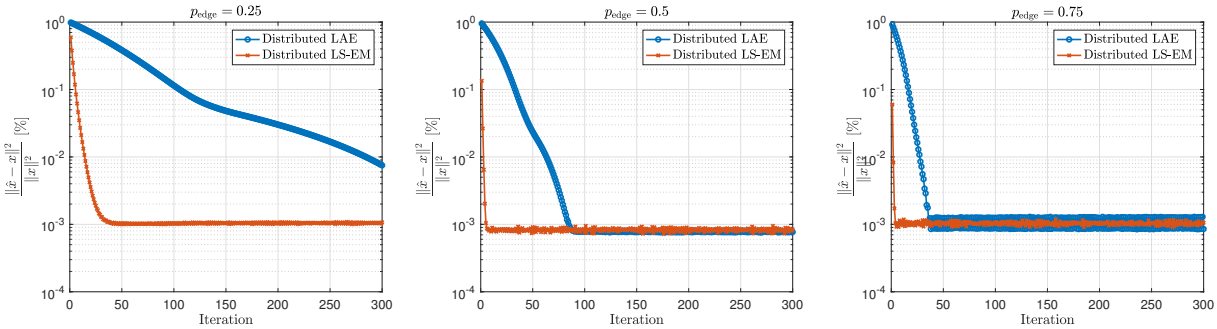

Fig. 10. NQE plotted against iteration count averaged over 50 trials for Distributed LAE and Distributed LS-EM Algorithm.

Proof of Theorem 3: From the definition of likelihood and using (18) we have

$$\begin{aligned} L(x,\alpha,\beta) &= \log \prod_{e\in\mathcal{E}} f(b_e|x,\alpha,\beta) \\ &= \sum_{e\in\mathcal{E}} \log f(b_e|x,\alpha,\beta) \\ &= \sum_{e\in\mathcal{E}} \log \sum_{z_e\in\{0,1\}} f(b_e,z_e|x,\alpha,\beta) \\ &= \sum_{e\in\mathcal{E}} \log \sum_{z_e\in\{0,1\}} \frac{f(b_e,z_e|x,\alpha,\beta)}{q(z_e)} q(z_e) \end{aligned}$$

for any map $q : \{0,1\} \to (0,1)$ such that $q(0) + q(1) = 1$. From Jensen's inequality we get $L(x,\alpha,\beta) \geq \sum_e \sum_{z_e=0,1} q(z_e) \log \frac{f(b_e,z_e|x,\alpha,\beta)}{q(z_e)}$. Let $\pi_e = q(z_e = 1) = 1 - q(z_e = 0)$. Therefore, we have

$$\begin{aligned} L(x,\alpha,\beta) \geq &\sum_{e\in\mathcal{E}} q(z_e=1) \log \frac{f(b_e,z_e=1|x,\alpha,\beta)}{q(z_e=1)} && (21) \\ &+ q(z_e=0) \log \frac{f(b_e,z_e=0|x,\alpha,\beta)}{q(z_e=0)} && (22) \\ &= \sum_{e\in\mathcal{E}} \pi_e \log \frac{\frac{p}{\sqrt{2\pi\beta^2}} \exp\left(-\frac{(b-Ax)_e^2}{2\beta^2}\right)}{\pi_e} \\ &+ \sum_{e\in\mathcal{E}} (1-\pi_e) \log \frac{\frac{1-p}{\sqrt{2\pi\alpha^2}} \exp\left(-\frac{(b-Ax)_e^2}{2\alpha^2}\right)}{1-\pi_e} \\ &= -\sum_{e\in\mathcal{E}} \left[(b-Ax)_e^2 \left(\frac{\pi_e}{2\beta^2} + \frac{1-\pi_e}{2\alpha^2} \right) - H(\pi_e) \right] \\ &+ \sum_{e\in\mathcal{E}} \left[\pi_e \log \frac{p}{\beta} + (1-\pi_e) \log \frac{1-p}{\alpha} \right] - \frac{|\mathcal{E}|}{2} \log(2\pi). \end{aligned}$$

This inequality is true for all $\pi_e \in [0,1]$ and $e \in \mathcal{E}$ since the function on the right-hand-side can be extended by continuity in $\pi_e \in \{0,1\}$. Therefore

$$\begin{aligned} &L(x,\alpha,\beta) \\ &\geq \max_{\pi\in[0,1]^{\mathcal{E}}} -\sum_{e\in\mathcal{E}} \left[(b-Ax)_e^2 \left(\frac{\pi_e}{2\beta^2} + \frac{1-\pi_e}{2\alpha^2} \right) - H(\pi_e) \right] \\ &+ \sum_{e\in\mathcal{E}} \left[\pi_e \log \frac{p}{\beta} + (1-\pi_e) \log \frac{1-p}{\alpha} \right] - \frac{|\mathcal{E}|}{2} \log(2\pi) \end{aligned}$$

Using the definition of function V in (13) we obtain

$$\begin{aligned} L(x,\alpha,\beta) &\geq \max_{\pi\in[0,1]^{\mathcal{E}}} -V(x,\pi,\alpha,\beta) - \frac{|\mathcal{E}|}{2} \log(2\pi) \\ &= -\min_{\pi\in[0,1]^{\mathcal{E}}} V(x,\pi,\alpha,\beta) - \frac{|\mathcal{E}|}{2} \log(2\pi) \end{aligned} \tag{23}$$

By differentiating $V(x,\pi,\alpha,\beta)$ with respect to π_e we write the optimality condition as

$$\begin{aligned} \frac{\partial V}{\partial \pi_e} &= -\log \frac{1-\pi_e}{\pi_e} - \log \frac{p}{\beta} + \frac{(b-Ax)_e^2}{2\beta^2} \\ &+ \log \frac{1-p}{\alpha} - \frac{(b-Ax)_e^2}{2\alpha^2} = 0 \end{aligned}$$

from which we obtain

$$\begin{aligned} \widehat{\pi}_e &= \frac{\frac{p}{\sqrt{2\pi\beta^2}} \exp\left(-\frac{(b-Ax)_e^2}{2\beta^2}\right)}{\frac{p}{\sqrt{2\pi\beta^2}} \exp\left(-\frac{(b-Ax)_e^2}{2\beta^2}\right) + \frac{1-p}{\sqrt{2\pi\alpha^2}} \exp\left(-\frac{(b-Ax)_e^2}{2\alpha^2}\right)} \\ &= f(z_e = 1|b_e,x,\alpha,\beta). \end{aligned}$$

Replacing $q(z_e = 1)$ and $q(z_e = 0)$ with $\widehat{\pi}_e = f(z_e = 1|b_e,x,\alpha,\beta)$ and $(1-\widehat{\pi}_e) = f(z_e = 0|b_e,x,\alpha,\beta)$ in (21) and (22), respectively, we get

$$\begin{aligned} &L(x,\alpha,\beta) \\ &\geq \sum_{e\in\mathcal{E}} \widehat{\pi}_e \log \frac{f(b_e,z_e=1|x,\alpha,\beta)}{\widehat{\pi}_e} \\ &+ (1-\widehat{\pi}_e) \log \frac{f(b_e,z_e=0|x,\alpha,\beta)}{1-\widehat{\pi}_e} \\ &\geq \sum_{e\in\mathcal{E}} f(z_e=1|b_e,x,\alpha,\beta) \log \frac{f(b_e,z_e=1|x,\alpha,\beta)}{f(z_e=1|b_e,x,\alpha,\beta)} \\ &+ f(z_e=0|b_e,x,\alpha,\beta) \log \frac{f(b_e,z_e=0|x,\alpha,\beta)}{f(z_e=0|b_e,x,\alpha,\beta)} \\ &= \sum_{e\in\mathcal{E}} \sum_{z_e\in\{0,1\}} f(z_e|b_e,x,\alpha,\beta) \log \frac{f(b_e,z_e|x,\alpha,\beta)}{f(z_e|b_e,x,\alpha,\beta)} \\ &= \sum_{e\in\mathcal{E}} \sum_{z_e\in\{0,1\}} f(z_e|b_e,x,\alpha,\beta) \log f(b_e|x,\alpha,\beta) \end{aligned}$$

from which we conclude that the inequality in (23) is actually an equality. Therefore

$$L(x,\alpha,\beta) = -\min_{\pi\in[0,1]^{\mathcal{E}}} V(x,\pi,\alpha,\beta) - \frac{|\mathcal{E}|}{2} \log(2\pi).$$

where the last expression is obtained using the definition of function V in (13). We conclude that

$$\max_{\alpha,\beta}\max_{x} L(x,\alpha,\beta) = -\min_{\alpha,\beta}\min_{x}\min_{\pi\in[0,1]^{\mathcal{E}}} V(x,\pi,\alpha,\beta)+c$$

with $c=-\frac{|\mathcal{E}|}{2}\log(2\pi)$. ■

B. Proof of Theorem 5: convergence of Algorithm 1

Lemma 7 (Monotonicity): The function $\tilde{V}$ defined in (15) is nonincreasing along the iterates $\zeta^{(t)} = (x^{(t)},\pi^{(t)},\alpha^{(t)},\beta^{(t)},\epsilon^{(t)})$.

Proof: Repeatedly applying Proposition 4 yields

$$\begin{aligned}\tilde{V}(\zeta^{(t+1)}) &= \tilde{V}(x^{(t+1)},\pi^{(t+1)},\alpha^{(t+1)},\beta^{(t+1)},\epsilon^{(t+1)})\\ &\le \tilde{V}(x^{(t+1)},\pi^{(t+1)},\alpha^{(t)},\beta^{(t)},\epsilon^{(t)})\\ &\le \tilde{V}(x^{(t+1)},\pi^{(t)},\alpha^{(t)},\beta^{(t)},\epsilon^{(t)})\\ &\le \tilde{V}(x^{(t)},\pi^{(t)},\alpha^{(t)},\beta^{(t)},\epsilon^{(t)}) = \tilde{V}(\zeta^{(t)})\end{aligned}$$

for every time t, proving the result. ■

The following lemma implies that Algorithm 1 converges numerically.

Lemma 8 (Asymptotic regularity): If $(x^{(t)})$ is the sequence generated by Algorithm 1, then $x^{(t+1)}-x^{(t)}\to 0$ as $t\to\infty$.

Proof: From their definitions we have

$$\alpha^{(t+1)} \ge \sqrt{\epsilon^{(t)}/|\mathcal{E}|} \qquad \beta^{(t+1)} \ge \sqrt{\epsilon^{(t)}/|\mathcal{E}|}.$$

Then, if $\alpha^{(t)}\to 0$ or $\beta^{(t)}\to 0$ as $t\to\infty$, we have $\epsilon^{(t)}\to 0$ and, consequently, $\|x^{(t+1)}-x^{(t)}\|_2\to 0$ and the assertion is verified. If instead neither α nor β converge to zero, then there exists a constant $K>0$ and a divergent sequence of integers t_ℓ such that $\min\{\alpha^{(t_\ell)},\beta^{(t_\ell)}\}>K$ for all $\ell\in\mathbb{N}$. It holds in general that

$$\begin{aligned}&\sum_{e\in\mathcal{E}}\pi_e^{(t)}\log\beta^{(t)} + \sum_{e\in\mathcal{E}}(1-\pi_e^{(t)})\log\alpha^{(t)} - |\mathcal{E}|\log 2\\ &\quad\le \tilde{V}(\zeta^{(t)}) \le \tilde{V}(\zeta^{(1)})\end{aligned} \tag{24}$$

where the last inequality follows from Lemma 7. Then, $\tilde{V}(\zeta^{(t_\ell)}) \ge (|\mathcal{E}|\log K - |\mathcal{E}|\log 2)$.

Since $x^{(t+1)} = \operatorname{argmin}_{x\in\mathbb{R}^{\mathcal{V}}} \tilde{V}(x,\pi^{(t)},\alpha^{(t)},\beta^{(t)},\epsilon^{(t)})$ we have

$$\tilde{V}(x^{(t+1)},\pi^{(t)},\alpha^{(t)},\beta^{(t)},\epsilon^{(t)}) \le \tilde{V}(x^{(t)},\pi^{(t)},\alpha^{(t)},\beta^{(t)},\epsilon^{(t)}) \tag{25}$$

and

$$\begin{aligned}&\nabla_x[\tilde{V}(x,\pi^{(t)},\alpha^{(t)},\beta^{(t)},\epsilon^{(t)})](x^{(t+1)})\\ &\quad= A^\top W^{(t)}Ax^{(t+1)} - A^\top W^{(t)}b\\ &\quad= L_{W^{(t)}}x^{(t+1)} - A^\top W^{(t)}b = 0\end{aligned} \tag{26}$$

where $W^{(t)} = \operatorname{diag}\left(\frac{1-\pi^{(t)}}{(\alpha^{(t)})^2}+\frac{\pi^{(t)}}{(\beta^{(t)})^2}\right)$ and $L_{W^{(t)}} = A^\top W^{(t)}A$. From (25) we then have

$$\begin{aligned}&\tilde{V}(x^{(t)},\pi^{(t)},\alpha^{(t)},\beta^{(t)},\epsilon^{(t)})\\ &\quad-\tilde{V}(x^{(t+1)},\pi^{(t+1)},\alpha^{(t+1)},\beta^{(t+1)},\epsilon^{(t+1)})\\ &\ge \tilde{V}(x^{(t)},\pi^{(t)},\alpha^{(t)},\beta^{(t)},\epsilon^{(t)})\\ &\quad-\tilde{V}(x^{(t+1)},\pi^{(t)},\alpha^{(t)},\beta^{(t)},\epsilon^{(t)})\\ &=\frac{1}{2}\sum_{e\in\mathcal{E}}\left((b-Ax^{(t)})_e^2+\frac{\epsilon^{(t)}}{|\mathcal{E}|}\right)\left(\frac{1-\pi_e^{(t)}}{(\alpha^{(t)})^2}+\frac{\pi_e^{(t)}}{(\beta^{(t)})^2}\right)\\ &\quad-\frac{1}{2}\sum_{e\in\mathcal{E}}\left((b-Ax^{(t+1)})_e^2+\frac{\epsilon^{(t)}}{|\mathcal{E}|}\right)\left(\frac{1-\pi_e^{(t)}}{(\alpha^{(t)})^2}+\frac{\pi_e^{(t)}}{(\beta^{(t)})^2}\right)\\ &=\frac{1}{2}(x^{(t)})^\top L_{W^{(t)}}x^{(t)} - (x^{(t)})^\top A^\top W^{(t)}b\\ &\quad-\frac{1}{2}(x^{(t+1)})^\top L_{W^{(t)}}x^{(t+1)} + (x^{(t+1)})^\top A^\top W^{(t)}b\\ &=\frac{1}{2}(x^{(t)}-x^{(t+1)})^\top L_{W^{(t)}}(x^{(t)}+x^{(t+1)})\\ &\quad-(x^{(t)}-x^{(t+1)})^\top A^\top W^{(t)}b\\ &=\frac{1}{2}(x^{(t)}-x^{(t+1)})^\top L_{W^{(t)}}x^{(t)}\\ &\quad+(x^{(t)}-x^{(t+1)})^\top(\frac{1}{2}L_{W^{(t)}}x^{(t+1)} - A^\top W^{(t)}b)\end{aligned}$$

From (26) we get

$$\begin{aligned}\tilde{V}(\zeta^{(t)})-\tilde{V}(\zeta^{(t+1)}) &\ge \frac{1}{2}(x^{(t)}-x^{(t+1)})^\top L_{W^{(t)}}x^{(t)}\\ &\quad-\frac{1}{2}(x^{(t)}-x^{(t+1)})^\top L_{W^{(t)}}x^{(t+1)}\\ &\ge \frac{1}{2}(x^{(t)}-x^{(t+1)})^\top L_{W^{(t)}}(x^{(t)}-x^{(t+1)})\\ &\ge \frac{1}{2}\min_{v:\mathbb{1}^\top v=0}\frac{v^\top(L_{W^{(t)}})v}{\|v\|^2}\|x^{(t)}-x^{(t+1)}\|^2.\end{aligned}$$

The last inequality is true since $L_{W^{(t)}}$ is positive semidefinite, the multiplicity of the eigenvalue 0 is equal to 1 and $\mathbb{1}^\top x^{(t)}=0$ for all $t\in\mathbb{N}$. We can thus define

$$\lambda^{(t)} := \frac{1}{2}\min_{v:\ \mathbb{1}^\top v=0}\frac{v^\top(L_{W^{(t)}})v}{\|v\|^2}.$$

We now prove that $\exists t_0\in\mathbb{N}$ such that $\lambda^{(t)}\ge c>0$ for all $t\ge t_0$. In fact, suppose by contradiction that there exists (t_j) such that $\lim_{j\to\infty}\lambda^{(t_j)}=0$. Then, there needs to exist a subsequence t_ℓ such that $\alpha^{(t_\ell)}$ or $\beta^{(t_\ell)}$ diverge. If $\beta^{(t_\ell)}\to\infty$, then (24) implies $\alpha^{(t_\ell)}\to 0$. From Step 8 in Algorithm 1 we obtain $\epsilon^{(t_\ell)}\to 0$ and $\kappa^{(t_\ell)}\to 1$. We deduce that there exists $\ell_0\in\mathbb{N}$ such that $\kappa^{(t_\ell)}=1$ for all $\ell>\ell_0$, from which we get the contradiction $\lambda^{(t_\ell)}>c>0$ for all $\ell>\ell_0$. The case $\alpha^{(t_\ell)}\to\infty$ is analogous.

We now compute for $\ell\in\mathbb{N}$

$$\begin{aligned}0\le\sum_{t=1}^{t_\ell-1}c\|x^{(t)}-x^{(t+1)}\|^2 &\le \sum_{t=1}^{t_\ell-1}\left(\tilde{V}(\zeta^{(t)})-\tilde{V}(\zeta^{(t+1)})\right)\\ &= \tilde{V}(\zeta^{(1)})-\tilde{V}(\zeta^{(t_\ell)})\\ &\le \tilde{V}(\zeta^{(1)})-(|\mathcal{E}|\log K-|\mathcal{E}|\log 2) = K'.\end{aligned}$$

By letting $\ell\to\infty$, we obtain that $\|x^{(t)}-x^{(t+1)}\|\to 0$. ■

Lemma 9: The sequence $(x^{(t)})_{t\in\mathbb{N}}$ is bounded.

Proof: If $(\alpha^{(t)})_{t\in\mathbb{N}}$ and $(\beta^{(t)})_{t\in\mathbb{N}}$ are both upper bounded by a constant $\chi > 0$, then

$$\begin{aligned} 0 &\le \|b - Ax^{(t)}\|_2^2 \\ &\le \sum_{e\in\mathcal{E}} (1-\pi_e^{(t)})[\alpha^{(t)}]^2 + \sum_{e\in\mathcal{E}} \pi_e^{(t)}[\beta^{(t)}]^2 \le \chi^2|\mathcal{E}|, \end{aligned}$$

which guarantees that $x^{(t)}$ is bounded as well. Next, we will show that if either $\alpha^{(t)}$ or $\beta^{(t)}$ were unbounded, $x(t)$ would actually be convergent and thus bounded.

To this purpose, we start by observing from (15) that

$$\begin{aligned} &\sum_{e\in\mathcal{E}} \pi_e^{(t)} \log\beta^{(t)} + \sum_{e\in\mathcal{E}} (1-\pi_e^{(t)}) \log\alpha^{(t)} - |\mathcal{E}|\log 2 \\ &\quad \le \tilde{V}(x^{(t)}, \pi^{(t)}, \alpha^{(t)}, \beta^{(t)}, \epsilon^{(t)}) \\ &\quad \le \tilde{V}(x^{(1)}, \pi^{(1)}, \alpha^{(1)}, \beta^{(1)}, \epsilon^{(1)}). \end{aligned} \tag{27}$$

Suppose now that $\beta^{(t)}$ is not upper bounded. Then, there exists a subsequence $(t_\ell)_{\ell\in\mathbb{N}}$ such that $\lim_{\ell\to\infty}\beta^{(t_\ell)} = \infty$. Then, inequality (27) implies that there are two cases: either we have $\pi^{(t_\ell)} \to 0$ for $\ell \to \infty$, or $\alpha^{(t_\ell)} \to 0$. In the former case, we have $x^{(t_\ell)} \to L^\dagger A^\top b$ (where L is the limit of $L_{W^{(t_\ell)}}$) implying that $x^{(t)}$ is bounded by asymptotic regularity. In the latter case, there exists $e \in \mathcal{E}$ such that $\pi_e^{(t_\ell)} \neq 0$, implying that $\lim_{\ell\to\infty}\alpha^{(t_\ell)} = 0$. From Steps 7 and 8 in Algorithm 1 we get that $\epsilon^{(t_\ell)} \to 0$ and consequently $\kappa^{(t_\ell)} \to 1$. Being $\kappa^{(t)}$ an integer, there exists $\ell_0 \in \mathbb{N}$ such that $\kappa^{(t_\ell)} = 1$ for all $\ell > \ell_0$. Since

$$\alpha^{(t_\ell)} = \sqrt{\frac{\sum_{e\in\mathcal{E}}(1-\pi_e^{(t_\ell)})|b_e - (Ax^{(t_\ell)})_e|^2 + \epsilon^{(t_\ell)}}{\sum_{e\in\mathcal{E}}(1-\pi_e^{(t_\ell)})}},$$

we have that if there exist $\epsilon > 0$ and $\{t_{\ell_j}\}_{j\in\mathbb{N}}$ such that $|b_e - (Ax^{(t_{\ell_j})})_e| > \epsilon$ then $\pi_e^{(t_{\ell_j})} \to 1$ as $j \to \infty$. On the other hand, if $|b_e - (Ax^{(t_\ell)})_e| \to 0$ then $\pi_e^{(t_\ell)} \to 0$. This means that

$$\lim_{j\to\infty} \pi_e^{(t_{\ell_j})} = \begin{cases} 1 & \text{if } e \in \Delta \\ 0 & \text{otherwise} \end{cases},$$

where the set Δ is defined as follows

$$\Delta = \{e \in \mathcal{E} : \exists\epsilon > 0 \text{ and } (t_{\ell_j})_j \text{ s.t } |b_e - (Ax^{(t_{\ell_j})})_e| > \epsilon\}.$$

Observe that the relative complement $\Delta^{\mathrm{c}} = \mathcal{E} \setminus \Delta$ has cardinality not smaller than s: using this notation, we can deduce that

$$\begin{aligned} \lim_{j\to\infty} A^\top W^{(t_{\ell_j})} A x^{(t_{\ell_j})} &= \lim_{j\to\infty} A^\top W^{(t_{\ell_j})} b \\ A_{\Delta^{\mathrm{c}}}^\top A_{\Delta^{\mathrm{c}}} \lim_{j\to\infty} x^{(t_{\ell_j})} &= \lim_{j\to\infty} A_{\Delta^{\mathrm{c}}}^\top b_{\Delta^{\mathrm{c}}}. \end{aligned}$$

Since with $\kappa^{(t_{\ell_j})} = 1$, the sequence $(x^{(t_{\ell_j})})_{j\in\mathbb{N}}$ converges

$$\lim_{j\to\infty} x^{(t_{\ell_j})} = (A_{\Delta^{\mathrm{c}}}^\top A_{\Delta^{\mathrm{c}}})^\dagger A_{\Delta^{\mathrm{c}}}^\top b_{\Delta^{\mathrm{c}}}.$$

and so does $x^{(t)}$ by asymptotic regularity.

Similarly, the case of $\alpha^{(t)}$ unbounded leads to two cases: either $\pi^{(t_\ell)} \to 1$ or $\beta \to 0$. The former case is actually forbidden by the presence of at least s components equal to zero. The latter case is treated in analogous way as the case $\alpha \to 0$ above: we omit its detailed discussion. ■

Lemma 10: Any accumulation point of $\zeta^{(t)}$ is a fixed point of Algorithm 1 and satisfies equalities (16a)-(16e).

Proof: If $(x^\sharp, \pi^\sharp, \alpha^\sharp, \beta^\sharp, \epsilon^\sharp)$ is an accumulation point of the sequence $(x^{(t)}, \pi^{(t)}, \alpha^{(t)}, \beta^{(t)}, \epsilon^{(t)})_{t\in\mathbb{N}}$, then there exists a subsequence $(x^{(t_\ell)}, \pi^{(t_\ell)}, \alpha^{(t_\ell)}, \beta^{(t_\ell)}, \epsilon^{(t_\ell)})_{\ell\in\mathbb{N}}$ that converges to $(x^\sharp, \pi^\sharp, \alpha^\sharp, \beta^\sharp, \epsilon^\sharp)$ as $\ell \to \infty$. We now show (16c), since the other conditions are immediate by continuity. In order to verify (16c), we need to prove that for all $i \in \mathrm{supp}(\pi^\sharp)$

$$\pi_i^\sharp = \frac{\exp(-\frac{|b_i-(Ax^\sharp)_i|^2}{2|\beta^\sharp|^2})\frac{p}{\beta^\sharp}}{\frac{1-p}{\alpha^\sharp}\exp(-\frac{|b_i-(Ax^\sharp)_i|^2}{2|\alpha^\sharp|^2}) + \frac{p}{\beta^\sharp}\exp(-\frac{|b_i-(Ax^\sharp)_i|^2}{2|\beta^\sharp|^2})} \tag{28}$$

and for any $i \notin \mathrm{supp}(\pi^\sharp)$ and $j \in \mathrm{supp}(\pi^\sharp)$

$$\begin{aligned} &\frac{\exp(-\frac{|b_i-(Ax^\sharp)_i|^2}{2|\beta^\sharp|^2})\frac{p}{\beta^\sharp}}{\frac{1-p}{\alpha^\sharp}\exp(-\frac{|b_i-(Ax^\sharp)_i|^2}{2|\alpha^\sharp|^2}) + \frac{p}{\beta^\sharp}\exp(-\frac{|b_i-(Ax^\sharp)_i|^2}{2|\beta^\sharp|^2})} \\ &\quad \le \frac{\exp(-\frac{|b_j-(Ax^\sharp)_j|^2}{2|\beta^\sharp|^2})\frac{p}{\beta^\sharp}}{\frac{1-p}{\alpha^\sharp}\exp(-\frac{|b_j-(Ax^\sharp)_j|^2}{2|\alpha^\sharp|^2}) + \frac{p}{\beta^\sharp}\exp(-\frac{|b_j-(Ax^\sharp)_j|^2}{|\beta^\sharp|^2})}. \end{aligned} \tag{29}$$

Since $\lim_{\ell\to\infty}\pi^{(t_\ell)} = \pi^\sharp$, then there exists ℓ_0 such that, $\forall\ell > \ell_0$ and $\forall i \in \mathrm{supp}(\pi^\sharp)$, $\pi_i^{(t_\ell)} \neq 0$ and $\pi_i^{(t_\ell)} = \xi_i^{(t_\ell)} \to \pi_i^\sharp$, so that (28) is verified.

If $i \notin \mathrm{supp}(\pi^\sharp)$, then we have to distinguish the following two cases: either (a) $\pi_i^{(t_\ell)}$ is zero eventually or (b) $\pi_i^{(t_\ell)}$ converges to zero asymptotically. In case (a), there exists $\ell_0 \in \mathbb{N}$ such that $\forall\ell > \ell_0$, $\pi_i^{(t_\ell)} = 0$, from which $\xi_i^{(t_\ell)} < \xi_j^{(t_\ell)}$, $\forall j \in \mathrm{supp}(\pi^\sharp)$ and (29) is satisfied. In case (b), there exists a strictly positive sub-sequence $(\ell_q)_{q\in\mathbb{N}}$ such that $\pi_i^{(t_{\ell_q})} = \xi_i^{(t_{\ell_q})} \to \pi_i^\sharp = 0$. Since at the same time $\xi_j^{(t_{\ell_q})}$ converges to $\pi_i^\sharp > 0$ for all $j \in \mathrm{supp}(\pi^\sharp)$, there exists $q_0 \in \mathbb{N}$ such that $\forall q > q_0$ we have $\xi_i^{(t_{\ell_q})} < \xi_j^{(t_{\ell_q})}$, and, by letting $q \to \infty$,

$$\begin{aligned} &\frac{\exp(-\frac{|b_i-(Ax^\sharp)_i|^2}{2|\beta^\sharp|^2})\frac{p}{\beta^\sharp}}{\frac{1-p}{\alpha^\sharp}\exp(-\frac{|b_i-(Ax^\sharp)_i|^2}{2|\alpha^\sharp|^2}) + \frac{p}{\beta^\sharp}\exp(-\frac{|b_i-(Ax^\sharp)_i|^2}{2|\beta^\sharp|^2})} \\ &\quad \le \frac{\exp(-\frac{|b_j-(Ax^\sharp)_j|^2}{2|\beta^\sharp|^2})\frac{p}{\beta^\sharp}}{\frac{1-p}{\alpha^\sharp}\exp(-\frac{|b_j-(Ax^\sharp)_j|^2}{2|\alpha^\sharp|^2}) + \frac{p}{\beta^\sharp}\exp(-\frac{|b_j-(Ax^\sharp)_j|^2}{|\beta^\sharp|^2})} \end{aligned}$$

We conclude that for all $i \notin \mathrm{supp}(\pi^\infty)$

$$\pi_i^\sharp = 0 = \mathcal{P}_s\left(\frac{\exp(-\frac{|b-Ax^\sharp|^2}{2|\beta^\sharp|^2})\frac{p}{\beta^\sharp}}{\frac{1-p}{\alpha^\sharp}\exp(-\frac{|x^\sharp|^2}{2|\alpha^\sharp|^2}) + \frac{p}{\beta^\sharp}\exp(-\frac{|x^\sharp|^2}{2|\beta^\sharp|^2})}\right)_i.$$

■

Since the sequence $x^{(t)}$ is bounded (see Lemma 9), there exists a subsequence $x^{(t_j)}$ such that $x^{(t_j)} \to x^\infty$. Moreover, $\alpha^{(t_j)} \to \alpha^\infty$, $\beta^{(t_j)} \to \beta^\infty$, $\pi^{(t_j)} \to \pi^\infty$, and $\varepsilon^{(t_j)} \to \varepsilon^\infty$. From Lemma 8, we get $\lim_{t\to\infty} x^{(t_j+1)} = \lim_{t\to\infty} x^{(t_j)} = x^\infty$, proving convergence. Finally, Lemma 10 ensures that $\zeta^{(t)}$ converges to a fixed point.

C. Proof of Theorem 6: convergence of Algorithm 2

Let us consider the function $V : \mathbb{R}^{\mathcal{V}} \times [0,1]^{\mathcal{E}} \to \mathbb{R}$ defined from (13) by fixing the variables α and β, together with a surrogate function $V^{\mathsf{S}} : \mathbb{R}^{\mathcal{V}} \times \mathbb{R}^{\mathcal{V}} \times [0,1]^{\mathcal{E}} \to \mathbb{R}$

$$V^{\mathsf{S}}(x,z,\pi) = V(x,\pi) + \frac{1}{2\tau}(x-z)^\top (I - \tau L_W)(x-z), \quad (30)$$

where

$$W = \operatorname{diag}\left(\frac{1-\pi}{\alpha^2} + \frac{\pi}{\beta^2}\right)$$

The following two lemmas are stated without proof.

Lemma 11 (Partial minimizations): If

$$\widehat{x} = \operatorname*{argmin}_{x \in \mathbb{R}^{\mathcal{V}}} V^{\mathsf{S}}(x,z,\pi) \qquad \text{and} \qquad \widehat{\pi} = \operatorname*{argmin}_{\pi \in [0,1]^{\mathcal{E}}} V(x,\pi),$$

then

$$\widehat{x} = (I - \tau L_W) z + \tau A^\top W b$$

$$\widehat{\pi}_e = \frac{\frac{p}{\beta} \mathrm{e}^{-\frac{|b_e - (Ax)_e|^2}{2\beta^2}}}{\frac{p}{\beta} \mathrm{e}^{-\frac{|b_e - (Ax)_e|^2}{2\beta^2}} + \frac{(1-p)}{\alpha} \mathrm{e}^{-\frac{|b_e - (Ax)_e|^2}{2\alpha^2}}} \quad \forall\, e \in \mathcal{E}$$

Lemma 12 (Monotonicity): The function V defined in this section is nonincreasing along the iterates $\zeta^{(t)} = (x^{(t)}, \pi^{(t)})$.

We are now able to show that Algorithm 2 converges numerically.

Lemma 13: If $x^{(t)}$ is the sequence generated by Algorithm 2, then $x^{(t+1)} - x^{(t)} \to 0$ as $t \to \infty$.

Proof: Define $\mu = \max_t \|A^\top W^{(t)} A\| \leq \|A\|^2/\alpha$ and $\|A\|$ is the spectral norm. Since from assumption $\tau < \alpha/\|A\|^2 < \mu^{-1}$ we have

$$\begin{aligned} 0 &\leq \frac{1}{2\tau}(1-\tau\mu)\|x^{(t)} - x^{(t+1)}\|^2 \\ &\leq \frac{1}{2\tau}(1 - \tau\|A^\top W^{(t)} A\|)\|x^{(t)} - x^{(t+1)}\|^2 \qquad (31) \\ &\leq \frac{1}{2\tau}(x^{(t)} - x^{(t+1)})^\top (I - \tau A^\top W^{(t)} A)(x^{(t)} - x^{(t+1)}). \end{aligned}$$

If we take the sum until T, then

$$\begin{aligned} 0 &\leq \sum_{t=1}^{T} \frac{1}{2\tau}(x^{(t)} - x^{(t+1)})^\top (I - \tau A^\top W^{(t)} A)(x^{(t)} - x^{(t+1)}) \\ &= \sum_{t=1}^{T} \left[V^{\mathsf{S}}(x^{(t+1)}, x^{(t)}, \pi^{(t)}) - V(x^{(t+1)}, \pi^{(t)}) \right] \qquad (32) \end{aligned}$$

Since $\pi^{(t+1)} = \operatorname*{argmin}_{\pi} V(x^{(t+1)}, \pi)$ then we have $V(x^{(t+1)}, \pi^{(t+1)}) \leq V(x^{(t+1)}, \pi^{(t)})$ and, combining with (31) and (32), we get

$$\begin{aligned} 0 &\leq \frac{1}{2\tau}(1-\tau\mu)\|x^{(t)} - x^{(t+1)}\|^2 \\ &\leq \sum_{t=1}^{T} \left[V^{\mathsf{S}}(x^{(t+1)}, x^{(t)}, \pi^{(t)}) - V(x^{(t+1)}, \pi^{(t+1)}) \right] \\ &\leq \sum_{t=1}^{T} \left[V^{\mathsf{S}}(x^{(t)}, x^{(t)}, \pi^{(t)}) - V(x^{(t+1)}, \pi^{(t+1)}) \right] \\ &= \sum_{t=1}^{T} \left[V(x^{(t)}, \pi^{(t)}) - V(x^{(t+1)}, \pi^{(t+1)}) \right] \end{aligned}$$

where the last inequality follows from the fact $x^{(t+1)} = \operatorname*{argmin}_x V^{\mathsf{S}}(x, x^{(t)}, \pi^{(t)})$. Finally, we observe that the truncated series is telescopic, from which

$$\begin{aligned} 0 &\leq \frac{1}{2\tau}(1-\tau\mu)\|x^{(t)} - x^{(t+1)}\|^2 \\ &\leq V(x^{(1)}, \pi^{(1)}) - V(x^{(T+1)}, \pi^{(T+1)}) \\ &\leq V(x^{(1)}, \pi^{(1)}) - \lambda N(\log \max\{\frac{p}{\beta}, \frac{1-p}{\alpha}\} - \log 2) = C' \end{aligned}$$

This last inequality holds for any $T \in \mathbb{N}$, then by letting $T \to \infty$, we obtain that the series is convergent, from which we deduce that as $t \to \infty$

$$\frac{1}{2\tau}(x^{(t)} - x^{(t+1)})^\top (I - \tau A^\top W^{(t)} A)(x^{(t)} - x^{(t+1)}) \to 0$$

and by inequality (31) the claim is proved. ■

Lemma 14: The sequence $(x^{(t)})_{t \in \mathbb{N}}$ is bounded.

Proof: Since $\mathbb{1}^\top x^{(t)} = 0$ for all t, Lemma 11 implies

$$\begin{aligned} &\|x^{(t+1)}\| \\ &= \|(I - \tau A^\top W^{(t)} A)\big(I - \frac{1}{N}\mathbb{1}\mathbb{1}^\top\big)x^{(t)} + \tau A^\top W^{(t)} b\| \\ &\leq \|(I - \tau\big(A^\top W^{(t)} A) - \frac{1}{N}\mathbb{1}\mathbb{1}^\top\big)\|\|x^{(t)}\| + \|\tau A^\top W^{(t)} b\| \\ &\leq (1 - \tau\mu_2)\|x^{(t)}\| + \tau\gamma \end{aligned}$$

where $\mu_2 = \min_t \|A^\top W^{(t)} A - \frac{1}{N}\mathbb{1}\mathbb{1}^\top\| > 0$ and $\gamma = \max_t \|A^\top W^{(t)} b\|$ (notice that $W^{(t)}$ belongs to a finite set of matrices). We conclude that

$$\lim_{t\to\infty} \|x^{(t)}\| \leq \lim_{t\to\infty} (1-\tau\mu_2)^t \|x^{(0)}\| + \sum_{s=0}^{\infty} (1-\tau\mu_2)^s \gamma\tau$$

which in turn is no larger than $\frac{\gamma}{\mu_2}$. ■

By Lemma 14, the sequence $x^{(t)}$ is bounded and then there exists a subsequence $x^{(t_j)}$ such that $x^{(t_j)} \to x^\infty$, $\alpha^{(t_j)} \to \alpha^\infty$, $\beta^{(t_j)} \to \beta^\infty$, and $\pi^{(t_j)} \to \pi^\infty$. From Lemma 13, we get $\lim_{t\to\infty} x^{(t_j+1)} = \lim_{t\to\infty} x^{(t_j)} = x^\infty$. Since $\pi_e(x) = f(z_e = 1 | x, \alpha, \beta)$ is a continuous function of x, then also $\pi^{(t)} \to \pi^\infty$ and (x^∞, π^∞) is a fixed point.

References

[1] P. Barooah and J. P. Hespanha, "Estimation from relative measurements: Electrical analogy and large graphs," *IEEE Transactions on Signal Processing*, vol. 56, no. 6, pp. 2181–2193, 2008.

[2] ——, "Estimation from relative measurements: Algorithms and scaling laws," *IEEE Control Systems Magazine*, vol. 27, no. 4, pp. 57–74, 2007.

[3] X. Jiang, L.-H. Lim, Y. Yao, and Y. Ye, "Statistical ranking and combinatorial Hodge theory," *Mathematical Programming*, vol. 127, no. 1, pp. 203–244, 2011.

[4] B. Osting, C. Brune, and S. J. Osher, "Optimal data collection for improved rankings expose well-connected graphs," *Journal of Machine Learning Research*, vol. 15, pp. 2981–3012, 2014.

[5] A. P. Dempster, N. M. Laird, and D. B. Rubin, "Maximum likelihood from incomplete data via the EM algorithm," *Journal of the Royal Statistical Society. Series B*, vol. 39, no. 1, pp. 1–38, 1977.

[6] T. K. Moon, "The expectation-maximization algorithm," *IEEE Signal Processing Magazine*, vol. 13, no. 6, pp. 47–60, 1996.

[7] A. Giridhar and P. R. Kumar, "Distributed clock synchronization over wireless networks: Algorithms and analysis," in *IEEE Conference on Decision and Control*, San Diego, CA, USA, Dec. 2006, pp. 4915–4920.

[8] S. Bolognani, S. Del Favero, L. Schenato, and D. Varagnolo, "Consensus-based distributed sensor calibration and least-square parameter identification in WSNs," *International Journal of Robust and Nonlinear Control*, vol. 20, no. 2, pp. 176–193, 2010.

[9] W. S. Rossi, P. Frasca, and F. Fagnani, "Distributed estimation from relative and absolute measurements," *IEEE Transactions on Automatic Control*, vol. 62, no. 12, pp. 6385–6391, 2017.

[10] A. Carron, M. Todescato, R. Carli, and L. Schenato, "An asynchronous consensus-based algorithm for estimation from noisy relative measurements," *IEEE Transactions on Control of Network Systems*, vol. 1, no. 3, pp. 283–295, 2014.

[11] N. M. Freris and A. Zouzias, "Fast distributed smoothing of relative measurements," in *IEEE Conference on Decision and Control*, Maui, HI, USA, Dec. 2012, pp. 1411–1416.

[12] C. Ravazzi, P. Frasca, R. Tempo, and H. Ishii, "Ergodic randomized algorithms and dynamics over networks," *IEEE Transactions on Control of Network Systems*, vol. 2, no. 1, pp. 78–87, 2015.

[13] P. Frasca, H. Ishii, C. Ravazzi, and R. Tempo, "Distributed randomized algorithms for opinion formation, centrality computation and power systems estimation," *European Journal of Control*, vol. 24, pp. 2–13, 2015.

[14] M. Todescato, A. Carron, R. Carli, A. Franchi, and L. Schenato, "Multi-robot localization via GPS and relative measurements in the presence of asynchronous and lossy communication," in *2016 European Control Conference (ECC)*, June 2016, pp. 2527–2532.

[15] L. Carlone, A. Censi, and F. Dellaert, "Selecting good measurements via ℓ_1 relaxation: A convex approach for robust estimation over graphs," in *IEEE/RSJ Int. Conf. on Intelligent Robots & Systems*, Sep. 2014, pp. 2667–2674.

[16] N. Bof, M. Todescato, R. Carli, and L. Schenato, "Robust distributed estimation for localization in lossy sensor networks," in *6th IFAC Workshop on Distributed Estimation and control in Networked Systems (NecSys16)*. IFAC, 2016, pp. 250–255.

[17] 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.

[18] F. Fagnani, S. M. Fosson, and C. Ravazzi, "A distributed classification/estimation algorithm for sensor networks," *SIAM Journal on Control and Optimization*, vol. 52, no. 1, pp. 189–218, 2014.

[19] F. Fagnani, S. Fosson, and C. Ravazzi, "Consensus-like algorithms for estimation of Gaussian mixtures over large scale networks," *Mathematical Models and Methods in Applied Sciences,*, pp. 1–21, 2014.

[20] V. Kekatos and G. Giannakis, "Distributed robust power system state estimation," *IEEE Transactions on Power Systems*, vol. 28, no. 2, pp. 1617–1626, 2013.

[21] R. D. 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.

[22] M. Rabbat and R. Nowak, "Distributed optimization in sensor networks," in *Third International Symposium on Information Processing in Sensor Networks, 2004. IPSN 2004*, April 2004, pp. 20–27.

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

[24] D. Wang, L. Kaplan, and T. F. Abdelzaher, "Maximum likelihood analysis of conflicting observations in social sensing," *ACM Transactions on Sensor Networks*, vol. 10, no. 2, p. 30, 2014.

[25] D. Wang, T. F. Abdelzaher, and L. Kaplan, *Social Sensing: Building Reliable Systems on Unreliable Data*. Morgan Kaufmann, 2015.

[26] C. M. Bishop, *Pattern Recognition and Machine Learning*. Springer, 2006.

[27] M. R. Gupta and Y. Chen, "Theory and use of the EM algorithm," *Foundations and Trends® in Signal Processing*, vol. 4, no. 3, pp. 223–296, 2011.

[28] F. Fagnani and P. Frasca, *Introduction to Averaging Dynamics over Networks*, ser. Lecture Notes in Control and Information Sciences. Springer Nature, 2017.

[29] W. S. Rossi, P. Frasca, and F. Fagnani, "Transient and limit performance of distributed relative localization," in *IEEE Conference on Decision and Control*, Maui, HI, USA, Dec. 2012, pp. 2744–2748.

[30] P. Bloomfield and W. Steiger, *Least Absolute Deviations: Theory, Applications, and Algorithms*. Springer, 2012.

[31] E. Candés and T. Tao, "The Dantzig selector: Statistical estimation when p is much larger than n," *The Annals of Statistics*, vol. 35, no. 6, pp. 2313–2351, 2007.

[32] S. Boyd and L. Vandenberghe, *Convex Optimization*. New York, NY, USA: Cambridge University Press, 2004.

[33] I. Daubechies, R. DeVore, M. Fornasier, and C. S. Güntürk, "Iteratively reweighted least squares minimization for sparse recovery," *Communications on Pure and Applied Mathematics*, vol. 63, no. 1, pp. 1–38, 2010.

[34] P. Barooah and J. P. Hespanha, "Error scaling laws for linear optimal estimation from relative measurements," *IEEE Transactions on Information Theory*, vol. 55, no. 12, pp. 5661–5673, 2009.

[35] W. S. Rossi, P. Frasca, and F. Fagnani, "Limited benefit of cooperation in distributed relative localization," in *IEEE Conference on Decision and Control*, Florence, Italy, Dec. 2013, pp. 5427–5431.

[36] D. E. Ba, B. Babadi, P. L. Purdon, and E. N. Brown, "Convergence and stability of iteratively re-weighted least squares algorithms," *IEEE Transactions on Signal Processing*, vol. 62, no. 1, pp. 183–195, 2014.

[37] C. Ravazzi and E. Magli, "Gaussian mixtures based IRLS for sparse recovery with quadratic convergence," *IEEE Transactions on Signal Processing*, vol. 63, no. 13, pp. 3474–3489, 2015.

[38] N. P. Chan, "On analysis and design of algorithms for robust estimation from relative measurements," Master's thesis, University of Twente, Enschede, the Netherlands, Mar. 2016. [Online]. Available: http://essay.utwente.nl/69425/

Chiara Ravazzi (M'13) received the Ph.D. degree in Mathematics for Engineering Sciences from Politecnico di Torino, Turin, Italy, in 2011. From 2011 to 2016, she held post-doctoral positions with the Departments of Mathematics and of Electronics and Telecommunications, Politecnico di Torino. Since 2017, she has been a Tenured Researcher with the CNR-IEIIT, working in the Systems Modeling & Control Group. Her research interests include signal processing, optimization and control of network systems.

Nelson P. K. Chan (S'18) obtained the BSc. degree in Mechanical Engineering from the Anton de Kom University of Suriname in Paramaribo, Suriname in 2013, the MSc. degree in Systems & Control from the University of Twente in Enschede, NL in 2016. Currently, he is working towards his Ph.D. degree within the Faculty of Science and Engineering at the University of Groningen, NL. His research interests cover control of multi-agent systems with a focus on the safety aspect.

Paolo Frasca (M'13, SM'18) received the Ph.D. degree from Politecnico di Torino, Turin, Italy, in 2009. After post-doctoral appointments at the CNR-IAC, Rome and in Torino, he has been an Assistant Professor at the University of Twente in Enschede, Netherlands, from 2013 to 2016. Since October 2016, he has been a CNRS Researcher with GIPSA-lab, Grenoble, France. His research interests are in the theory of network systems and cyber-physical systems, with applications to robotic, sensor, infrastructural, and social networks. On these topics, he has (co)authored more than sixty journal and conference papers and the book *Introduction to Averaging Dynamics over Networks*, published by Springer. Dr. Frasca has served or is serving as Associate Editor in the Editorial Boards of several conferences and journals, including the IEEE CONTROL SYSTEMS LETTERS.