J. Webster (ed.), *Wiley Encyclopedia of Electrical and Electronics Engineering*

SIGNAL DETECTION AND PROCESSING

Detection in signal processing is often referred to as determining the number of signal sources in a noisy environment. It has many engineering applications, ranging from military surveillance to mobile communications. As an example, the engine vibration of an underwater submarine is a source of signals. The determination of the number of such sources is of interest in sonar signal processing.

Detection is also important in other areas of signal processing, such as the estimation of certain parameters: frequency spectrum and direction of arrival. The spectral peaks obtained from the Fourier transform of the data were used initially to estimate these parameters, but the spectral resolution was generally poor due to practical limitations such as the time length of the data record. In the last 20 years, model-based parameter estimation has been an area of active research. Many high-resolution approaches have been developed, but they require certain prior knowledge. Among them the number of signal sources is often the most crucial. It is thus clear that signal detection plays an important role in parameter estimation, system modeling and identification, and stochastic realization.

Detection problems are generally classified into two categories: determination of the number of signals, each having different frequencies; and determination of signals, each coming from different locations. While the former problem utilizes a single measurement device such as a broadband sensor, the latter generally uses an array of sensors to gain spatial information. The latter models are more complicated. Without loss of generality, only the array signal formulation will be considered here.

The earlier use of the Fourier (frequency or spatial) spectrum to determine the number of signals by observing the number of peaks lacks resolution in that when two sources are closely spaced, their spectral peaks may merge into one. In this situation signals may be inadvertently missed. The recent development of information-theoretic approaches has greatly improved the determination.

Detection is complicated in practice as ideal conditions rarely occur. Rather than ideal narrow-band signals, wide-band signals may be encountered; rather than ideal uncorrelated signals, there may be coherent signals; rather than Gaussian stationary noise, colored non-Gaussian nonstationary noise may be the case. In order to obtain the necessary theoretic results only ideal conditions will be addressed; the nonideal situations remain for future research.

Assumption and Signal Model

Suppose that we have an array of two identical sensors, and a plane-wave signal from the far field impinges on this array so that the normal of the plane wave front makes an angle ϕ with the line joining the sensors in the array. The signal arriving at the second sensor $x_2(t)$ is a delayed version of the signal arriving at the first sensor $x_1(t)$. That is to say, if $x_1(t) = s(t)$, then $x_2(t) = s(t - \tau)$. If two sensors are spaced a distance d apart, then the time delay τ between two sensors is $\tau = d \sin \phi / v = (2\pi d/\lambda_c \omega_c) \sin \phi$, where v and λ_c are the speed and wavelength of the plane-wave signal, respectively. Now we further assume that q uncorrelated plane-wave signals simultaneously impinge on an array of p sensors.

The following assumptions are made throughout the entire article unless being otherwise stated.

- The p array sensors are identical to each other. The number of sensors is greater than the number of signal sources ($p > q$).
- The q signal wave fronts s_k are narrow-band (compared with the center wavelength λ_c) plane waves (far-field).
- The observation noise n_i at each sensor is additive complex white Gaussian noise with zero mean and variance $\sigma^2/2$ for independent real and imaginary parts), independent from sensor to sensor and from the signals.

With the preceding generic assumptions, the signal arriving at the ith sensor located at (x_i, y_i) (for $i = 1, \ldots, p$) at time t is

$$\begin{aligned} x_i(t) &= \sum_{k=1}^{q} e^{j(2\pi/\lambda_c)(x_i \sin\phi_k + y_i \cos\phi_k)} s_k(t) + n_i(t) \\ &= \left(e^{j(2\pi/\lambda)(x_i \sin\phi_1 + y_i \cos\phi_1)} \cdots e^{j(2\pi/\lambda)(x_i \sin\phi_q + y_i \cos\phi_q)}\right) \begin{pmatrix} s_1(t) \\ \vdots \\ s_q(t) \end{pmatrix} + n_i(t) \end{aligned} \tag{1}$$

for $i = 1, \ldots, p$, where $s_k(t)$ is the kth narrow-band signal (with center wavelength λ_c) arriving at an angle ϕ_k. The signals arriving at all the sensors at time t are

$$x(t) = \begin{pmatrix} e^{j(2\pi/\lambda)(x_1 \sin\phi_1 + y_1 \cos\phi_1)} & \cdots & e^{j(2\pi/\lambda)(x_1 \sin\phi_q + y_1 \cos\phi_q)} \\ \vdots & \ddots & \vdots \\ e^{j(2\pi/\lambda)(x_p \sin\phi_1 + y_p \cos\phi_1)} & \cdots & e^{j(2\pi/\lambda)(x_p \sin\phi_q + y_p \cos\phi_q)} \end{pmatrix} \begin{pmatrix} s_1(t) \\ \vdots \\ s_q(t) \end{pmatrix} + \begin{pmatrix} n_1(t) \\ \vdots \\ n_p(t) \end{pmatrix}$$

or in matrix notation,

$$\boldsymbol{x}(t) = A(\Phi)\boldsymbol{s}(t) + \boldsymbol{n}(t) \tag{2}$$

where $A(\Phi) = [A(\phi_1), \ldots, A(\phi_q)]$ is referred to as the direction of arrivals (DOA) matrix, $\boldsymbol{n}(t)$ is the receiver's noise vector, and q is assumed unknown but is less than p. For statistical analysis, it is assumed that $\{A(\phi_1), \ldots, A(\phi_q)\}$ are linearly independent and the signals $\boldsymbol{s}(t)$ are distributed as a complex multivariate with zero mean vector 0 and nonsingular covariance matrix Ψ. The definition of the complex multivariate distribution can be found in many books on signal processing, for example, Ref. 1. It is further assumed that $\boldsymbol{s}(t)$ and $\boldsymbol{n}(t)$ are independent. Then the covariance matrix Σ of $\boldsymbol{x}(t)$ is given by

$$\Sigma = E\{\boldsymbol{x}(t)\boldsymbol{x}(t)^*\} = A\Psi A^* + \sigma^2 I \tag{3}$$

where E denotes the mathematical expectation and the superscript $*$ denotes the transpose complex conjugate. Let $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_p$ denote the eigenvalues of Σ. Due to the fact that Ψ is a $q \times q$ matrix, the eigenvalues may be arranged as $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_q > \lambda_{q+1} = \cdots = \lambda_p = \sigma^2 > 0$. It is this basic fact that there are $p - q$ equal smallest eigenvalues on which most methods of detection are based. The subspace spanned by $A(\Phi)$ is referred

to as the signal subspace and its orthogonal complement is called the noise subspace. Suppose that there are n independent observations $\boldsymbol{x}(t_1), \boldsymbol{x}(t_2), \ldots, \boldsymbol{x}(t_n)$ available. Let the sample covariance matrix be $S = (1/n)\Sigma^n{}_{i=1}$ $\boldsymbol{x}(t_i)\boldsymbol{x}^*(t_i)$ and the eigenvalues of S be $\delta_1 \geq \delta_2 \geq \ldots \geq \delta_p$. Note that the assumptions made in this section are supported by real data.

Hypotheses Testing

A statistical hypotheses testing procedure was developed to find the multiplicity (or the number of equal eigenvalues) of eigenvalues (2), which is equivalent to finding the number of signals by Eq. (3). It consists of a set of nested hypotheses as given by

H_p	All the eigenvalues of Σ are equal
H_{p-1}	Only the last $p-1$ eigenvalues of Σ are equal $\cdots$
H_k	Only the k eigenvalues of Σ are equal $\cdots$

To test H_k, the following χ^2 statistics with $(k^2 + k - 2)/2$ degrees of freedom is used:

$$\chi^2 = -\left\{ n - (p-k) - \tfrac{1}{6}(2k + 1 + 2/k) + \hat{\sigma}^2 \sum_{j=1}^{p-k} \frac{1}{(\delta_j - \hat{\sigma})^2} \right\} \log N_k$$

where

$$N_k = \prod_{j=p-k+1}^{p} \delta_j \Bigg/ \left(\frac{1}{k} \sum_{j=p-k+1}^{p} \delta_j \right)^k$$

$$\hat{\sigma} = \sum_{j=p-k+1}^{p} \frac{\delta_j}{k}$$

If $H_p, \ldots, H_{k+1}$ are rejected and H_k are accepted, then an estimate $\hat{q}$ of q is given as the corresponding value $p - k$.

Another hypotheses test, called the predicted eigenthreshold approach (3), of the multiplicity of the eigenvalues σ^2 is as follows: Define

$$l_i = \frac{1}{p-i+1} \sum_{j=i}^{p} \delta_j, \qquad i = k+1, \ldots, p$$

$$d_{p-m} = \left[(m+1) \frac{1 + t(n(m+1))^{-1/2}}{1 - t(nm)^{-1/2}} - m \right] l_{p-m+1}$$

where t is the critical value of standard normal for the two-tailed test with an α level of significance.

Assuming that the multiplicity of σ^2 is m, let the two hypotheses be

$$\begin{array}{ll} H_0 & k < p - m \\ H_a & k = p - m \end{array}$$

Starting with $m = 1$, accept H_a or H_0 according to whether $\delta_{p-m} > d_{p-m}$ or $\delta_{p-m} \le d_{p-m}$. If H_a is accepted, then an estimate of q is $\hat{\theta} = p - m$. Otherwise increase m by 1 and continue the test until either H_a is accepted or $m = p$.

In both of the preceding estimates, a subjective threshold is required. Since the exact joint distribution of the eigenvalues is difficult to compute and the asymptotic distribution depends fundamentally on the eigenstructure of Σ, the performances of both tests are unclear due to the fact that the rejection region is decided by the limiting distribution of the eigenvalues in practice. The advantage for both tests is that the probability of underfitting is somehow under control.

Information-Theoretic Criteria

White-Noise Case. The general idea of the information-theoretic criteria is based on either minimizing a type of "closeness" between two probability functions or minimizing the length of coding of a data set from a model while being penalized for using a more complex model.

The Akaike information criterion (*AIC*) is a well-known information-theoretic criterion of the first type and the minimum description length (*MDL*) criterion is a well-known information theoretic criterion of the second type. Consider the following model of k signals:

$$\text{Model } M_k: \quad f_k(\cdot|\boldsymbol{\theta}), \quad 0 \le k < p \tag{4}$$

with θ^* being the true parameter. Let $\{\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_n\}$ be an independent identically distributed data sequence according to a probability function $f_{k_0}(\cdot|\theta^*|)$. Let $\hat{\mathrm{M}}_{q_k}$ be the maximum likelihood estimate (*MLE*) of θ^* based on $\{\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_n\}$ under the assumption that the true model is the model M_k. Denote the number of free parameters in the model M_k by $|M_k|$.

The AIC (4) is to choose the model $\hat{\lambda}$ so that

$$\hat{q} = \arg\min_{0 \le k < p} \left\{ -2 \sum_{i=1}^{n} \log f(\boldsymbol{x}_i|\hat{\boldsymbol{\theta}}_k) + 2|M_k| \right\} \tag{5}$$

while the MDL criterion (5) is to choose the model $\hat{\sigma}^2_v$ so that

$$\hat{q} = \arg\min_{0 \le k < p} \left\{ -2 \sum_{i=1}^{n} \log f(\boldsymbol{x}_i|\hat{\boldsymbol{\theta}}_k) + \frac{|M_k|}{2} \log n \right\}. \tag{6}$$

Apply these criteria to the data set of Eq. (2) and use the model M_k in Eq. (4) with parameter vector $\theta = (\lambda_1, \ldots, \lambda_k, \sigma^2, \mathrm{v}^T{}_1, \ldots, \mathrm{v}^T{}_k)^T$, where $\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_k$ are, respectively, the eigenvectors of $\lambda_1, \lambda_2, \ldots, \lambda_k$. Leaving out terms

not involving θ, the logarithmic likelihood function is given as

$$L(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n|\hat{\boldsymbol{\theta}}_k) = (p-k)n\left[\sum_{i=k+1}^{p} \frac{1}{p-k} \log \delta_i - \log\left(\frac{1}{p-k}\sum_{i=k+1}^{p} \delta_i\right)\right] \quad (7)$$

and the number of free parameters is $k(2p - k) + 1$. Discarding terms not involving k, the two criteria for estimating the number of signals q are given as

$$\hat{q} = \arg\min_{0\leq k<p} \left\{-2L(X|\hat{\boldsymbol{\theta}}_k) + 2[k(2p-k+1)+1]\right\}, \qquad \text{AIC} \qquad (8)$$

$$\hat{q} = \arg\min_{0\leq k<p} \left\{-L(X|\hat{\boldsymbol{\theta}}_k) + \tfrac{1}{2}k[(2p-k+1)+1]\log n\right\}, \qquad \text{MDL} \qquad (9)$$

It was shown that MDL yields a consistent estimate of q in the sense that the probability of an incorrect detection goes to zero as n goes to infinity, while the AIC overestimates the number of signals with positive probability (6). In fact, MDL will give the true number of signals for large n with probability 1 (7) and the upper bound on the probability of wrong detection converges to zero at an exponential rate (8).

The additive term, which is an increasing function of $|M_k|$, is called a penalty term, as it penalizes for the use of a more complex model. The greater the number of free parameters in a model, the larger the penalty. For example the AIC in Eq. (8) the penalty term $2[k(2p - k +1)+1]$ is an increasing function of k where k reflects the complexity of the model. Comparing the AIC in Eq. (8) with the MDL criterion in Eq. (9), the difference is essentially the two penalty terms. This results in a completely different performance for the two criteria. The AIC tends to overestimate while the MDL criterion is strongly consistent but is penalized because of the possibility of underestimation for small sample size. A class of strongly consistent information-theoretic criteria is proposed in Ref. 7, which are obtained by adjusting the penalty term as follows:

$$\hat{q} = \arg\min_{0\leq k<p} \left\{-2L(X|\hat{\boldsymbol{\theta}}_k) + k[(2p-k+1)+1]C_n\right\} \qquad (10)$$

where C_n is a function of n satisfying

$$\lim_{n\to\infty} \frac{C_n}{n} = 0, \qquad \lim_{n\to\infty} \frac{C_n}{\log\log n} = \infty \qquad (11)$$

Arguing that the set of free parameters should not include the eigenvectors because of the fact that the eigenvectors are invariant under a similarity transformation (9), a smaller parameters set was employed using the following criteria:

$$\hat{q} = \arg\min_{0\leq k<p} \left\{-\log g(X|\hat{\theta}_k) + \tfrac{1}{2}k\log n\right\} \qquad (12)$$

where

$$-\log g(X|\hat{\theta}_k) = n\left(\sum_{i=1}^{k}\frac{\delta_i}{\hat{\lambda}_i} + \sum_{i=k+1}^{p}\frac{\delta_i}{\hat{\sigma}_v^2}\right) + (n-p+1)\sum_{i=1}^{k}\log\hat{\lambda}_i + (p-k)(n-k)\log\hat{\sigma}_v^2$$
$$+ \sum_{\substack{i,j=1\\ i<j}}^{k}\log(\hat{\lambda}_i - \hat{\lambda}_j) + \sum_{i=1}^{k}(p-k)\log(\hat{\lambda}_i - \hat{\sigma}_v^2) - \sum_{\substack{i,j=k+1\\ i<j}}^{p}\log(\delta_i - \delta_j)$$
$$+ \frac{k}{2}(2p-k-1)\log n - \sum_{i=p-k+1}^{p}\log\Gamma(i)$$

$\hat{\sigma}_i$ and $\bar{\sigma}_v^2$ are solutions to the system of equations

$$\hat{\lambda}_i = \delta_i - \frac{\hat{\lambda}_i}{n}\sum_{\substack{j=1\\ j\neq i}}^{k}\frac{\hat{\lambda}_j}{\hat{\lambda}_i - \hat{\lambda}_j} - \frac{p-k}{n}\frac{\hat{\lambda}_i\hat{\sigma}_v^2}{\hat{\lambda}_i - \hat{\sigma}_v^2}, \quad i = 1, 2\ldots, k$$
$$\hat{\sigma}_v^2 = \frac{1}{p-k}\sum_{j=k+1}^{p}\delta_j + \frac{1}{n}\sum_{i=1}^{k}\frac{\hat{\lambda}_i\hat{\sigma}_v^2}{\hat{\lambda}_i - \hat{\sigma}_v^2}$$

For large n, this criterion is approximately equal to

$$-\log g(X|\hat{\theta}_k) \cong -nL(X|\hat{\theta}_k) + \sum_{\substack{i,j=1\\ i<j}}^{k}\log(\delta_i - \delta_j)^2 + \sum_{i=1}^{k}\log(\delta_i - \bar{\sigma}_v^2)^{p-k}$$
$$+\frac{k}{2}(2p-k-1)\log n - \sum_{i=p-k+1}^{p}\log\Gamma(i) \qquad (13)$$

where $\bar{\lambda}_i = [1/(p-k)]\ \Sigma^p{}_{i=k+1}\ \delta_i$. While this criterion is not known to be consistent due to the difficulty in the analysis because some of the logarithmic terms could be infinite, some criteria having the spirit of Eqs. (12)

and (13) are proposed in Ref. 10, such as letting the function $-\log g(X|\bar{\sigma}^2_{\hat{v}_k})$ in Eq. (12) be given as

$$-\log g(X|\hat{\theta}_k) = n\left(\sum_{i=1}^{k}\frac{\delta_i}{\bar{\lambda}_i}+\sum_{i=k+1}^{p}\frac{\delta_i}{\bar{\sigma}_v^2}\right)+(n-p+1)\sum_{i=1}^{k}\log\bar{\lambda}_i+(p-k)(n-k)\log\bar{\sigma}_v^2$$
$$+\sum_{\substack{i,j=1\\ i<j}}^{k}\log(\bar{\lambda}_i-\bar{\lambda}_j+\xi)+\sum_{i=1}^{k}(p-k)\log(\bar{\lambda}_i-\bar{\sigma}_v^2+\xi)-\sum_{\substack{i,j=k+1\\ i<j}}^{p}\log(\delta_i-\delta_j+\xi)$$
$$+\frac{k}{2}(2p-k-1)\log n-\sum_{i=p-k+1}^{p}\log\Gamma(i)$$

where $\hat{\Phi}$ and $\hat{\Gamma}$ are solutions to the system of equations:

$$\bar{\lambda}_i=\delta_i-\frac{\bar{\lambda}_i}{n}\sum_{\substack{j=1\\ j\neq i}}^{k}\frac{\bar{\lambda}_j}{\bar{\lambda}_i-\hat{\lambda}_j+\xi}-\frac{p-k}{n}\frac{\bar{\lambda}_i\bar{\sigma}_v^2}{\bar{\lambda}_i-\hat{\sigma}_v^2+\xi},\quad i=1,2\ldots,k$$
$$\bar{\sigma}_v^2=\frac{1}{p-k}\sum_{j=k+1}^{p}\delta_j+\frac{1}{n}\sum_{i=1}^{k}\frac{\bar{\lambda}_i\bar{\sigma}_v^2}{\bar{\lambda}_i-\bar{\sigma}_v^2+\xi}$$
$$\bar{\lambda}_i-\bar{\lambda}_j|\leq\rho\xi,\ \ |\bar{\lambda}_i|\leq n^{2/5}\log n,\quad i,j=1,2,\ldots,k+1$$

and $0<\rho<1$ is a constant. It was proved in Ref. 10 that the system has a unique solution for large n and for $\rho\xi>\lambda_1-\sigma^2$. These criteria are shown to be strongly consistent in Ref. 10.

Even though the consistent detection criteria discussed above can be applied to the non-normal noise case while the consistency of estimates still holds, their original derivations are based on the likelihood function of the complex normal distribution. A nonparametric detection criterion was proposed in Ref. 11 which uses a type of function, called *r*-regular, in place of the logarithmic likelihood function.

Definition. *Let f be a real-valued function defined on finite sequences of $\mathscr{M}$, an open subset of the real numbers. f is said to be r-regular if the following conditions are satisfied*

(1) If $x_1=x_2=\cdots=x_m$, then $f(x_1, x_2, \cdots, x_m)=f(x_1)$.
(2) If $x_1, x_2, \cdots, x_m$ are not identical, then $f(x_1, x_2, \cdots, x_m)>f(x_1)$.
(3) For each k, then restriction of f on $\mathscr{M}^k$, denoted by f_k, belongs to $C^{r+1}(\mathscr{M}^k)$, where $\mathscr{M}^k$ is the k Cartesian product of $\mathscr{M}$, and the partial derivatives of f_k up to order r are all zero on the set $\{(x_1, x_2, \cdots, x_m)\in\mathscr{M}^k | x_1=x_2=\cdots=x_k\}$.

An example of a 3-regular function is

Definition. *Let f be a real-valued function defined on finite sequences of $\mathscr{M}$, an open subset of the real numbers. f is said to be r-regular if the following conditions are satisfied:*

An example of a 3-regular function is

$$f(x_1, x_2, \cdots, x_m) = (1/m)\left(\sum_{i=1}^{m} x_i - (1/m)\sum_{i=1}^{m} x_i\right)^4$$

and $\mathcal{M} = R$, the set of real numbers.

Suppose that a sequence of p values $\{l_1^{(n)}, l_2^{(n)}, \ldots, l_p^{(n)} | n = 1, 2, \cdots\}$ satisfies the following:

(1) $|l_i^{(n)} - \lambda_i| = O(\alpha_n)$ a.s. for $i = 1, 2, \ldots, p$, $0 < \alpha_n \to 0$ as $n \to \infty$
(2) $0 < \beta_n \to 0$ and ${\alpha_n}^{r+1}/\beta_n \to 0$ as $n \to \infty$

where a.s. stands for "almost surely". For a r-regular function f the following criterion gives a strongly consistent estimate of q:

$$\hat{q} = \arg \min_{0 \le k < p} \{ f(l_{k+1}^{(n)}, l_{k+2}^{(n)}, \cdots, l_p^{(n)}) + k\beta_n \}.$$

The advantage of this method is that it is not necessary to compute the likelihood ratio test statistic and to count the number of parameters. It seems that the larger the r, the faster the convergence.

For an example of this type of estimate, let $l_i^{(n)}$ be the eigenvalues δ_i of the covariance matrix S. Then $|l_i^{(n)} - \lambda_i| = O\hat{E}$ a.s. or $\alpha_n = \widehat{q^l}$. β_n is chosen to satisfy

$$\beta_n (n/\log\log n)^{(r+1)/2} \to 0. \tag{14}$$

For instance, $\beta_n = c\ (\log n / n)^{(r+1)/2}$, where c is a constant, will satisfy the requirement. Choose an r-function f with $r \ge 2$, then the error bound of wrong detection

$$P(\hat{q} \ne q) \le pP\left(\beta_n \le \sum_{i=1}^{p} |l_i^{(n)} - \lambda_i|^{r+1}\right) \tag{15}$$

tends to zero (11) faster than that of Eq. (10).

Color-Noise Case. When the noise is colored, i.e., the noise covariance matrix is not equal to $\sigma^2 I$, the criteria discussed above are not valid. Consider the case for two independent samples, one for the noise and the other for the received signals with arbitrary noise added. A type of pre-white-noise procedure is performed first. In particular, let S_1 be a sample covariance matrix estimate for noise covariance matrix Σ_1 and let S_2 be the independent estimate of Σ_2, where

$$\Sigma_2 = A\Psi A^* + \Sigma_1.$$

Then ${\Sigma_1}^{-1/2}\Sigma_2{\Sigma^{-1/2}}_1 = {\Sigma^{1/2}}_1 A\Psi A^* {\Sigma^{-1/2}}_1 + \sigma^2 I$ is similar to Eq. (3). In this case a strongly consistent information-theoretic criterion (12) is given as

$$\hat{q} = \arg\min_{0\le k<p}\left\{\log\prod_{i=k+1}^{p}\left(\frac{[(n_2\delta_i + n_1\hat{\lambda}_{k0})/n]^n}{\delta_i^{n_2}\hat{\lambda}_{k0}^{n_1}}\right) + \nu(p,k)C_n\right\} \tag{16}$$

where $\delta_1 \ge \delta_2 \ge \cdots \ge \delta_p$ are the eigenvalues of $S_2{S_1}^{-1}$, $\left(\sqrt{\log\log n/n}\right)_{k0}$ satisfies the following equation:

$$p - k = \sum_{j=k+1}^{p}\frac{n\hat{\lambda}_{k0}}{n_1\hat{\lambda}_{k0} + n_2\delta_j}$$

n_1 and n_2 are, respectively, the sample sizes for S_1 and S_2, and $n = n_1 + n_2$. The terms $\nu(p, k) = k(2p - k + 1)/2$ and C_n satisfy Eq. (11).

For the case when only one sample of the output on the receivers is available, a method employing two well-separated arrays, with, respectively, p_1 and p_2 sensors to receive the q signals, is formulated (13,14). Suppose that both p_1 and p_2 are larger than q and that $p_1 \le p_2$. As in the case of Eq. (2), these outputs may be written as

$$\boldsymbol{x}_i(t) = A_i(\Phi_i)\boldsymbol{s}(t) + \boldsymbol{n}_i(t), \quad i = 1, 2$$

The covariance matrix of $(\boldsymbol{x}_1(t), \boldsymbol{x}_2(t))^T$ (where the superscript T denotes transpose) is equal to

$$\Sigma = \begin{pmatrix}\Sigma_{11} & \Sigma_{12}\\ \Sigma_{21} & \Sigma_{22}\end{pmatrix}$$

where $\Sigma_{ii} = A_i(\Phi_i)\,\Psi{A_i}^*(\Phi_i) + \Sigma_i$, $i = 1{,}2$ and $\Sigma_{12} = {\Sigma_{21}}^* = A_1(\Phi_1)\,\Psi{A_2}^*(\Phi_2)$.

Suppose that there is a set of independent observations $({\boldsymbol{x}^T}_1(t_1), {\boldsymbol{x}^T}_2(t_1))^T, \cdots, ({\boldsymbol{x}^T}_1(t_n), {\boldsymbol{x}^T}_2(t_n))^T$ on the two arrays, where ${\boldsymbol{x}^T}_1(t_1)$ denotes the transpose of $\boldsymbol{x}_1(t_1)$ and $\boldsymbol{x}_1(t_i)$ and $\boldsymbol{x}_2(t_i)$ are, respectively, the outputs on the two arrays. With the assumption that the noises impinging on the two arrays are independent, and with model M_k of Eq. (4) the logarithmic likelihood statistics are equivalent to

$$L_k(X) = -n\log\prod_{i=1}^{p_1}(1 - \gamma_i^2)$$

where ${\gamma^2}_1 \ge {\gamma_2}^2 \ge \cdots \ge {\gamma_{p1}}^2$ are the eigenvalues of ${S_{11}}^{-1/2}\,S_{12}{S_{22}}^{-1}S_{21}{S_{11}}^{-1/2}$ and

$$S_{jj} = \frac{1}{n}\sum_{i=1}^{n}\boldsymbol{x}_1(t_i)\boldsymbol{x}_1^*(t_i), \quad j = 1, 2; \quad S_{12} = S_{21}^* = \frac{1}{n}\sum_{i=1}^{n}\boldsymbol{x}_1(t_i)\boldsymbol{x}_2^*(t_i) \tag{17}$$

Given an increasing function $\nu(k)$, a strongly consistent estimate$\sqrt{\log \log n/n}$ of q (14) is given as

$$\hat{q} = \arg \min_{0 \le k < p_1} \{L_k(X) + \nu(k) C_n\} \tag{18}$$

where C_n may be chosen as a function satisfying Eq. (11).

Coherent Signals

The case with fully correlated signals is referred to as the coherent signal case. In this situation, it is usually assumed that $A(\Phi)$ is of full rank, which is supported by real data. It is of interest to estimate the rank q of $A(\Phi)$ and the true number of signals, which is less than q. The signals are assumed to be either random or nonrandom. If the signals are assumed to be random, the true number of signals q' is the same as the rank of the covariance matrix of the signals. The previous approaches are usually not applicable to the coherent case since Ψ will be singular and hence λ_q will not be greater than λ_{q+1}. However, an information-theoretic criterion for a uniform linear array of sensors can be formulated. It uses a preprocessing scheme that involves stacking the data sequence so that the new Ψ matrix will be nonsingular (15).

For the case of white noise, the method in Ref. 11 still works with the use of differently defined $l_i^{(n)}$, and an example is given in Ref. 11. Another example is given by the differential residues (16). For a given model M_k of Eq. (4), a $p \times p$ matrix $A(\Phi^{(k)})$ as a function of $\Phi^{(k)}$ can be computed as:

$$A(\Phi^{(k)}) = [A(\phi_1), \ldots, A(\phi_k)]$$

where $\Phi^{(k)}$ denotes the vector $(\phi_1, \ldots, \phi_{k)}$ and the projection matrix $P(\Phi^{(k)})$ onto the noise subspace is equal to

$$P(\Phi^{(k)}) = I - A(\Phi^{(k)})[A(\Phi^{(k)})]^*[A(\Phi^{(k)})]^{-1} A(\Phi^{(k)})^*.$$

Define the set $\{l_k^{(n)}, k = 0,1,\cdots, K\}$ as

$$l_k^{(n)} = \begin{cases} \min\limits_{\Phi^{(k)}} \operatorname{tr} P(\Phi^{(k)}) S - (p-k-1) l_K & k < K \\ \frac{1}{p-K} \min\limits_{\Phi^{(K)}} \operatorname{tr} P(\Phi^{(K)}) S & k = K \end{cases}$$

where tr stands for the trace of a matrix, $K = \lfloor p/2 \rfloor$, the largest integer less than or equal to $p/2$ and q is assumed to be bounded above by $p/2$. Three r-functions were used:

$$f_1(k) = -\sum_{i=k}^{K} \log l_i^{(n)} + (K+1-k)\log\left(\frac{1}{K+1-k}\sum_{i=k}^{K} l_i^{(n)}\right)$$

$$f_2(k) = \frac{1}{K+1-k}\sum_{i=k}^{K}\left(l_i^{(n)} - \frac{1}{K+1-k}\sum_{i=k}^{K} l_i^{(n)}\right)^4$$

$$f_3(k) = \exp\left(\left|l_k^{(n)} - \frac{1}{K+1-k}\sum_{i=k}^{K} l_i^{(n)}\right|\right).$$

An estimate $\hat{q}$ of q is then given as

$$\hat{q} = \arg\min_{0\le k\le K}\{f_i(k) + k\beta_n\}, \quad i = 1, 2, 3$$

The f_1 and f_2 are, respectively, 1- and 3-regular functions while f_3 is a 1-regular function. The sequence β_n can be taken as the one given in Eq. (14). The probability of error bound in Eq. (15) holds for these three criteria.

Another criterion based on the MDL principle is given in Ref. 17. Let $l_1(\Phi^{(k)}) \ge \cdots \ge l_{p-k}(\Phi^{(k)})$ denote the nonzero eigenvalues of the $p \times p$ matrix $P(\Phi^{(k)})SP(\Phi^{(k)})$. Suppose that

$$\hat{\Phi}^{(k)} = \arg\min_{\Phi^{(k)}} \sum_{i=1}^{p-k} l_i(\Phi^{(k)})$$

Then the MDL based criterion is given as

$$\hat{q} = \arg\min_{0\le k<p}\{LC(X|\hat{\Phi}^{(k)}) + \tfrac{1}{2}k(2p-k+1)\log n\} \tag{19}$$

where

$$LC(X|\hat{\Phi}^{(k)}) = n(p-k)\left\{\log\frac{1}{p-k}\sum_{i=1}^{p-k} l_i(\hat{\Phi}^{(k)}) - \frac{1}{p-k}\sum_{i=1}^{p-k}\log(\hat{\Phi}^{(k)})\right\}$$

Note the similarity of this function and that of the logarithmic likelihood function in Eq. (7). This is no coincidence in that they both involve eigenvalues for the noise subspace. Improving on the estimate of the $\hat{\Phi}^{(k)}$ and with a smaller set of free parameters, another MDL type estimator was derived in Ref. 18:

$$\hat{q} = \arg\min_{0\le k<p}\{LD(X|\hat{\Phi}^{(k)}) + \tfrac{1}{2}k(k+1)\log n\} \tag{20}$$

where

$$LD(X|\hat{\Phi}(k)) = n \log \left[\prod_{i=1}^{k} l_i^S(\hat{\Phi}^{(k)}) \left(\frac{1}{p-k} \sum_{i=1}^{p-k} l_i^N(\hat{\Phi}^{(k)}) \right)^{p-k} \right]$$

$$\hat{\Phi}^{(k)} = \arg \min_{\Phi^{(k)}} \left[\prod_{i=1}^{k} l_i^S(\Phi^{(k)}) \left(\frac{1}{p-k} \sum_{i=1}^{p-k} l_i^N(\Phi^{(k)}) \right)^{p-k} \right]$$

$l_1{}^S(\Phi^{(k)}) \geq \cdots \geq l_k{}^S(\Phi^{(k)})$ are the nonzero eigenvalues of the matrix $[I - P(\Phi^{(k)})]S[I - P(\Phi^{(k)})]$ and $l_1{}^N(\Phi^{(k)}) \geq \cdots \geq l_{p-k}{}^N(\Phi^{(k)})$ denote the nonzero eigenvalues of the matrix $P(\Phi^{(k)})SP(\Phi^{(k)})$. As pointed out in Ref. 18, this MDL criterion seems to perform better than the previous one in Eq. (19), notwithstanding more computation on the parameter $\hat{\Phi}^{(k)}$.

Reasoning that the presence of coherent signals can be detected by the "distance" between signal subspace and the array manifold, a subspace-fitting method is formulated (19). Suppose that the model M_k of Eq. (4) is true. Using the eigendecomposition of the sample covariance matrix S, let $\hat{\Gamma}_s = \text{diag}(\delta_1, \ldots, \delta_k)$ and $\hat{E}_s = (\boldsymbol{v}_1, \ldots, \boldsymbol{v}_k)$ where $\boldsymbol{v}_1, \ldots, \boldsymbol{v}_k$ are, respectively, the eigenvectors of $\delta_1, \ldots, \delta_k$. Suppose that

$$W = \tilde{\Gamma}^2 \hat{\Gamma}_s^{-1}, \quad \tilde{\Gamma} = \hat{\Gamma}_s - \hat{\sigma}^2 I$$

where $\hat{\sigma}^2 = [1/(p-k)] \Sigma^p{}_{i=k+1} \delta_i$. Define

$$\hat{\Phi}^{(k)} = \arg \min_{\Phi^{(k)}} \text{tr}\{P(\Phi^{(k)}) \hat{E}_s W \hat{E}_s^*\}$$

$$g(k) = \frac{2n}{\hat{\sigma}^2} \text{tr}\{P(\hat{\Phi}^{(k)}) \hat{E}_s W \hat{E}_s^*\}$$

Starting with an estimate $\widehat{q'}$ of q', such as found by applying the MDL criteria for white noise, and with C_n satisfying Eq. (11), a test of hypotheses for a given significance level α proceeds as follows:

(1) Set $k = \widehat{q'}$.

(2) Compute the critical value γ for a right-tailed area α of χ^2 distribution with $2\widehat{q'}(p-k) - k$ degrees of freedom.

(3) If $g(k) \leq \gamma C_n$, then let $\hat{q} = k$; otherwise increase k by 1 and go to step 2.

With properly chosen C_n, this procedure gives a strongly consistent estimate of q.

Using the model of two well-separated arrays for the case of colored noise, a method based on a test of hypotheses is given in Ref. 20. Assuming the model M_k, let γ_i, $i = 1, 2, \ldots, k$ be the singular values of $S_{11}{}^{-1/2} S_{12} S_{22}{}^{-1/2}$ and U be the matrix of the left singular vectors. The canonical vector matrix is $L = S^{-1/2}{}_{11} U$. Let

L_s be the first k columns of L and also R_s be the first k columns of ${S_{11}}^{1/2}\,U$. Define

$$A_1(\Phi^{(k)}) = (A_1(\phi_1), \cdots, A_1(\phi_k))$$
$$f(k) = \mathrm{tr} S_{11}^{-1} R_s \Gamma^2 R_s^* - \max_{\Phi^{(k)}} \mathrm{tr} A_1(\Phi^{(k)})[A_1^*(\Phi^{(k)}) S_{11}^{-1} A_1(\Phi^{(k)})]^{-1} A_1^*(\Phi^{(k)}) L_s \Gamma^2 L_s^*$$

where $\Gamma = \mathrm{diag}(\gamma_1, \ldots, \gamma_k)$. Then $2nf(k)$ has an asymptotic χ^2 distribution of $2q'(p_1 - k) - k$ degrees of freedom. In order to apply the test, an estimate of q' is required and a method is given in Ref. 21. Denote the estimate of q' by $\widehat{q'}$. The test then proceeds as follows:

(1) Set $k = \widehat{q'}$.
(2) Compute $2n\,f(k)$.
(3) Find the critical value $c(k)$ for a χ^2 distribution with $2\widehat{q'}(p_1 - k) - k$ degrees of freedom for a given α significance level.
(4) If $2nf(k) \le c(k)$, set $\hat{q} = k$. Otherwise increase k by 1 and go to step 2.

Analysis

The accuracy of detection is measured by the probability of error, $P(\hat{q} \neq q | M_q)$. It is clear that this quantity depends on how a detection criterion is constructed. When the criterion is based on the sample eigenvalues, the accuracy of the detection depends on how accurately the eigenvalues of the sample covariance matrix approximate those of a true covariance matrix. When two or more signals have nearly equal angles of arrival, the matrix A will have two or more nearly equal columns. This will in term imply that the some of eigenvalues, other than those of the last $p - q$ are almost equal. With small n, this may affect the accuracy of the detection.

For two Hermitian $p \times p$ matrices A and B with their ordered eigenvalues $\lambda_1 \ge \lambda_2 \ge \cdots \ge \lambda_p$ and $\gamma_1 \ge \gamma_2 \ge \cdots \ge \gamma_p$, respectively, it can be shown that

$$|\lambda_i - \gamma_i| \le p \max_{i,j} |a_{ij} - b_{ij}|$$

where $A = (a_{ij})$ and $B = (b_{ij})$. Under certain conditions, $\Sigma - S = O\,(\sqrt{\log \log n / n}$ a.s.). Thus it is no surprise that the probability of wrong detection depends on sample size and also on the signal-to-noise ratio (SNR). The most general type of result about this probability (8,14,22) is of the following form:

$$P(\hat{q} \neq q | M_q) \le \sum_{i,j} P(|s_{ij} - \sigma_{ij}| \ge \alpha)$$

where α satisfies certain conditions. This inequality holds for MDL criteria of Eq. (9) and criteria of Eqs. (10), (16), and (18) under less stringent conditions such as (i) $\nu\,(p, k)$ or $\nu\,(k)$ are increasing, (ii) C_n satisfies Eq. (11), (iii) the signal $\boldsymbol{x}(t)$ has at least second moments, and (iv) $\alpha = \alpha_n \to 0$, $C_n/(n\alpha^2) \to 0$ as $n \to \infty$.

If signals are such that $E\{\exp[\kappa|\boldsymbol{x}(t)|^2]\} < \infty$ for some $\kappa > 0$, then an exponential decrease of the probability of error is possible:

$$P(\hat{q} \neq q|M_q) \leq ce^{(-bn^{1-\epsilon})}$$

where b, c, and ϵ are some positive numbers. Analysis of the performance of signal detection may also include quantification of its SNR threshold under which the detection algorithms fail (23).

Discussion

When the sample size is small, the penalty term in the criteria significantly affects the performance. Even for a small change such as a different constant factor, the results could be drastically different. It is then reasonable to make the penalty also depend on the data set, and an example of such a criterion was mentioned in Ref. 14. More studies of this type may be of great value.

For the information-theoretic criterion, an accurate evaluation of the information provided by the data set should yield an improved determination. As an example, in the derivation of the MDL, a smaller parameter set would require a shorter encoding string and thus a smaller description length. Such is the case of the criterion in Eq. (12). Another possibility is to use the Bayesian approach (24).

Many detection criteria were derived with the assumption that the noise is complex normally distributed [denoted as condition (N)]. A criterion may still be consistent even without condition (N), but its performance for a small sample is questionable. Note also that the complex normality assumption is at most an approximation to reality. It is often the case that the noise is not complex normal or it is far from being complex normal. For example, the noise may be a mixture of two independent noises, where one represents the nominal background noise and the other could be an impulsive component. Even through both noises are complex normal, the mixture may not be normal. Furthermore, the assumption of the independence of the observations may be a problem [which was relaxed to include the case when the sequence of observations is a mixing sequence (7,11)]. Thus there may be a need to look into robust criteria.

Other recent research on signal detection includes (but is not limited to) many specific topics, such as the multiple sources within a cluster (25); the signals in unknown noise fields (26, 27); and adaptive algorithms (28).

BIBLIOGRAPHY

1. S. M. Kay *Fundamentals of Statistical Signal Processing*, Englewood Cliffs, NJ: Prentice-Hall, 1998.
2. R. J. Muirhead *Aspects of Multivariate Statistical Theory*, New York: Wiley, 1982.
3. W. G. Chen K. M. Wong J. P. Reilly Detection of the number of signals: A predicted eigen-threshold approach, *IEEE Trans. Signal Process.* **39**: 1088–1098, 1991.
4. Y. Sakamoto M. Ishiguro G. Kitagawa *Akaike Information Criterion Statistics*, Hingham, MA: Reidel, 1986.
5. J. Rissanen *Stochastic Complexity in Statistical Inquiry*, Teaneck, NJ: World Scientific, 1989.
6. M. Wax T. Kailath Detection of signals by information theoretic criteria, *IEEE Trans. Acoust. Speech, Signal Process.*, **33**: 387–392, 1985.
7. L. C. Zhao P. R. Krishnaiah Z. D. Bai On determination of the number of signals in presence of white noise, *J. Multivariate Anal.* **20**: 1–25, 1986.
8. Z. D. Bai P. R. Krishnaiah L. C. Zhao On rates of convergences of efficient detection criteria in signal processing, *IEEE Trans. Inf. Theory*, **35**: 380–388, 1989.

9. K. M. Wong *et al.* On information theoretic criteria for determining the number of signals in higher resolution array processing, *IEEE Trans. Acoust. Speech, Signal Process.* **38**: 1959–1971, 1990.
10. K. W. Tam Y. Wu On rate of convergence of some consistent estimates of the number of signals, *Proc. 8th IEEE-SP Workshop Statistical Signal Array Process.*, Greece, 1996, pp. 44–47.
11. Y. Q. Yin P. R. Krishnaiah On some nonparametric methods for detection of the number of signals, *IEEE Trans. Signal Process.* **35**: 1533–1538, 1987.
12. L. C. Zhao P. R. Krishnaiah Z. D. Bai On determination of the number of signals when the noise covariance matrix is arbitrary, *J. Multivariate Anal.* **20**: 26–49, 1986.
13. P. Stoica K. M. Wong Q. Wu On a nonparametric detection method for array signal processing, *IEEE Trans. Signal Process.* **44**: 1030–1033, 1996.
14. Y. Wu K. W. Tam On determination of the number of signals in spatially correlated noise, *IEEE Trans. Signal Process.* **46**: 3023–3029, 1998.
15. H. Krim J. H. Cozzens data-based enumeration techniques for fully correlated signals, *IEEE Trans. Signal Process.* **42**: 1662–1668, 1994.
16. Q. Wu D. R. Fuhrman A parametric method for determining the number of signals in narrow-band direction finding, *IEEE Trans. Signal Process.* **39**: 1848–1857, 1991.
17. M. Wax I. Zizkind Detection of the number of coherent signals by the MDL principle, *IEEE Trans. Acoust., Speech, Signal Process.* **37**: 1190–1196, 1989.
18. M. Wax Detection and localization of multiple sources via the stochastic signals model, *IEEE Trans. Signal Process.* **39**: 2450–2456, 1991.
19. M. Viberg B. Ottersten T. Kailath Detection and estimation in sensor arrays using weighted subspace fitting, *IEEE Trans. Signal Process.*, **39**: 2436–2448, 1991.
20. Q. Wu K. M. Wong Determination of the number of signals in unknown noise environments—PARADE, *IEEE Trans. Signal Process.*, **43**: 362–365, 1995.
21. W. G. Chen J. P. Reilly K. M. Wong Detection of the number of signals in the presence of noise with unknown banded covariance matrices, *ICASSP' 92*, 1992, pp. 377–380.
22. K. W. Tam Y. Wu On rates of convergence of general information theoretic criteria in signal processing when the noise covariance matrix is arbitrary, *IEEE Trans. Inf. Theory*, **37**: 1667–1671, 1991.
23. H. Lee Fu Li Quantification of the difference between detection and resolution thresholds for multiple closely-spaced emitters, *IEEE Trans. Signal Process.*, **41**: 2274–2277, 1993.
24. P. M. Djuric Asymptotic MAP criteria for model selection, *IEEE Trans. Signal Process.*, **46**: 2726–2735, 1998.
25. H. Lee Fu Li An eigenvector technique for detecting the number of emitters in a cluster, *IEEE Trans. Signal Process.*, **42**: 2380–2388, 1994.
26. P. Stoica M. Cedervall Detection tests for array processing in the unknown noise correlated noise fields, *IEEE Trans. Signal Process.*, **45**: 2351–2362, 1997.
27. M. Wax J. Sheinvald A. Weiss Detection and localization in colored noise via generalized least squares, *IEEE Trans. Signal Process.*, **44**: 1734–1743, 1996.
28. X. Yu I. Reed Adaptive detection of signals with linear feature mappings and representations, *IEEE Trans. Signal Process.*, **43**: 2953–2963, 1995.

FU LI
Portland State University
KWOK-WAI TAM
Portland State University
YUEHUA WU
York University