SEQUENTIAL M-ESTIMATION

Duc Son Pham, Yee Hong Leung *

Communications and Signal Processing Group,
Curtin University of Technology,
GPO Box U1987, WA 6845, Australia
dspham@ieee.org

Abdelhak Zoubir, Ramon Brcic

Signal Processing Group,
Darmstadt University of Technology,
Merckstrasse 25, D-64283 Darmstadt, Germany
zoubir@ieee.org

ABSTRACT

We propose a sequential M-estimation algorithm as an alternative to sequential least squares. Being an approximation of the exact M-estimator, the proposed technique is robust to non-Gaussian processes and outperforms sequential least squares. Simulation results demonstrate the power of the proposed sequential M-estimator.

1. INTRODUCTION

Consider the linear signal model,

$$\boldsymbol{y} = \boldsymbol{H}\boldsymbol{\theta} + \boldsymbol{x}, \tag{1}$$

where $\boldsymbol{y} \in \mathbb{R}^N$ is the observation vector, $\boldsymbol{H} \in \mathbb{R}^{N\times P}$ is the system matrix, $\boldsymbol{\theta} \in \mathbb{R}^P$ is the vector of P unknown parameters to be estimated, and $\boldsymbol{x} \in \mathbb{R}^N$ is the noise vector, with a non-Gaussian density function $f(x)$. Specifically, we are interested in the case where $f(x)$ can be described by the ε-contaminated model

$$f(x) = (1-\varepsilon) f_G(x;\nu^2) + \varepsilon\mathcal{I}(x), \tag{2}$$

where ε represents the contamination percentage, ν^2 is the variance of the Gaussian background, f_G is the zero-mean Gaussian distribution, and $\mathcal{I}(x)$ is some unknown symmetric function representing the impulsive part of the noise.

In many situations, the observations $y_1, y_2, \ldots, y_n$ arrive sequentially. It is of practical interest to compute the estimate of the unknown parameter vector for a small number of observations and then update the estimate whenever a new observation arrives. The goal of this work is to derive the update of the parameter estimates for non-Gaussian noise. The signal model for the nth observation becomes

$$\boldsymbol{y}[n] = \boldsymbol{H}[n]\boldsymbol{\theta} + \boldsymbol{x}[n], \tag{3}$$

and for the next observation

$$\boldsymbol{y}[n+1] = [y_1, y_2, \ldots, y_n, y_{n+1}]^T = \begin{bmatrix} \boldsymbol{y}[n] \\ y_{n+1} \end{bmatrix}, \tag{4}$$

$$\boldsymbol{H}[n+1] = \begin{bmatrix} \boldsymbol{h}_1^T \\ \boldsymbol{h}_2^T \\ \vdots \\ \boldsymbol{h}_{n+1}^T \end{bmatrix} = \begin{bmatrix} \boldsymbol{H}[n] \\ \boldsymbol{h}_{n+1}^T \end{bmatrix} = \begin{bmatrix} n \times P \\ 1 \times P \end{bmatrix}. \tag{5}$$

When the noise is purely Gaussian, i.e. $\varepsilon = 0$, the following sequential LS algorithm is optimum in the maximum likelihood sense [5].

$$\hat{\boldsymbol{\theta}}[n+1] = \hat{\boldsymbol{\theta}}[n] + \boldsymbol{K}[n+1]\left(y[n+1] - \boldsymbol{h}_{n+1}^T\hat{\boldsymbol{\theta}}[n]\right), \tag{6}$$

$$\boldsymbol{K}[n+1] = \frac{\boldsymbol{\Sigma}[n]\boldsymbol{h}_{n+1}}{\nu_n^2 + \boldsymbol{h}_{n+1}^T\boldsymbol{\Sigma}[n]\boldsymbol{h}_{n+1}}, \tag{7}$$

$$\boldsymbol{\Sigma}[n+1] = \left(\boldsymbol{I} - \boldsymbol{K}[n+1]\boldsymbol{h}_{n+1}^T\right)\boldsymbol{\Sigma}[n], \tag{8}$$

$$\boldsymbol{\Sigma}[n_0] = \left(\boldsymbol{H}^T[n_0]\boldsymbol{C}^{-1}[n_0]\boldsymbol{H}[n_0]\right)^{-1} \tag{9}$$

where $\boldsymbol{C}[n_0] = \mathrm{diag}(\nu_1^2, \nu_2^2, \ldots, \nu_{n_0}^2)$, and $\nu_1^2, \nu_2^2, \ldots, \nu_n^2$ are the variances of the noise at observations $1, 2, \ldots, n$. It is normally assumed $\nu_1^2 = \nu_2^2 = \ldots = \nu_n^2 = \nu^2$. However, in the presence of non-Gaussian noise, Gaussian-optimised solutions are *not robust* [2, 4, 9]. In robust statistics, instead of using the quadratic LS cost function, a less increasing cost function $\rho(x)$ is used. Under regularity conditions, the robust estimate of the unknown parameter $\boldsymbol{\theta}$ for the signal model (3) is found by solving the M-equations

$$\boldsymbol{H}[n]^T\psi\left(\boldsymbol{y}[n] - \boldsymbol{H}[n]\boldsymbol{\theta}\right) = \boldsymbol{0}_P, \tag{10}$$

where $\psi(x) = \partial\rho(x)/\partial x$ is the score function, and $\psi\left(\boldsymbol{y}[n] - \boldsymbol{H}[n]\boldsymbol{\theta}\right)$ represents a vector whose ith entry is $\psi(y_i - \boldsymbol{h}_i^T\boldsymbol{\theta})$, and $\boldsymbol{h}_i^T$ is the ith row of $\boldsymbol{H}[n]$ [4, 9]. Designing score functions is the key to robust estimation. Note that the Gaussian-optimised solution can be derived from (10) with the score function $\psi(x) = x/\nu^2$. One well-known score function for the ε-contaminated noise model is the minimax score function [2]

$$\psi(x) = \begin{cases} \frac{x}{\nu^2} & \text{for } |x| \le k\nu^2 \\ k\,\mathrm{sign}(x) & \text{for } |x| > k\nu^2, \end{cases} \tag{11}$$

where k is dependent on ε and ν via

$$\frac{\phi(k\nu)}{k\nu} - Q(k\nu) = \frac{\varepsilon}{2(1-\varepsilon)}, \tag{12}$$

in which $\phi(x) = \frac{1}{\sqrt{2\pi}}e^{-x^2/2}$ and $Q(x) = \frac{1}{\sqrt{2\pi}}\int_x^\infty e^{-t^2/2}dt$ [9]. When ε and ν are known, the minimax score function can be constructed.

In Section 2, we derive a robust sequential algorithm for updating the estimate of the vector of unknown parameters $\boldsymbol{\theta}$ in (1) when the noise is non-Gaussian under the framework of robust statistics. In Section 3, we present some simulation results to illustrate the use of this technique. Section 4 concludes the paper.

*This research work was partly supported by the Australian Telecommunications CRC.

2. SEQUENTIAL M-ESTIMATION

2.1. The algorithm

The robust estimates of $\boldsymbol{\theta}$ at the nth and $(n+1)$th observations satisfy the following M-equations

$$\boldsymbol{H}^T[n]\psi\left(\boldsymbol{y}[n]-\boldsymbol{H}[n]\hat{\boldsymbol{\theta}}[n]\right)=\mathbf{0}_P \tag{13}$$

and

$$\boldsymbol{H}^T[n+1]\psi\left(\boldsymbol{y}[n+1]-\boldsymbol{H}[n+1]\hat{\boldsymbol{\theta}}[n+1]\right)=\mathbf{0}_P. \tag{14}$$

Denoting

$$\boldsymbol{g}_{\boldsymbol{\theta}}[n+1]=\boldsymbol{H}^T[n+1]\psi\left(\boldsymbol{y}[n+1]-\boldsymbol{H}[n+1]\boldsymbol{\theta}\right), \tag{15}$$

and using Taylor series expansion up to the first order of $\boldsymbol{g}_{\boldsymbol{\theta}}[n+1]$ around $\hat{\boldsymbol{\theta}}[n]$, we have

$$\boldsymbol{g}_{\boldsymbol{\theta}}[n+1]=\boldsymbol{g}_{\hat{\boldsymbol{\theta}}[n]}[n+1]+\left.\frac{\partial \boldsymbol{g}_{\boldsymbol{\theta}}[n+1]}{\partial \boldsymbol{\theta}}\right|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}[n]}\left(\boldsymbol{\theta}-\hat{\boldsymbol{\theta}}[n]\right). \tag{16}$$

Noting that $\boldsymbol{g}_{\hat{\boldsymbol{\theta}}[n+1]}[n+1]=\mathbf{0}_P$ and rearranging (16), we have

$$\hat{\boldsymbol{\theta}}[n+1]=\hat{\boldsymbol{\theta}}[n]-\boldsymbol{J}[n+1]\boldsymbol{g}_{\hat{\boldsymbol{\theta}}[n]}[n+1], \tag{17}$$

where

$$\boldsymbol{J}[n+1]=\left(\left.\frac{\partial \boldsymbol{g}_{\boldsymbol{\theta}}[n+1]}{\partial \boldsymbol{\theta}}\right|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}[n]}\right)^{-1}. \tag{18}$$

To derive the update for $\boldsymbol{J}[n+1]$, given $\boldsymbol{J}[n]$, we define $z_i(\boldsymbol{\theta})=y_i-\boldsymbol{h}_i^T\boldsymbol{\theta},\ \ i=1,2,\ldots,n,n+1$, and start from the definition of the gradient of a row vector [8]

$$\begin{aligned}
\frac{\partial \boldsymbol{g}_{\boldsymbol{\theta}}[n+1]}{\partial \boldsymbol{\theta}} &= \boldsymbol{H}^T[n+1]\frac{\partial}{\partial \boldsymbol{\theta}}\begin{bmatrix}\psi\left(y_1-\boldsymbol{h}_1^T\boldsymbol{\theta}\right)\\ \psi\left(y_2-\boldsymbol{h}_2^T\boldsymbol{\theta}\right)\\ \vdots\\ \psi\left(y_{n+1}-\boldsymbol{h}_{n+1}^T\boldsymbol{\theta}\right)\end{bmatrix}\\
&= \boldsymbol{H}^T[n+1]\begin{bmatrix}\nabla_{\boldsymbol{\theta}}^T\psi(z_1(\boldsymbol{\theta}))\\ \nabla_{\boldsymbol{\theta}}^T\psi(z_2(\boldsymbol{\theta}))\\ \vdots\\ \nabla_{\boldsymbol{\theta}}^T\psi(z_{n+1}(\boldsymbol{\theta}))\end{bmatrix}\\
&= \boldsymbol{H}^T[n+1]\begin{bmatrix}\boldsymbol{h}_1^T\gamma(z_1(\boldsymbol{\theta}))\\ \boldsymbol{h}_2^T\gamma(z_2(\boldsymbol{\theta}))\\ \vdots\\ \boldsymbol{h}_{n+1}^T\gamma(z_{n+1}(\boldsymbol{\theta}))\end{bmatrix}\\
&= \boldsymbol{H}^T[n+1]\boldsymbol{C}_{\boldsymbol{\theta}}[n+1]\boldsymbol{H}[n+1],
\end{aligned} \tag{19}$$

where $\nabla_{\boldsymbol{\theta}}\psi(x)=\left[\frac{\partial}{\partial\theta_1}\psi(x),\ldots,\frac{\partial}{\partial\theta_P}\psi(x)\right]^T$ denotes the gradient of a scalar function $\psi(x)$ with respect to $\boldsymbol{\theta}$ and

$$\begin{aligned}
\boldsymbol{C}_{\boldsymbol{\theta}}[n+1] &= \operatorname{diag}\Big(\gamma(z_1(\boldsymbol{\theta})),\gamma(z_2(\boldsymbol{\theta})),\ldots,\gamma(z_{n+1}(\boldsymbol{\theta}))\Big)\\
&= \begin{bmatrix}\gamma(z_1(\boldsymbol{\theta})) & 0 & \ldots & 0\\ 0 & \gamma(z_2(\boldsymbol{\theta})) & \ldots & 0\\ \vdots & \vdots & \vdots & \vdots\\ 0 & 0 & \ldots & \gamma(z_{n+1}(\boldsymbol{\theta}))\end{bmatrix},
\end{aligned} \tag{20}$$

and we have assumed that $\gamma(x)=-\partial\psi(x)/\partial x$ exists. Define

$$\begin{aligned}
\boldsymbol{F}_{\boldsymbol{\theta}}[n+1] &= \left(\boldsymbol{H}^T[n+1]\boldsymbol{C}_{\boldsymbol{\theta}}[n+1]\boldsymbol{H}[n+1]\right)^{-1}\\
&= \left(\begin{bmatrix}\boldsymbol{H}[n]\\ \boldsymbol{h}_{n+1}^T\end{bmatrix}^T\begin{bmatrix}\boldsymbol{C}_{\boldsymbol{\theta}}[n] & \mathbf{0}\\ \mathbf{0}^T & \gamma(z_{n+1}(\boldsymbol{\theta}))\end{bmatrix}\begin{bmatrix}\boldsymbol{H}[n]\\ \boldsymbol{h}_{n+1}^T\end{bmatrix}\right)^{-1}\\
&= \left(\boldsymbol{H}^T[n]\boldsymbol{C}_{\boldsymbol{\theta}}[n]\boldsymbol{H}[n]+\gamma(z_{n+1}(\boldsymbol{\theta}))\boldsymbol{h}_{n+1}\boldsymbol{h}_{n+1}^T\right)^{-1}.
\end{aligned} \tag{21}$$

Noting that $z_{n+1}(\boldsymbol{\theta})=y_{n+1}-\boldsymbol{h}_{n+1}^T\boldsymbol{\theta}$, and using Woodbury's identity [5], we have

$$\boldsymbol{F}_{\boldsymbol{\theta}}[n+1]=\boldsymbol{F}_{\boldsymbol{\theta}}[n]-\frac{\gamma\left(y_{n+1}-\boldsymbol{h}_{n+1}^T\boldsymbol{\theta}\right)\boldsymbol{F}_{\boldsymbol{\theta}}[n]\boldsymbol{h}_{n+1}\boldsymbol{h}_{n+1}^T\boldsymbol{F}_{\boldsymbol{\theta}}[n]}{1+\gamma\left(y_{n+1}-\boldsymbol{h}_{n+1}^T\boldsymbol{\theta}\right)\boldsymbol{h}_{n+1}^T\boldsymbol{F}_{\boldsymbol{\theta}}[n]\boldsymbol{h}_{n+1}}. \tag{22}$$

Further, with

$$\boldsymbol{J}[n+1] = \boldsymbol{F}_{\hat{\boldsymbol{\theta}}[n]}[n+1], \tag{23}$$

$$\boldsymbol{J}[n] = \boldsymbol{F}_{\hat{\boldsymbol{\theta}}[n-1]}[n], \tag{24}$$

and the possibility of inferring $\boldsymbol{F}_{\hat{\boldsymbol{\theta}}[n]}[n+1]$ from $\boldsymbol{F}_{\hat{\boldsymbol{\theta}}[n]}[n]$ using Equation (22), the task then is to find a relation between $\boldsymbol{F}_{\hat{\boldsymbol{\theta}}[n-1]}[n]$ and $\boldsymbol{F}_{\hat{\boldsymbol{\theta}}[n]}[n]$. Continuing, we have

$$\boldsymbol{F}_{\hat{\boldsymbol{\theta}}[n]}[n] = \left(\boldsymbol{H}^T[n]\boldsymbol{C}_{\hat{\boldsymbol{\theta}}[n]}[n]\boldsymbol{H}[n]\right)^{-1}, \tag{25}$$

$$\boldsymbol{F}_{\hat{\boldsymbol{\theta}}[n-1]}[n] = \left(\boldsymbol{H}^T[n]\boldsymbol{C}_{\hat{\boldsymbol{\theta}}[n-1]}[n]\boldsymbol{H}[n]\right)^{-1}, \tag{26}$$

where

$$\boldsymbol{C}_{\hat{\boldsymbol{\theta}}[n-1]}[n]=\operatorname{diag}\left(\gamma(z_1(\hat{\boldsymbol{\theta}}[n-1])),\ldots,\gamma(z_n(\hat{\boldsymbol{\theta}}[n-1]))\right) \tag{27}$$

and

$$\boldsymbol{C}_{\hat{\boldsymbol{\theta}}[n]}[n]=\operatorname{diag}\left(\gamma(z_1(\hat{\boldsymbol{\theta}}[n])),\ldots,\gamma(z_n(\hat{\boldsymbol{\theta}}[n]))\right). \tag{28}$$

Denoting

$$\delta\boldsymbol{C}_{\hat{\boldsymbol{\theta}}[n]}=\boldsymbol{C}_{\hat{\boldsymbol{\theta}}[n]}[n]-\boldsymbol{C}_{\hat{\boldsymbol{\theta}}[n-1]}[n], \tag{29}$$

we have

$$\begin{aligned}
\boldsymbol{F}_{\hat{\boldsymbol{\theta}}[n]}[n] &= \left(\boldsymbol{H}^T[n]\boldsymbol{C}_{\hat{\boldsymbol{\theta}}[n]}\boldsymbol{H}[n]\right)^{-1}\\
&= \left(\boldsymbol{H}^T[n]\boldsymbol{C}_{\hat{\boldsymbol{\theta}}[n-1]}\boldsymbol{H}[n]+\boldsymbol{H}^T[n]\delta\boldsymbol{C}_{\hat{\boldsymbol{\theta}}[n]}\boldsymbol{H}[n]\right)^{-1}\\
&= \left(\boldsymbol{F}_{\hat{\boldsymbol{\theta}}[n-1]}^{-1}[n]+\boldsymbol{H}^T[n]\delta\boldsymbol{C}_{\hat{\boldsymbol{\theta}}[n]}\boldsymbol{H}[n]\right)^{-1}.
\end{aligned} \tag{30}$$

To proceed, we will make use of the following result [1]

$$\boldsymbol{B}^{-1}=\boldsymbol{A}^{-1}-\boldsymbol{B}^{-1}(\boldsymbol{B}-\boldsymbol{A})\boldsymbol{A}^{-1}. \tag{31}$$

Assuming that the difference between $\boldsymbol{B}^{-1}$ and $\boldsymbol{A}^{-1}$ is negligible, we propose to modify the above identity to obtain the following approximation

$$\boldsymbol{B}^{-1}\approx\boldsymbol{A}^{-1}-\boldsymbol{A}^{-1}(\boldsymbol{B}-\boldsymbol{A})\boldsymbol{A}^{-1}. \tag{32}$$

Substituting

$$\boldsymbol{B}=\boldsymbol{H}^T[n]\boldsymbol{C}_{\hat{\boldsymbol{\theta}}[n]}[n]\boldsymbol{H}[n], \tag{33}$$

$$\boldsymbol{A}=\boldsymbol{H}^T[n]\boldsymbol{C}_{\hat{\boldsymbol{\theta}}[n-1]}[n]\boldsymbol{H}[n], \tag{34}$$

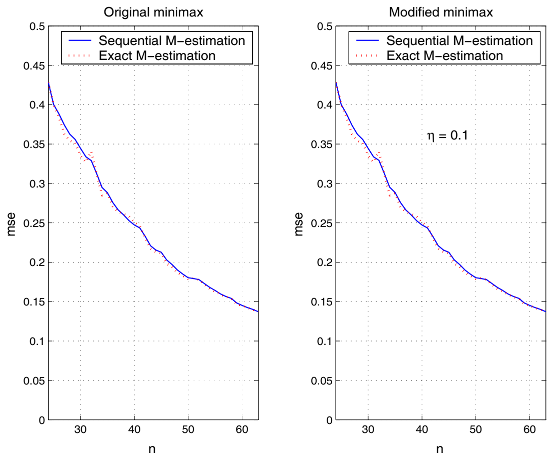

Fig. 1. Minimax and modified minimax performance.

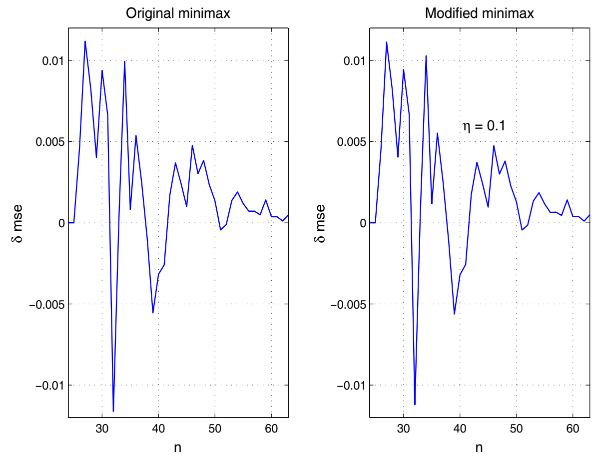

Fig. 2. Relative minimax and modified minimax performance.

and noting from the result above that

$$\boldsymbol{B} - \boldsymbol{A} = \boldsymbol{H}^T[n]\delta\boldsymbol{C}_{\hat{\boldsymbol{\theta}}[n]}\boldsymbol{H}[n], \tag{35}$$

$$\boldsymbol{F}_{\hat{\boldsymbol{\theta}}[n-1]}[n] = \left(\boldsymbol{H}^T[n]\boldsymbol{C}_{\hat{\boldsymbol{\theta}}[n-1]}[n]\boldsymbol{H}[n]\right)^{-1}, \tag{36}$$

we have

$$\begin{aligned}\boldsymbol{F}_{\hat{\boldsymbol{\theta}}[n]}[n] &= \left(\boldsymbol{H}^T[n]\boldsymbol{C}_{\hat{\boldsymbol{\theta}}[n]}\boldsymbol{H}[n]\right)^{-1} \\ &= \boldsymbol{F}_{\hat{\boldsymbol{\theta}}[n-1]}[n]\left(\boldsymbol{I} - \left(\boldsymbol{H}^T[n]\delta\boldsymbol{C}_{\hat{\boldsymbol{\theta}}[n]}\boldsymbol{H}[n]\right)\boldsymbol{F}_{\hat{\boldsymbol{\theta}}[n-1]}[n]\right),\end{aligned} \tag{37}$$

which is the relation of interest. Finally, it is noted that

$$\begin{aligned}\boldsymbol{g}_{\hat{\boldsymbol{\theta}}[n]}[n+1] &= \begin{bmatrix}\boldsymbol{H}^T[n] \\ \boldsymbol{h}_{n+1}\end{bmatrix}^T \psi\left(\begin{bmatrix}\boldsymbol{y}[n] \\ y_{n+1}\end{bmatrix} - \begin{bmatrix}\boldsymbol{H}[n] \\ \boldsymbol{h}_{n+1}^T\end{bmatrix}\hat{\boldsymbol{\theta}}[n]\right) \\ &= \boldsymbol{h}_{n+1}\psi(y_{n+1} - \boldsymbol{h}_{n+1}^T\hat{\boldsymbol{\theta}}[n]),\end{aligned} \tag{38}$$

since by definition $\boldsymbol{H}^T[n]\psi(\boldsymbol{y}[n] - \boldsymbol{H}[n]\hat{\boldsymbol{\theta}}[n]) = \boldsymbol{0}$.

Hence, the sequential M-estimation algorithm can be summarised as follow.

- *Compute* $\delta\boldsymbol{C}_{\hat{\boldsymbol{\theta}}[n]} = \boldsymbol{C}_{\hat{\boldsymbol{\theta}}[n]}[n] - \boldsymbol{C}_{\hat{\boldsymbol{\theta}}[n-1]}[n]$, *where the diagonal matrices are defined in (27) and (28).*
- *Compute*

$$\boldsymbol{F}_{\hat{\boldsymbol{\theta}}[n]}[n] = \boldsymbol{J}[n]\left(\boldsymbol{I} - \left(\boldsymbol{H}^T[n]\delta\boldsymbol{C}_{\hat{\boldsymbol{\theta}}[n]}\boldsymbol{H}[n]\right)\boldsymbol{J}[n]\right).$$

- *Update the vector*

$$\begin{aligned}\boldsymbol{J}[n+1] &= \boldsymbol{F}_{\hat{\boldsymbol{\theta}}[n]}[n] \\ &- \frac{\gamma\left(y_{n+1} - \boldsymbol{h}_{n+1}^T\hat{\boldsymbol{\theta}}[n]\right)\boldsymbol{F}_{\hat{\boldsymbol{\theta}}[n]}[n]\boldsymbol{h}_{n+1}\boldsymbol{h}_{n+1}^T\boldsymbol{F}_{\hat{\boldsymbol{\theta}}[n]}[n]}{1 + \gamma\left(y_{n+1} - \boldsymbol{h}_{n+1}^T\hat{\boldsymbol{\theta}}[n]\right)\boldsymbol{h}_{n+1}^T\boldsymbol{F}_{\hat{\boldsymbol{\theta}}[n]}[n]\boldsymbol{h}_{n+1}}.\end{aligned}$$

- *Update the parameter estimates*

$$\hat{\boldsymbol{\theta}}[n+1] = \hat{\boldsymbol{\theta}}[n] - \boldsymbol{J}[n+1]\boldsymbol{h}_{n+1}\psi(y_{n+1} - \boldsymbol{h}_{n+1}^T\hat{\boldsymbol{\theta}}[n])$$

To use the sequential updates, one must start with some *robust* estimate of the unknown parameter for a given number of observations n_0, i.e. by solving the following M-equations

$$\boldsymbol{g}_{\boldsymbol{\theta}}[n_0] = \boldsymbol{H}^T[n_0]\psi\left(\boldsymbol{y}[n_0] - \boldsymbol{H}[n_0]\hat{\boldsymbol{\theta}}[n_0]\right) = \boldsymbol{0}_P. \tag{39}$$

Numerical procedures for solving M-equations can be found, for example, in [4, 7, 6, 9]. For identifiability reason [3, 7], $n_0 > P$, where P is the number of parameters. If the total number of observations $N >> P$, it is recommended to select $n_0 \geq 4P$ [3].

2.2. Remarks

We note that while the sequential LS algorithm yields the *exact* updates, the proposed sequential M-estimation technique only provides *approximate* updates. This is because of the difficulty when dealing with the nonlinearity of the score function $\psi(x)$. In particular, we have used Taylor series expansion approximation and a matrix inversion approximation. By doing so, we have also implicitly assumed that the estimates are reasonably close to the true value in a sense that these approximations are satisfactory. Finally, it has been assumed in the derivation that $\gamma(x)$ exists. For the minimax score function this assumption does not hold at the corner points. It can be resolved by smoothing the minimax score function as described in [9]. The modified minimax score function is reproduced here for convenience.

$$\psi(x) = \begin{cases}\frac{x}{\nu^2} & \text{if } |x| \leq (k-\eta)\nu^2 \\ k - \eta + \eta\tanh\left(\frac{x-(k-\eta)\nu^2}{\eta\nu^2}\right) & \text{if } x > (k-\eta)\nu^2 \\ -(k-\eta) + \eta\tanh\left(\frac{x+(k-\eta)\nu^2}{\eta\nu^2}\right) & \text{if } x < -(k-\eta)\nu^2\end{cases}$$

where $\eta << k$ is a small number. However, it is found in the example considered that there is not much difference between the performance of the modified and the original minimax score functions.

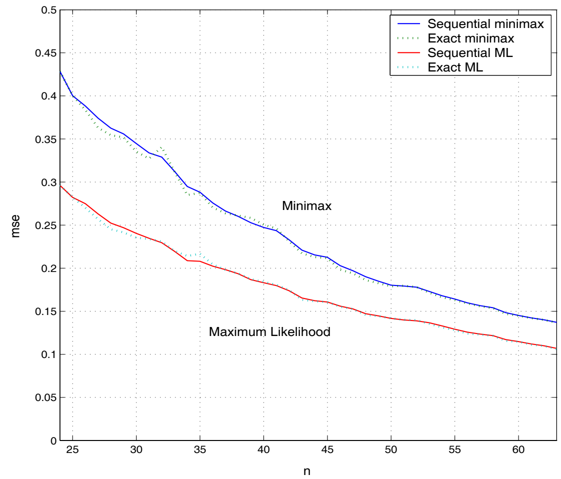

Fig. 3. Performance of the minimax and ML algorithms.

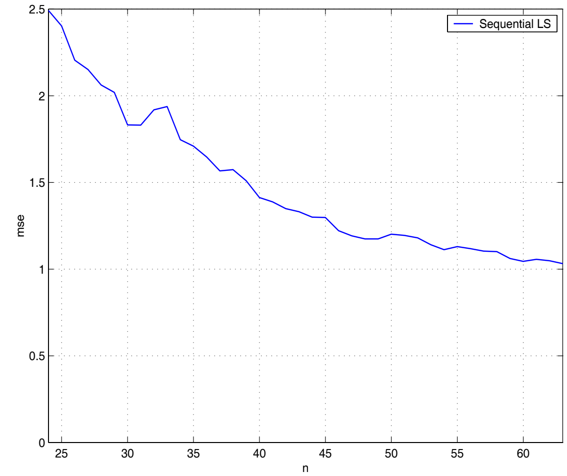

Fig. 4. Performance of the sequential LS technique.

3. SIMULATION RESULTS

In this section, we give one example to illustrate the proposed sequential M-estimation algorithm. The impulsive part of the noise $\mathcal{I}(x)$ is modelled by another Gaussian function $\mathcal{I}(x) = f_G(x; \kappa\nu^2)$ where $\kappa >> 1$. This is particularly relevant to non-Gaussian interference encountered in radio communications [9]. In the simulation, we select $\varepsilon = 0.1$, and $\kappa = 100$. The number of unknown parameters $P = 6$, the total number of observations $N = 63$, and we choose the starting point of the sequential M-estimation algorithm to be $n_0 = 24$ according to the guideline discussed above.

In this Monte-Carlo simulation, we measure the average of $(\hat{\boldsymbol{\theta}}[n] - \boldsymbol{\theta})^T(\hat{\boldsymbol{\theta}}[n] - \boldsymbol{\theta})$ where $\boldsymbol{\theta}$ is the true value of the parameter vector. We also compare this with the one obtained by directly solving the M-equations at each observation. The simulated detector schemes include the minimax detector, the modified minimax detector with $\eta = 0.1$, and the maximum likelihood (ML) detector with the score function being $\psi(x) = -f'(x)/f(x)$ [9]. To illustrate the robustness issue, we also include the performance of the sequential LS algorithm.

As shown in Fig. 1, with the current settings, there is no apparent difference between the performance of the two minimax detectors. This can be intuitively explained that the probability for the residuals to fall in the corner regions is relatively small with the current choice of η. Another observation is that the sequential M-estimate is very close to the exact M-estimates obtained directly from solving the M-equations. Fig. 2 magnifies the difference between $(\hat{\boldsymbol{\theta}}[n] - \boldsymbol{\theta})^T(\hat{\boldsymbol{\theta}}[n] - \boldsymbol{\theta})$ of the sequential and the exact ones for the minimax and the modified minimax detectors. Note that since we start with a *robust* estimate, the difference is 0 at $n_0 = 24$. It is of interest to notice that the difference between the sequential and exact M-estimates becomes smaller as more observations are available.

In Fig. 3 and Fig. 4, we show the performance of the sequential M-estimation algorithm for the minimax and ML detectors, and the sequential LS algorithm. As can be shown, even though the sequential LS estimate converges in a global sense, its error is significantly higher than those of the robust sequential M-estimates. This can be theoretically explained because the exact M-estimate has smaller error than the LS estimate of the full model, i.e. when $n = N$, in non-Gaussian noise, the sequential versions should also behave similarly.

4. CONCLUSION

A sequential M-estimation algorithm has been presented. Even though it serves as an approximation of the exact M-estimation one, the accuracy is desirable. It is also concluded that the original minimax detector can still be used ignoring the discontinuity problem of $\gamma(x)$ at the corner points with nearly no loss in performance. Under the framework of M-estimation, the algorithm proposed here has been demonstrated to be robust against non-Gaussian noise.

5. REFERENCES

[1] G.H. Golub and C.F. Van Loan. *Matrix Computations*. Johns Hopkins University Press, Baltimore and London, $2nd$ edition, 1989.

[2] P.J. Huber. Robust estimation of a location parameter. *Annals of Mathematical Statistics*, 35(1):73–101, 1964.

[3] P.J. Huber. Robust regression: Asymptotics, Conjectures and Monte Carlo. *Annals of Statistics*, 1(5):799–821, 1973.

[4] P.J Huber. *Robust Statistics*. Wiley, New York, 1981.

[5] S.M. Kay. *Fundamentals of Statistical Signal Processing, Part I: Estimation Theory*. Prentice Hall, New York, 1993.

[6] D.S. Pham, Y.H. Leung, A.M. Zoubir, and R.F. Brcic. On the computational aspects of robust multiuser detection. In *Proceedings of Information Technology and Signal Processing Symposium ISSPIT'03*, to appear.

[7] D.S. Pham, A.M. Zoubir, and R.F. Brcich. Adaptive ML signal detection in non-Gaussian channels. In *Proceedings of IEEE Workshop on Statistical Signal Processing SSP'03*, pages 45–48, St. Louis, Missouri, USA, 2003.

[8] L.L. Scharf. *Statistical Signal Processing: Detection Estimation and Time Series Analysis*. Addison Wesley, New York, 1995.

[9] X. Wang and H.V. Poor. Robust Multiuser Detection in Non-Gaussian Channels. *IEEE Transactions on Signal Processing*, 47(2):289–305, 1999.