**A Weighted Least Squares Criterion for Estimation of Signal Parameters in Wavefields**

D. Maiwald, D. Kraus, G. Bugla, and J.F. Böhme

Department of Electrical Engineering,
Ruhr University Bochum,
4630 Bochum, Germany.

## 1. Introduction

The wave parameter estimation problem is of considerable interest in geophysical applications. Earthquakes give rise to a number of different waves, e.g. pressure waves, shear waves, or surface waves. These waves are received by an array of sensors. In order to determine the direction of arrival (DOA) the beamformer is used frequently and gives good results if the data contains only one type of wave. Because the pressure wave has the highest velocity of all types of waves and arrives at the array at first the beamformer can be used for determining its DOA. Other waves can arrive at the same time at the array for regional events with a distance of some 100 km between the earthquake and the array due to different reasons. The beamformer fails to separate the different waves in a wavenumber plot in such a situation, and we are faced with a resolution problem concerning velocity and direction. In this contribution we approach this problem by a parametric method. We fit a parametric model of the spectral density matrix of the sensor outputs to a nonparametric estimate of the spectral density matrix by minimizing a quadratic criterion. For the narrowband case least squares fits of this kind have been proposed in [3]. We extend these results to the broadband case and show the good asymptotic behaviour of our estimates.
The outline of the paper is as follows. In section 2 the data model and the parameter structure are introduced. The parameter estimates are developed in section 3. In section 4 we report on numerical experiments with simulated and measured seismic data.

## 2. Data Model

A conventional model is used. Earthquakes generate signals which are transmitted by a wavefield. We assume that $m = 1,\ldots,M$ different wave types arrive at the array. The outputs of the sensors $x_n(t)$ at positions $\underline{r}_n$ $(n = 1,\ldots,N)$ are Fourier-transformed:

$$\underline{X}_T(\omega) = \sum_{t=0}^{T-1} w(\frac{t}{T})\underline{x}(t)\exp^{-j\omega t}, \tag{1}$$

with a smooth window $w(s)$, where $\sum_{t=0}^{T-1} w^2(\frac{t}{T}) = 1$. The reception-propagation situation is described by a $(N\times M)$ matrix $\mathbf{H}(\omega) = [\underline{d}_1,\ldots,\underline{d}_M]$ with the phase vectors $\underline{d}_i = [e^{-j\underline{k}_i'\underline{r}_1},\ldots,e^{-j\underline{k}_i'\underline{r}_N}]'$. $\underline{k}_i = \frac{\omega}{V_i}[\cos\phi_i\cos\alpha_i, \cos\phi_i\sin\alpha_i, \sin\phi_i]'$ is the wavenumber vector of a wave at frequency $\omega$ with velocity $V_i$, and seen at the origin of the array at azimuth $\alpha_i$ and elevation $\phi_i$. The wavenumber vectors $\underline{k}_i$ may be written as $\underline{k}_i = \omega\underline{\xi}_i$ $(i = 1,\ldots,M)$ where $\underline{\xi}_i = (\xi_{ix},\xi_{iy},\xi_{iz})'$ is the so called slowness vector. Let us collect all $\underline{\xi}_i$ in a 3M-dimensional vector $\underline{\eta} = (\underline{\xi}_1',\ldots,\underline{\xi}_M')'$. The $(N\times N)$ spectral density matrix $\mathbf{C}_{\underline{X}}(\omega)$ of the array output can be expressed by

$$\mathbf{C}_{\underline{X}}(\omega,\underline{\theta}(\omega_i)) = \mathbf{H}(\omega,\underline{\eta})\mathbf{C}_{\underline{S}}(\omega)\mathbf{H}^*(\omega,\underline{\eta}) + \nu_0(\omega)\mathbf{I}, \tag{2}$$

where $\mathbf{C}_{\underline{S}}(\omega)$ is the spectral matrix of the signals and $\nu_0(\omega)$ is the spectral parameter of sensor noise. $\underline{\theta}(\omega_i) = (\underline{\eta}', \text{vec}\mathbf{C}_{\underline{S}}(\omega_i)', \nu_0(\omega_i))'$ is the p-dimensional parameter vector ($p = 3M + M^2 + 1$) of model (2). $\mathbf{I}$ is the identity matrix and the $*$ indicates the hermitian operation.

We need nonparametric estimates of $\mathbf{C}_{\underline{X}}(\omega_i)$ to develop a least squares fit in the next section. For stationary sensor outputs $\underline{x}(t)$ the asymptotic properties of $\underline{X}_T(\omega)$ are well known [1]. Under certain regularity conditions if the window length T is large and $0 < \omega_1 < \ldots < \omega_p < \pi$, then:

> $\underline{X}_T(\omega_1), \ldots, \underline{X}_T(\omega_p)$ are independent complex normally distributed random vectors with zero mean and covariance matrices $\mathbf{C}_{\underline{X}}(\omega_1), \ldots, \mathbf{C}_{\underline{X}}(\omega_p)$, respectively.

Using this property a consistent estimate is given by

$$\hat{\mathbf{C}}_{\underline{X}}(\omega) = \frac{1}{B_T T} \sum_{t \neq 0 \,(\text{mod}\, T)} W((\omega - \omega_t)/B_T) \underline{X}_T(\omega_t) \underline{X}_T(\omega_t)^*,$$

where $\omega_t = \frac{2\pi t}{T}$ and $B_T = O(T^{-\gamma})$ with $\gamma > \frac{1}{5}$ and $B_T T \to \infty$ for $T \to \infty$. The spectral window $W(\lambda)$ is a real valued and even function satisfying the conditions $W(\lambda) = 0$ if $|\lambda| > \pi$,

$$\int W(\lambda) d\lambda = 1, \qquad \int |W(\lambda)| d\lambda < \infty, \qquad \text{and} \int |\lambda|^2 |W(\lambda)| d\lambda < \infty.$$

Let $\mathbf{C}_{\underline{X}}(\omega)$ have bounded derivatives up to second order, then

$$\sqrt{K_T}\, \text{vec}(\hat{\mathbf{C}}_{\underline{X}}(\omega) - \mathbf{C}_{\underline{X}}(\omega)) \overset{a.s.}{\sim} N^c_{N^2}(\underline{0}, \Sigma(\omega)), \tag{3}$$

where $(\Sigma(\omega))_{kl} = (\mathbf{C}_{\underline{X}}(\omega))_{im} (\mathbf{C}_{\underline{X}}(\omega))_{nj}$ with $k = N(j-1) + i$ and $l = N(n-1) + m$, cf. [1]. $K_T$ is defined by $K_T = B_T T / \int W(\lambda)^2 d\lambda$.

## 3. Parameter Estimates

We use the following least squares fit,

$$q_T(\underline{\theta}) = \sum_{\omega_i \in \mathcal{B}} q_T(\underline{\eta}, \nu_0(\omega_i), \mathbf{C}_{\underline{S}}(\omega_i)) = \sum_{\omega_i \in \mathcal{B}} \text{tr}\left\{ \left[ \hat{\mathbf{C}}_{\underline{X}}(\omega_i) - \mathbf{C}_{\underline{X}}(\omega_i, \underline{\theta}) \right] \mathbf{W}(\omega_i) \right\}^2. \tag{4}$$

The symbol $\mathcal{B}$ denotes the frequency bands in which the signal is present significantly. The $\mathbf{W}(\omega_i)$ are positively definite (p.d.) weighting matrices.

We initialize the weighting matrices by $\mathbf{W}(\omega_i) = \mathbf{I}$. The minimization of $q_T(\underline{\theta})$ leads to an estimate $\underline{\theta}_1$. Using this parameter value we calculate $\mathbf{C}_{\underline{X}}(\omega_i, \underline{\theta}_1)$ and update the weighting matrix by $\mathbf{W}(\omega_i) = \mathbf{C}_{\underline{X}}^{-1}(\omega_i, \underline{\theta}_1)$. Then we start again the iteration. The following theorem justifies this procedure.

*Theorem:*

Let the regularity conditions such that (3) holds be satisfied. Minimizing of $q_T(\underline{\theta})$ supplies for $\mathbf{W}(\omega_i)$ positive definite $\sqrt{K_T}$ consistent estimates of the true parameter value $\underline{\theta}_0$, i.e.

$$\sqrt{K_T}(\theta_T - \underline{\theta}_0) \to 0 \quad \text{in probability,}$$

where

$$\underline{\theta}_T = \arg\min_{\underline{\theta}} q_T(\underline{\theta})$$

and $\underline{\theta}_0$ denotes the true parameter vector.
Furthermore, if $\mathbf{W}(\omega_i) = \mathbf{C}_{\underline{X}}^{-1}(\omega_i, \underline{\theta})$ then $\sqrt{K_T}(\underline{\theta}_T - \underline{\theta}_0)$ is asymptotically normally distributed with zero mean and covariance matrix $\mathbf{J}^{-1}$ where the elements of $\mathbf{J}$ are given by:

$$(\mathbf{J})_{i,j} = \sum_{\omega_i \in B} \mathrm{tr}\left[\frac{\partial \mathbf{C}_{\underline{X}}(\omega_i)}{\partial \theta_i} \mathbf{C}_{\underline{X}}^{-1}(\omega_i) \frac{\partial \mathbf{C}_{\underline{X}}(\omega_i)}{\partial \theta_j} \mathbf{C}_{\underline{X}}^{-1}(\omega_i)\right] \quad (i, j = 1, \ldots, N).$$

The proof is given in the appendix.

The implementation of the parameter estimates differs from the procedure described above in some details. We first try to obtain explicit solutions for the spectral parameters. Minimization of $q_T(\underline{\eta}, \nu_0(\omega_i), \mathbf{C}_{\underline{S}}(\omega_i))$ over the spectral parameters without restrictions yields

$$\tilde{\mathbf{C}}_{\underline{S}}(\omega_i, \underline{\eta}) = \mathbf{H}^{\#}(\omega_i, \underline{\eta}) \left[\check{\mathbf{C}}_{\underline{X}}(\omega_i) - \nu_0(\omega_i)\mathbf{I}\right] \mathbf{H}^{\#*}(\omega_i, \underline{\eta}), \tag{5}$$

where

$$\mathbf{H}^{\#}(\omega_i, \underline{\eta}) = \left[\mathbf{H}^*(\omega_i, \underline{\eta})\mathbf{W}(\omega_i)\mathbf{H}(\omega_i, \underline{\eta})\right]^{-1} \mathbf{H}^*(\omega_i, \underline{\eta})\mathbf{W}(\omega_i), \tag{6}$$

and

$$\tilde{\nu}_0(\omega_i) = \frac{\mathrm{tr}\left\{(\mathbf{P}(\omega_i, \underline{\eta})\mathbf{P}^*(\omega_i, \underline{\eta}) - \mathbf{I})\mathbf{W}(\omega_i)[\check{\mathbf{C}}_{\underline{X}}(\omega_i) - \mathbf{P}(\omega_i, \underline{\eta})\check{\mathbf{C}}_{\underline{X}}(\omega_i)\mathbf{P}^*(\omega_i, \underline{\eta})]\mathbf{W}(\omega_i)\right\}}{\mathrm{tr}\{(\mathbf{P}(\omega_i, \underline{\eta})\mathbf{P}^*(\omega_i, \underline{\eta}) - \mathbf{I})\mathbf{W}(\omega_i)\}^2}, \tag{7}$$

with

$$\mathbf{P}(\omega_i, \underline{\eta}) = \mathbf{H}(\omega_i, \underline{\eta})\mathbf{H}^{\#}(\omega_i, \underline{\eta}). \tag{8}$$

We replace the spectral density matrix of the sources $\mathbf{C}_{\underline{S}}(\omega_i)$ in (4) by the estimate (5) and get the following criterion

$$\begin{aligned} q_T(\underline{\eta}) &= \sum_{\omega_i \in B} q_T(\omega_i, \underline{\eta}) = \sum_{\omega_i \in B} \mathrm{tr}\{[\mathbf{W}(\omega_i)(\check{\mathbf{C}}_{\underline{X}}(\omega_i) - \tilde{\nu}_0(\omega_i)\mathbf{I})]^2 - \\ &\mathbf{W}(\omega_i)\mathbf{P}(\omega_i, \underline{\eta})\mathbf{W}(\omega_i)(\check{\mathbf{C}}_{\underline{X}}(\omega_i) - \tilde{\nu}_0(\omega_i)\mathbf{I})\mathbf{P}^*(\omega_i, \underline{\eta})\mathbf{W}(\omega_i)\mathbf{P}(\omega_i, \underline{\eta})(\check{\mathbf{C}}_{\underline{X}}(\omega_i) - \tilde{\nu}_0(\omega_i)\mathbf{I})\} \end{aligned} \tag{9}$$

The spectral parameter $\tilde{\nu}_0(\omega_i)$ is given by (7). The criterion has to be iteratively minimized over all elements of $\underline{\eta}$ where the weighting matrices $\mathbf{W}(\omega_i)$ are choosen according to the iteration procedure described before.

## 4. Numerical Experiments

In order to investigate the proposed algorithm we use simulated and real data. In both cases a circular array with 25 sensors which are distributed on 4 circles is used. The diameter of the array is about 3 km and the vertical aperture is about 200 m.

We simulated two sources with $\underline{\xi}_1 = [0.166, 0, 0]'$s/km and $\underline{\xi}_2 = [0.191, 0, 0]'$s/km impinging on the array with an SNR of -3 dB. The consistent estimate is used with 40 degrees of freedom. We used 3 frequency bands in the criterion (4). For the experiment $3 \times 512$ pseudo-random matrices have been generated. The results are depicted in Figure 1. The sources are $\mid \Delta\underline{\xi} \mid = 0.025$s/km distant from each other. The beamformer cannot resolve them. If we use $\mathbf{W}(\omega_i) = \mathbf{C}_{\underline{X}}^{-1}(\omega_i, \underline{\theta})$ the estimates have a higher accuracy than the ones obtained in the case $\mathbf{W}(\omega_i) = \mathbf{I}$.
The real seismic data was recorded by a corresponding array in the Bavarian Forest. In Figure 2 sensor outputs of an earthquake are shown. 512 sample points were taken at a sampling frequency of 40 Hz. Three frequency bands centred at 2 Hz, 2.6 Hz, and 3.2 Hz were used for the criterion (4). We smoothed the periodogram over 7 frequencies for obtaining an improved spectral density matrix esimate. The left part of Figure 3 shows the output of the beamformer. We also plotted the weighted least squares criterion with $\mathbf{W}(\omega_i) = \mathbf{I}$. The iteration procedure gives $\underline{\xi}_1 = [-0.139, 0.203, -0.026]'$s/km and $\underline{\xi}_2 = [-0.110, 0.341, -0.081]'$s/km. Indeed, the proposed algorithm can resolve two sources where the beamformer fails. Although we cannot make a statement on the accuracy of our estimates another analysis using the beamformer and more data samples has given similiar results.

## 5. Concluding Remarks

In this paper we extended the least squares fit for the narrowband case [3] to the wideband case. The good asymptotic properties of the wave parameter estimates have been proven. Numerical experiments with simulated data have shown the accuracy and stability of the proposed iteration procedure. The successful application of the estimates to real seismic data has been presented. The improvement of the estimates by taking account of the transience of the seismic signals is currently under research.

## 6. Acknowledgment

The authors wish to thank the Department of Geophysics of Ruhr University Bochum for access to the seismic data and the German Science Foundation (DFG) for supporting these investigations.

## 7. Appendix

To proof the theorem we use a technique to design simplified estimators with nice asymptotic properties as presented, for example in [4].
*Assumption 1:* There exists an $\sqrt{K_T}$ consistent estimator $\underline{\theta}_1$ of the true parameter vector $\underline{\theta}_0$ and a p-dimensional random vector $\Phi_{T,\underline{\theta}_0}$ depending on the values of the parameter vector $\underline{\theta}_0$ such that

$$\Phi_{T,\underline{\theta}_1} - \Phi_{T,\underline{\theta}_0} + \mathbf{J}_{\underline{\theta}_0}\sqrt{K_T}(\underline{\theta}_1 - \underline{\theta}_0) \rightarrow 0$$

in probability where $\mathbf{J}_{\underline{\theta}}$ is a nonsingular matrix with non random entries.
*Assumption 2:* There exists a $(p \times p)$-matrix $\mathbf{D}_1$ with random elements which is a $\sqrt{K_T}$ consistent estimator of the matrix $\mathbf{D}_{\underline{\theta}_0} = \mathbf{J}_{\underline{\theta}_0}$.

*Assumption 3:* The distribution of the random vector $\Phi_{T,\underline{\theta}_0}$ as $T \to \infty$ approaches a p-dimensional normal distribution $\mathcal{N}(\underline{0}, \mathbf{J})$ where $\mathbf{J}$ is a fixed covariance matrix.
Under the above stated assumptions the following lemma is valid.
*Lemma:* Let

$$\underline{\theta}_2 = \underline{\theta}_1 + (\sqrt{K_T}\mathbf{D}_1)^{-1}\Phi_{T,\underline{\theta}_1} \tag{10}$$

where

$$\sqrt{K_T}\Phi_{T,\underline{\theta}_0} - \mathbf{J}_{\underline{\theta}_0}\sqrt{K_T}(\underline{\theta}_1 - \underline{\theta}_0) \to 0,$$

in probability.
The random vector $\sqrt{K_T}(\underline{\theta}_2 - \underline{\theta}_0)$, as $T \to \infty$, possesses a normal distribution $\mathcal{N}(\underline{0}, \mathbf{D}_{\underline{\theta}_0}^{-1}\mathbf{J}\mathbf{D}_{\underline{\theta}_0}^{-1})$.

If the weighting matrices are given by $\mathbf{W}(\omega_i) = \mathbf{I}$ the minimization of $q_T(\underline{\theta})$ supplies $\sqrt{K_T}$ consistent estimates $\underline{\theta}_1$ (cf. [3]). We choose $\Phi_{T,\underline{\theta}} = \nabla q_T(\underline{\theta}) \mid_{\mathbf{W}(\omega_i)\text{p.d.}}$ where $\nabla q_T(\underline{\theta})$ is the gradient of the criterion (4) with entries

$$\frac{\partial q_T(\underline{\theta})}{\partial \theta_i} = -2 \sum_{\omega_i \in B} \operatorname{tr}\left[\frac{\partial \mathbf{C}_{\underline{X}}(\omega_i, \underline{\theta})}{\partial \theta_i}\mathbf{W}(\omega_i)(\check{\mathbf{C}}_{\underline{X}}(\omega_i) - \mathbf{C}_{\underline{X}}(\omega_i, \underline{\theta}))\mathbf{W}(\omega_i)\right].$$

It can be shown using (3) that in probability as $T \to \infty$

$$\sqrt{K_T}\Phi_{T,\underline{\theta}_1} - \sqrt{K_T}\Phi_{T,\underline{\theta}_0} - \mathbf{J}_{\underline{\theta}_0}\sqrt{K_T}(\underline{\theta}_1 - \underline{\theta}_0) \to 0, \tag{11}$$

where $\mathbf{J}_{\underline{\theta}_0}$ is defined by

$$(\mathbf{J}_{\underline{\theta}_0})_{i,j} = \sum_{\omega_i \in B} 2 \operatorname{tr}\left[\frac{\partial \mathbf{C}_{\underline{X}}(\omega_i, \underline{\theta}_0)}{\partial \theta_i}\mathbf{W}(\omega_i)\frac{\partial \mathbf{C}_{\underline{X}}(\omega_i, \underline{\theta}_0)}{\partial \theta_j}\mathbf{W}(\omega_i)\right]\Bigg|_{\mathbf{W}(\omega_i)\text{p.d.}}.$$

Therefore assumption 1 is satisfied. The matrix $\mathbf{D}_1$ is the asymptotic expected value of the matrix of second derivatives $\frac{\partial^2 q_T(\underline{\theta})}{\partial \theta_i \partial \theta_j}$ of the criterion, where

$$(\mathbf{D}_1)_{i,j}(\underline{\theta}) = \sum_{\omega_i \in B} 2\left[\frac{\partial \mathbf{C}_{\underline{X}}(\omega_i, \underline{\theta})}{\partial \theta_i}\mathbf{W}(\omega_i)\frac{\partial \mathbf{C}_{\underline{X}}(\omega_i, \underline{\theta})}{\partial \theta_j}\mathbf{W}(\omega_i)\right]. \tag{12}$$

$\mathbf{D}_1(\underline{\theta}_1)$ is a $\sqrt{K_T}$ consistent estimator of the matrix $\mathbf{J}_{\underline{\theta}_0}$ because it is a continuous function of $\underline{\theta}$. Thus, assumption 2 is satisfied. Lengthy calculations lead to

$$\sqrt{K_T}\Phi_{T,\underline{\theta}_0} - \sum_{\omega_i \in B} \mathbf{Q}(\omega_i, \underline{\theta}_0)\sqrt{K_T}\operatorname{vec}(\check{\mathbf{C}}_{\underline{X}}(\omega_i) - \mathbf{C}_{\underline{X}}(\omega_i, \underline{\theta}_0)) \to 0 \quad \text{in probability}, \tag{13}$$

where $\mathbf{Q}(\omega_i, \underline{\theta}_0) = (\operatorname{vec}\mathbf{Q}_1(\omega_i, \underline{\theta}_0), \ldots, \operatorname{vec}\mathbf{Q}_p(\omega_i, \underline{\theta}_0))'$ with

$$\operatorname{vec}\mathbf{Q}_i(\omega_i, \underline{\theta}_0) = \operatorname{vec}\left[-2\mathbf{W}(\omega_i)\frac{\partial \mathbf{C}_{\underline{X}}(\omega_i, \underline{\theta}_0)}{\partial \theta_i}\mathbf{W}(\omega_i)\right]\Bigg|_{\mathbf{W}(\omega_i)\text{p.d.}}.$$

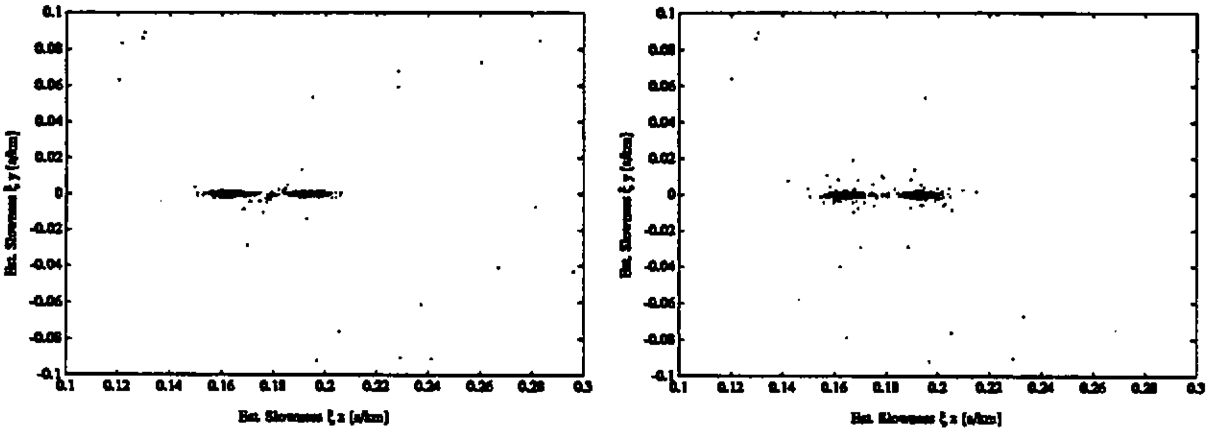

Figure 1: Monte Carlo simulations assuming stationary data, results for $\mathbf{W}(\omega_i) = \mathbf{I}$ on the left side and for $\mathbf{W}(\omega_i) = \mathbf{C}_{\underline{X}}^{-1}(\omega_i, \underline{\theta})$ on the right side.

The sum of (13) is a linear transformation of an asymptotically normally distributed random vector. Exploiting the asymptotic independence for different frequency bins we obtain

$$\lim_{T\to\infty} \sqrt{K_T}\Phi_{T,\underline{\theta}_0} \overset{a.s.}{\sim} N_p(\underline{0}, \mathbf{J}) \tag{14}$$

where $\mathbf{J} = \sum_{\omega_i \in B} \mathbf{J}(\omega_i)$ with typical entries

$$(\mathbf{J})_{ij}(\omega_i) = \operatorname{tr}\left\{\mathbf{Q}_i(\omega_i)\mathbf{C}_{\underline{X}}(\omega_i, \underline{\theta}_0)\mathbf{Q}_j(\omega_i)\mathbf{C}_{\underline{X}}(\omega_i, \underline{\theta}_0)\right\}\big|_{\mathbf{W}(\omega_i)\text{p.d.}} . \tag{15}$$

Using now $\mathbf{W}(\omega_i) = \mathbf{C}_{\underline{X}}^{-1}(\omega_i, \underline{\theta}_0)$ in (12) and (15) and applying the lemma gives $\sqrt{K_T}(\underline{\theta}_2 - \underline{\theta}_0) \overset{a.s.}{\sim} N_p(\underline{0}, \mathbf{J}^{-1})$.

## References


[1] D.R. Brillinger: Time Series: Data Analysis and Theory, Holden-Day, San Francisco 1981

[2] M.A. Cameron, E.J. Hannan: Transient Signals, Biometrika, 1979, 66, 2, pp. 243 - 258

[3] D. Kraus, G. Schmitz, J.F. Böhme: Least Squares Estimates for Source Location and Asymptotic Behaviours, Proc. EUSIPCO, Barcelona 1990, pp. 649-652

[4] K. Dzhaparidze: Parameter Estimation and Hypothesis Testing in Spectral Analysis of Stationary Time Series, Springer–Verlag, New York 1986

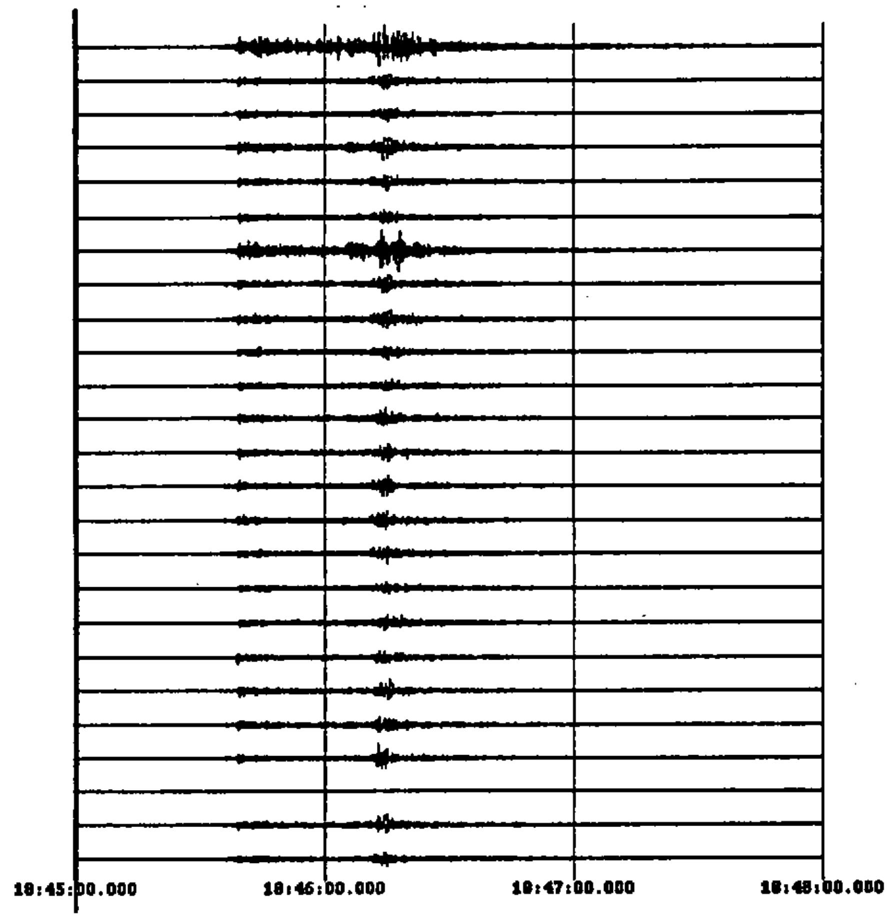


Figure 2: 25 traces of seismic data

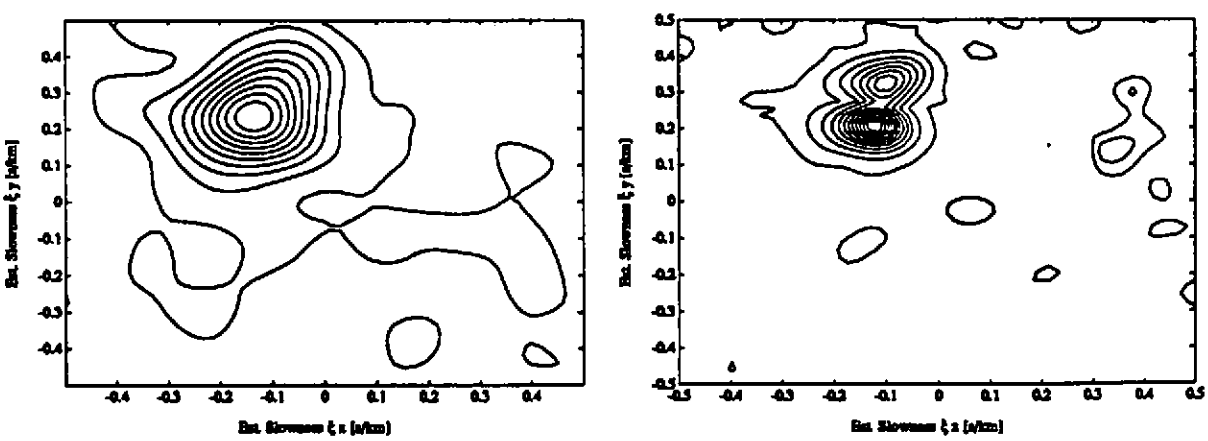

Figure 3: Beamformer on the left side, weighted least squares criterion on the right side.