# MULTIPLE PARAMETER ESTIMATION OF COMPOSITE SIGNALS IN UNCHARACTERIZED IMPULSIVE NOISE

*A. Mertins, N. Fliege*

Telecommunications Group, Technical University of Hamburg
Eissendorfer Str. 40, D-2100 Hamburg 90, Germany

## ABSTRACT

The problem of estimating amplitudes and arrival times of composite signals in non-Gaussian noise is addressed. First the maximum likelihood estimator (MLE) in Gaussian mixture noise with statistical independent noise samples will be presented, and it will be shown that the MLE is related to a nonlinear form of the orthogonality principle. Some important properties of the multiple parameter estimation problem will be extracted from Fisher's information matrix. The problem of the MLE is that the noise density is seldom known in practice. In this paper a new nonlinear nonparametric iterative method which does not require the knowledge of the noise density will be presented and compared to the MLE. The basic difference between the nonparametric method and the MLE is that the nonlinear manipulations are based on the actual input samples and therefore the nonparametric method is applicable in non-stationary noise processes.

## 1 INTRODUCTION

Composite signals appear in many important applications such as data communication systems or measurement problems. The assumption of Gaussian noise is very popular because of the simplicity of the estimation algorithms and their computational efficiency.

In certain problems the noise is characterized by a low background noise and some interfering sharp spikes. Examples are given by atmospheric noise and noise encountered in underwater detection problems. Several robust solutions to the detection problem have been presented in [4] - [7], and in [3], [5] some methods for robust estimation are presented.

This paper addresses the problem of multiple parameter estimation in non-Gaussian noise. As an example for a typical estimation problem, Fig. 1 shows a mass spectrometer signal. The concentration of mass is proportional to the amplitudes of impulses and has to be estimated.

Since the noise processes are related to physical phenomena which may change their characteristics, the processes are often non-stationary, and in practice, their characteristics are never known a priori.

Throughout this paper, the noise process is modelled as $\varepsilon$- contaminated noise with mutually statistically independent samples and density is given by

$$f_n(x) = (1-\varepsilon) f_{n_G}(x) + \varepsilon f_{n_I}(x), \qquad (1)$$

which is a special case of "Middleton Class A Noise" [4]. The densities $f_{n_G}(x)$ of background noise and $f_{n_I}(x)$ of interfering noise are assumed to be Gaussian with mean $\mu_G = \mu_I = 0$ and variances $\sigma^2_{n_G}$ and $\sigma^2_{n_I}$. In practice, typical values are often $\varepsilon = 0.01$ - $0.1$ and $\sigma^2_{n_I}/\sigma^2_{n_G} = 100$. The variance of the mixture process is given by

$$\sigma^2_n = (1-\varepsilon)\sigma^2_{n_G} + \varepsilon\sigma^2_{n_I}, \qquad (2)$$

so that the interfering noise contains more than 90% of the whole noise energy for $\varepsilon = 0.1$ and $\sigma^2_{n_I}/\sigma^2_{n_G} = 100$. An estimation algorithm which is optimized under the Gaussian assumption will perform poorly under these conditions.

The signals considered are linear combinations of basic signals of the form:

$$r(kT) = n(kT) + \sum_{i=1}^{I} a_i \cdot s(kT - (i-1)\tau - D). \qquad (3)$$

The basic signal $s(t)$ and the dalay $\tau$ are assumed to be known, while $a_1, \ldots, a_I$ and $D$ are the unknown parameters. The delay $\tau$ does not have to be an integer multiple of the sampling rate $T$. We can write (3) in vector notation as

$$\mathbf{r}(kT) = \mathbf{S}(kT - D) \cdot \mathbf{a} + \mathbf{n}(kT), \qquad (4)$$

where $\mathbf{r}(kT)$ includes $m$ samples of the received signal, $\mathbf{a}$ contains the amplitudes, and the $m \times I$- dimensional signal matrix $\mathbf{S}$ contains the samples of the delayed versions of the signal $s(t)$.

In order to estimate the delay $D$ from samples of the received signal we use the decomposition

$$D = \Delta + \delta, \quad |\delta| \leq T/2, \qquad (5)$$

where $\Delta$ is an integer multiple of $T$. For $kT = \Delta$ we can rewrite (4) as

$$\mathbf{r}(\Delta) = \mathbf{S}(\delta) \cdot \mathbf{a} + \mathbf{n}(\Delta). \qquad (6)$$

## 2 MAXIMUM LIKELIHOOD ESTIMATION

The maximum likelihood estimates $\hat{\mathbf{a}}, \hat{D}$ have to be found by solving the likelihood equations [9]

$$\left.\frac{\partial \ln f_{\mathbf{n}}(\mathbf{r}-\mathbf{S}\mathbf{a})}{\partial \mathbf{a}}\right|_{\mathbf{a}=\hat{\mathbf{a}},\, \Delta=\hat{\Delta},\, \delta=\hat{\delta}} = 0 \quad (7)$$

$$\left.\frac{\partial \ln f_{\mathbf{n}}(\mathbf{r}-\mathbf{S}\mathbf{a})}{\partial \delta}\right|_{\mathbf{a}=\hat{\mathbf{a}},\, \Delta=\hat{\Delta},\, \delta=\hat{\delta}} = 0 \quad (8)$$

Because of the statistical independence of noise samples we can write $\ln f$ as

$$\ln f_{\mathbf{n}}(\mathbf{r}-\mathbf{B}\mathbf{b}) = \sum_{i=0}^{m-1} \ln f_n(r_i - \mathbf{e}_i^T \mathbf{S}\mathbf{a}), \quad (9)$$

where $\mathbf{e}_i$, $i = 0, \ldots, m-1$ are unit vectors. Straightforward derivation leads to

$$\frac{\partial \ln f_{\mathbf{n}}(\mathbf{r}-\mathbf{S}\mathbf{a})}{\partial \mathbf{a}} = \mathbf{S}^T \cdot g(\mathbf{r}-\mathbf{S}\mathbf{a}), \quad (10)$$

$$\frac{\partial \ln f_{\mathbf{n}}(\mathbf{r}-\mathbf{S}\mathbf{a})}{\partial \delta} = \mathbf{a}^T \dot{\mathbf{S}}^T \cdot g(\mathbf{r}-\mathbf{S}\mathbf{a}), \quad (11)$$

where $g(x)$ is a memoryless nonlinearity given by

$$g(x) = x\ \frac{\frac{1-\varepsilon}{\sigma_{n_G}^2} f_{n_G}(x) + \frac{\varepsilon}{\sigma_{n_I}^2} f_{n_I}(x)}{(1-\varepsilon) f_{n_G}(x) + \varepsilon\ f_{n_I}(x)}. \quad (12)$$

The likelihood equations (7), (8) become

$$\left.\mathbf{S}^T \cdot g(\mathbf{r}-\mathbf{S}\mathbf{a})\right|_{\mathbf{a}=\hat{\mathbf{a}},\, \Delta=\hat{\Delta},\, \delta=\hat{\delta}} = 0 \quad (13)$$

$$\left.\mathbf{a}^T \dot{\mathbf{S}}^T \cdot g(\mathbf{r}-\mathbf{S}\mathbf{a})\right|_{\mathbf{a}=\hat{\mathbf{a}},\, \Delta=\hat{\Delta},\, \delta=\hat{\delta}} = 0, \quad (14)$$

and they can be interpreted as a nonlinear version of the orthogonality principle. The nonlinear manipulated estimated noise $\mathbf{r}-\mathbf{S}\hat{\mathbf{a}}$ has to be orthogonal to all signals in $\mathbf{S}$ and to $\dot{\mathbf{S}}\hat{\mathbf{a}}$. For the parameters $\varepsilon = 0$ and $\varepsilon = 1$ $g(x)$ is linear, and we receive the well known orthogonality principle [2], [8].

**Iteration:**
For solving (13), (14), the Newton-Raphson search can be used, but it has two disadvantages: a) the Hessian has to be built up in each iteration step; b) it finds the next maximum <u>or</u> minimum of the target function, and the initial values have to be very close to the optimum. Therefore we use a modified Newton-Raphson search, replacing the negative of the Hessian by its expected value in the optimum [1], which is Fisher's information matrix $\mathbf{J}$ in our application. The modified algorithm finds the next maximum, and the computational effort is minimal because $\mathbf{J}^{-1}$ has to be calculated only once.

**Building up $\dot{\mathbf{S}}$:**
In order to find unbiased estimates, the matrices $\mathbf{S}, \dot{\mathbf{S}}$ should achieve the condition

$$\dot{\mathbf{S}}^T(\delta) \cdot \mathbf{S}(\delta) = 0 \quad (15)$$

for any $\delta$. We propose the following method to build up $\dot{\mathbf{S}}$:

- Build up a matrix $\dot{\mathbf{S}}_A$ analogous to $\mathbf{S}$ from the samples of $\frac{\partial}{\partial \delta} s(t - [i-1]\tau - \Delta - \delta)$
- Orthogonal projection:

$$\dot{\mathbf{S}} = \left[\mathbf{I} - \mathbf{S}\left[\mathbf{S}^T\mathbf{S}\right]^{-1}\mathbf{S}^T\right]\dot{\mathbf{S}}_A, \quad (16)$$

so that (15) holds.

**The Gaussian Assumption:**
Under the Gaussian assumption, $g(x)$ becomes linear, and (13) can be solved in a closed form. By substituting the solution of (13) given by

$$\hat{\mathbf{a}} = \left[\mathbf{S}^T\mathbf{S}\right]^{-1}\mathbf{S}^T\mathbf{r} \quad (17)$$

for $\mathbf{a}$ in (14), we find the scalar target function

$$Q(D) = Q(\Delta, \delta) = \left.\hat{\mathbf{a}}^T \dot{\mathbf{S}}^T \mathbf{r} - \hat{\mathbf{a}}^T \dot{\mathbf{S}}^T \mathbf{S}\hat{\mathbf{a}}\right|_{\Delta=\hat{\Delta},\, \delta=\hat{\delta}} = 0 \quad (18)$$

for the time delay estimation problem.

## 3 FISHER'S INFORMATION MATRIX

A straightforward derivation leads to the following equations for the elements of the information matrix $\mathbf{J}$ where the elements of $\mathbf{S}$ are denoted by $S_{li}$

$$\begin{aligned} J_{i,j} &= E\left\{\sum_{l=0}^{m-1} S_{li} S_{lj}\right\} E\left\{g^2(x)\right\} \\ J_{i,I+1} &= E\left\{\sum_{l=0}^{m-1} S_{li} \mathbf{e}_l^T \dot{\mathbf{S}}\mathbf{a}\right\} E\left\{g^2(x)\right\} \\ J_{I+1,I+1} &= E\left\{\sum_{l=0}^{m-1} \left(\mathbf{e}_l^T \dot{\mathbf{S}}\mathbf{a}\right)^2\right\} E\left\{g^2(x)\right\} \\ & \qquad i, j = 1, \ldots, I \end{aligned} \quad (19)$$

By assuming orthogonal signals in $\mathbf{S}$, the information matrix becomes diagonal, and the lower bounds for variances $\sigma_{\hat{\mathbf{a}}}^2, \sigma_{\hat{D}}^2$ can be found to be:

$$\sigma_{\hat{\mathbf{a}}}^2 \geq \frac{1}{E_s\ E\left\{g^2(x)\right\}}, \quad (20)$$

$$\sigma_{\hat{D}}^2 \geq \frac{1}{E_{\dot{s}}\ E\left\{\mathbf{a}^T\mathbf{a}\right\}\ E\left\{g^2(x)\right\}}, \quad (21)$$

where $E_s$ denotes the energy of the basic signal, and $E_{\dot{s}}$ denotes the energy of its derivate.

The variance $\sigma_{\hat{a}}^2$ depends on the signal energy while $\sigma_{\hat{D}}^2$ depends on the energy of the derivate and of $E\{a_i^2\}$. It should be pointed out that $\sigma_{\hat{D}}^2$ decreases if the number of basic signals increases. If the expected energy of all amplitudes is equal $(E\{\mathbf{a}^T\mathbf{a}\} = I \cdot E\{a^2\})$, we get the property

$$\sigma_{\hat{D}}^2 \sim \frac{1}{I}, \tag{22}$$

which says that we should use as many basic signals as possible to achieve a low variance $\sigma_{\hat{D}}^2$.

## 4 NONPARAMETRIC METHOD

If we want to use the maximum likelihood estimation described in section 2, the parameters of the noise process have to be known completely. The idea of the nonparametric method is to replace the received signal $\mathbf{r}$ successively by a pseudo-received signal $\tilde{\mathbf{r}}$ while outliers should be rejected during the procedure. The procedure uses orthogonal projections to extract some properties of the noise samples included in $\mathbf{r}$, followed by nonlinear manipulations depending on the extracted parameters.

Procedure:

- Start:

$$\tilde{\mathbf{r}} := \mathbf{r} \tag{23}$$

- Orthogonal Projection:

$$\mathbf{x} := \mathbf{S}\left[\mathbf{S}^T\mathbf{S}\right]^{-1}\mathbf{S}^T\tilde{\mathbf{r}} \tag{24}$$

- Tolerance:

$$q(\tilde{\mathbf{r}}) := q_0 \cdot m^{-\frac{1}{2}}\,||\tilde{\mathbf{r}} - \mathbf{x}|| \tag{25}$$

  The tolerance $q$ is an estimate of the standard error multiplied by a factor ($q_0$), which is responsible for the convergence of the procedure.

- Nonlinear Manipulation:

$$\tilde{r}_i := \begin{cases} \mathbf{e}_i^T\mathbf{x} + q & \text{, if } \tilde{r}_i - \mathbf{e}_i^T\mathbf{x} > +q \\ \mathbf{e}_i^T\mathbf{x} - q & \text{, if } \tilde{r}_i - \mathbf{e}_i^T\mathbf{x} < -q \\ \tilde{r}_i & \text{else} \end{cases} \tag{26}$$

$$i = 0, \ldots, m-1$$

- Steps 2 - 4 of the procedure can be applied several times.

- Linear Estimation:

$$\hat{\mathbf{a}} = \left[\mathbf{S}^T\mathbf{S}\right]^{-1}\mathbf{S}^T\tilde{\mathbf{r}} \tag{27}$$

To illustrate the effect of nonlinear manipulation, Fig. 1 shows an example of the received signal and the nonlinear manipulated signal. As can easily be seen, all outliers could be reduced to very small values.

The arrival time can be estimated by first manipulating the received signal and then finding the solution of the target function $Q(D)$ using $\tilde{\mathbf{r}}$ instead of $\mathbf{r}$. Figure 2 shows some examples of $Q(D)$. We see that the nonparametric method can still achieve good estimates when the linear estimator is unable to find the correct time delay.

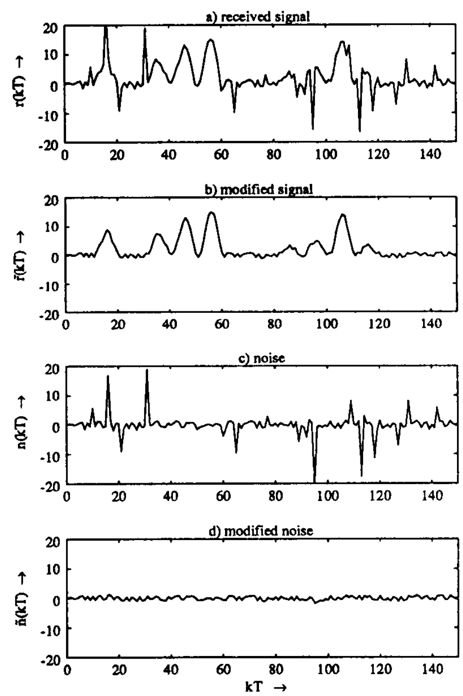


Fig. 1: Nonlinear manipulation with parameter $q_0 = 1.6$ and ten iterations.

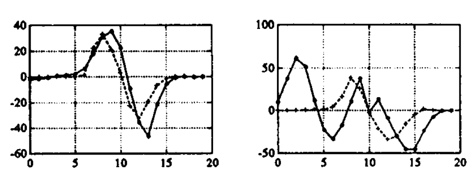

Fig. 2: Target function $Q(D)$ using the received vector $\mathbf{r}$ (—) and the manipulated vector $\tilde{\mathbf{r}}$ (- - -). True delay: $D = 10$

## 5 SIMULATION RESULTS

Simulations were made for a single signal with unknown arrival time and unknown amplitude. Figure 3 shows the variances of the different procedures in relation to the mixture parameter $\varepsilon$, while the variances $\sigma^2_{n_G}, \sigma^2_{n_I}$ were kept constant. For the MLE the noise density was assumed to be known, while the nonparametric method was applied with ten iterations and $q_0 = 1.6$.

As could be expected, ignoring the non-Gaussianity of noise leads to the highest variances. In the case of the MLE, the initial values are very important for the accuracy of the results. Without knowledge of the noise density, the variance of the nonparametric method can be lower than of the MLE, and therefore it can be referred to as a robust method. The MLE is probably the best method if the nonparametric method is used to produce initial values.

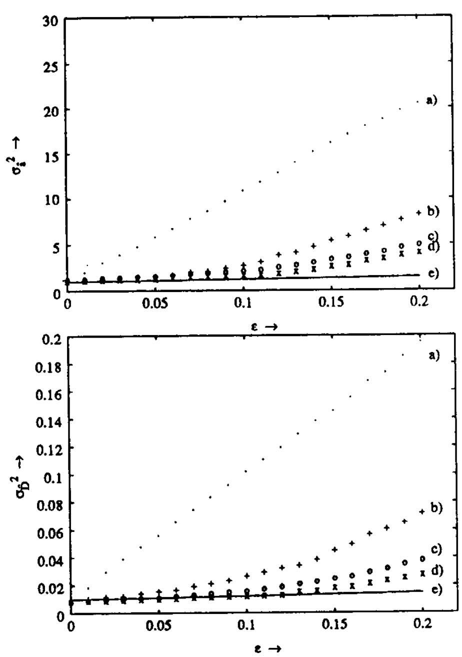


Fig. 3: Variances $\sigma^2_{\hat{a}}$, $\sigma^2_{\hat{D}}/T^2$; Parameters: $I = 1, E\{a^2\} = 100$, $\sigma^2_{n_G} = 1$, $\sigma^2_{n_I} = 100$; a) Gaussian assumption; b) MLE with initial values from (a); c) nonparametric algorithm; d) MLE with initial values from (c); e) Cramer Rao Bound

## 6 CONCLUSION

A new iterative nonlinear algorithm and the maximum likelihood estimator for the estimation of amplitudes and arrival times of composite signals in non-Gaussian noise have been presented and compared to linear methods and their Cramer Rao bounds. The problems of the MLE are the required knowledge of noise density and the generation of initial values, and under realistic conditions, the nonparametric method can lead to better results. The basic difference between the nonparametric method and the MLE is that the nonlinear manipulations are based on the actual input samples and therefore the nonparametric method is applicable in non-stationary noise processes.

## REFERENCES

[1] Burkhardt, H.; Moll, H.: A Modified Newton-Raphson-Search for the Model-Adaptive Identification of Delays. In: Isermann, R. (ed.): Identification and System Parameter Estimation, pp. 1279 - 1286. Pergamon 1979.

[2] Goldberger, A.S.: Econometric Theory. New York: John Wiley & Sons, Inc. 1964.

[3] Huber, P.J.: Robust Statistics. New York: John Wiley & Sons, Inc. 1981.

[4] Kassam, S.A.: Signal Detection in Non-Gaussian Noise. New York: Springer 1988.

[5] Kassam, S.A.; Poor, H.V.: Robust Techniques for Signal Processing, A Survey. Proceedings of the IEEE, Vol 73, No 3 (1985), pp. 433 - 481.

[6] Kay, S.M.: Asymptotically Optimal Detection in Incompletely Characterized Non-Gaussian Noise. IEEE Transactions on Acoustics, Speech, and Signal Processing. Vol. 37, No 5 (1989), pp. 627 - 633.

[7] Miller, J. H.; Thomas, J. B.: Detectors for discrete-time signals in Non-Gaussian Noise. IEEE Transactions on Inform. Theory. Vol. IT-18, Mar. 1972, pp. 755 - 771.

[8] Sage, A.P.; Melsa, J.L.: Estimation Theory with Applications to Communications and Control. New York: Mc Graw Hill 1971.

[9] Van Trees, H.L.: Detection, Estimation, and Modulation Theory, Part I. New York: Wiley 1968.