

# Subspace-based estimation of time of arrival and Doppler shift for a signal of known waveform

V.V. LATYSHEV

*The subspace-based technique is used for the estimation of the time of arrival and Doppler shift of a signal of known waveform. The tool to find required subspaces is a special orthogonal decomposition of received data. It allows one to concentrate Fisher information on the desired parameter in just a few of the first terms of the decomposition. This approach offers a low-dimensional vector of sufficient statistics. It leads to computationally efficient Bayesian estimation. Besides, it results in expansion of the signal-to-noise ratio (SNR) range for effective maximum likelihood (ML) estimation. Finally, we can obtain independent time arrival and Doppler shift estimations based on generalized eigenvectors.*

**Keywords:** Time of arrival, Doppler shift estimates, subspace-based technique

Received 2 February 2009; Revised 2 March 2009; first published online 15 May 2009

## I. INTRODUCTION

Dimension reduction techniques are often applied as a data pre-processing step or as a part of the data analysis to simplify a data model. This typically involves identification of a suitable low-dimensional representation for the original high-dimensional data set or a subspace of an observation space. By working with this reduced representation, tasks such as classification or estimation can often yield more accurate and readily interpretable results, while computational costs may be significantly reduced.

In mathematical terms, the problem we consider can be stated as follows: given the  $N$ -dimensional random variable  $\mathbf{x} = (x_1, x_2, \dots, x_N)^T$ , find a lower dimensional representation of it,  $\mathbf{y} = (y_1, y_2, \dots, y_m)^T$  with  $m < N$ , that captures the content in the original data, according to some criterion.

Among the linear dimension reduction techniques, principal component analysis (PCA) is the best in the mean-square error sense [1]. In various fields, it is also known as the singular value decomposition (SVD) and the Karhunen–Loeve transform. In essence, PCA seeks to reduce the dimension of the data by finding a few orthogonal linear combinations (the PCs) of the original variables. The first PC  $y_1$  is the linear combination with the largest variance. The second PC  $y_2$  is the linear combination with the second largest variance and orthogonal to the first PC, and so on. There are as many PCs as the number of original variables. For many data sets, the first several PCs explain most of the variance, so that the rest can be disregarded with minimal loss of information.

Here we would like to use dimension reduction techniques in a signal parameter estimation problem. We hope to obtain a similar PCA technique with a different criterion.

## II. DATA DIMENSION REDUCTION

The estimation problem, which we consider here, can be phrased as follows: given  $x(t) = s(t, \theta) + w(t)$ , where  $s(t, \theta)$ ,  $0 \leq t \leq T$  is a signal of the known waveform to be observed in the presence of an additive Gaussian noise process  $w(t)$  with zero mean. We have an *a priori* probability density  $p(\theta)$  in the random parameter estimation problem. In all of our discussions we assume that  $p(\theta)$  is known. We need an estimation parameter  $\theta$  of a signal. In general, the variable  $\theta$  appears in a signal in a nonlinear manner.

We assume that the observation space corresponds to the set of  $N$  observations:  $x_1, x_2, \dots, x_N$ . Thus, each set can be thought of as a point in an  $N$ -dimensional space and can be denoted by a column vector  $\mathbf{x} = \mathbf{s}(\theta) + \mathbf{w}$ , where  $\mathbf{s}(\theta) \in \mathbb{R}^N$  and  $\mathbf{w} \in \mathbb{R}^N$  are the  $N$ -dimensional vectors of a signal and a noise correspondingly. Vector  $\mathbf{w}$  has a nonsingular covariance matrix  $\mathbf{R}_w$ . Then the probability density of  $\mathbf{x}$  is

$$p(\mathbf{x}|\theta) = ((2\pi)^N |\mathbf{R}_w|)^{-1/2} \exp \left[ -\frac{1}{2} (\mathbf{x} - \mathbf{s}(\theta))^T \mathbf{R}_w^{-1} (\mathbf{x} - \mathbf{s}(\theta)) \right]. \quad (1)$$

To obtain the  $m$ -dimensional vector  $\mathbf{y}$  with  $m < N$ , we use linear transformation  $\mathbf{y} = \mathbf{C}\mathbf{x}$  with the transformation matrix  $\mathbf{C}$ . We need such a matrix  $\mathbf{C}$  that guarantees minimal loss of estimation accuracy of a parameter  $\theta$ , using vector  $\mathbf{y}$ . In addition to the foregone requirement we try to represent  $\mathbf{x}$  in a new coordinate system in which the components are statistically independent random variables:  $\mathbf{C}\mathbf{R}\mathbf{C}^T = \mathbf{I}$ , where  $\mathbf{I}$  is a diagonal identity matrix. It is convenient to write the transformation matrix as  $\mathbf{C} = \mathbf{A}\mathbf{R}_w^{-1/2}$ . Here  $\mathbf{R}_w^{-1/2}$  is a symmetric square root from  $\mathbf{R}_w^{-1}$  ( $\mathbf{R}_w^{-1/2} \mathbf{R}_w^{-1/2} = \mathbf{R}_w^{-1}$ ),

$\mathbf{A}^T = (\mathbf{a}_1, \dots, \mathbf{a}_m)$ ,  $\mathbf{A}\mathbf{A}^T = \mathbf{I}$ . So we have

$$\mathbf{y} = \mathbf{A}\mathbf{R}_w^{-1/2} \mathbf{x}, \quad \mathbf{y} \in \mathbb{R}^m. \quad (2)$$

Moscow Aviation Institute (State Technical University), Moscow, Russia.

**Corresponding author:**

V.V. Latyshev

Email: lvv@mai.ru

First of all, recall that the variance of any unbiased estimate of an arbitrary parameter  $\theta$  is determined from the Cramer–Rao inequality [2]:

$$\sigma_\theta^2 \geq \left( \left[ \frac{\partial \ln p(\mathbf{x}|\theta)}{\partial \theta} \right]^2 \right)^{-1} = \left( \left[ -\frac{\partial^2 \ln p(\mathbf{x}|\theta)}{\partial \theta^2} \right] \right)^{-1}. \quad (3)$$

The inequalities are defined if the derivatives involved are existent and absolutely integrable. Recall that the denominator of the right-hand sides of CRI is usually referred to as the Fisher information on parameter  $\theta$ :

$$I(\theta) = \left[ \frac{\partial \ln p(\mathbf{x}|\theta)}{\partial \theta} \right]^2 = -\frac{\partial^2 \ln p(\mathbf{x}|\theta)}{\partial \theta^2}. \quad (4)$$

Inserting (1) in (3) we have

$$I_N(\theta) = (\mathbf{s}'(\theta))^T \mathbf{R}_w^{-1} \mathbf{s}'(\theta), \quad (5)$$

where subscript  $N$  is used to distinguish the initial dimension of the observation from a new reducing dimension  $m$ , and  $\mathbf{s}'(\theta) = ((\partial s_1(\theta)/\partial \theta), \dots, (\partial s_N(\theta)/\partial \theta))^T$  is the column vector of derivatives.

The Fisher information in the vector  $\mathbf{y}$  [3] is

$$I_m(\theta) = \sum_{k=1}^m [\mathbf{a}_k^T \mathbf{R}_w^{-1/2} \mathbf{s}'(\theta)]^2.$$

The loss of the Fisher information is

$$\begin{aligned} \Delta I(\theta) &= I_N(\theta) - I_m(\theta) \\ &= (\mathbf{s}'(\theta))^T \mathbf{R}_w^{-1} \mathbf{s}'(\theta) - \sum_{k=1}^m [\mathbf{a}_k^T \mathbf{R}_w^{-1/2} \mathbf{s}'(\theta)]^2. \end{aligned}$$

The mean of the loss of the Fisher information is

$$\Delta I = E_\theta \left\{ (\mathbf{s}'(\theta))^T \mathbf{R}_w^{-1} \mathbf{s}'(\theta) - \sum_{k=1}^m [\mathbf{a}_k^T \mathbf{R}_w^{-1/2} \mathbf{s}'(\theta)]^2 \right\}, \quad (6)$$

where  $E_\theta$  denotes an expectation over the random variable  $\theta$ . Thus, we need the transformation matrix (2) which provides minimal value of  $\Delta I$ .

*Theorem:* The linear transformation with the matrix  $\mathbf{A} \mathbf{R}_w^{-1/2}$  provides minimal mean of the loss of Fisher information  $\Delta I$ , if the column vectors  $\mathbf{a}_1, \dots, \mathbf{a}_m$  of  $\mathbf{A}^T$  are the orthonormal eigenvectors of

$$\mathbf{B} = \mathbf{R}_w^{-1/2} E_\theta \{ \mathbf{s}'(\theta) (\mathbf{s}'(\theta))^T \} \mathbf{R}_w^{-1/2}, \quad (7)$$

corresponding to  $m$  largest eigenvalues. At the same time

$$\Delta I_{\min} = \lambda_{m+1} + \dots + \lambda_N, \quad (8)$$

where  $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_N$  are the eigenvalues of  $\mathbf{B}$ .

*Proof:* Rewrite (3) in the next form

$$\Delta I = E_\theta \{ (\mathbf{s}'(\theta))^T \mathbf{R}_w^{-1} \mathbf{s}'(\theta) \} - E_\theta \left\{ \sum_{k=1}^m [\mathbf{a}_k^T \mathbf{R}_w^{-1/2} \mathbf{s}'(\theta)]^2 \right\}.$$

The first term does not depend on  $\mathbf{a}_k$ . Therefore, we find the minimal value of  $\Delta I$  if the subtrahend is maximal. Denote it as  $H(\mathbf{a}_1, \dots, \mathbf{a}_m)$ . Inverting averaging with summation and taking into account the equality

$$[\mathbf{a}_k^T \mathbf{R}_w^{-1/2} \mathbf{s}'(\theta)]^2 = \mathbf{a}_k^T \mathbf{R}_w^{-1/2} \mathbf{s}'(\theta) (\mathbf{s}'(\theta))^T \mathbf{R}_w^{-1/2} \mathbf{a}_k,$$

we have

$$H(\mathbf{a}_1, \dots, \mathbf{a}_m) = \sum_{k=1}^m \mathbf{a}_k^T [\mathbf{R}_w^{-1/2} E_\theta \{ \mathbf{s}'(\theta) (\mathbf{s}'(\theta))^T \} \mathbf{R}_w^{-1/2}] \mathbf{a}_k.$$

The expression in brackets is a symmetric matrix:

$$\mathbf{B} = \mathbf{R}_w^{-1/2} E_\theta \{ \mathbf{s}'(\theta) (\mathbf{s}'(\theta))^T \} \mathbf{R}_w^{-1/2}.$$

In compliance with the theorem about eigenvalues and eigenvectors [4], the maximal value of  $H(\mathbf{a}_1, \dots, \mathbf{a}_m)$  takes place if  $\mathbf{a}_1, \dots, \mathbf{a}_m$  are the orthonormal eigenvectors of the matrix  $\mathbf{B}$ , corresponding to  $m$  largest eigenvalues  $\lambda_1 > \lambda_2 > \dots > \lambda_m$  and

$$\max H(\mathbf{a}_1, \dots, \mathbf{a}_m) = \sum_{k=1}^m \lambda_k.$$

The equality  $\mathbf{R}_w^{-1/2} \mathbf{R}_w^{-1/2} = \mathbf{R}_w^{-1}$  implies

$$\text{tr } \mathbf{B} = E_\theta \{ (\mathbf{s}'(\theta))^T \mathbf{R}_w^{-1} \mathbf{s}'(\theta) \}.$$

On the other hand,  $\text{tr } \mathbf{B} = \sum_{k=1}^N \lambda_k$ . It implies  $\Delta I_{\min} = \sum_{k=m+1}^N \lambda_k$ .

Note that we can assert that a subspace spanned by the column vectors  $\mathbf{a}_1, \dots, \mathbf{a}_m$  is the  $m$ -dimensional subspace of an observation space with maximal Fisher information content concerning the parameter  $\theta$  among any other  $m$ -dimensional subspaces.

### III. NUMERICAL EXAMPLE

To illustrate the advantages of the dimension reduction technique in signal parameter estimation, let us consider the Doppler shift problem. We model the next signal

$$x(t) = s(t, f_D) + w(t). \quad (9)$$

Here  $s(t, f_D)$  is a signal of known form to be observed in the presence of an additive white Gaussian noise process  $w(t)$  with zero mean and variance  $\sigma_w^2$ . The time delay is known and we wish to estimate the Doppler frequency shift  $f_D$  only. Note that for the radar case, the standard narrowband assumption is employed here, i.e. the frequency offset  $f_D$  is a narrowband approximation to the stretching or shrinking of the frequency axis due to the Doppler effect induced by the relative motion of the reflecting target.

The quantity  $f_D$  is a random parameter with a uniform *a priori* density  $p_o(f_D)$ ,  $f_D \in [-F; F]$ , so we may use the Bayesian estimation procedure. However, it is very complicated to put it into practice. In this example, our goal is to simplify the Bayesian estimate for the cases when a cost function

is the square of the error. For the problem at hand, we have to capture Fisher information concerning Doppler shift after data dimension reduction.

To illustrate our results we use the FM signal with  $T \cdot \Delta f = 32$ :

$$s(t, F_D) = E \cos\left((\omega_0 - 2\pi F_D)t + \frac{\pi \Delta f}{T} t^2\right), \quad (10)$$

$$-\frac{T}{2} \leq t \leq \frac{T}{2}.$$

A set of  $N$  observations,  $x_1, x_2, \dots, x_N$ , consists of statistically independent values  $x_k$  with variance  $\sigma_w^2$ . So we have a column vector  $\mathbf{x} = \mathbf{s}(f_D) + \mathbf{w} \in R^N$  with the nonsingular covariance matrix  $\mathbf{R}_w = \sigma_w^2 \mathbf{I}$ .

The Bayesian estimate uses the *a posteriori* density:

$$p(f_D|\mathbf{x}) = \frac{p(\mathbf{x}|f_D)p_0(f_D)}{\int_{-F}^F p(\mathbf{x}|f_D)p_0(f_D) df_D}. \quad (11)$$

The estimate is the mean of the *a posteriori* density (or the conditional mean):

$$f_D(\mathbf{x}) = \int_{-F}^F f_D p(f_D|\mathbf{x}) df_D. \quad (12)$$

The simplification of the Bayesian estimate may be achieved using linear transformation with matrix  $\mathbf{A}$  to provide minimal mean loss of Fisher information about Doppler shift. From the foregoing theorem the column vectors  $\mathbf{a}_1, \dots, \mathbf{a}_m$  of  $\mathbf{A}^T$  are the orthonormal eigenvectors of the matrix

$$\mathbf{B} = \frac{1}{\sigma_w^2} E_\theta \{ \mathbf{s}'(f_D) (\mathbf{s}'(f_D))^T \},$$

corresponding to  $m$  largest eigenvalues. Here  $\mathbf{s}'(f_D)$  is the column vector of derivatives with respect to  $f_D$ .

If we use the vector  $\mathbf{y} = \mathbf{A}\mathbf{x}$ ,  $\mathbf{y} \in R^m$ , the Bayesian estimate is the same as in (11) and (12), where  $\mathbf{x}$  is replaced by  $\mathbf{y}$ .

For the simulations  $T = 1$ ,  $N = 512$ ,  $f_D$  is a random parameter with the uniform *a priori* density  $f_D \in [-0.4; 0.4]$ . The mean squared error (MSE) of the Doppler shift estimate was calculated for each of the algorithms based on 10 000 Monte Carlo trials for various signal-to-noise ratio (SNR) values. The results are plotted in Fig. 1 together with the appropriate Cramer–Rao bound (CRB). The demonstrated results correspond to the Bayesian estimate using initial 512-dimensional vector  $\mathbf{x}$  and results based on five-dimensional and 10-dimensional vectors  $\mathbf{y}$ .

As can be seen from Fig. 1, the 512-dimensional and the 10-dimensional data sets are equivalent with respect to estimation accuracy. At the same time the computational complexity of the Bayesian algorithm is two orders of magnitude less in this example. This is due to the fact that we capture Fisher information concerning Doppler shift. Further, the effect of dimension reduction from 10 to 5 degrades accuracy noticeably.

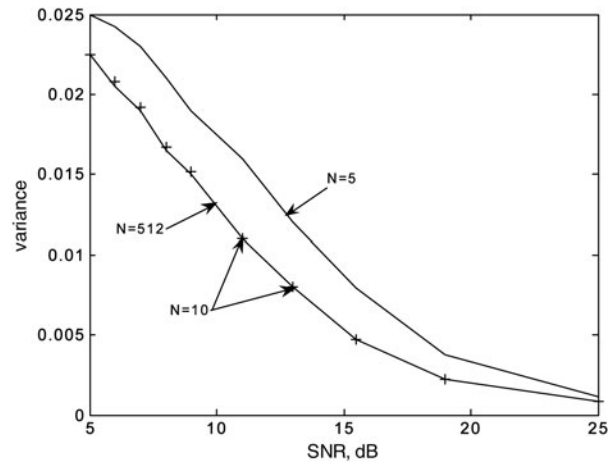


Fig. 1. MSE of the Bayesian estimate using initial (512-lower thin line) and reduced dimensional vectors (5, 10).

#### IV. SUBSPACE-BASED MAXIMUM-LIKELIHOOD (ML) ESTIMATION

The theorem that is proved above allows us to reveal an  $m$ -dimensional subspace of the observation space containing Fisher information on the arbitrary parameter  $\theta$ . If the informational contents of  $m$ -dimensional vector  $\mathbf{y}$  and  $N$ -dimensional  $\mathbf{x}$  are equal to one another, we may regard  $\mathbf{y}$  as a vector of sufficient statistics in a sense that an estimate of  $\theta$  based on  $\mathbf{y}$  is characterized by the same CRB as the estimate based on the initial observation. Otherwise  $\mathbf{y}$  may be considered as an approximate vector of sufficient statistics. The degree of approximation is determined by the sum (8).

Next, we consider an approach to the synthesis of subspace-based ML estimation algorithms providing for a higher accuracy of the results. Note that the estimate as obtained according to the ML principle in nonlinear estimation problems is asymptotically optimal for large SNR.

As an illustration, let us consider estimation of the time of arrival. Instead of (9) now we have

$$x(t) = s(t - \tau) + w(t)$$

with

$$s(t) = 1 - \cos\left(\frac{2\pi}{T}t\right), \quad 0 \leq t \leq T. \quad (13)$$

The unknown parameter is a signal delay  $\tau$  with respect to the point on the time axis that is chosen as an original. It is necessary to estimate  $\tau$  from the observation  $x(t)$ .

We are taking into account that an estimation problem is usually solved either after a signal has been detected or simultaneously with signal detection. Therefore, below we assume that an expected signal has been detected and that its approximate position on the time axis is known. It is necessary to specify the signal's position in the presence of distortions. This assumption fits the real operation conditions of tracking equipment containing a discriminator that in each measurement cycle is tuned to a given reference value of a parameter to be measured.

Let us compare the information content of the vector  $\mathbf{y}$  with different dimensions for this problem. We can take into account the results of calculation of Fisher information from [5, 6] for signal (10) to be observed in the presence of an additive white Gaussian noise process  $w(t)$ . Normalized graphs are shown in Figs. 2–4 for different dimensions of vector  $\mathbf{y}$  and  $T = 1$ .

It is noteworthy that the first eigenvector  $\mathbf{a}_1$  coincides with the pulse response of the ML-filter which follows from the ML equation to within a constant [5]. Therefore, the solution of the ML equation corresponds to the first coordinate of the vector  $\mathbf{y}$ . Its informational content is shown in Fig. 2. The graph in Fig. 2 can be interpreted as an informational explanation of the asymptotic optimal property of the ML estimate. The informational content of the observation  $\mathbf{x}$  and of the one-dimensional subspace are close to each other for small  $\tau$  only. In an estimation problem the estimate is close to the true value  $\tau$  in case of high SNR and approximate position on the time axis is known within narrow *a priori* range ( $\tau_{min}; \tau_{max}$ ). In general, if the SNR is not high, the 1D subspace is associated with large losses of Fisher information concerning the time delay. So the first coordinate of the vector  $\mathbf{y}$  cannot be considered as sufficient statistic for the time delay estimation. Figure 2 shows the losses of Fisher information using the gray shading.

The rest of the coordinates of vector  $\mathbf{y}$  yield an additional information channel that can be used to enhance the accuracy of estimation. In Figs. 3 and 4 we see informational contents of 2D and 3D subspaces, respectively. The losses are noticeably

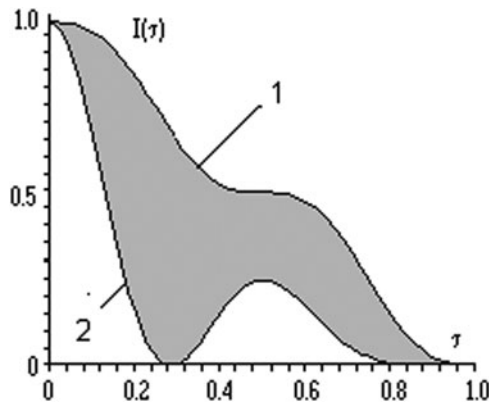


Fig. 2. Fisher information in observation (1) and in the output signal of an ML filter (2).

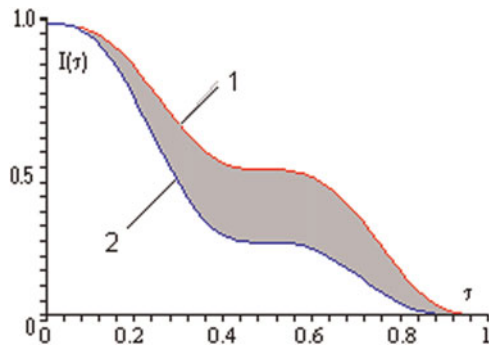


Fig. 3. Fisher information in observation (1) and in 2D vector (2).

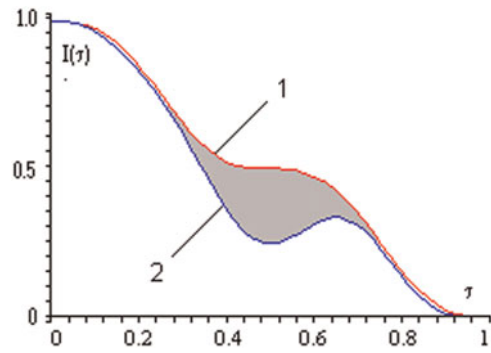


Fig. 4. Fisher information in observation (1) and in 3D vector (2).

less. Therefore, we can consider 2D or 3D vectors  $\mathbf{y}$  as an approximate vector of sufficient statistics within larger *a priori* range ( $\tau_{min}; \tau_{max}$ ).

As long as we have an approximate vector of sufficient statistics, each of its components can be used for an independent estimation of the signal delay. In practice, this estimation procedure necessitates a multi-channel device providing for the formation of the corresponding vector components in its channels. Each of the channels contains a linear filter with an impulse response  $\mathbf{a}_k$  as the corresponding eigenvector of (7).

This approach actually deals with an  $m$ -channel device, so that parameter  $\tau$  is estimated in each channel using the ML estimator (MLE) (see Fig. 5).

Therefore, the results in channels can be integrated in a weight adder (WA). Since an estimation procedure in each channel is nonlinear, it is difficult to exactly calculate its statistical characteristics. However, potential characteristics can be employed. As shown above, all channels are characterized by the mean of the Fisher information  $\lambda_k$ . Hence, the estimates in the channels can be characterized by the different CRBs for the error variance. The optimal integration in WA involves calculation of the average of the estimates obtained in the channels with the weights proportional to  $\sqrt{\lambda_k}$ .

Simulation of this approach is illustrated in Fig. 6. For comparison, Fig. 6 shows the error variances obtained on the bases of 1D, 2D and 3D approximate vectors of sufficient statistics. Recall that the estimate for the case  $m = 1$  coincides with the classic ML estimate. Each point of the plots has been obtained as a result of 10 000 estimation cycles with various white noise realizations. The initial dimension of the observation is equal to  $N = 1024$ . Here it is assumed that the approximate value of the delay is known within an *a priori* range  $(-T/4; T/4)$ ,  $T = 1$ . Thus, the domain of uncertainty is the half-duration of the signal.

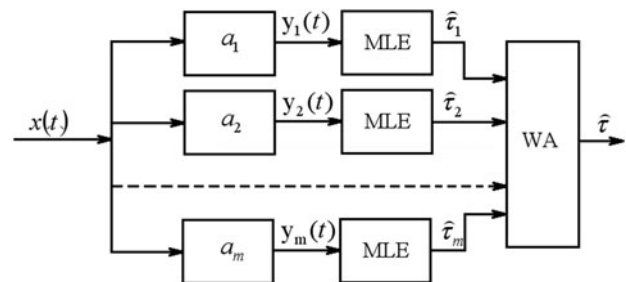


Fig. 5. A multi-channel estimator of the time of arrival.

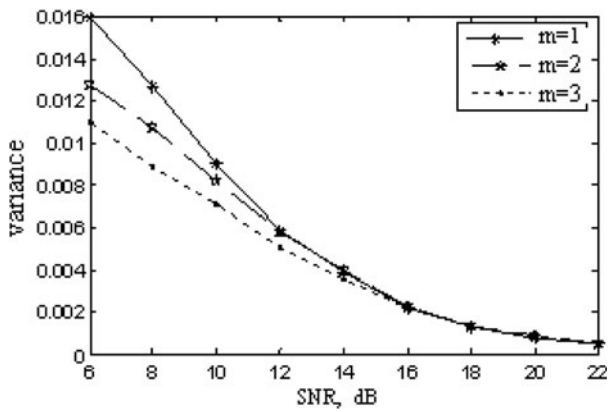


Fig. 6. MSE of the estimates based on approximate vectors of sufficient statistics.

The simulation results show that the pulse's time of arrival may be estimated more accurately than that in the case when the classical ML approach is applied ( $m = 1$ ). In the example considered, this effect is the result of averaging independent estimates of the time of arrivals in independent channels. As is known from estimation theory, the averaging effectively enhances the accuracy of the determination of various signal parameters. The method proposed in this paper allows application of the aforementioned effect of averaging.

## V. INDEPENDENT TIME ARRIVAL AND DOPPLER SHIFT ESTIMATIONS

Consider the widespread Gaussian observation model. Let the  $N$ -dimensional data column vector be an additive mixture of the deterministic component and the distortion vector

$$\mathbf{x} = \mathbf{s}(\tau, f_D) + \mathbf{w}, \quad (14)$$

where signal  $\mathbf{s}(\tau, f_D)$  depends on the two *a priori* unknown parameters. Assume that both parameters are statistically mutually independent. The parameter vector  $\mathbf{v} = (\tau, f_D)$  has a bounded domain of variation, where 2D probability density  $p_o(\tau, f_D)$  is specified. The distortion vector  $\mathbf{w}$  consists of independent sample values, described by the Gaussian probability distribution with the zero mean and a nonsingular covariance matrix  $\mathbf{R}_w = \sigma_w^2 \mathbf{I}$ .

Assume that it is necessary to estimate  $\tau$  from the observed vector  $\mathbf{x}$ . In such a case we can consider  $f_D$  to be a nuisance parameter. To find an independent estimate of  $\tau$ , it is desirable to obtain the statistics in the form of linear functions of vector  $\mathbf{x}$  that are not affected by the nuisance parameter. We consider these statistics as invariant to variations of the nuisance parameter. It is also necessary to minimize a possible deterioration of the estimation accuracy, if  $\tau$  is estimated using these statistics.

The problem formulated above may be solved on the basis of the following ideas. If the data or the statistics obtained from these data do not contain Fisher information on a certain parameter, it is impossible to estimate this parameter by the use of this statistic. This situation may be interpreted as an invariance to the changes of this parameter. On the

other hand, when the observations are used to form a statistic or data set concentrating the complete Fisher information on a desired parameter, this parameter will be estimated with the same accuracy as that provided by all of the observations. Hence, it is necessary to have a tool that ensures control of the Fisher information content in the statistics obtained via linear transformations of the observations. The theorem proved above allows one to reveal the subspaces of an observation space containing Fisher information on specific scalar parameters  $\tau$  or  $f_D$ . We use the next two matrices, which are similar to (7):

$$\mathbf{B}_\tau = E_v \{ \mathbf{s}'_\tau(\tau, f_D) (\mathbf{s}'_\tau(\tau, f_D))^T \} / \sigma_w^2, \quad (15)$$

$$\mathbf{B}_f = E_v \{ \mathbf{s}'_f(\tau, f_D) (\mathbf{s}'_f(\tau, f_D))^T \} / \sigma_w^2. \quad (16)$$

Here, differentiation is performed with respect to the parameters indicated by the indexes  $\tau$  or  $f_D$ .  $E_v$  denotes an expectation over the vector random variable  $\mathbf{v} = (\tau, f_D)$ .

In order to eliminate Fisher information on the nuisance parameter  $f_D$  from an observation, it is possible to obtain a decomposition of the observed signal using the eigenvectors of  $\mathbf{B}_f$ . Then we remove the terms with nonzero eigenvalues, and sum the remaining terms.

A drawback of this approach is that the Fisher information on desired parameter  $\tau$  is beyond transformations. We need the eliminated statistics with minimum Fisher information concerning  $\tau$ . Therefore, retaining the idea of the method applied to find approximate invariant statistics by means of elimination of some terms from the orthogonal decomposition, we modify the approach. Note that the considered transformation of the observed vector  $\mathbf{x}$  to the vector of decomposition coefficients actually converts  $\mathbf{B}_\tau$  into a diagonal matrix. Denoting the matrix that consists of the eigenvectors of (15) by  $\mathbf{A}$ , we have  $\mathbf{A} \mathbf{B}_\tau \mathbf{A}^T = \mathbf{D}$ , where  $\mathbf{D}$  is the diagonal matrix with the eigenvalues of  $\mathbf{B}_\tau$ . It is known that two matrices  $\mathbf{B}_\tau$  and  $\mathbf{B}_f$  can be diagonalized simultaneously. According to [4], this can be done using the single matrix  $\mathbf{A}$  so that

$$\mathbf{A} \mathbf{B}_f \mathbf{A}^T = \mathbf{I}, \quad \mathbf{A} \mathbf{B}_\tau \mathbf{A}^T = \mathbf{D}, \quad (17)$$

where  $\mathbf{D}$  and  $\mathbf{A}^T$  are chosen to be the matrices of the eigenvalues and the eigenvectors of the matrix  $\mathbf{B}_f^{-1} \mathbf{B}_\tau$ , respectively. Simultaneously with performance of the above transformation, it is necessary to separate Fisher information on parameter  $\tau$  from that on parameter  $f_D$ .

Following [4], let us introduce the auxiliary matrix

$$\mathbf{Q} = \mathbf{B}_f + \mathbf{B}_\tau. \quad (18)$$

Using matrix  $\mathbf{A}$ , we can choose a linear transformation so that  $\mathbf{Q}$  and  $\mathbf{B}_\tau$  are diagonalized simultaneously and the following conditions are fulfilled:

$$\mathbf{A} \mathbf{Q} \mathbf{A}^T = \mathbf{I}, \quad \mathbf{A} \mathbf{B}_\tau \mathbf{A}^T = \mathbf{D}^{(1)}, \quad (19)$$

where  $\mathbf{D}^{(1)}$  is the diagonal matrix with ordered entries  $\lambda_1^{(1)} \geq \lambda_2^{(1)} \geq \dots \geq \lambda_N^{(1)}$ , which are the eigenvalues of matrix  $\mathbf{B}_\tau$ .

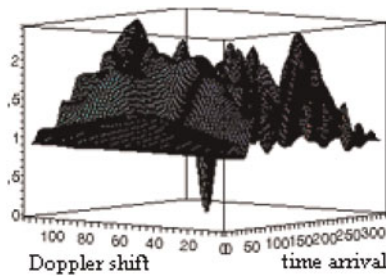


Fig. 7. Logarithm of the likelihood function for 7 bits  $M$ -sequences.

The substitution (18) into (19) yields

$$\mathbf{I} = \mathbf{A}(\mathbf{B}_\tau + \mathbf{B}_f)\mathbf{A}^T = \mathbf{D}^{(1)} + \mathbf{A}\mathbf{B}_f\mathbf{A}^T$$

or

$$\mathbf{A}\mathbf{B}_f\mathbf{A}^T = \mathbf{I} - \mathbf{D}^{(1)}. \tag{20}$$

Hence, it follows that matrices  $\mathbf{B}_\tau$  and  $\mathbf{B}_f$  are diagonalized simultaneously under the above conditions. Besides, the eigenvalues of both matrices satisfy the relationship  $\lambda_k^{(2)} = 1 - \lambda_k^{(1)}$ ,  $k = 1, \dots, N$ . Thus, when  $\mathbf{B}_\tau$  and  $\mathbf{B}_f$  are diagonalized simultaneously, the greatest eigenvalues of the first matrix correspond to the least eigenvalues of the second matrix and vice versa. An advantage of this approach is that the orthogonal decomposition based on matrix  $\mathbf{A}$  yields an orthogonal series, whose terms are ranked simultaneously according to the degree of decrease of Fisher information. However, these terms are arranged in reciprocal order: Fisher information on the desired parameter is concentrated in the first decomposition terms and Fisher information on the nuisance parameter is concentrated in the last terms. Naturally, in this situation, elimination of terms with the greatest content of information on the nuisance parameter guarantees both minimal loss of Fisher information on the parameter to be estimated and minimal loss in the accuracy of estimation.

In such a way, on the basis of the generalized eigenvectors of the matrix pair  $(\mathbf{B}_\tau, \mathbf{B}_f)$  we can divide the observation space into two mutually orthogonal subspaces containing Fisher information about  $\tau$  and  $f_D$ , respectively. For example, to obtain an independent estimate of the time of arrival we have to remove Fisher's information about Doppler shift. For this purpose, we use the subspace with Fisher's information about  $\tau$  only. Figures 7 and 8 illustrate this possibility for 7 bits  $M$ -sequences. Both figures show relief of  $|l(\tau, f_D)|$ , where  $l(\tau, f_D)$  is the logarithm of the likelihood function. Figure 7 corresponds to observation data without noise. We see here the local gap near the true values of  $\tau$  and  $f_D$ . Figure 8 corresponds to  $|l(\tau, f_D)|$  using the orthogonal projection of observation data onto the subspace with Fisher information concerning  $\tau$  only. Here we see that the narrow canyon is definitely parallel to  $f_D$  axes. It implies that the true value of  $\tau$  may be obtained independently from  $f_D$ . If  $\tau$  is estimated, the next step is to estimate  $f_D$  using the subspace with Fisher information concerning  $f_D$ .

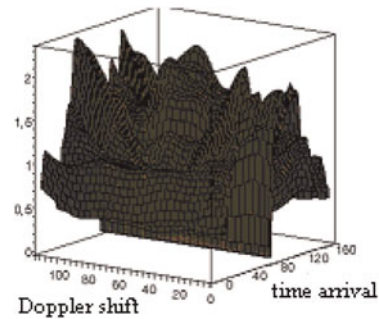


Fig. 8. Logarithm of the likelihood function in invariant to  $f_D$  subspace.

## VI. CONCLUSIONS

Based on the Fisher information approach, a linear dimension reduction technique has been proposed. It is similar to PCA, but it makes good use of the minimal loss of Fisher information as a criterion. As a result, we obtain subspaces of an observation space concentrating Fisher information concerning interesting parameters of a signal. The projections of an observation onto these subspaces are the vectors of the sufficient statistics. Simulation performed with the use of the test signals has shown that the accuracy of the estimation is based on the ML method and can be improved if we use these vectors. The generalized eigenvectors of the matrix pair allow obtaining mutually orthogonal subspaces for independent estimation of two different parameters.

## REFERENCES

- [1] Jolliffe, I.T.: *Principal Component Analysis*, Springer-Verlag, 1986.
- [2] Van Trees, H.L.: *Detection, Estimation, and Modulation Theory, Part 1*, Wiley, New York, 2001, ch. 2.
- [3] Latyshev, V.: The optimum decrease of the dimensionality of data for estimating signal parameters. *Sov. J. Commun. Technol. Electron.*, **33** (1988), 172–175.
- [4] Fukunaga, K.: *Introduction to Statistical Pattern Recognition*, San Diego, Academic Press, 1990, ch. 2.
- [5] Latyshev, V.: Informational analysis of statistics in time delay estimation problem, in *Proc. IRS2007*, Cologne, Germany, pp. 169–173.
- [6] Latyshev, V.: Subspace-based estimation of time of arrival and Doppler shift for a signal of known waveform, in *Proc. ESAV'08*, Capri, Italy, pp. 94–99.



**Viacheslav Latyshev** received the diploma in electrical engineering from the Moscow Aviation Institute in 1971. In 1979, he received the Ph.D. degree in electrical engineering and D.Sc. degree in 1999. From 1971 to 1979, he has held research positions at the Department of Electrical Engineering, Moscow Aviation Institute.

In 1985, he was appointed Professor of Signal Processing, where he is working currently. His research interests include stochastic signal processing, fast methods for signal parameter estimation, higher order statistics.