DISTRIBUTED TLS ESTIMATION UNDER RANDOM DATA FAULTS

*Silvana Silva Pereira**, *Alba Pagès-Zamora** *and Roberto López-Valcarce*†

*SPCOM Group, Universitat Politècnica de Catalunya-Barcelona Tech, Spain
†GPSC, Universidade de Vigo, Spain

ABSTRACT

This paper addresses the problem of distributed estimation of a parameter vector in the presence of noisy input and output data as well as data faults, performed by a wireless sensor network in which only local interactions among the nodes are allowed. In the presence of unreliable observations, standard estimators become biased and perform poorly in low signal-to-noise ratios. We propose two different distributed approaches based on the Expectation-Maximization algorithm: in the first one the regressors are estimated at each iteration, whereas the second one does not require explicit regressor estimation. Numerical results show that the proposed methods approach the performance of a clairvoyant scheme with knowledge of the random data faults.

Index Terms— Diffusion, distributed estimation, expectation-maximization, sensor networks, total least squares.

1. INTRODUCTION

We study the problem of distributed estimation of an unknown parameter vector when both the observations (*output data*) and the regressors (*input data*) are assumed noisy. The estimation is performed by a wireless sensor network (WSN) where the nodes are allowed to communicate within a small neighborhood only, and where some nodes may be subject to random transducer faults, in which case, the sensor is assumed to observe only noise [1, 2]. When dealing with noisy input and output data, a good alternative is the Total Least Squares (TLS) solution [3, 4], for which distributed implementations have been reported in the literature [5–7]. However, in the presence of data faults in the output data, the standard TLS estimation becomes biased, with the consequent penalty in the Mean Square Error (MSE). Treating the random data faults as hidden random variables, we derive two distributed estimators based on the Expectation-Maximization (EM) algorithm, a numerical method to compute the maximum-likelihood (ML) estimator in the presence of incomplete observations [9, 10]. As in previous contributions, the information is spread across the network by means of diffusion strategies [11], where a slower term for information diffusion is combined with a faster term for information averaging [7, 8]. The estimation is performed based on a single data snapshot collected by each node, i.e., there is no new data streaming in. The two implementations proposed differ in that, in the first one, the regressors are estimated at each iteration of the EM algorithm, whereas in the second they are not.

This work has been partially funded by the Spanish "Ministerio de Economía y Competitividad", ERDF funds (TEC2013-41315-R, TEC2013-47020-C2-1-R, TACTICA), Catalan Government (2014 SGR 60 AGAUR), and the Galician Government (CN 2012/260 AtlantTIC, GRC2013/009).

2. PROBLEM STATEMENT

Consider the problem of estimating a parameter vector $\boldsymbol{x} \in \mathbb{R}^L$ based on a set of noisy observations

$$y_i = a_i \boldsymbol{h}_i^T \boldsymbol{x} + w_i, \tag{1}$$

$$\boldsymbol{z}_i = \boldsymbol{h}_i + \boldsymbol{v}_i \tag{2}$$

where $w_i \sim \mathcal{N}(0, \sigma^2)$ and $\boldsymbol{v}_i \sim \mathcal{N}(\boldsymbol{0}, r^2\sigma^2 \boldsymbol{I}_L)$ for all $i = 1, \dots, N$. The random variables $\{a_i\}$ are i.i.d. Bernoulli distributed with $\Pr(a_i = 1) = p$, and reflect whether a sensor has been subjected to a data fault: $a_i = 1$ shows that the i-th sensor correctly acquired the corresponding observation, whereas $a_i = 0$ shows that only noise was sensed. We assume that $\{a_i, w_j, \boldsymbol{v}_k\}$ are statistically independent for all $\{i, j, k\}$. Note that in case of a data fault at the i-th sensor, only (1) is affected, but not (2). The regressors $\boldsymbol{h}_i$ are only observed through the noisy estimates $\boldsymbol{z}_i$, and the ratio r^2 of the regressor noise variance to the output noise variance is assumed known[1]. Whereas $\boldsymbol{x}$ is the parameter of interest, p, σ^2 and $\{\boldsymbol{h}_i\}$ are regarded as deterministic, unknown *nuisance* parameters. Let us define $\boldsymbol{y} \triangleq [y_1 \cdots y_N]^T$, $\boldsymbol{H} \triangleq [\boldsymbol{h}_1^T; \cdots; \boldsymbol{h}_N^T]$, $\boldsymbol{Z} \triangleq [\boldsymbol{z}_1^T \cdots \boldsymbol{z}_N^T]$ and $\boldsymbol{A} \triangleq \mathrm{diag}(\boldsymbol{a})$ with $\boldsymbol{a} \triangleq [a_1 \cdots a_N]^T$. When knowledge of $\boldsymbol{A}$ is available, the *clairvoyant* Ordinary Least Squares (CV-OLS) estimator can be computed as

$$\hat{\boldsymbol{x}}_{\text{CV-OLS}} = \arg\min_{\boldsymbol{x}} \|\boldsymbol{y} - \boldsymbol{A}\boldsymbol{Z}\boldsymbol{x}\|^2 = (\boldsymbol{Z}^T\boldsymbol{A}\boldsymbol{Z})^{-1}\boldsymbol{Z}^T\boldsymbol{A}\boldsymbol{y}. \tag{3}$$

This is the ML estimator of $\boldsymbol{x}$ when $r^2 = 0$. However, this estimator is biased if the input data is noisy ($r^2 > 0$). For known $\boldsymbol{A}$, the ML estimators of $\boldsymbol{x}$ and $\{\boldsymbol{h}_i\}$ in (1)-(2) are the solutions to

$$\min_{\boldsymbol{x}, \boldsymbol{H}} \; r^2\|\boldsymbol{y} - \boldsymbol{A}\boldsymbol{H}\boldsymbol{x}\|^2 + \|\boldsymbol{Z} - \boldsymbol{H}\|_F^2 \tag{4}$$

[1] Introducing different unknown variances would result in an overparameterized problem, yielding lack of identifiability.

where $\|\cdot\|_F$ denotes the Frobenius norm. Problem (4) is a generalized TLS problem [4], and its solution (clairvoyant TLS) is given by

$$\begin{bmatrix} \hat{\boldsymbol{x}}_{\text{CV-TLS}} \\ -1/r \end{bmatrix} = \text{least eigenvector of} \begin{bmatrix} \boldsymbol{Z}^T \\ r\boldsymbol{y}^T \end{bmatrix} \boldsymbol{A} \begin{bmatrix} \boldsymbol{Z} & r\boldsymbol{y} \end{bmatrix}. \quad (5)$$

If one assumes $\boldsymbol{A} = \boldsymbol{I}$ in (4), the standard TLS solution $\hat{\boldsymbol{x}}_{\text{TLS}}$ is obtained. However, $\hat{\boldsymbol{x}}_{\text{TLS}}$ becomes biased when data faults are present. Therefore, in the sequel we discuss a means to compute the ML estimator for the problem (1)-(2) when the $\{a_i\}$ are *unknown*. Two solutions based on the EM algorithm are derived, which differ in their treatment of the regressor matrix $\boldsymbol{H}$.

3. ML ESTIMATION VIA THE EM ALGORITHM

Let $\boldsymbol{\theta} = [\boldsymbol{x}^T \quad \boldsymbol{h}_1^T \cdots \boldsymbol{h}_N^T \quad \sigma^2 \quad p]^T$. Due to the independence of the data, the likelihood function of $\boldsymbol{\theta}$ is given by

$$f(\boldsymbol{y}, \boldsymbol{Z} \,|\, \boldsymbol{\theta}) = \prod_{i=1}^{N} f(y_i \,|\, \boldsymbol{\theta}) \cdot \prod_{j=1}^{N} f(\boldsymbol{z}_j \,|\, \boldsymbol{\theta}) \quad (6)$$

where

$$f(y_i \,|\, \boldsymbol{\theta}) = \frac{1}{\sqrt{2\pi\sigma^2}} \left[p\, \mathrm{e}^{-\frac{(y_i - \boldsymbol{h}_i^T \boldsymbol{x})^2}{2\sigma^2}} + (1-p)\mathrm{e}^{-\frac{y_i^2}{2\sigma^2}} \right] \quad \text{(7a)}$$

$$f(\boldsymbol{z}_j \,|\, \boldsymbol{\theta}) = \frac{1}{(2\pi r^2 \sigma^2)^{\frac{L}{2}}} \mathrm{e}^{-\frac{\|\boldsymbol{z}_j - \boldsymbol{h}_j\|^2}{2r^2\sigma^2}} \quad \text{(7b)}$$

Maximizing (6) w.r.t. $\boldsymbol{\theta}$ in closed form is not possible, and one has to resort to numerical methods. The EM algorithm is particularly well suited to problems like the one at hand in which hidden random variables (the $\{a_i\}$ in this case) are present. We denote $\{\boldsymbol{y}, \boldsymbol{Z}\}$ as the *incomplete* data set and $\{\boldsymbol{y}, \boldsymbol{Z}, \boldsymbol{a}\}$ as the *complete* data set. Assume for the moment a centralized approach in which a single node has access to the set $\{\boldsymbol{y}, \boldsymbol{Z}\}$. Then, at iteration t of the EM algorithm one performs the following:

1. *E-step:* given an estimate $\hat{\boldsymbol{\theta}}_t$ and a trial value $\tilde{\boldsymbol{\theta}}$ of $\boldsymbol{\theta}$, compute the conditional expectation
$$Q(\tilde{\boldsymbol{\theta}}\,;\, \hat{\boldsymbol{\theta}}_t) = \mathbb{E}_{\boldsymbol{a}} \left\{ \log f(\boldsymbol{y}, \boldsymbol{Z}, \boldsymbol{a} \,|\, \tilde{\boldsymbol{\theta}}) \;\middle|\; \hat{\boldsymbol{\theta}}_t, \boldsymbol{y}, \boldsymbol{Z} \right\}. \quad (8)$$
2. *M-step:* obtain the estimate for the next iteration as
$$\hat{\boldsymbol{\theta}}_{t+1} = \arg\max_{\tilde{\boldsymbol{\theta}}} Q(\tilde{\boldsymbol{\theta}}\,;\, \hat{\boldsymbol{\theta}}_t). \quad (9)$$

Since $\boldsymbol{Z}$ and $\boldsymbol{a}$ are statistically independent we have

$$f(\boldsymbol{y}, \boldsymbol{Z}, \boldsymbol{a} \,|\, \tilde{\boldsymbol{\theta}}) = f(\boldsymbol{y} \,|\, \tilde{\boldsymbol{\theta}}, \boldsymbol{a}) \cdot f(\boldsymbol{Z} \,|\, \tilde{\boldsymbol{\theta}}) \cdot f(\boldsymbol{a} \,|\, \tilde{\boldsymbol{\theta}}). \quad (10)$$

After some algebra we get

$$f(\boldsymbol{y}, \boldsymbol{Z}, \boldsymbol{a} \,|\, \tilde{\boldsymbol{\theta}}) = \frac{1}{(2\pi r^2 \tilde{\sigma}^2)^{\frac{NL}{2}}} \cdot \mathrm{e}^{-\frac{1}{2r^2\tilde{\sigma}^2}\left(\|\boldsymbol{Z} - \boldsymbol{H}\|_F^2\right)} \cdot \frac{1}{(2\pi\tilde{\sigma}^2)^{\frac{N}{2}}} \cdot \mathrm{e}^{-\frac{\|\boldsymbol{y} - \boldsymbol{A}\tilde{\boldsymbol{H}}\tilde{\boldsymbol{x}}\|^2}{2\tilde{\sigma}^2}} \cdot \prod_{i=1}^{N} \tilde{p}^{\,a_i} \,(1-\tilde{p})^{1-a_i} \quad (11)$$

and taking the logarithm yields

$$\log f(\boldsymbol{y}, \boldsymbol{Z}, \boldsymbol{a} \,|\, \tilde{\boldsymbol{\theta}}) \propto -\frac{N(L+1)}{2} \log \tilde{\sigma}^2 - \frac{1}{2r^2\tilde{\sigma}^2} \|\boldsymbol{Z} - \tilde{\boldsymbol{H}}\|_F^2 - \frac{1}{2\tilde{\sigma}^2} \left(\|\boldsymbol{y} - \boldsymbol{A}\tilde{\boldsymbol{H}}\tilde{\boldsymbol{x}}\|^2 \right) + \sum_{i=1}^{N} [a_i \log \tilde{p} + (1-a_i)\log(1-\tilde{p})].$$

Further, let $\hat{a}_{i,t} = \mathbb{E}_{\boldsymbol{a}}[a_i \,|\, \hat{\boldsymbol{\theta}}_t, \boldsymbol{y}, \boldsymbol{Z}]$ and $\hat{\boldsymbol{A}}_t = \mathrm{diag}(\hat{\boldsymbol{a}}_t)$ with $\hat{\boldsymbol{a}}_t \triangleq [\hat{a}_1 \cdots \hat{a}_N]^T$. Then, $Q(\tilde{\boldsymbol{\theta}}\,;\, \hat{\boldsymbol{\theta}}_t)$ can be expressed as

$$\begin{aligned} Q(\tilde{\boldsymbol{\theta}}\,;\, \hat{\boldsymbol{\theta}}_t) \propto & -\frac{N(L+1)}{2} \log \tilde{\sigma}^2 - \frac{1}{2r^2\tilde{\sigma}^2} \Big(\|\boldsymbol{Z} - \tilde{\boldsymbol{H}}\|_F^2 \\ & + r^2(\boldsymbol{y} - \tilde{\boldsymbol{H}}\tilde{\boldsymbol{x}})^T \hat{\boldsymbol{A}}_t (\boldsymbol{y} - \tilde{\boldsymbol{H}}\tilde{\boldsymbol{x}}) + r^2 \boldsymbol{y}^T(\boldsymbol{I} - \hat{\boldsymbol{A}}_t)\boldsymbol{y} \Big) \\ & + \hat{S}_t \log \tilde{p} + (N - \hat{S}_t)\log(1-\tilde{p}) \end{aligned} \quad (12)$$

where $\hat{S}_t = \sum_{i=1}^{N} \hat{a}_{i,t}$. Note that, since a_i is Bernoulli, one has $\hat{a}_{i,t} = \Pr(a_i = 1 | \hat{\boldsymbol{\theta}}_t, \boldsymbol{y}, \boldsymbol{Z})$. Making use of Bayes' theorem, these a posteriori probabilities can be obtained as follows

$$\hat{a}_{i,t} = \frac{\hat{p}_t \rho_{i,t}^1}{(1-\hat{p}_t)\rho_{i,t}^0 + \hat{p}_t \rho_{i,t}^1} \quad (13)$$

with

$$\rho_{i,t}^a \triangleq \exp\left(-\frac{(y_i - a\hat{\boldsymbol{h}}_{i,t}^T \hat{\boldsymbol{x}}_t)^2}{2\hat{\sigma}_t^2} \right), \quad a \in \{0, 1\}. \quad (14)$$

The new set of estimates is found by maximizing (12) w.r.t. $\tilde{\boldsymbol{x}}, \tilde{\boldsymbol{h}}, \tilde{\sigma}^2$ and $\tilde{p}$. Maximizing first (12) w.r.t. $\tilde{p}$ yields

$$\hat{p}_{t+1} = \tfrac{1}{N}\hat{S}_t. \quad (15)$$

For the remaining parameters, two different ways can be followed at this point: either estimating $\{\boldsymbol{h}_i\}$ at every iteration, or replacing its expression in the log-likelihood function (LLF), as shown below.

Cyclic EM – Estimating $\boldsymbol{H}$: The estimates for $\hat{\boldsymbol{x}}_{t+1}$ and $\hat{\boldsymbol{H}}_{t+1}$ can be found solving

$$\min_{\tilde{\boldsymbol{x}}, \tilde{\boldsymbol{H}}} \|\boldsymbol{Z} - \tilde{\boldsymbol{H}}\|_F^2 + r^2(\boldsymbol{y} - \tilde{\boldsymbol{H}}\tilde{\boldsymbol{x}})^T \hat{\boldsymbol{A}}_t (\boldsymbol{y} - \tilde{\boldsymbol{H}}\tilde{\boldsymbol{x}}). \quad (16)$$

Although (16) resembles a TLS problem, the presence of the diagonal weighting matrix $\hat{\boldsymbol{A}}_t$ precludes a closed-form expression for the solution. Equating to zero the gradients of (16) w.r.t. $\tilde{\boldsymbol{x}}$ and $\tilde{\boldsymbol{H}}$, and evaluating the resulting conditions at $\tilde{\boldsymbol{x}} = \hat{\boldsymbol{x}}_{t+1}$ and $\tilde{\boldsymbol{H}} = \hat{\boldsymbol{H}}_{t+1}$, one respectively obtains

$$\hat{\boldsymbol{H}}_{t+1}^T \hat{\boldsymbol{A}}_t \hat{\boldsymbol{H}}_{t+1} \hat{\boldsymbol{x}}_{t+1} = \hat{\boldsymbol{H}}_{t+1}^T \hat{\boldsymbol{A}}_t \boldsymbol{y}, \quad (17)$$

$$\hat{\boldsymbol{H}}_{t+1} + r^2 \hat{\boldsymbol{A}}_t \hat{\boldsymbol{H}}_{t+1} \hat{\boldsymbol{x}}_{t+1} \hat{\boldsymbol{x}}_{t+1}^T = \boldsymbol{Z} + r^2 \hat{\boldsymbol{A}}_t \boldsymbol{y} \hat{\boldsymbol{x}}_{t+1}^T. \quad (18)$$

From (17), if $\hat{\boldsymbol{H}}_{t+1}$ was available, then $\hat{\boldsymbol{x}}_{t+1}$ could be readily obtained by solving an $L \times L$ linear system. Conversely, if $\hat{\boldsymbol{x}}_{t+1}$ was available, then $\hat{\boldsymbol{H}}_{t+1}$ could also be explicitly obtained, as the following result shows:

Lemma 1. *The solution $\hat{\boldsymbol{H}}_{t+1}$ to* (18) *is a rank-1 perturbation of $\boldsymbol{Z}$ given by $\hat{\boldsymbol{H}}_{t+1} = \boldsymbol{Z} + \boldsymbol{u}_{t+1}\hat{\boldsymbol{x}}_{t+1}^T$, with*

$$\boldsymbol{u}_{t+1} \triangleq r^2 \left(\boldsymbol{I} + r^2\|\hat{\boldsymbol{x}}_{t+1}\|^2\hat{\boldsymbol{A}}_t\right)^{-1} \hat{\boldsymbol{A}}_t\left(\boldsymbol{y} - \boldsymbol{Z}\hat{\boldsymbol{x}}_{t+1}\right). \quad (19)$$

The proof follows immediately by substituting $\hat{\boldsymbol{H}}_{t+1} = \boldsymbol{Z} + \boldsymbol{u}_{t+1}\hat{\boldsymbol{x}}_{t+1}^T$ in the left-hand side of (18). Note that $\boldsymbol{I} + r^2\|\hat{\boldsymbol{x}}_{t+1}\|^2\hat{\boldsymbol{A}}_t$ is a diagonal matrix, and hence its inversion is computationally cheap. In order to obtain $\hat{\boldsymbol{x}}_{t+1}$ and $\hat{\boldsymbol{H}}_{t+1}$, we propose a *cyclic-minimization* (CM) procedure [13]: the expression in (16) is minimized w.r.t. one of the variables assuming that the other is fixed, and the process is repeated until convergence. Starting with some initial $\hat{\boldsymbol{x}}^{(0)}$, and for $k = 1, \ldots, K$, compute

$$\hat{\boldsymbol{H}}^{(k)} = \boldsymbol{Z} + r^2\left(\boldsymbol{I} + r^2\left\|\hat{\boldsymbol{x}}^{(k-1)}\right\|^2\hat{\boldsymbol{A}}_t\right)^{-1} \times \hat{\boldsymbol{A}}_t\left(\boldsymbol{y} - \boldsymbol{Z}\hat{\boldsymbol{x}}^{(k-1)}\right)\left(\hat{\boldsymbol{x}}^{(k-1)}\right)^T, \quad (20)$$

and then solve for $\hat{\boldsymbol{x}}^{(k)}$ in

$$\left(\hat{\boldsymbol{H}}^{(k)}\right)^T\hat{\boldsymbol{A}}_t\hat{\boldsymbol{H}}^{(k)}\hat{\boldsymbol{x}}^{(k)} = \left(\hat{\boldsymbol{H}}^{(k)}\right)^T\hat{\boldsymbol{A}}_t\boldsymbol{y}. \quad (21)$$

We then take $\hat{\boldsymbol{x}}_{t+1} = \hat{\boldsymbol{x}}^{(K)}$ and $\hat{\boldsymbol{H}}_{t+1} = \hat{\boldsymbol{H}}^{(K)}$. The estimate for the variance is found in terms of those of $\boldsymbol{x}$ and $\{\boldsymbol{h}_i\}$:

$$\hat{\sigma}_{t+1}^2 = \frac{r^2(\boldsymbol{y}^T\boldsymbol{y} - \hat{\boldsymbol{\psi}}_t^T\hat{\boldsymbol{x}}_{t+1}) + \|\boldsymbol{Z} - \hat{\boldsymbol{H}}_{t+1}\|_F^2}{r^2N(L+1)} \quad (22)$$

where $\hat{\boldsymbol{\psi}}_t = \hat{\boldsymbol{H}}_{t+1}^T\hat{\boldsymbol{A}}_t\boldsymbol{y}$. We propose to embed the CM iteration just described within the EM iterative procedure, yielding the modification of the centralized EM algorithm (CEM). In essence, at each outer EM iteration, only one inner CM iteration is performed, i.e., we take $\hat{\boldsymbol{x}}_t^{(0)} = \hat{\boldsymbol{x}}_t$ and $K = 1$. Although the parameter trajectories of this modified version do not necessarily coincide with those of the true EM algorithm, the sets of fixed points of both schemes are the same.

Blind EM – Substituting $\boldsymbol{H}$ in the LLF: Next we propose an alternative approach in which explicit estimation of $\boldsymbol{h}_i$ is not needed. To this end, first we show how to obtain the scalar products $\hat{\boldsymbol{h}}_{i,t}^T\hat{\boldsymbol{x}}_t$ featuring in (14). From Lemma 1, it follows that $\hat{\boldsymbol{H}}_t = \boldsymbol{Z} + \boldsymbol{u}_t\hat{\boldsymbol{x}}_t^T$. Therefore, after some algebra, one finds

$$\hat{\boldsymbol{H}}_t\hat{\boldsymbol{x}}_t = \boldsymbol{y} - (\boldsymbol{I} + r^2\|\hat{\boldsymbol{x}}_t\|^2\hat{\boldsymbol{A}}_{t-1})^{-1}(\boldsymbol{y} - \boldsymbol{Z}\hat{\boldsymbol{x}}_t). \quad (23)$$

Using this in (14), one obtains

$$\rho_{i,t}^a = \exp\left(-\frac{\left(y_i(1-a) + \frac{a(y_i - \boldsymbol{z}_i^T\hat{\boldsymbol{x}}_t)}{1 + r^2\|\hat{\boldsymbol{x}}_t\|^2\hat{a}_{i,t-1}}\right)^2}{2\hat{\sigma}_t^2}\right). \quad (24)$$

Now note that using again Lemma 1, the cost in (16) can be minimized w.r.t. $\tilde{\boldsymbol{H}}$. Substituting the optimum value of $\tilde{\boldsymbol{H}}$ (which depends on $\tilde{\boldsymbol{x}}$) in (16) results in the following problem after some algebraic manipulations:

$$\hat{\boldsymbol{x}}_{t+1} = \arg\min_{\tilde{\boldsymbol{x}}}\left(\boldsymbol{y} - \boldsymbol{Z}\tilde{\boldsymbol{x}}\right)^T\hat{\boldsymbol{D}}_t(\tilde{\boldsymbol{x}})\left(\boldsymbol{y} - \boldsymbol{Z}\tilde{\boldsymbol{x}}\right) \quad (25)$$

where $\hat{\boldsymbol{D}}_t(\tilde{\boldsymbol{x}}) = (\boldsymbol{I} + r^2\|\tilde{\boldsymbol{x}}\|^2\hat{\boldsymbol{A}}_t)^{-1}\hat{\boldsymbol{A}}_t$. Note that the cost in (25) is not quadratic due to the fact that the diagonal weighting matrix $\hat{\boldsymbol{D}}_t(\tilde{\boldsymbol{x}})$ depends on $\tilde{\boldsymbol{x}}$. We propose to replace $\hat{\boldsymbol{D}}_t \triangleq \hat{\boldsymbol{D}}_t(\hat{\boldsymbol{x}}_t) \approx \hat{\boldsymbol{D}}_t(\tilde{\boldsymbol{x}})$ in (25) in order to obtain a quadratic problem (weighted least squares), whose solution is

$$\hat{\boldsymbol{x}}_{t+1} = (\boldsymbol{Z}^T\hat{\boldsymbol{D}}_t\boldsymbol{Z})^{-1}\boldsymbol{Z}^T\hat{\boldsymbol{D}}_t\boldsymbol{y}. \quad (26)$$

The estimate of p is as before and given by (15), whereas the new estimate of σ^2 is computed as follows

$$\hat{\sigma}_{t+1}^2 = \frac{(\boldsymbol{y} - \boldsymbol{Z}\hat{\boldsymbol{x}}_{t+1})^T\hat{\boldsymbol{D}}_t(\boldsymbol{y} - \boldsymbol{Z}\hat{\boldsymbol{x}}_{t+1}) + \boldsymbol{y}^T(\boldsymbol{I} - \hat{\boldsymbol{A}}_t)\boldsymbol{y}}{N(L+1)} \quad (27)$$

We refer to this scheme as the blind EM (BEM) estimator. Inspired by previous distributed approaches [2, 5–7, 11], we develop distributed implementations of the centralized schemes presented in this section, which are based on the diffusion principle of [8]. The key observation is the fact that the global quantities featuring in the centralized versions can be written as summations over local quantities.

4. DISTRIBUTED SOLUTIONS

Assume that each node $i = 1, \cdots, N$ only has access to its own measurements $\{y_i, \boldsymbol{z}_i\}$ and can only communicate with a small subset of neighbors. At each node i and at time k,[2] a local copy of the parameters $\{\hat{\boldsymbol{x}}_{i,k},\ \hat{\sigma}_{i,k}^2,\ \hat{p}_{i,k}\}$ (and also of $\hat{\boldsymbol{h}}_{i,k}^T$ for the CEM) is updated in terms of the information gathered from neighboring nodes. Let $\boldsymbol{W} \in \mathbb{R}^{N\times N}$ denote a *weight matrix* with a nonzero $\{ij\}^{th}$ entry W_{ij} only if nodes i and j can communicate with each other. We assume that the network is connected, i.e., there is a path between any pair of nodes $\{i, j\}$. $\boldsymbol{W}$ is assumed symmetric and satisfies $\boldsymbol{W}\boldsymbol{1} = \boldsymbol{1}$, $\rho(\boldsymbol{W} - \frac{\boldsymbol{1}\boldsymbol{1}^T}{N}) < 1$, where $\boldsymbol{1}$ is an all-ones vector and $\rho(\cdot)$ denotes spectral radius. The general steps of the distributed EM estimator are summarized in Table 1, whereas the intermediate variables are specified below for each method.

Distributed (D)-CEM: For the D-CEM approach, the a posteriori probability $\hat{a}_{i,k}$ at node i and at time k in (28) is computed using

$$\rho_{i,k}^a = \exp\left(-\frac{(y_i - a\hat{\boldsymbol{h}}_{i,k}^T\hat{\boldsymbol{x}}_{i,k})^2}{2\hat{\sigma}_{i,k}^2}\right), \quad a \in \{0, 1\} \quad (33)$$

and the intermediate variables are given by

$$\boldsymbol{F}(j,k) = \hat{a}_{j,k}\hat{\boldsymbol{h}}_{j,k}\hat{\boldsymbol{h}}_{j,k}^T \quad (34)$$

$$\boldsymbol{f}(j,k) = \hat{a}_{j,k}y_j\hat{\boldsymbol{h}}_{j,k} \quad (35)$$

$$f_\sigma(j,k) = r^2(y_j^2 - y_j\hat{a}_{j,k}\hat{\boldsymbol{h}}_{j,k}^T\hat{\boldsymbol{x}}_{j,k+1}) + \|\boldsymbol{z}_j - \hat{\boldsymbol{h}}_{j,k}\|^2 \quad (36)$$

[2]We denote by k (rather than t) the iteration index for the distributed approaches, in order to emphasize the difference w.r.t. the centralized schemes.

For $i = 1, \cdots, N$

1. Initialize $\hat{a}_{i,0}$ and the local estimates $\hat{\boldsymbol{\theta}}_{i,1}$. Initialize the intermediate variables $\boldsymbol{F}(j,k)$, $\boldsymbol{f}(j,k)$ and $f_\nu(j,k)$, $\forall\, \nu \in \{\sigma^2, a, 1\}$.

For $k \geqslant 1$,

2. *E-Step*: given $\hat{\boldsymbol{\theta}}_{i,k}$ compute

$$\hat{a}_{i,k} = \frac{\rho^1_{i,k}\hat{p}_{i,k}}{\rho^1_{i,k}\hat{p}_{i,k}+\rho^0_{i,k}(1-\hat{p}_{i,k})} \tag{28}$$

3. *M-Step*: for every $\nu \in \{\sigma^2, a, 1\}$, compute the intermediate variables

$$\phi_\nu(i,k)=\sum_{j=1}^N W_{ij}\left((1-\beta_k)\phi_\nu(j,k-1)+\alpha_k f_\nu(j,k)\right) \tag{29}$$

$$\boldsymbol{\Phi}(i,k)=\sum_{j=1}^N W_{ij}\left((1-\beta_k)\boldsymbol{\Phi}(j,k-1)+\alpha_k \boldsymbol{F}(j,k)\right) \tag{30a}$$

$$\boldsymbol{\varphi}(i,k)=\sum_{j=1}^N W_{ij}\left((1-\beta_k)\boldsymbol{\varphi}(j,k-1)+\alpha_k \boldsymbol{f}(j,k)\right) \tag{30b}$$

where $f_1(j,k) = 1$, $f_a(j,k) = \hat{a}_{j,k}$, $\forall j,k$ and

$$\alpha_k = \tfrac{1}{k}, \quad \beta_k = \tfrac{1}{k^\delta}, \quad 0 < \delta < 1, \quad k = 1, 2, \cdots \tag{31}$$

Solve for $\hat{\boldsymbol{x}}_{i,k+1}$ in the $L \times L$ linear system

$$\boldsymbol{\Phi}(i,k)\hat{\boldsymbol{x}}_{i,k+1} = \boldsymbol{\varphi}(i,k) \tag{32}$$

and update

$$\hat{\sigma}^2_{i,k+1} = \tfrac{1}{r^2(L+1)}\tfrac{\phi_\sigma(i,k)}{\phi_1(i,k)}, \; \hat{p}_{i,k+1} = \tfrac{\phi_a(i,k)}{\phi_1(i,k)}.$$

4. *Repeat* steps 2 and 3 until convergence.

Table 1. Diffusion-Based Distributed EM Algorithm

Distributed (D)-BEM: For the D-BEM approach, the a posteriori probabilities in (28) are computed as

$$\rho^a_{i,k} = \exp\left(-\frac{1}{2\hat{\sigma}^2_{i,k}}\left(\frac{y_i - a\boldsymbol{z}_i^T\hat{\boldsymbol{x}}_{i,k}}{1+ar^2\|\hat{\boldsymbol{x}}_{i,k}\|^2\hat{a}_{i,k}}\right)^2\right), \; a \in \{0,1\} \tag{37}$$

whereas the intermediate variables are given by

$$\boldsymbol{F}(j,k)=\hat{d}_{j,k}\boldsymbol{z}_j\boldsymbol{z}_j^T \tag{38}$$

$$\boldsymbol{f}(j,k)=\hat{d}_{j,k}y_j\boldsymbol{z}_j \tag{39}$$

$$f_\sigma(j,k)=r^2(\hat{d}_{j,k}(y_j - \boldsymbol{z}_j^T\hat{\boldsymbol{x}}_{j,k+1})^2 + (1-\hat{a}_{j,k})y_j^2), \tag{40}$$

with $\hat{d}_{j,k} = \hat{a}_{j,k}/(1+r^2\|\hat{\boldsymbol{x}}_{j,k}\|^2\hat{a}_{j,k})$.

5. NUMERICAL RESULTS

Computer simulations of the proposed EM-based estimators have been performed in a network composed of $N = 100$ nodes randomly deployed on a unit square with an average connectivity radius $r_c = 0.25$. We set $L = 5$, $p = 0.8$ and $r = 1$, and generate the entries of $\boldsymbol{H}$ as zero-mean i.i.d. Gaussian random variables of unit variance. The distributed schemes are run with a Metropolis weight matrix $\boldsymbol{W}$ [14] and $\delta = 0.8$. The local estimates are initialized as follows: $\hat{a}_{i,0} = \hat{p}_{i,1} = 1/2$, $\hat{\boldsymbol{x}}_{i,1} = y_i\boldsymbol{z}_i/\boldsymbol{z}_i^T\boldsymbol{z}_i$ and $\hat{\sigma}^2_{i,1} = y_i^2(1-\hat{a}_{i,0})$. Conditioned on $\boldsymbol{H}$, the signal-to-noise (SNR) is given by

$$\mathrm{SNR} = \frac{\boldsymbol{x}^T\boldsymbol{H}^T\boldsymbol{H}\boldsymbol{x}+\mathrm{tr}(\boldsymbol{H}^T\boldsymbol{H})}{N(L+1)\sigma^2} \leqslant \frac{(1+\|\boldsymbol{x}\|^2)\|\boldsymbol{H}\|_F^2}{N(L+1)\sigma^2}. \tag{41}$$

For the simulations, we take the upper bound in (41) as the SNR, as it only depends on $\|\boldsymbol{x}\|^2$ and $\|\boldsymbol{H}\|_F^2$. The performance metric considered is the normalized MSE, defined as

$$\mathrm{NMSE}\{\hat{\boldsymbol{x}}(k)\} = \frac{1}{N\|\boldsymbol{x}\|^2}\sum_{i=1}^N \mathbb{E}\left[\|\hat{\boldsymbol{x}}_i(k) - \boldsymbol{x}\|_2^2\right]. \tag{42}$$

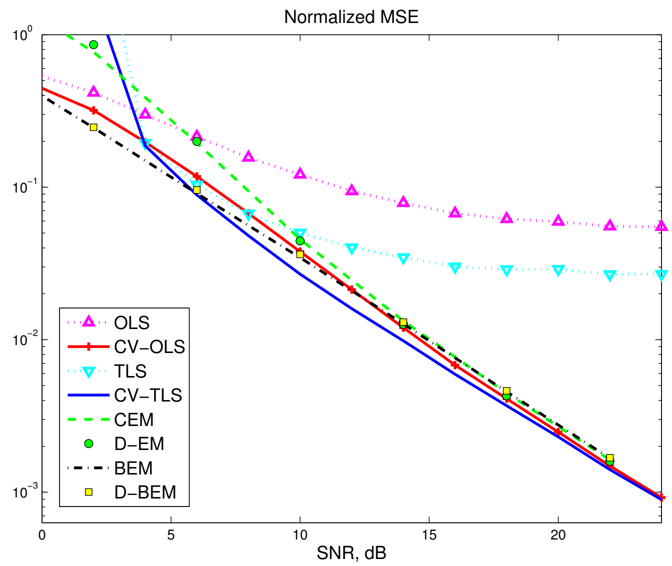

Fig. 1. Normalized MSE vs. SNR in dB for the proposed EM-based estimators, the TLS, the OLS, and their respective clairvoyant versions CV-TLS and CV-OLS.

Fig. 1 depicts the NMSE vs. SNR for all methods, for SNR values in the range $[0, 24]$ dB. Note that both CV-TLS and TLS exhibit a worse performance in the low SNR regime, caused by numerical problems with the eigenvalue decomposition, while all EM-based methods approach the curve for the CV-TLS as the SNR increases. As expected, TLS and OLS exhibit a flooring effect caused by the random data faults, while CEM and BEM approach the clairvoyant solutions. Of the two methods proposed, BEM seems to degrade more gracefully as the SNR is decreased. Therefore, if estimation of the regressors is not needed, the D-BEM estimator offers a good alternative at low complexity. Summing up, the EM-based proposed methods constitute a better alternative than the TLS or the OLS when knowledge of the data faults is not available.

6. REFERENCES

[1] Y. Zhang and X. Rong Li, ”Detection and diagnosis of sensor and actuator failures using IMM estimator,” *IEEE Trans. Aerosp. Electron. Syst.,* vol. 34, no. 4, pp. 1293–1313, 1998.

[2] Q. Zhou, S. Kar, L. Huie and S. Cui, ”Distributed Estimation in Sensor Networks with Imperfect Model Information: An Adaptive Learning-Based Approach,” *Proc. ICASSP 2012*.

[3] G.H. Golub and C. Van Loan, *Matrix Computations.* The Johns Hopkins University Press, 2nd ed., 1989.

[4] I. Markovsky and S. Van Huffel, “Overview of Total Least Squares methods,” *Signal Process.,* vol. 87, no. 10, pp. 2283–2302, 2007.

[5] A. Bertrand, M. Moonen, ”Consensus-based distributed Total Least Squares estimation in *ad hoc* wireless sensor networks,” *IEEE Trans. Signal Process.,* vol. 59, no. 5, pp. 2320–2330, May 2011.

[6] A. Bertrand, M. Moonen, ”Low-complexity distributed Total Least Squares estimation in ad hoc sensor networks,” *IEEE Trans. Signal Process.,* vol. 60, no. 8, pp. 4321–4333, Aug. 2012.

[7] R. López-Valcarce, S. Silva Pereira, A. Pagès-Zamora, “Distributed Total Least Squares Estimation over Networks,” *IEEE ICASSP*, pp. 7580–7582, May 2014.

[8] S. Silva Pereira, R. López-Valcarce, A. Pagès-Zamora, “A Diffusion-Based EM Algorithm for Distributed Estimation in Unreliable Sensor Networks,” *IEEE Signal Processing Letters.*, vol. 20, no. 6, pp. 595–598, June 2013.

[9] A. P. Dempster, N. M. Laird, and D. B. Rubin, “Maximum likelihood from incomplete data via the EM algorithm,” *J. Royal Stat. Soc., Series B*, vol. 39, no. 1, pp. 1–38, 1977.

[10] T. K. Moon, ”The Expectation-Maximization algorithm,” *IEEE Signal Process. Mag.,* vol. 13, no. 6, pp. 47–60, 1996.

[11] A. H. Sayed, ”Diffusion adaptation over networks,” to appear in *E-Reference Signal Processing,* R. Chellapa and S. Theodoridis, eds., Elsevier, 2013. Preprint: http://arxiv.org/pdf/1205.4220v1.pdf.

[12] A. M. Zoubir, V. Koivunen, Y. Chakhchoukh, and M. Muma, “Robust Estimation in Signal Processing: A Tutorial-Style Treatment of Fundamental Concepts,” *IEEE Signal Process. Mag.*, vol. 29, no. 4, pp. 61–80, July 2012.

[13] P. Stoica, Y. Selén, ”Cyclic minimizers, majorization techniques, and the expectation–maximization algorithm: a refresher,” *IEEE Signal Process. Mag.,* vol. 21, no. 1, pp. 112–114, Jan. 2004.

[14] L. Xiao, S. Boyd, S. Lall, ”A scheme for asynchronous distributed sensor fusion based on average consensus,” in *Proc. 4th Int. Symp. Inf. Process. Sensor Netw.,* pp. 63–70, 2005.