Distributed Sequential Joint Detection and Estimation for non-Gaussian Noise

Dominik Reinhard and Abdelhak M. Zoubir
Signal Processing Group
Technische Universität Darmstadt
64283 Darmstadt, Germany
{reinhard,zoubir}@spg.tu-darmstadt.de

***Abstract*—The problem of jointly testing a hypothesis and estimating a random parameter in non-Gaussian noise is investigated in a sequential and distributed setup. The non-Gaussian noise is modeled by a mixture of a completely known Gaussian distribution and an unknown contaminating distribution. Starting from the consensus+innovations approach, we present two robust communication schemes that are insensitive to the contaminating distribution. After deriving upper bounds for the variances of the estimators, a sequential scheme is designed at every sensor such that i) detection and estimation errors are limited for all possible contaminating distributions ii) the resulting scheme uses a minimum number of samples on average. A numerical example validates the proposed method.**

***Index Terms*—joint detection and estimation, sequential analysis, distributed inference, statistical robustness**

I. Introduction

In many applications, detection and estimation tasks arise in a coupled problem, i.e., one has to decide among several models and estimate simultaneously one or more parameters of the selected model. The problem of joint detection and estimation dates back to Middleton and Esposito [1] who investigated the problem in a Bayesian context. Fredriksen *et al.* [2] extended that framework to multiple hypotheses. More recently, this problem has been investigated in the context of communications [3], radar [4], change point detection [5] and biomedical engineering [6], to mention a few.

Sequential analysis dates back to the groundbreaking work of Wald [7]. In the sequential framework, one is interested in performing inference, e.g., detection and/or estimation, with a minimum number of samples while controlling the quality of the inference. Especially for applications where time and energy consumption is critical, sequential procedures should be preferred over conventional ones.

Jointly inferring the true hypothesis and model parameters while using a minimum amount of samples was first studied by Yılmaz *et al.* In [8], [9], the number of used samples is minimized while keeping a combined detection and estimation cost function below a certain level. In [10], we developed a Bayesian framework in which the individual error probabilities and the mean-squared errors (MSEs) are controlled and the resulting scheme uses a minimum number of samples on average. The same framework was then applied to joint signal detection and signal-to-noise power ratio estimation [11]. Subsequently, we provided an extension to multiple hypotheses and applied it to joint symbol decoding and noise power estimation [12].

Many modern signal processing applications rely on multiple sensors, which form a sensor network. To avoid a single point of failure and to minimize the communication load in the network, inference should be performed in a distributed fashion. That is, instead of sending all information to a fusion center, the sensors perform inference based on their interaction with their neighbors.

The development of signal processing algorithms often relies on the exact knowledge of the data generating model. Assumptions made throughout the design process might be violated, and can thus cause the inference to fail. Especially the assumption of Gaussian noise does not always hold in practice when the measured signal contains gross outliers. Instead, robust methods are designed to not break down under model deviations while working well under the nominal model. An overview on robust signal processing is given in [13].

In this work, we consider the problem of sequential joint detection and estimation in a distributed sensor network for non-Gaussian noise. Herein, the non-Gaussian noise is modeled by a mixture of a completely known Gaussian process and an unknown contaminating distribution, which allows to, for example, represent gross outliers. Since treating the detection and the estimation separately does not result in an overall optimal performance [14], the true hypothesis and the parameter have to be inferred *jointly*. Moreover, each sensor should use on average a minimum number of samples, which leads to a *sequential* setup. We develop a sequential scheme that jointly infers the true hypothesis and the value of the random parameter while controlling the detection and estimation errors at nominal levels, irrespective of the true contaminating distribution.

This work is structured as follows. In Section II, we formulate the problem and review the fundamentals of sequential joint detection and estimation. The robust communication schemes are introduced in Section III, which is followed by a statistical characterization of the resulting state variable and a derivation of upper bounds on its variance. Based on these properties, we describe the test design in Section IV. A numerical example concludes the work.

The work of Dominik Reinhard was supported by the German Research Foundation (DFG) under grant number 390542458.

II. Problem Formulation

Consider a network of K sensors, which can be modeled by a simple, connected and undirected graph $\mathcal{G} = (\mathcal{E}, \mathcal{V})$ with nodes $\mathcal{E}$ and edges $\mathcal{V}$. The open and closed neighborhoods of node k are given by $\mathcal{N}_k^{\mathrm{o}} = \{l \in \mathcal{E} \,|\, (k,l) \in \mathcal{V}\}$ and $\mathcal{N}_k^{\mathrm{c}} = \mathcal{N}_k^{\mathrm{o}} \cup \{k\}$, respectively. Each sensor $k \in \{1, \ldots, K\}$ observes a sequence of random variables $\mathcal{X}_N^k = (X_1^k, \ldots, X_N^k)$ generated according to the linear model

$$x_n^k = \theta + v_n^k\,, \quad n = 1, \ldots, N\,, \tag{1}$$

where θ is the random parameter of interest and v_n^k represents measurement noise. The random parameter θ could have been generated under two different hypotheses H_0 and H_1 with prior probabilities $p(\mathrm{H}_0)$ and $p(\mathrm{H}_1)$, respectively. The underlying hypothesis only affects the distribution of the random parameter θ, i.e.,

$$\begin{aligned} \mathrm{H}_0: &\quad \theta \sim p_0(\theta)\,, \\ \mathrm{H}_1: &\quad \theta \sim p_1(\theta)\,, \end{aligned} \tag{2}$$

where the priors $p_0(\theta)$ and $p_1(\theta)$ have a disjoint support to ensure that the hypotheses are mutually exclusive.

We model the non-Gaussianity of the noise process by ε-contamination [15]. That is, a fraction $1 - \varepsilon$ of the noise samples follows a Gaussian distribution, while the remaining fraction follows an unknown contaminating distribution. Mathematically, this can be written as

$$v_n^k \overset{\text{iid}}{\sim} F = (1-\varepsilon)F_{\text{nom}} + \varepsilon F_{\text{cont}}(1-\varepsilon)\mathcal{N}(0, \sigma_0^2) + \varepsilon F_{\text{cont}}\,,$$

where $0 \leq \varepsilon < 1$ is a *known* contamination ratio and $F_{\text{cont}} \in \mathcal{F}_{\text{cont}}$ is an *unknown* contaminating distribution. We restrict the class of contaminating distributions $\mathcal{F}_{\text{cont}}$ to zero-mean symmetric distributions.

Each sensor shall infer the true hypotheses H_0 or H_1 and estimate the parameter θ. To perform this inference task, each sensor has access to its own measurements, as well as to information from its neighbors. We assume that all sensors observe the same realizations of H_i and θ, i.e., the exchanged measurements contain information about the same phenomenon. Furthermore, it is assumed that these realizations, i.e., the quantities of interest, remain constant during the observation period.

Prior to providing a more technical problem formulation, we revise some fundamentals on sequential joint detection and estimation. The stopping and decision rules for sensor k at time n are denoted by $\Psi_n^k \in \{0,1\}$ and $\delta_n^k \in \{0,1\}$, respectively. The estimator under H_i at time n and sensor k is denoted by $\hat{\theta}_{i,n}^k$. The collection π^k, referred to as policy, comprises the stopping rule, the decision rule and the estimators at node k. The number of samples observed by sensor k is

$$\tau^k = \min\{n : \Psi_n^k = 1\}\,.$$

To evaluate the performance at node k, the average sample number $\mathsf{E}[\tau^k]$ along the error probabilities and the MSEs are considered, which are defined as

$$\alpha_i^k = \mathsf{E}\Big[\mathbb{1}\{\delta_{\tau^k}^k \neq i\} \,\big|\, \mathrm{H}_i\Big]\,, \quad i \in \{0,1\}\,,$$

$$\beta_i^k = \mathsf{E}\Big[\mathbb{1}\{\delta_{\tau^k}^k = i\}\big(\theta - \hat{\theta}_{i,\tau^k}^k\big)^2 \,\big|\, \mathrm{H}_i\Big]\,, \quad i \in \{0,1\}\,.$$

In the equations above, $\mathbb{1}\{\cdot\}$ denotes the indicator function.

In this work, we consider truncated sequential schemes, i.e., schemes that use at most N samples, thereby implying that $\Psi_N^k = 1$ for all $k = 1, \ldots, K$. Our inference scheme limits the detection and estimation errors to certain levels for each node and all $F_{\text{cont}} \in \mathcal{F}_{\text{cont}}$. For tractability reasons, the problem is solved in two steps. First, a contamination-insensitive communication scheme has to be developed. Then, for a given communication scheme, we can proceed as in [16] and solve

$$\begin{aligned} &\min_{\pi^k} \; \mathsf{E}\big[\tau^k\big], \quad \Psi_N^k = 1 \\ &\text{s.t.} \quad \alpha_i^k \leq \bar{\alpha}_i\,, \quad \beta_i^k \leq \bar{\beta}_i\,, \quad i = 0, 1\,, \end{aligned} \tag{3}$$

where $\bar{\alpha}_0, \bar{\alpha}_1 \in (0,1)$ and $\bar{\beta}_0, \bar{\beta}_1 \in (0, \infty)$ are nominal detection and estimation errors, respectively. As long as N is large enough and $\mathsf{E}[\theta^2 \,|\, \mathrm{H}_i] < \infty$ for all $i \in \{0,1\}$, Eq. (3) always admits a solution.

III. Robust Neighborhood Communication

To design the test, we use the sample mean as low-dimensional representation of the data as in [16]. Although the sample mean only serves as a sufficient statistic for Gaussian distributed data, we show later that this also holds for the proposed non-Gaussian case. The states and the measurements of the sensors are shared within their neighborhood and the states are updated via the consensus+innovations [17] approach. Let $\mathbf{t}_n = [t_n^1, \ldots, t_n^K]^\top$ be the vector containing all states at time n with initial state $t_0^k = 0$ for all $k \in \{1, \ldots, K\}$. The consensus+innovations update is given by

$$\mathbf{t}_{n+1} = (n+1)^{-1}\Big(n\mathbf{W}\mathbf{t}_n + \mathbf{W}\mathbf{x}_n^k\Big)\,, \tag{4}$$

where $\mathbf{x}_n = [x_n^1, \ldots, x_n^k]^\top$ is the vector collecting all measurements at time n and $\mathbf{W}$ is a suitable weighting matrix. Similarly to [16], [18], [19], we assign equal weights to the information of the closed neighborhood of a node. Thus, the last term in Eq. (4) is equal to the sample mean, the optimal estimator for the location parameter of Gaussian data. However, the sample mean is non-robust, i.e., a single outlier can cause its breakdown [13]. We therefore replace the sample mean by robust estimators to deal with outliers [18], [19]. However, simply replacing the sample mean in the update equation by a robust estimator as in [18], [19] does not solve the present problem, i.e., the procedure remains non-robust. As shown in the remainder of this section, the statistical properties of robust and non-robust estimators differ significantly. The test design has to account for these differing statistical properties.

Let $\hat{x}_n^k$ denote an estimator of the location parameter at node k and time n. By defining $\hat{\mathbf{x}}_n = \big[\hat{x}_n^1, \ldots, \hat{x}_n^K\big]^\top$, Eq. (4) turns into

$$\mathbf{t}_{n+1} = (n+1)^{-1}\Big(n\mathbf{W}\mathbf{t}_n + \hat{\mathbf{x}}_n\Big)\,. \tag{5}$$

Then, the conditional first and second order moments of the state variable are given by [16]:

$$\mathsf{E}\Big[t_n^k \,\big|\, \mathrm{H}_i, \theta\Big] = \mathsf{E}[\hat{x}_n^k] \tag{6}$$

$$\mathsf{Var}\Big[t_n^k \,|\, \mathrm{H}_i, \theta\Big] = \frac{1}{n^2}\sum_{i=1}^{n} \mathbf{e}_k^\top \mathbf{W}^{i-1}\mathbf{\Sigma}\Big(\mathbf{W}^{i-1}\Big)^\top \mathbf{e}_k \qquad (7)$$

The kth column of the identity matrix is denoted by $\mathbf{e}_k$ and $\mathbf{\Sigma}$ is the covariance matrix of the estimator, i.e.,

$$\mathbf{\Sigma} = \mathsf{E}\Big[\hat{\mathbf{x}}_n\hat{\mathbf{x}}_n^\top\Big] - \mathsf{E}\Big[\hat{\mathbf{x}}_n\Big]\,\mathsf{E}\Big[\hat{\mathbf{x}}_n^\top\Big]\,. \qquad (8)$$

If the sample mean is applied to uncontaminated data, Eq. (6) and Eq. (8) turn into those in [16]

$$\mathsf{E}\Big[t_n^k \,|\, \mathrm{H}_i, \theta\Big] = \theta \quad \text{and} \quad \mathbf{\Sigma}_{\text{mean}} = \sigma_0^2 \mathbf{W}\mathbf{W}^\top\,. \qquad (9)$$

In what follows, we investigate the median and the M-estimator, two robust estimators for a location parameter of a symmetric distribution.

A. Measurement Combination via Median

The median $\hat{x}_{n,\text{med}}^k$ is asymptotically normally distributed, i.e., for a large size of the neighborhood, it holds that [20]

$$\hat{x}_{n,\text{med}}^k \,|\, \mathrm{H}_i, \theta \,\dot{\sim}\, \mathcal{N}\bigg(\theta, \frac{\sigma_{\text{med}}^2}{|\mathcal{N}_k^{\text{c}}|}\bigg)\,,$$

where $\dot{\sim}$ means *approximately distributed as*, and σ_{med}^2 is given by [20, Eq. (2.26)]

$$\sigma_{\text{med}}^2 = \Big(2\tilde{f}(\xi)\Big)^{-2}\,.$$

In the previous equation, ξ and $\tilde{f}(\cdot)$ denote the population median and the probability density function (pdf) of $x_n^k \,|\, \mathrm{H}_i, \theta_i$, respectively. For symmetrically distributed noise, the population median ξ and the population mean coincide. For zero-mean noise and the measurement model in Eq. (1), it further holds that $\tilde{f}(\xi) = \tilde{f}(\theta) = f(0)$. However, the pdf of the noise is not completely known, but strongly affects σ_{med}^2 that is crucial for the subsequent test design. To ensure a given performance for all feasible contaminating distributions, an upper bound of the variance has to be found. The variance can be expressed as

$$\sigma_{\text{med}}^2 = \Big(2f(0)\Big)^{-2} = \Big(2(1-\varepsilon)f_{\text{nom}}(0) + 2\varepsilon f_{\text{cont}}(0)\Big)^{-2}$$

which attains its maximum for $f_{\text{cont}}(0) = 0$. Thus, the upper bound for the variance is

$$\sigma_{\text{med}}^2 \leq \Big(2(1-\varepsilon)f_{\text{nom}}(0)\Big)^{-2} = \frac{\sigma_0^2\pi}{2(1-\varepsilon)^2}\,.$$

Hence, apart from a constant scaling factor in the variance, the asymptotic distribution of the sample median is equal to the one of the sample mean.

For the covariance matrix of the estimators, which describes the spatial correlation in the network, we use the one from Eq. (9), but with a different scaling, i.e.,

$$\mathbf{\Sigma}_{\text{med}} \approx \frac{\sigma_0^2\pi}{2(1-\varepsilon)^2}\mathbf{W}\mathbf{W}^\top\,.$$

B. Measurement Combination via M-Estimators

Next, we study the M-estimator [15]. The M-estimator $\hat{x}_{n,\text{M}}^k$ for a location parameter and known scale σ_0 is obtained by solving

$$\sum_{i\in\mathcal{N}_k^{\text{c}}} \psi\bigg(\frac{x_n^i - \hat{x}_{n,\text{M}}^k}{\sigma_0}\bigg) = 0\,, \qquad (10)$$

where ψ denotes the score function. Here, we use Huber's score function, which is given by [13, Eq. (1.21)]

$$\psi_{\text{Hub}}(x) = \begin{cases} x & |x| \leq c_{\text{Hub}}\,, \\ c_{\text{Hub}}\,\text{sign}(x) & |x| > c_{\text{Hub}}\,, \end{cases}$$

where $\text{sign}(x)$ denotes the sign of x and c_{Hub} is a tuning constant which trades off efficiency and robustness. In general, the following derivations are not restricted to Huber's score function, but are valid for any odd score function. The M-estimator can be calculated according to [13, Algorithm 1] when replacing their scale estimate by the scale of the nominal noise σ_0. According to [20, Section 2.2.2], the M-estimator is asymptotically normally distributed, i.e.,

$$\hat{x}_{n,\text{M}}^k \,|\, \mathrm{H}_i, \theta \,\dot{\sim}\, \mathcal{N}\Big(\theta, \frac{\sigma_{\text{M},F}^2}{|\mathcal{N}_k^{\text{c}}|}\Big)\,.$$

If the data is distributed according to probability measure F, the variance $\sigma_{\text{M},F}^2$ can be calculated via [20, Eq. (2.6.3)]

$$\sigma_{\text{M},F}^2 = \sigma_0^2\,\mathsf{E}_F\Big[\big(\psi(\frac{x}{\sigma_0})\big)^2\Big]\Big(\mathsf{E}_F\big[\psi'(\frac{x}{\sigma_0})\big]\Big)^{-2}\,, \qquad (11)$$

where $\psi'(x)$ denotes the derivative of $\psi(x)$. However, the data generating distribution is not completely specified. Hence, we provide an upper bound for Eq. (11). This bound is given by

$$\sigma_{\text{M}}^2 = \max_{F_{\text{cont}}\in\mathcal{F}_{\text{cont}}} \sigma_0^2\,\mathsf{E}_F\Big[\big(\psi(\frac{x}{\sigma_0})\big)^2\Big]\Big(\mathsf{E}_F\big[\psi'(\frac{x}{\sigma_0})\big]\Big)^{-2}\,. \qquad (12)$$

This problem can be solved using a projected gradient ascent whose gradient is given by

$$\nabla_{f_{\text{cont}}}\sigma_{\text{M},F}^2 = \sigma_0^2\varepsilon\frac{\psi^2(\frac{x}{\sigma_0})\,\mathsf{E}_F[\psi'(\frac{x}{\sigma_0})] - 2\psi'(\frac{x}{\sigma_0})\,\mathsf{E}_F[\psi^2(\frac{x}{\sigma_0})]}{\Big(\mathsf{E}_F[\psi'(\frac{x}{\sigma_0})]\Big)^3}\,.$$

The projection operator is given by

$$\Pr(f_{\text{cont}}) = \max\{0, c f_{\text{cont}}\}\,,$$

where the constant c has to be chosen such that the result is a valid density. As long as the initial density is zero-mean and symmetric, the projection onto the set of feasible densities is also zero-mean and symmetric.

For the covariance matrix, we proceed similarly as for the median case and approximate it by

$$\mathbf{\Sigma}_{\text{M}} \approx \sigma_{\text{M}}^2\mathbf{W}\mathbf{W}^\top\,.$$

IV. Design of the Optimal Test

As shown in the previous section, the statistical properties of the state variable are similar to those of the non-robust scheme proposed in [16], except for the constant scaling factor of the variance. Hence, the design problem boils down to designing a non-robust test with larger variance and we can proceed as in [10], [16]. However, this holds only for the design and not for the implementation of the test. For the sake of completeness, we briefly summarize the design steps.

We first convert Eq. (3) to an unconstrained problem, i.e.,

$$\min_{\pi^k}\bigg\{\mathsf{E}\big[\tau^k\big] + \sum_{i=0}^{1}p(\mathrm{H}_i)\Big(\lambda_i^k\alpha_i^k + \mu_i^k\beta_i^k\Big)\bigg\}\,, \qquad (13)$$

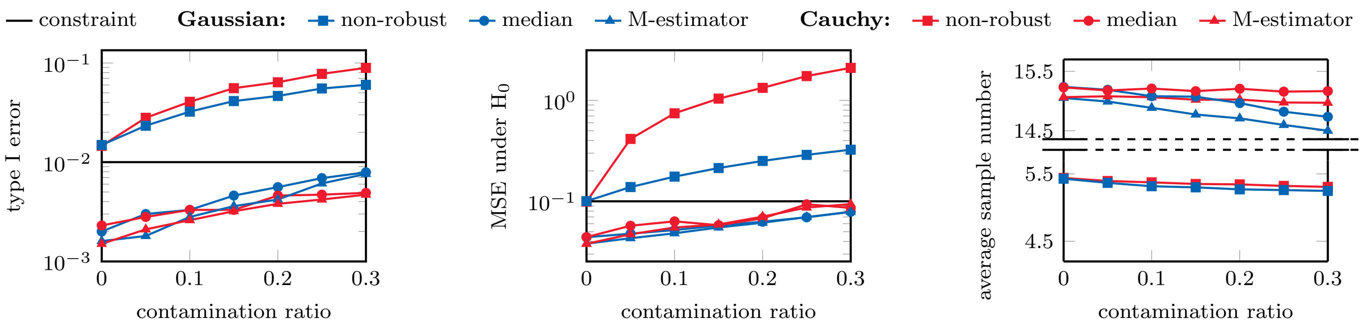

Figure 1: Performance measures for three different communication schemes (non-robust [16], median, Huber's M-estimator) and two different contaminating distributions (Gaussian with variance $10\sigma_0^2$, standard Cauchy distribution).

where λ_i^k and μ_i^k, $k = 1, \ldots, K$, $i = 0, 1$, are non-negative and finite cost coefficients. We first solve Eq. (13) with respect to the decision rule and then with respect to the estimators. The optimal stopping problem becomes then

$$\min_{\Psi^k} \mathsf{E}\left[\Phi_n^k\left(n + g^k(t_n^k)\right)\right], \tag{14}$$

where $\Phi_n^k = \Psi_n^k \prod_{i=0}^{n-1}(1 - \Psi_i^k)$ is the shorthand notation for stopping at time n and not before. The instantaneous cost for stopping at time n is given by

$$g^k(t_n^k) = \min\left\{ D_{0,n}^k(t_n^k), D_{1,n}^k(t_n^k) \right\}.$$

The cost for stopping is the minimum between the cost for stopping and deciding in favor of the null hypothesis $D_{0,n}^k$ and the cost for stopping and deciding in favor of the alternative $D_{1,n}^k$. The cost for stopping and deciding in favor of H_i is for $i \in \{0, 1\}$ defined as

$$D_{i,n}^k(t_n^k) = \lambda_{1-i}^k p(\mathrm{H}_{1-i} \,|\, t_n^k) + \mu_i^k p(\mathrm{H}_i \,|\, t_n^k) \, \mathsf{Var}[\theta \,|\, \mathrm{H}_i, t_n^k],$$

where the first part penalizes wrong decisions and the second part penalizes the estimation inaccuracy.

For fixed λ_i^k and μ_i^k, $i = 0, 1$, the optimal stopping problem in Eq. (14) can be solved by means of dynamic programming and its solution can be characterized as, see, e.g., [10], [21],

$$\rho_n^k(t_n^k) = \min\left\{ g^k(t_n^k), \, d_n^k(t_n^k) \right\}, n < N,$$
$$\rho_N^k(t_N^k) = g^k(t_N^k),$$

with the cost for taking a new sample

$$d_n^k(t_n^k) = 1 + \mathsf{E}[\rho_{n+1}^k(t_{n+1}^k) \,|\, t_n^k]. \tag{15}$$

The expected value in Eq. (15) can either be evaluated by means of numerical integration or can be approximated by Markov Chain Monte Carlo integration. For details, refer to [16]. The optimal policy at node k is summarized below.

$$\begin{aligned} \delta_n^k(t_n^k) &= \mathbb{1}\{D_{0,n}^k(t_n^k) > D_{1,n}^k(t_n^k)\} \\ \Psi_n^k(t_n^k) &= \mathbb{1}\{\rho_n^k(t_n^k) = g^k(t_n^k)\} \\ \hat{\theta}_{i,n}^k &= \mathsf{E}[\theta \,|\, \mathrm{H}_i, t_n^k], \quad i = 0, 1 \end{aligned} \tag{16}$$

Since this policy depends on λ_i^k and μ_i^k, $i = 0, 1$, it is only optimal with respect to these coefficients. Hence, we have to chose λ_i^k and μ_i^k, $k = 1, \ldots, K$, $i = 0, 1$, such that Eq. (16) also solves Eq. (3). By exploiting a strong connection between the performance measures and the cost function [10, Theorem 4.2] [21, Theorem 3.2], the final optimization problem can be formulated as a linear program, which can be solved efficiently by various off-the-shelf solvers

$$\begin{aligned} \max_{\lambda^k \geq 0, \mu^k \geq 0, \rho_n^k} \quad & \left\{ \rho_0^k(t_0^k) - \sum_{i=0}^{1} p(\mathrm{H}_i)(\lambda_i^k \bar{\alpha}^i + \mu_i^k \bar{\beta}^i) \right\} \\ \text{s.t.} \quad & \rho_n^k \leq D_{i,n}^k, \; i \in \{0, 1\}, \; n = 0, \ldots, N, \\ & \rho_n^k \leq 1 + \mathsf{E}[\rho_{n+1}^k(t_{n+1}^k) \,|\, t_n^k], \; n < N. \end{aligned} \tag{17}$$

V. Numerical Results

Consider a network of $K = 20$ nodes with coordinates uniformly sampled on $[0, 1] \times [0, 1]$. The neighborhood of a node k is the set of all nodes within a normalized communication radius of 0.3. The generation of the network is repeated until the graph is connected. Each sensor should jointly test the two equal probable hypotheses

$$\begin{aligned} \mathrm{H}_0 &: \; -\theta \sim \mathrm{Gam}(1.8, 1) \\ \mathrm{H}_1 &: \quad \theta \sim \mathrm{Gam}(1.8, 1) \end{aligned}$$

and infer the value of the random parameter θ. The Gamma distribution with shape a and scale b is denoted by $\mathrm{Gam}(a, b)$. The variance of the nominal noise is set to $\sigma_0^2 = 4$ and a contamination rate of $\varepsilon = 0.3$ is used during the design process. Each sensor is allowed to use at most $N = 100$ samples. The detection and estimation errors are limited to 0.01 and 0.1, respectively. To design the test, the spaces of t_n^k and θ are discretized on the interval $[-6, 6]$ with 1001 points and on the interval $[-12, 12]$ with 8000 points, respectively. The linear program in Eq. (17) is solved by the Gurobi optimizer [22] which is called via the Matlab cvx interface [23], [24]. The expectation in Eq. (15) is approximated by Markov Chain Monte Carlo integration with $5 \cdot 10^4$ samples, which are generated via a Metropolis-Hastings algorithm similar to [16, Table 1].

We design the tests for three different communication schemes: measurement combination via median, measurement combination via Huber's M-estimator and the non-robust one from [16]. For Huber's M-estimator, the tuning constant is set to $c_{\text{Hub}} = 1.345$ so that it has an asymptotic relative efficiency of 0.95 under the nominal model. To validate the performance of the tests, Monte Carlo simulations with 10^5 runs are performed. Each communication scheme is evaluated for two different distributions of the contaminating noise, a zero-mean Gaussian distribution with variance $10\sigma_0^2$ as in [18], [19] and a standard Cauchy distribution. The Cauchy distribution does not have finite moments and is hence a challenging candidate for robust schemes due to its heavy tails. For the Monte Carlo simulations, the contaminating ratio ε is varied from 0 to 0.3, i.e. from no contamination to the contamination ratio that was assumed during the design process.

The results are summarized in Fig. 1. Note that the detection and estimation errors are only shown for H_0, since the problem is symmetric and leads to similar errors under H_1. Recall that the goal was to design a sequential scheme that fulfills constraints on the detection and estimation errors irrespective of the contaminating distribution. It can be seen that the non-robust approach fulfills the constraints, within the range of numerical inaccuracies, for the non-contaminated case only. Even for a small contamination ratio, the constraints are severely violated. As expected, this effect is even more severe for Cauchy distributed contaminations. On the other hand, the robust schemes keep the empirical detection and estimation errors below the nominal levels. As expected, the detection and estimation errors of the robust schemes increase with increasing contamination ratio. There are no significant differences in the estimation error results for the robust schemes for different noise processes, but the robust schemes have smaller detection errors for high contamination rates and Cauchy distributed noise. The non-robust scheme provides the lowest average sample number. This comes, however, with the cost of violating the constraints on the detection and estimation errors. All robust schemes have similar average sample numbers, which are around factor 3 compared to the non-robust scheme, even for $\varepsilon = 0$. This effect is due to the worst case approximations of the variances in Sections III-A and III-B. However, the performance of the M-estimator could be tuned further by choosing different values of the tuning constant c_{Hub}.

VI. Conclusion

We have investigated the problem of sequential joint detection and estimation for a random location parameter in non-Gaussian noise in a distributed sensor network. The noise has been modeled by the ε-contamination model. Based on the consensus+innovations approach, we presented two robust communication schemes that are insensitive to the introduced model uncertainties. For both schemes, it has been shown that the resulting states follow a Gaussian distribution. In addition, we have derived upper bounds for the variances of states that serve as worst-case approximations. Based on these approximations, we have designed the test such that the detection and estimation errors are kept below pre-defined levels for all feasible contaminating noise distributions. Numerical results validate the designed tests in the sense that they fulfill the constraints even for very heavy-tailed contaminating distributions, such as the Cauchy distribution. These distributions cause the existing non-robust scheme to violate the nominal error levels. This restoration of the validity in terms of error levels comes at the cost of larger average sample numbers.

References

[1] D. Middleton and R. Esposito, "Simultaneous optimum detection and estimation of signals in noise," *IEEE Trans. Inf. Theory*, vol. 14, no. 3, pp. 434–444, May 1968.

[2] A. Fredriksen, D. Middleton, and V. VandeLinde, "Simultaneous signal detection and estimation under multiple hypotheses," *IEEE Trans. Inf. Theory*, vol. 18, no. 5, pp. 607–614, 1972.

[3] Z. Wei, W. Hu, D. Han, M. Zhang, B. Li, and C. Zhao, "Simultaneous channel estimation and signal detection in wireless ultraviolet communications combating inter-symbol-interference," *Opt. Express*, vol. 26, no. 3, pp. 3260–3270, 2018.

[4] A. Tajer, G. H. Jajamovich, X. Wang, and G. V. Moustakides, "Optimal joint target detection and parameter estimation by MIMO radar," *IEEE J. Sel. Topics Signal Process.*, vol. 4, no. 1, pp. 127–145, Feb 2010.

[5] S. Boutoille, S. Reboul, and M. Benjelloun, "A hybrid fusion system applied to off-line detection and change-points estimation," *Inform. Fusion*, vol. 11, no. 4, pp. 325–337, 2010.

[6] L. Chaari, T. Vincent, F. Forbes, M. Dojat, and P. Ciuciu, "Fast joint detection-estimation of evoked brain activity in event-related fMRI using a variational approach," *IEEE Trans. Med. Imag.*, vol. 32, no. 5, pp. 821–837, 2013.

[7] A. Wald, *Sequential Analysis*. New York: Wiley, 1947.

[8] Y. Yılmaz, G. V. Moustakides, and X. Wang, "Sequential joint detection and estimation," *Theory Probab. Appl.*, vol. 59, no. 3, pp. 452–465, 2015.

[9] Y. Yılmaz, S. Li, and X. Wang, "Sequential joint detection and estimation: Optimum tests and applications," *IEEE Trans. Signal Process.*, vol. 64, no. 20, pp. 5311–5326, Oct 2016.

[10] D. Reinhard, M. Fauß, and A. M. Zoubir, "Bayesian sequential joint detection and estimation," *Sequential Anal.*, vol. 37, no. 04, pp. 530–570, 2018.

[11] D. Reinhard, M. Fauß, and A. M. Zoubir, "Bayesian sequential joint signal detection and signal-to-noise ratio estimation," in *Proc. 27th Eur. Signal Process. Conf. (EUSIPCO)*, September 2019.

[12] ——, "Sequential joint detection and estimation with an application to joint symbol decoding and noise power estimation," in *Proc. 45th IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, May 2020.

[13] A. M. Zoubir, V. Koivunen, E. Ollila, and M. Muma, *Robust Statistics for Signal Processing*. Cambridge University Press, 2018.

[14] G. V. Moustakides, G. H. Jajamovich, A. Tajer, and X. Wang, "Joint detection and estimation: Optimum tests and applications," *IEEE Trans. Inf. Theory*, vol. 58, no. 7, pp. 4215–4229, 2012.

[15] P. Huber, "Robust estimation of a location parameter," *Ann. Math. Stat.*, vol. 35, no. 1, pp. 73–101, 1964.

[16] D. Reinhard, M. Fauß, and A. M. Zoubir, "Distributed joint detection and estimation: A sequential approach," in *Proc. 54th Annu. Conf. Inf. Sci. Syst. (CISS)*, Princeton, USA, Mar. 2020.

[17] S. Kar and J. M. F. Moura, "Consensus + innovations distributed inference over networks: Cooperation and sensing in networked systems," *IEEE Signal Process. Mag.*, vol. 30, no. 3, pp. 99–109, May 2013.

[18] M. R. Leonard and A. M. Zoubir, "Robust sequential detection in distributed sensor networks," *IEEE Trans. Signal Process.*, vol. 66, no. 21, pp. 5648–5662, Nov. 2018.

[19] W. Hou, M. R. Leonard, and A. M. Zoubir, "Robust distributed sequential detection via robust estimation," in *Proc. 25th Eur. Signal Process. Conf. (EUSIPCO)*, Aug. 2017, pp. 2051–2055.

[20] R. A. Maronna, D. Martin, and V. J. Yohai, *Robust statistics : Theory and methods*. John Wiley & Sons, 2006.

[21] M. Fauß and A. M. Zoubir, "A linear programming approach to sequential hypothesis testing," *Sequential Anal.*, vol. 34, no. 2, pp. 235–263, 2015.

[22] I. Gurobi Optimization, "Gurobi optimizer reference manual," 2016. [Online]. Available: http://www.gurobi.com

[23] M. Grant and S. Boyd, "Graph implementations for nonsmooth convex programs," in *Recent Advances in Learning and Control*, ser. Lecture Notes in Control and Information Sciences, V. Blondel, S. Boyd, and H. Kimura, Eds. Springer-Verlag Limited, 2008, pp. 95–110, http://stanford.edu/~boyd/graph_dcp.html.

[24] ——, "CVX: Matlab software for disciplined convex programming, version 2.1," http://cvxr.com/cvx, Mar. 2017.