ELSEVIER

Signal Processing 60 (1997) 281–288

SIGNAL PROCESSING

Convergence rates for decentralized consistent location parameter estimation in the presence of Gaussian outliers

Hakan Deliç*

Signal and Image Processing Laboratory, Department of Electrical and Electronics Engineering, Boğaziçi University, Bebek 80815 Istanbul, Turkey

Received 22 November 1996; revised 7 April 1997

Abstract

We consider a distributed system where sensors make location parameter estimates using their observations. A central processor collects the local estimates and declares a final estimate based on them. We present a simple study of the convergence properties of three structures where empirical mean and M-estimates are used in various combinations. It is shown that when occasional outliers exist, decentralized estimators that provide robustness at stages where data corruption occurs perform superiorly. © 1997 Elsevier Science B.V.

Zusammenfassung

Wir betrachten ein verteiltes System, in dem mittels Beobachtungen einzelner Sensoren Schätzungen von Lageparametern durchgeführt werden. Eine zentrale Verarbeitungseinheit sammelt die lokalen Schätzwerte und bestimmt daraus einen endgültigen Schätzwert. Wir präsentieren eine einfache Analyse der Konvergenzeigenschaften von drei Strukturen, die empirische Mittelwerte und M-Schätzwerte in unterschiedlichen Kombinationen verwenden. Es wird gezeigt, daß bei gelegentlich vorhandenen Ausreißern dezentralisierte Schätzer, welche in den Bereichen gestörter Daten Robustheit aufweisen, leistungsfähiger sind. © 1997 Elsevier Science B.V.

Résumé

Nous considérons un système distribué dans lequel les senseurs estiment les paramètres de position en utilisant leurs observations. Un processeur central collecte les estimations locales et en déduit une estimation finale. Nous présentons une étude des propriétés de convergence de trois structures où la moyenne empirique et les M-estimateurs sont utilisés dans diverses combinaisons. Il est montré qu'en présence d'outliers, les estimateurs décentralisés qui sont robustes là où la corruption des données a lieu, marchent mieux. © 1997 Elsevier Science B.V.

Keywords: Multiple sensor estimation; Gaussian outlier; M-estimate; Asymptotics

*Tel.: 90 212 263 1540: fax: 90 212 287 2465; e-mail: delic@busim.ee.boun.edu.tr.

PII S0165-1684(97)00078-9

1. Introduction

An important aspect of extracting information from a set of data is the estimation of a parameter. In radar systems, for example, typical parameters requiring estimation are the time delay, signal amplitude, direction of arrival, and possibly other parameters that are characteristic of the target. Classical parameter estimation theory has evolved around the fundamental assumption that all data are available at a central processor. However, because of the constraints on the communication bandwidth and computing power of the central processor, centralized systems may not be practical.

An alternative to centralized systems is decentralized estimation based on multiple-sensor data. These systems consist of local processors which collect observations, and subsequently transmit transformed forms of the data to a central processor for the final estimate. The transformation serves the purpose of data compression. For example, in [2], local stations send the sufficient statistics to the central processor with no loss of information. Decentralized systems have several other advantages over the centralized ones, such as improved computational efficiency, increased fault-tolerance, etc.

In this paper, we assume that the local processors estimate the parameter of interest based on the data they collect. The central processor then makes its own estimate upon receiving the local estimates. This form of a decentralized system allows us to build hybrid structures; that is, different types of estimators at the sensor and central levels. We concentrate on the case where observations are occasionally corrupted by outlier data, and investigate the robustness properties of several distributed estimators, as manifested by their convergence rate and breakdown point performance.

The multiple-sensor location parameter estimation problem may be part of a broader hypothesis testing setting. That is, the estimated parameters can be used in detector design. To keep the discussion clear, we assume perfect, noiseless transmission between the sensors and the central processor. Robot navigation using a sonar ring [4] is a well-known application of the scenario considered here.

The organization of the paper is as follows. Section 2 introduces the system and outlier-based noise model. Section 3 has convergence properties of two consistent location parameter estimators. Convergence rates are derived for decentralized multiple-sensor estimation in Section 4. Robustness results are also verified through the determination of breakdown points in Section 5. Section 6 has conclusions and a generalization to nonidentical noise characteristics across sensors.

2. System and noise model

Throughout the paper, we will focus on the estimation of a location parameter θ. This translates to the well-studied problem of estimating the amplitude of a known, constant signal in additive noise. For symmetric noise densities, θ coincides with the mean of the density.

We assume that the observations are generated by stationary and memoryless processes. This implies independent and identically distributed (i.i.d.) data so that

$$f_j(\boldsymbol{x}_j|\theta) = \prod_{i=1}^{N} f_j(x_{ji} - \theta), \tag{1}$$

where $\boldsymbol{x}_j = \{x_{j1}, \ldots, x_{jN}\}$ denotes the data record observed by jth sensor, $j = 1, \ldots, L$, and f_j is the data generating density function for sensor j.

Each sensor calculates an estimate $\hat{\theta}_j = \hat{\theta}_j(\boldsymbol{x}_j)$ based on its own observations. The central processor receives the set of local estimates $\{\hat{\theta}_j\}_{j=1}^{L}$, and then declares the final, or global, estimate $\hat{\theta}_c = \hat{\theta}_c(\hat{\theta})$, where $\hat{\theta} = \{\hat{\theta}_1, \ldots, \hat{\theta}_L\}$. Note that the global estimate by the central processor implicitly depends on the local observations. We assume that the sensors are jointly independent.

In many applications, the noise density in Eq. (1) is assumed to be Gaussian. This assumption is invalid for underwater sonar detection in shallow waters or in the Arctic region, for example, where the noise distribution is typically heavy-tailed. Furthermore, even in situations where it is reasonable to assume Gaussian noise density, occasional high-amplitude outliers may corrupt the observations. In this paper, we will use the following mixture model where the nominal Gaussian noise is occasionally corrupted by independent, symmetric and

Gaussian outliers:

$$f_j(x_{ji}) = (1-\varepsilon)\phi(x_{ji}) + \frac{\varepsilon}{2}[\phi(x_{ji}+\mu) + \phi(x_{ji}-\mu)], \tag{2}$$

for $j = 1, \ldots, L$, where $0 < \varepsilon < 1$ and $\mu > 0$. Throughout the paper, $\phi(x)$ and $\Phi(x)$ will denote the density function and the cumulative distribution function of the zero-mean, unit-variance Gaussian random variable at point x, respectively. The constant ε can be regarded as the a priori probability of departure from the nominal Gaussian assumption, and it is unknown.

A physically realistic, but complex noise model is Middleton's Class-A noise [8], which consists of an infinite expansion of Gaussian densities with different variances and identical means. Generally, the first two terms of the expansion are kept so that the noise becomes the mixture of two Gaussian densities with zero means and nonidentical variances [1, 9]. The model in Eq. (2) puts more weight on the tails of the nominal Gaussian density than a Gaussian–Gaussian mixture, and it is more tractable than Middleton's model.

For the sake of simplicity, in the sequel, we will assume that the noise characteristics are identical for all sensors and drop the sensor index j from the density expressions. The extension of our calculations to cases where each sensor monitors observations generated by a different noise density with nonidentical ε and μ values in Eq. (2) is straightforward, but more messy.

3. Consistent location parameter estimates

It is well-known from estimation theory that a consistent estimate of location is the empirical mean estimate

$$\hat{\theta}^{\mathrm{E}}(\boldsymbol{x}) = \frac{1}{N}\sum_{i=1}^{N} x_i. \tag{3}$$

For Gaussian data, this also coincides with the maximum likelihood estimate, and hence it is the optimal choice. However, the linear empirical mean estimate is very sensitive to the existence of outliers; i.e., it is nonrobust.

A robust class of location parameter estimates for stationary and memoryless processes is defined by

$$\hat{\theta}^{\mathrm{M}}(\boldsymbol{x})\colon \frac{1}{N}\sum_{i=1}^{N} \psi[x_i - \hat{\theta}^{\mathrm{M}}(\boldsymbol{x})] = 0, \tag{4}$$

where $\psi(x)$ is a continuous, real and monotonically nondecreasing function such that $\psi(-\infty) < 0$ and $\psi(\infty) > 0$. If $\psi(x) = x$, then we obtain the empirical mean estimate in Eq. (3). The class of estimates in Eq. (4), generated by various choices of the function ψ, has been called M-estimates by Huber [6]. For the ψ function in Eq. (4), let us define the following:

$$m(\lambda) = \int_R \psi(x-\lambda) f(x)\,\mathrm{d}x,$$
$$\Sigma^2(\lambda) = \int_R \psi^2(x-\lambda) f(x)\,\mathrm{d}x, \quad \lambda \in R. \tag{5}$$

The asymptotic properties of the two estimates above are stated in the form of two theorems whose proofs can be found in [7].

Theorem 1. *Let the data be generated by a real, stationary and ergodic process. Then $\hat{\theta}^{\mathrm{E}}$ is a consistent estimate of the mean of the process. If the data generating process is stationary and memoryless with mean θ and variance σ^2, then the random variable $\hat{\theta}^{\mathrm{E}}$ is also asymptotically Gaussian with mean θ and variance σ^2/N.*

Theorem 2. *Let the data be generated by a real, stationary and memoryless process, whose first-order density function is $f(x)$. Let $\psi(x)$ be a real and scalar function that is monotonically nondecreasing and is such that $\psi(-\infty) < 0$ and $\psi(\infty) > 0$. Let some λ_0 exist such that $m(\lambda_0) = 0$, where $m(\lambda)$ is given in Eq. (5). Suppose that $m(\lambda)$ and $\Sigma^2(\lambda)$ are both differentiable at $\lambda = \lambda_0$, and $-\infty < \dot{m}(\lambda_0) < 0$, and $\Sigma^2(\lambda) < \infty$, where $\dot{m} = \partial m/\partial\lambda$. Then, the* M-*estimate $\hat{\theta}^{\mathrm{M}}$ in Eq. (2) is such that $\sqrt{N}\hat{\theta}^{\mathrm{M}}$ is asymptotically Gaussian with mean $\sqrt{N}\lambda_0$ and variance $\Sigma^2(\lambda_0)/[\dot{m}(\lambda_0)]^2$. If λ_0 is the mean of the data generating process, then $\hat{\theta}^{\mathrm{M}}$ is a consistent location parameter estimate.*

Now, consider the multiple-sensor model described in the previous section. That is, let each sensor process its own data record to make an estimate, and let the central processor fuse the local estimates into a final estimate.

Ideally, each sensor and the central processor would employ the maximum likelihood estimation procedure. However, for the noise model in Eq. (2), the maximum likelihood solution is very complex and requires iterative techniques such as the expectation-maximization algorithm [3]. Therefore, we shall limit ourselves to the availability of simpler estimators such as the empirical mean and the M-estimates. In particular, we will let $\psi(x)$ in Eq. (4) be the following function:

$$\psi(x) = \begin{cases} d & \text{if } x \geqslant d, \\ x & \text{if } -d \leqslant x \leqslant d, \\ -d & \text{if } x \leqslant -d, \end{cases} \qquad (6)$$

where $d > 0$. The variable d is the trimming parameter, whose value should depend on the severity of the outliers.

A number of combinations of estimators between the sensors and the central processor are possible, including some hybrid structures. We define them below.

EE-estimator. This system deploys the empirical mean estimate at both sensor and central processor levels. Thus,

$$\hat{\theta}_j^{\mathrm{E}} = \frac{1}{N} \sum_{i=1}^{N} x_{ji}, \quad j = 1, \ldots, L, \qquad (7)$$

$$\hat{\theta}_c^{\mathrm{EE}} = \frac{1}{L} \sum_{j=1}^{L} \hat{\theta}_j^{\mathrm{E}}. \qquad (8)$$

ME-estimator. The sensors use the M-estimate defined by Eqs. (4) and (6), and the central processor employs the empirical mean estimate. Then,

$$\hat{\theta}_j^{\mathrm{M}}\colon \frac{1}{N} \sum_{i=1}^{N} \psi(x_{ji} - \hat{\theta}_j^{\mathrm{M}}) = 0, \quad j = 1, \ldots, L,$$

$$\hat{\theta}_c^{\mathrm{ME}} = \frac{1}{L} \sum_{j=1}^{L} \hat{\theta}_j^{\mathrm{M}}.$$

EM-estimator. The sensors use the empirical mean estimate and the central processor uses the M-estimate, defined by Eqs. (4) and (6). Thus, the local estimates are as in Eq. (7), and the global estimate is of the form

$$\hat{\theta}_c^{\mathrm{EM}}\colon \frac{1}{L} \sum_{j=1}^{L} \psi(\hat{\theta}_j^{\mathrm{E}} - \hat{\theta}_c^{\mathrm{EM}}) = 0,$$

where $\psi(x)$ is defined in Eq. (6).

EE- and ME-estimators are the least and the most complex of the three structures, respectively, since the implementation of the M-estimator is more involved than the simple empirical averager.

4. Convergence rates of decentralized estimates

Since all decentralized estimates in Section 3 are consistent, as will be shown in this section, one good measure of comparison is the rate of convergence; that is, the asymptotic ($N \to \infty, L \to \infty$) value of the variance. The location parameter problem for a density function can be formulated as estimating the mean of the density. Although this mean is zero for the density in Eq. (2), the convergence results in this study hold regardless of the particular value of the mean (in other words, convergence rates are invariant to signal amplitude). The mean and variance of the density function in Eq. (2) are as follows, respectively:

$$\begin{aligned} &E[x_{ji} | f_j] = 0, \quad j = 1, \ldots, L, \\ &\sigma^2 = E[x_{ji}^2 | f_j] = 1 + \varepsilon\mu^2, \quad j = 1, \ldots, L. \end{aligned} \qquad (9)$$

Proposition 1. *The rate of convergence of the* EE-*estimate is*

$$R_{\mathrm{EE}} = (1 + \varepsilon\mu^2)/LN.$$

Proof. From Eqs. (7) and (8),

$$\hat{\theta}_c^{\mathrm{EE}} = \frac{1}{LN} \sum_{j=1}^{L} \sum_{i=1}^{N} x_{ji}.$$

By the asymptotic normality of the empirical mean estimate (Theorem 1) and the independence of the sensors,

$$R_{\mathrm{EE}} = \sigma^2/LN.$$

The final result follows from Eq. (9). □

Proposition 2. *The* ME-*estimate is consistent and its minimum rate of convergence is*

$$R^*_{\mathrm{ME}} = \eta_\varepsilon(d^*)/LN < \frac{\pi}{2(1-\varepsilon)^2}\frac{1}{LN},$$

where

$$d^* = \arg\min_d \eta_\varepsilon(d),$$

$$\eta_\varepsilon(d) = \left.\frac{\Sigma^2(0)}{[\dot m(0)]^2}\right|_{\mathrm{ME}}$$

$$\approx \frac{((2-\varepsilon)/(1-\varepsilon))d^2 - 1 - 2(d^2-1)\Phi(d) - 2d\phi(d)}{(1-\varepsilon)[1-2\Phi(d)]^2}, \tag{10}$$

assuming that $\mu \gg d$ *and* $\varepsilon < 0.5$.

Proof. For the noise model in Eq. (2),

$$m(0) = 0. \tag{11}$$

In addition, the M-estimates have zero-mean by Eq. (11) and Theorem 2. Therefore, the ME-estimator produces consistent estimates for the density in Eq. (2). Under the assumption that $\mu \gg d$, $\Phi(d+\mu) \approx \Phi(\mu)$ and $\Phi(d-\mu) \approx \Phi(-\mu)$. Then, due to the asymptotic normality of M-estimates and the independence of the sensors, the asymptotic variance of the estimate $\hat\theta_c^{\mathrm{ME}}$ is equal to $\eta_\varepsilon(d)/LN$, where $\eta_\varepsilon(d)$ is defined in Eq. (10).

For $\varepsilon < 0.5$, $\eta_\varepsilon(d)$ is convex in d. Moreover,

$$\min_d \eta_\varepsilon(d) < \eta_\varepsilon(0) = \frac{\pi}{2(1-\varepsilon)^2}.$$

Thus, assuming that the parameter d is selected equal to d^*, the estimate $\hat\theta_c^{\mathrm{ME}}$ will converge to zero fastest, with rate proportional to $\eta_\varepsilon(d^*)$.

Notice that Proposition 2 essentially furnishes a recipe for optimally choosing the trimming parameter d by minimizing the rate of convergence. □

Proposition 3. *The* EM-*estimate is consistent with minimum rate of convergence*

$$R^*_{\mathrm{EM}} = (1+\varepsilon\mu^2)/LN.$$

Proof. By Theorem 1, it is known that $\hat\theta_j^{\mathrm{E}}$ is asymptotically normally distributed with zero mean and variance σ_{E}^2, where

$$\sigma_{\mathrm{E}}^2 = \sigma^2/N = (1+\varepsilon\mu^2)/N.$$

Since the normal density with zero mean is an even function and $\psi(x)$ in Eq. (6) is odd, it is easy to see that $m(0) = 0$, proving the consistency of the estimate $\hat\theta_c^{\mathrm{EM}}$.

Define

$$\gamma_\varepsilon(d) = \left.\frac{\Sigma^2(0)}{[\dot m(0)]^2}\right|_{\mathrm{EM}}$$

$$= \frac{2d^2 - \sigma_{\mathrm{E}}^2 - 2(d^2-\sigma_{\mathrm{E}}^2)\Phi(d/\sigma_{\mathrm{E}}) - 2\sigma_{\mathrm{E}} d\phi(d/\sigma_{\mathrm{E}})}{[1-2\Phi(d/\sigma_{\mathrm{E}})]^2}. \tag{12}$$

Then,

$$R^*_{\mathrm{EM}} = \min_d \gamma_\varepsilon(d)/L = \lim_{d\to\infty} \gamma_\varepsilon(d)/L = (1+\varepsilon\mu^2)/LN,$$

and the result follows. □

The EM-estimate performs best for $d \to \infty$, that is, when it is equivalent to the empirical mean estimate $\hat\theta_{\mathrm{EE}}$. For other values of d, the EE-estimate has better convergence properties despite the central-level robustness of the EM-estimate. We conclude that employing robust operations after the data has already been processed is useless and results in further performance deterioration. Indeed, because the local estimates are asymptotically Gaussian, the empirical mean estimate is expected to converge faster than the M-estimate at the central-level.

From Propositions 1 and 3, we see that both R_{EE} and R^*_{EM} are directly proportional to μ^2. Thus, in cases where extreme outliers occur with nonzero probability, the convergence rates of the EE- and EM-estimates become asymptotically slow, eventually resulting in divergence. On the other hand, R_{ME} is always below the quantity $\pi/2(1-\varepsilon)^2$, regardless of the value of μ, for $\mu \gg d$. Consistent and relatively fast convergence is achieved by resisting the outliers as soon as data are collected.

The lemma below formally proves the superiority of the ME-estimate over the other two estimates, in the presence of Gaussian outliers as in Eq. (2).

Lemma 1. *Let $\eta_\varepsilon(d)$ and $\gamma_\varepsilon(d)$ be defined as in Eqs.* (10) *and* (12), *respectively, and let $N \to \infty$. Then, assuming $\varepsilon \ll 1$,*

$$\eta_\varepsilon(d)/N < \gamma_\varepsilon(d)$$

for all d such that $d \ll \mu$.

Proof. From Eqs. (10) and (12), it can be shown that

$$\gamma_\varepsilon(d) \approx \sigma_{\mathrm{E}}^2 \eta_\varepsilon(d/\sigma_{\mathrm{E}})$$

for $\varepsilon \ll 1$. Furthermore,

$$d^* = \arg\min_d \gamma_\varepsilon(d) = \arg\min_d \eta_\varepsilon(d).$$

It follows that

$$\gamma_\varepsilon(d^*) \approx \sigma_{\mathrm{E}}^2 \eta_\varepsilon(d^*/\sigma_{\mathrm{E}}) \geqslant (1 + \varepsilon\mu^2)\eta_\varepsilon(d^*)/N > \eta_\varepsilon(d^*)/N,$$

where the first inequality is due to the convexity of $\eta_\varepsilon(d)$. □

The next corollary summarizes the convergence results we have obtained so far, and it is a direct consequence of Propositions 1–3 and Lemma 1.

Corollary 1. *Let $\varepsilon \ll 1$ and $d \ll \mu$. Then,*

$$R_{\mathrm{ME}}^* < R_{\mathrm{EM}}^* = R_{\mathrm{EE}},$$

where R_{EE} is the convergence rate of the EE-*estimate, and R_{ME}^* and R_{EM}^* are the minimum rates of convergence of the* ME- *and* EM-*estimates, respectively.*

We conclude this section with a table that illustrates the magnitude of advantage the ME-estimate yields compared to the other two estimates. Table 1 displays the R_{ME}^*, R_{EE} and R_{EM}^* values for various choices of ε. As expected, the optimal trimming parameter d^* is very small and inversely proportional to ε. For $\mu < 3$, the deviation from Gaussianity is very minor so that the EE- and EM-estimators produce slightly better performance. However for $\mu > 5$, the ME-estimate converges much faster than its competition. Hence, the performance guarantee of the ME-estimator in the presence of extreme Gaussian outliers comes at the expense of comparably negligible increase in the convergence rate under nominal or near-nominal conditions. This favorable trade-off is typical of robust operations. Furthermore, R_{ME}^* appears to be less sensitive to ε, which is an additional advantage since the exact value of ε is unknown.

Further evidence supporting the robustness of the ME-estimate is given in Section 5.

5. Breakdown points

One measure to study the degree of robustness of an estimate is the breakdown point, introduced by Hampel [5]. Given a parameter estimate, its breakdown point ε^* is the supremum of the frequency of extreme amplitude outliers that the estimate can tolerate, while still converging to a function of the true parameter value [7]. In this section, we find the

Table 1
Convergence rates versus the outlier frequency ε. While R_{ME}^* is insensitive to outlier amplitudes, R_{EE} and R_{EM}^* are calculated for a range of μ values (all numbers are normalized by LN)

			$R_{\mathrm{EE}} = R_{\mathrm{EM}}^*$				
ε	d^*	R_{ME}^*	$\mu = 1$	$\mu = 2$	$\mu = 5$	$\mu = 10$	$\mu = 100$
0.001	2.632880	1.00955	1.001	1.004	1.025	1.1	11
0.01	1.945110	1.06524	1.01	1.04	1.25	2	101
0.10	1.140170	1.48985	1.1	1.4	3.5	11	1001
0.20	0.861592	2.04553	1.2	1.8	6	21	2001
0.30	0.684476	2.82145	1.3	2.2	8.5	31	3001
0.40	0.549156	3.99583	1.4	2.6	11	41	4001
0.49	0.446892	5.68266	1.49	2.96	13.25	50	4901

breakdown points of the estimation schemes introduced earlier.

Suppose that with probability $1-\varepsilon$, the data are generated by a stationary, memoryless and unit-variance Gaussian process whose mean is θ. Let independent outliers of constant amplitude y occur with probability ε.

Consider, first, the decentralized EE-estimator where both sensors and the central processor use empirical mean estimates. The estimate $\hat{\theta}_j^{\mathrm{E}}$ converges asymptotically ($N \to \infty$) to the value $(1-\varepsilon)\theta + \varepsilon y$. Thus,

$$\lim_{N\to\infty} \hat{\theta}_c^{\mathrm{EE}} = (1-\varepsilon)\theta + \varepsilon y. \tag{13}$$

When the outlier amplitude y tends to $\pm\infty$, $\hat{\theta}_c^{\mathrm{EE}}$ converges to εy, which is not a function of θ, for every nonzero value of ε. As a result, the breakdown point $\varepsilon_{\mathrm{EE}}^*$ of the EE-estimate is

$$\varepsilon_{\mathrm{EE}}^* = 0.$$

The breakdown point of the ME-estimate is the same as that of the M-estimate. Therefore,

$$\varepsilon_{\mathrm{ME}}^* = 0.5,$$

as shown in [7].

Finally, from Eq. (13),

$$\lim_{N\to\infty} \hat{\theta}_c^{\mathrm{EM}} = \psi[(1-\varepsilon)\theta + \varepsilon y],$$

where ψ is as in Eq. (6). As $y \to \pm\infty$, we observe that $\hat{\theta}_c^{\mathrm{EM}} \to \pm d$. Thus, the EM-estimate does not converge to a function of θ for any value of ε, and

$$\varepsilon_{\mathrm{EM}}^* = 0.$$

Comparing the breakdown points of the decentralized estimates discussed in this paper, we notice that the EE- and EM-estimates have no tolerance to outliers. Both estimates fail to converge to a function of the true parameter when the amplitude of the outlier asymptotically dominates the main process. The robust ME-estimator is able to produce acceptable estimates as long as $\varepsilon < 0.5$.

6. Discussion

We have provided a simple study of the convergence and robustness properties of some consistent multiple-sensor location parameter estimates. The results indicate that robust procedures such as the M-estimate should be employed at stages where the observations are suspected to be corrupted by outliers. Systems that fail to combat outliers in a timely fashion display extremely poor performance in terms of the breakdown point, and tend not to converge.

We have neglected the MM-estimator that utilizes M-estimates at both sensor- and central-levels in our investigations. Although the rate of convergence and breakdown point derivations for the MM-estimate are more involved, we can extrapolate our findings to this structure. In particular, since the M-estimate does not perform as well as the empirical mean estimate for Gaussian data, it can be safely assumed that the MM-estimate converges slower than $\hat{\theta}_c^{\mathrm{ME}}$ when the transmission between sensors and the central processor is noiseless. This is in agreement with the well-known fact that robust procedures result in performance deterioration when there is no need for them. On the other hand, in situations where the transmission channels between the sensors and the central processor are noisy, the MM-estimator is the only truely robust structure, with nonzero breakdown point and guaranteed convergence.

Finally, in the event that sensor observations are corrupted by varying frequencies of outliers, a weighted combination of local estimates should be computed. If sensor j operates under the mixture model in Eq. (2) with outlier frequency ε_j, $j = 1, \ldots, L$ (where ε_j are distinct), then the optimal linear fusion rule becomes

$$\hat{\theta}_c^{\mathrm{ME}} = \frac{1}{L}\sum_{j=1}^{L} c_j \hat{\theta}_j^{\mathrm{M}},$$

where

$$c_j = \begin{cases} \dfrac{1-\varepsilon_j}{\sum_{j=1}^{L}(1-\varepsilon_j)} & \text{if } \varepsilon_j < 0.5, \\ 0 & \text{if } \varepsilon_j \geqslant 0.5, \end{cases}$$

and sensors supply the pairs $\{\hat{\theta}_j^{\mathrm{M}}, c_j\}_{j=1}^{L}$ to the central processor. The weight c_j reflects the confidence of the central processor in sensor j's estimate $\hat{\theta}_j^{\mathrm{M}}$, for all j. In practice, it is very hard to know the outlier frequencies, and guesses may have to be made based on empirical observations.

Acknowledgements

The author would like to thank Dr. Dimitri Kazakos for helpful discussions.

References

[1] M. Bouvet, S.C. Schwartz, Comparison of adaptive and robust receivers for signal detection in ambient underwater noise, IEEE Trans. Acoust. Speech Signal Process. 37 (5) (1989) 621–626.

[2] D.A. Castanon, D. Teneketzis, Distributed estimation algorithms for nonlinear systems, IEEE Trans. Automat. Control 30 (5) (1985) 418–425.

[3] A.P. Dempster, N.M. Laird, D.B. Rubin, Maximum-likelihood from incomplete data via the EM algorithm, J. Roy. Statist. Soc. Ser. B. 39 (1977) 1–17.

[4] A. Elfes, Sonar-based real world mapping and navigation, IEEE J. Robotics Automat. 3 (3) (1987) 249–265.

[5] F.R. Hampel, A general qualitative definition of robustness, Ann. Math. Statist. 42 (6) (1971) 1887–1895.

[6] P.J. Huber, Robust estimation of a location parameter, Ann. Math. Statist. 35 (2) (1964) 73–101.

[7] D. Kazakos, P. Papantoni-Kazakos, Detection and Estimation, Computer Science Press, New York, NY, 1990.

[8] D. Middleton, Statistical–physical models of electromagnetic interference, IEEE Trans. Electromagnetic Compatibility 19 (3) (1977) 106–127.

[9] P.F. Swaszek, A.J. Efron, D.W. Tufts, Detection based on the mixture model, Proc. 19th Ann. Conf. Information Sciences and Systems, Baltimore, Maryland, March 1985, pp. 120–125.