An Information Theoretic Approach to Source Enumeration in Array Signal Processing

Shahrokh Valaee, *Senior Member, IEEE,* and Peter Kabal

Abstract—In this paper, a new information theoretic algorithm is proposed for signal enumeration in array processing. The approach is based on predictive description length (PDL) that is defined as the length of a predictive code for the set of observations. We assume that several models, with each model representing a certain number of sources, will compete. The PDL criterion is computed for the candidate models and is minimized over all models to select the best model and to determine the number of signals. In the proposed method, the correlation matrix is decomposed into two orthogonal components in the signal and noise subspaces. The maximum likelihood (ML) estimates of the angles-of-arrival are used to find the projection of the sample correlation matrix onto the signal and noise subspaces. The summation of the ML estimates of these matrices is the ML estimate of the correlation matrix. This method can detect both coherent and noncoherent signals. The proposed method can be used online and can be applied to time-varying systems and target tracking.

Index Terms—Adaptive signal detection, array signal processing, information theoretic signal resolution, information theory, maximum likelihood detection, minimum description length, modeling, predictive description length, signal detection, signal resolution, stochastic complexity, time-varying systems.

I. INTRODUCTION

ARRAY signal processing involves signal enumeration and source localization. Array processing can be applied to many applications ranging from radar and sonar to mobile communications. In various applications, the objective might be to determine the number of sources. This process is often called *enumeration*. An enumerator uses a set of hypotheses to estimate the number of signals. Each hypothesis represents a fixed number of signals, and the best hypothesis is selected by optimizing an appropriate cost function.

There exist a number of different approaches to signal enumeration [1]. Recently, much attention has been given to information theoretic criteria; see, for instance, [2]–[9]. Two such criteria are the *Akaike information criterion* (AIC) [10] and the *minimum description length* (MDL) [11]. These methods minimize the Kullback–Leibler distance between the hypothesized model and the observed data. The number of sources is determined by computing any one of these criteria for all candidate

models and choosing the model with the smallest description. A general class of information theoretic criteria, evolving from MDL, has been discussed in [4].

The MDL criterion is based on minimizing the length of the code required to describe data. Codelength minimization is appropriate for model selection since the model that best fits the data is the one that gives the most information about it; having more information results in a smaller codelength. Coding of data in the MDL criterion is performed in two steps. First, the data is encoded using a uniquely decodable prefix code. Then, the parameter vector is encoded and added as a preamble to the codeword of data. The overparameterization term in the MDL criterion represents the number of digits required to encode the parameter vector to an optimal precision [12]. The MDL principle can take different forms [13], [14]. Here, we speak of MDL as a two-step coding scheme. This definition of MDL coincides with the *Baysian information criterion* (BIC) [15], which is a model selection criterion that is widely used in statistical community. In the sequel, we will refer to the "two-step MDL" simply as MDL since this term has frequently been used in array processing.

By *coherent* (fully correlated) sources, we refer to a setting in which the signals emitted by two or more sources are identical, except possibly for a multiplicative constant factor. *Noncoherent* signals are the ones that are not coherent. For coherent signals, the rank of the signal cross-correlation matrix is smaller than the number of signal [16]. The signal enumerator techniques that use the multiplicity of the noise eigenvalue cannot resolve the true number of coherent signals.

The application of information theoretic approach to signal enumeration in array signal processing was first proposed in [2]. Wax and Kailath [2] propose two enumerators using the AIC and MDL methods. However, their methods cannot resolve coherent sources. In [3], a signal enumerator has been proposed that can resolve coherent sources. The method uses the MDL principle and decomposes data into signal and noise components. The MDL descriptor is then computed for signal and noise components separately, and the results are added to obtain the total MDL cost. In this paper, we will use a similar approach, decompose data into signal and noise components, and apply our technique to these components. In [5], the performance of the information theoretic criteria has been studied, and the probability of error for the AIC and MDL techniques has been derived using the asymptotic distribution of the sample eigenvalues.

All information theoretic techniques discussed so far use eigenvalue decomposition. Wu and Fuhrmann [6] propose an alternative approach. They use "differential residues" to formulate an information theoretic algorithm. The differential residues are defined by a recursion of the noise eigenvalues.

S. Valaee is with the Department of Electrical and Computer Engineering, University of Toronto, Toronto, ON, M5S 3G4 Canada (e-mail: valaee@comm.utoronto.ca).

P. Kabal is with the Department of Electrical and Computer Engineering, McGill University, Montreal, QC, H3A 2A7 Canada.

The authors show that using differential residues instead of eigenvalues decreases the resolution *signal-to-noise ratio* (SNR) threshold and increases the computational complexity.

In [7], the MDL principle has been formulated for cyclostationary signals. It has been shown that the number of cyclostationary signals might be larger than the number of sensors. In fact, the specific structure of cyclostationary signals allows the number of signals grow beyond what is usually dictated by the number of sensors. The techniques discussed above assume that noise is spatially uncorrelated. In [8], a signal enumerator for spatially correlated noise has been proposed.

Green and Taylor [9] summarize the enumeration methods based on AIC, MDL, and the modified rank sequence method of [17] and then propose a stopping criterion to estimate the number of signals. The authors also assume that a variable number of antennas is available. Then, they use their algorithm to minimize the required number of antennas with respect to the number of signals. We will show in Section V that the method of [9] does not perform better than MDL for small SNR.

Restricting the coding method to a two-step scheme in the MDL approach increases the codelength. It has been shown that the MDL is consistent [4]. Both the AIC and MDL methods can only be applied to a batch of data; they cannot be used in a recursive manner. In this paper, we develop an algorithm based on the *predictive description length* (PDL) [12], [18]. The PDL criterion is the cumulative log-likelihood function of the observation vectors such that at each time instant, the maximum likelihood (ML) estimate of the parameter vector, based on the past data, is used in the probability distribution function. It has been shown that the PDL achieves the shortest codelength for data relative to the generating model class [18]. The algorithm is recursive, and its structure makes it suitable for online use.

The proposed technique exploits the unique structure of the signal and noise subspaces and can resolve coherent and noncoherent signals. We use the *alternating projection* algorithm [19] for the ML estimation. We choose to use alternating projection for its simplicity. Our simulation studies show that the proposed technique is very effective in detecting the number of incident signals and outperforms the MDL [3] and G&T [9] algorithms.

II. INFORMATION THEORETIC CRITERIA

Suppose that an array of p sensors is exposed to $q < p$ far-field sources. The signals from the sources can be partially or fully correlated. The fully correlated case (also called the coherent case) arises from multipath propagation or smart jamming and is of practical importance in signal processing. We assume narrowband signals with known center frequency. If the size of the array is much smaller than the distance between the array and the sources, the sources are assumed to be in the far-field of the array of sensors. With this modeling, the arriving wavefronts at the array are planar and do not carry any information about the range of the sources. It is also frequently assumed that the sources and the sensors are in the same plane. In such a case, the only information about the position of a source is in the *direction-of-arrival* (DOA) vector. Assume that the signals arrive at the array from distinct directions $\theta_1, \ldots, \theta_q$.

The objective is to find the number of sources q and their DOA $\boldsymbol{\theta}^q = [\theta_1, \ldots, \theta_q]$.

The complex envelope representation of the ith snapshot of the array output is

$$\mathbf{x}_i = \mathbf{A}(\theta^q)\mathbf{s}_i + \mathbf{n}_i \qquad (1)$$

where $\mathbf{x}_i$ is the $p \times 1$ observation vector, $\mathbf{s}_i$ is the $q \times 1$ signal vector, $\mathbf{n}_i$ is the $p \times 1$ noise vector, and $\mathbf{A}(\boldsymbol{\theta}^q) = [\mathbf{a}(\boldsymbol{\theta}_1) \ldots \mathbf{a}(\boldsymbol{\theta}_q)]$ is the $p \times q$ location matrix. The array manifold is defined by $\mathcal{A} = \{\mathbf{a}(\theta)|\theta \in \Theta\}$, where Θ is the region of search. It is assumed that the array manifold is known and that any p location vectors $\mathbf{a}(\theta_\ell), \ell = 1, \ldots, p$ with distinct θ_ℓ's are linearly independent. The methods that we present in this paper can be applied to general array manifold. However, for simplicity, we have used linear arrays in our simulations. For uniform linear arrays, the ℓth location vector is represented by $\mathbf{a}(\theta_\ell) = [1 \quad \mu_\ell \quad \mu_\ell^2 \quad \ldots \quad \mu_\ell^{p-1}]^T$ with $\mu_\ell = \exp(-j\omega_0 d \sin \theta_\ell / c)$, where d is the distance between two adjacent sensors, c is the wave speed, and ω_0 is the center frequency of the source signal.

The signal snapshots form an independent identically distributed (i.i.d.) sequence of complex circular Gaussian random vectors with an unknown covariance matrix $\mathbf{S}^q$. As we will discuss later, this is not a restrictive assumption. Indeed, similar to [3], we can use the proposed algorithm with the signal samples modeled as unknown constants. The noise snapshots form an i.i.d. sequence of complex circular Gaussian random vectors with unknown covariance matrix $\sigma^2 \mathbf{I}$, independent of the signal samples where prewhitening can be used to accommodate nonwhite noise. With these assumptions, the observation vectors will be samples of a complex circular Gaussian process with zero mean and the correlation matrix

$$\mathbf{R}^q = E[\mathbf{x}\mathbf{x}^H | q, \boldsymbol{\theta}^q, \sigma^2, \mathbf{S}^q] \qquad (2)$$
$$= \mathbf{A}(\boldsymbol{\theta}^q)\mathbf{S}^q\mathbf{A}^H(\boldsymbol{\theta}^q) + \sigma^2\mathbf{I}. \qquad (3)$$

The correlation matrix $\mathbf{R}^q$ is a function of the parameter set $\boldsymbol{\phi} = [q, \boldsymbol{\theta}^q, \sigma^2, \mathbf{S}^q]$. The conditional probability density function of the observation vector is given by

$$f(\mathbf{x}|\boldsymbol{\phi}) = \frac{1}{\pi^p |\mathbf{R}^q|} \exp\left\{-\mathbf{x}^H [\mathbf{R}^q]^{-1}\mathbf{x}\right\} \qquad (4)$$

where $|.|$ represents the determinant of a matrix.

The objective in an enumeration and localization problem is to estimate the number of sources q and their DOAs $\boldsymbol{\theta}^q$. It might be thought that the unknown parameters of $\mathbf{R}^q$ can be estimated by jointly maximizing the likelihood function. However, the maximum value of the likelihood function is an increasing function of the order of the system. In other words, direct ML estimation always gives the maximum allowable value for the number of sources q. For this reason, the MDL algorithm includes a second term that is added to the log-likelihood function to penalize overparameterization.

Mathematically, the MDL criterion is represented by

$$\text{MDL}(N) = -\log f(\mathbf{X}^N | \hat{\phi}_N) + \frac{k}{2} \log N \qquad (5)$$

where $\mathbf{X}^N$ is the $p \times N$ matrix of observations up to time N, $\hat{\boldsymbol{\phi}}_N$ is the ML estimate of the parameter vector based on N snap-

shots, k is the number of free elements of the parameter vector, and f is the generating model class (probability density function). In (5), the first term is the log-likelihood function of the observation vectors, and the second term is the overparameterization factor.

The PDL of the observation vectors $\mathbf{x}_i$, $i = 1, \ldots, N$, is defined as [18]

$$\text{PDL}(N) = -\sum_{i=1}^{N} \log f(\mathbf{x}_i | \hat{\boldsymbol{\phi}}_{i-1}) \qquad (6)$$

where $\hat{\boldsymbol{\phi}}_{i-1}$ is the ML estimate of the parameter vector with respect to the observations up to time $(i - 1)$. The PDL principle is based on the predictive encoding of data. At each time instant, the parameter vector is estimated using the past observation. Therefore, the ith term $-\log f(\mathbf{x}_i | \hat{\boldsymbol{\phi}}_{i-1})$ is indeed the codelength of the prediction error [18].

The initial point in the recursion (6) is chosen arbitrarily. In practice, $f(\mathbf{x}_1 | \hat{\boldsymbol{\phi}}_0)$ is obtained by selecting a $\hat{\boldsymbol{\phi}}_0$ from the uniform distribution $1/(\boldsymbol{\phi}_2 - \boldsymbol{\phi}_1)$ in $[\boldsymbol{\phi}_1, \boldsymbol{\phi}_2]$ for some $\boldsymbol{\phi}_1$ and $\boldsymbol{\phi}_2$ [20]. In this paper, we take an alternative approach. We collect p snapshots and form the sample correlation matrix using these snapshots. This sample correlation matrix is then used to estimate the parameter vector. The PDL is computed for all $i = p+1, p+2, \ldots, N$ and added to find the total code length.

III. SUBSPACE DECOMPOSITION

In order to handle coherent sources, we will make use of the subspace decomposition. Let us assume that the p-dimensional complex observation vector space is represented by $\mathcal{C}^p$. In the subspace decomposition approach, this vector space is decomposed into two orthogonal subspaces called the signal and the noise subspaces. The *signal subspace* is the subspace spanned by the column vectors of the location matrix $\mathbf{A}(\boldsymbol{\theta}^q)$. For noncoherent sources, the column span of $\mathbf{A}(\boldsymbol{\theta}^q)$ coincides with the span of the eigenvectors of $\mathbf{R}^q$ corresponding to the q largest eigenvalues [16]. Assuming a one-to-one relationship between $\mathbf{A}(\boldsymbol{\theta}^q)$ and $\boldsymbol{\theta}^q$, an estimate of the signal subspace can be obtained by estimating the DOAs. The ML estimate of the signal subspace for a model of order k at time instant i is represented by the column span of the matrix $\mathbf{A}(\hat{\theta}_i^k)$, where $\hat{\theta}_i^k$ is the ML estimate of the DOAs. The dimension of signal subspace is k. The *noise subspace* is the orthogonal complement of the signal subspace. The dimension of the noise subspace for a model with k signals is $(p - k)$. It is well known that the subspace decomposition is uniquely defined by the DOA angles, and vice versa [16]. We will use this fact in the sequel to devise a signal enumeration technique.

The projection matrices onto the signal and noise subspaces are denoted as

$$\mathbf{P}_s(\boldsymbol{\theta}^q) = \mathbf{A}(\boldsymbol{\theta}^q) \left(\mathbf{A}^H(\boldsymbol{\theta}^q)\mathbf{A}(\boldsymbol{\theta}^q)\right)^{-1} \mathbf{A}^H(\boldsymbol{\theta}^q) \qquad (7)$$
$$\mathbf{P}_n(\boldsymbol{\theta}^q) = \mathbf{I} - \mathbf{P}_s(\boldsymbol{\theta}^q). \qquad (8)$$

Using these matrices, the observation vector can be decomposed into two orthogonal components

$$\mathbf{x} = \mathbf{P}_s(\boldsymbol{\theta}^q)\mathbf{x} + \mathbf{P}_n(\boldsymbol{\theta}^q)\mathbf{x} \qquad (9)$$
$$= \mathbf{x}_s + \mathbf{x}_n. \qquad (10)$$

Note that the projection of the observation vector onto the noise subspace $\mathbf{x}_n$, is due to the additive noise only and is independent of the signal and the noise component in the signal subspace. Thus, the correlation matrix of the array output can be represented by

$$\mathbf{R}^q = \mathbf{R}_s^q + \mathbf{R}_n^q \qquad (11)$$

where

$$\begin{aligned} \mathbf{R}_s^q &= \mathbf{P}_s(\boldsymbol{\theta}^q)\mathbf{R}^q\mathbf{P}_s(\boldsymbol{\theta}^q) \\ &= \mathbf{P}_s(\boldsymbol{\theta}^q)\mathbf{S}^q\mathbf{P}_s(\boldsymbol{\theta}^q) + \sigma^2\mathbf{P}_s(\boldsymbol{\theta}^q) \end{aligned} \qquad (12)$$

and

$$\begin{aligned} \mathbf{R}_n^q &= \mathbf{P}_n(\boldsymbol{\theta}^q)\mathbf{R}^q\mathbf{P}_n(\boldsymbol{\theta}^q) \\ &= \sigma^2\mathbf{P}_n(\boldsymbol{\theta}^q) \end{aligned} \qquad (13)$$

are the projections of the correlation matrix onto the signal and the noise subspaces, respectively. Using this formulation, the ML estimate of the correlation matrix can be formed by adding the ML estimates of $\mathbf{R}_s^q$ and $\mathbf{R}_n^q$.

Note that $\mathbf{R}_s^q$ and $\mathbf{R}_n^q$ are orthogonal Hermitian matrices. The following lemma represents the determinant of a Hermitian matrix in terms of its orthogonal components.

Lemma 1: Let $\mathbf{A}, \mathbf{B}$ be $n \times n$ Hermitian matrices orthogonal to each other such that $\mathbf{A}^H\mathbf{B} = \mathbf{B}^H\mathbf{A} = \mathbf{0}$. If the matrix $\mathbf{C}$ is given by

$$\mathbf{C} = \mathbf{A} + \mathbf{B} \qquad (14)$$

where $\mathbf{C}$ is full rank, then

$$\zeta(\mathbf{C}) = \zeta(\mathbf{A})\zeta(\mathbf{B}) \qquad (15)$$

where $\zeta(.)$ represents the multiplication of the nonzero eigenvalues. Additionally, if $\mathbf{C}$ is full rank, then $|\mathbf{C}| = \zeta(\mathbf{C})$, where $|\mathbf{C}|$ is the determinant of $\mathbf{C}$.

Proof: Let Λ_a and Λ_b be the diagonal matrices of nonzero eigenvalues of $\mathbf{A}$ and $\mathbf{B}$, and let $\mathbf{V}_a$ and $\mathbf{V}_b$ denote the matrices of corresponding eigenvectors. Write $\mathbf{C}$ as

$$\begin{aligned} \mathbf{C} &= \mathbf{A} + \mathbf{B} \\ &= \mathbf{V}_a\Lambda_a\mathbf{V}_a^H + \mathbf{V}_b\Lambda_b\mathbf{V}_b^H \\ &= \mathbf{V}\begin{bmatrix} \Lambda_a & 0 \\ 0 & \Lambda_b \end{bmatrix}\mathbf{V}^H \end{aligned} \qquad (16)$$

where $\mathbf{V} = [\mathbf{V}_a \mathbf{V}_b]$. Since $\mathbf{A}^H\mathbf{B} = \mathbf{0}$ implies $\mathbf{V}_a^H\mathbf{V}_b = \mathbf{0}$, $\mathbf{V}$ is unitary (i.e. $\mathbf{V}^H\mathbf{V} = \mathbf{I}$). Therefore, (16) is an eigendecomposition of $\mathbf{C}$, and the Lemma follows. $\square$

Lemma 1 will be used in the following section to represent the determinant of the sample correlation matrix as the multiplication of nonzero eigenvalues of two orthogonal matrices.

IV. SOURCE ENUMERATION

An enumerator optimizes a certain cost function over a set of competing models. We represent each model by a number k selected from the set $K = \{0, 1, \ldots, p-1\}$, where p is the number of sensors. The number associated with each model indicates the number of sources in that model. Therefore, by restraining k into the set K, it is implicitly assumed that the maximum number of

sources is smaller than the total number of sensors. All vectors and matrices that belong to a model k will be represented by the corresponding superscript. Therefore, the DOA vector of model k will be denoted by θ^k, the location matrix will be represented by $\mathbf{A}(\theta^k)$, and the source correlation matrix will be shown as $\mathbf{S}^k$. In the sequel, we will minimize an appropriate cost function—derived from the concept of stochastic complexity—over all tentative models and choose the one with the smallest cost. In fact

$$\hat{q}_n = \arg\min_{k \in K} \mathrm{PDL}_k(n) \tag{17}$$

where $\hat{q}_n$ is the estimated number of signals at time n, and $\mathrm{PDL}_k(n)$ is the predictive stochastic complexity of the data over the window $[1, n]$.

We start by observing that for a full-rank matrix $\mathbf{S}^k$ (noncoherent sources), the column span of $\mathbf{A}(\theta^k)$ is the same as the span of the eigenvectors of the correlation matrix that correspond to the k largest eigenvalues. This is an important fact that begets the effectiveness of the subspace decomposition method. We use this concept to develop a PDL algorithm that can also detect coherent sources.

In the preceding section, we showed that the true correlation matrix of the array output can be represented as the sum of two orthogonal matrices $\mathbf{R}_s^q$ and $\mathbf{R}_n^q$. Since it is assumed that $\mathbf{R}^q$ is full rank, we can use Lemma 1 to prove that the determinant of $\mathbf{R}^q$ is equal to the multiplication of the nonzero eigenvalues of $\mathbf{R}_s^q$ and $\mathbf{R}_n^q$, that is

$$|\mathbf{R}^q| = \zeta(\mathbf{R}_s^q)\,\zeta(\mathbf{R}_n^q). \tag{18}$$

In practice, the true correlation matrix is unknown and is estimated by the sample correlation matrix. In the sequel, we will represent sample estimates by a "bar" $(\bar{\cdot})$ placed on top of the character and the ML estimates by a "hat" $(\hat{\cdot})$. In this paper, the sample correlation matrix is obtained by the recursion

$$\bar{\mathbf{R}}_i = \alpha\bar{\mathbf{R}}_{i-1} + (1-\alpha)\mathbf{x}_i\mathbf{x}_i^H \tag{19}$$

where $\alpha < 1$ is a real smoothing factor. In (19), the factor $1/(1-\alpha)$ is the effective length of the exponential window [21]. Similar to the true correlation matrix, we can project the sample correlation matrix $\bar{\mathbf{R}}_{i-1}$ onto the signal and noise subspaces. The projected correlation matrices for the kth model are represented by

$$\bar{\mathbf{R}}_{s_{i-1}}^k = \mathbf{P}_s(\theta^k)\bar{\mathbf{R}}_{i-1}\mathbf{P}_s(\theta^k) \tag{20}$$

$$\bar{\mathbf{R}}_{n_{i-1}}^k = \mathbf{P}_n(\theta^k)\bar{\mathbf{R}}_{i-1}\mathbf{P}_n(\theta^k) \tag{21}$$

where, like before, $\mathbf{P}_s(\theta^k)$ and $\mathbf{P}_n(\theta^k)$ are, respectively, the projection matrices onto the signal and noise subspaces; here, the signal subspace is generated by the DOA vector θ^k. We use these matrices to find the ML estimate of the correlation matrix for the kth model.

The ML estimate of the correlation matrix for the kth model and the $(i-1)$th snapshot can be represented by

$$\hat{\mathbf{R}}_{i-1}^k = \hat{\mathbf{R}}_{s_{i-1}}^k + \hat{\mathbf{R}}_{n_{i-1}}^k \tag{22}$$

where $\hat{\mathbf{R}}_{s_{i-1}}^k$ and $\hat{\mathbf{R}}_{n_{i-1}}^k$ are the ML estimates of the projection of the correlation matrix onto the signal and noise subspaces,

respectively. If $\hat{\theta}_{i-1}^k$ is the ML estimator of DOAs, we have [22], [23]

$$\hat{\mathbf{R}}_{s_{i-1}}^k = \bar{\mathbf{R}}_{s_{i-1}}^k. \tag{23}$$

Using a similar method, it is possible to show that $\hat{\mathbf{R}}_{n_{i-1}}^k$ has the same eigenvectors as $\bar{\mathbf{R}}_{n_{i-1}}^k$ and a single nonzero eigenvalue with multiplicity $(p-k)$ that can be found from

$$\hat{\sigma}\left(\hat{\theta}_{i-k}^k\right) = \frac{1}{p-k}\mathrm{tr}\bar{\mathbf{R}}_{n_{i-1}}^k \tag{24}$$

where tr stands for the trace of a matrix. Note that $\hat{\mathbf{R}}_{n_{i-1}}^k$ is obtained by applying a linear transformation $\mathbf{T}_i^k$ to $\bar{\mathbf{R}}_{n_{i-1}}^k$, that is

$$\hat{\mathbf{R}}_{n_{i-1}}^k = \mathbf{T}_{i-1}^k\bar{\mathbf{R}}_{n_{i-1}}^k \tag{25}$$

$$\mathbf{T}_{i-1}^k = \bar{\mathbf{V}}_{n,p-k}\mathrm{diag}\left[\frac{\hat{\sigma}\left(\hat{\theta}_{i-1}^k\right)}{\lambda_j\left(\bar{\mathbf{R}}_{n_{i-1}}\right)}\right]\bar{\mathbf{V}}_{n,p-k}^H \tag{26}$$

where $\lambda_j(\bar{\mathbf{R}}_{n_{i-1}}), j = 1, \ldots, p-k$ are the nonzero eigenvalues of $\bar{\mathbf{R}}_{n_{i-1}}$, and $\bar{\mathbf{V}}_{n,p-k}$ is the $p \times (p-k)$ matrix of corresponding eigenvectors; diag[.] denotes a diagonal matrix formed by the elements in the brackets.

The ML estimate for the determinant of the correlation matrix is obtained by multiplying the nonzero eigenvalues of its projected components

$$\left|\hat{R}_{i-1}^k\right| = \zeta\left(\hat{R}_{s_{i-1}}^k\right)\zeta\left(\hat{R}_{n_{i-1}}^k\right) \tag{27}$$

where from (23) and (24)

$$\zeta\left(\hat{R}_{s_{i-1}}^k\right) = \zeta\left(\bar{R}_{s_{i-1}}^k\right) \tag{28}$$

$$\zeta\left(\hat{R}_{n_{i-1}}^k\right) = \left(\hat{\sigma}\left(\hat{\theta}_{i-1}^k i\right)\right)^{p-k}. \tag{29}$$

Using these results, the PDL criterion is expressed as

$$\begin{aligned}
\mathrm{PDL}_k(N) = \sum_{i=1}^{N}\Big[&\log\zeta\left(\bar{\mathbf{R}}_{s_{i-1}}^k\right) + (p-k) \\
&\times\log\left(\frac{1}{p-k}\mathrm{tr}\bar{\mathbf{R}}_{n_{i-1}}^k\right) \\
&+\mathbf{x}_i^H\left(\bar{\mathbf{R}}_{s_{i-1}}^k + \mathbf{T}_{i-1}^k\bar{\mathbf{R}}_{n_{i-1}}^k\right)^{-1}\mathbf{x}_i\Big].
\end{aligned} \tag{30}$$

The computation of PDL depends on the estimation of the angles-of-arrival $\hat{\theta}_{i-1}^k$. In the conventional version of the PDL algorithm, the ML estimate of the parameter vector is used. The ML estimator of the DOAs—with the signal samples modeled as unknown constants—has been derived in [3] as

$$\begin{aligned}
\hat{\theta}_{i-1}^k &= \arg\min_{\psi^k}\left\{\log\left(\hat{\sigma}(\psi^k)\right)\right\} \\
&= \arg\min_{\psi^k}\mathrm{tr}\left[\mathbf{P}_n(\psi^k)\bar{\mathbf{R}}_{i-1}\right].
\end{aligned} \tag{31}$$

This is a multivariate nonlinear optimization problem. To reduce the computational complexity of the PDL algorithm, we use the *alternating projection* method [19].

In the alternating projection method, the optimization is performed in each step of the algorithm to determine the best value for one parameter element while keeping the rest of

the elements fixed. The ML estimation is hence decomposed into several one-variable nonlinear optimization problems. Let $\boldsymbol{\psi}^k = (\psi_1, \ldots, \psi_k)$, and for $1 \leq m \leq k$, define $\boldsymbol{\psi}^k_{(m)} \triangleq (\psi_1, \ldots, \psi_{m-1}, \psi_{m+1}, \ldots, \psi_k)$ in which the mth parameter ψ_m has been removed. The alternating projection algorithm uses the fact that

$$\mathbf{P}_n(\boldsymbol{\psi}^k) = \mathbf{P}_n\left(\boldsymbol{\psi}^k_{(m)}\right) + \mathbf{P}_n\left(\psi_m | \boldsymbol{\psi}^k_{(m)}\right) \qquad (32)$$

where $\mathbf{P}_n(\psi_m | \boldsymbol{\psi}^k_{(m)})$ is the projection onto the space spanned by the residual of $\mathbf{a}(\psi_m)$ when projected onto the space $\mathbf{A}(\boldsymbol{\psi}^k_{(m)})$, that is

$$
\begin{aligned}
&\mathbf{P}_n\left(\psi_m | \boldsymbol{\psi}^k_{(m)}\right) \\
&= \frac{\left(\mathbf{I} - \mathbf{P}_n\left(\boldsymbol{\psi}^k_{(m)}\right)\right) \mathbf{a}(\psi_m)\mathbf{a}^H(\psi_m) \left(\mathbf{I} - \mathbf{P}_n\left(\boldsymbol{\psi}^k_{(m)}\right)\right)}{\left\| \left(\mathbf{I} - \mathbf{P}_n\left(\boldsymbol{\psi}^k_{(m)}\right)\right) \mathbf{a}(\psi_{\mathbf{m}}) \right\|^2}.
\end{aligned} \qquad (33)
$$

We now use (32) in (31) to get

$$\hat{\theta}^k_{i-1} = \arg \min_{\boldsymbol{\psi}^k} \operatorname{tr}\left[\mathbf{P}_n\left(\boldsymbol{\psi}^k_{(m)}\right) \bar{\mathbf{R}}_{i-1} + \mathbf{P}_n(\boldsymbol{\psi}^k)\bar{\mathbf{R}}_{i-1} \right]. \qquad (34)$$

Note that the first term in (34) is independent of the angle ψ_m. Therefore, the ML estimator (31) can be decomposed into k one-dimensional optimization problems

$$\hat{\theta}^k_{i-1,m} = \arg \min_{\psi_m} \operatorname{tr}\left[\mathbf{P}_n\left(\psi_m | \boldsymbol{\psi}^k_{(m)}\right) \bar{\mathbf{R}}_{i-1} \right] \qquad (35)$$

where $\mathbf{P}_n(\psi_m | \boldsymbol{\psi}^k_{(m)})$ is given by (33). If we define

$$\mathbf{b}(\psi_m) \triangleq \frac{\left(\mathbf{I} - \mathbf{P}_n\left(\boldsymbol{\psi}^k_{(m)}\right)\right) \mathbf{a}(\psi_m)}{\left\| \left(\mathbf{I} - \mathbf{P}_n\left(\boldsymbol{\psi}^k_{(m)}\right)\right) \mathbf{a}(\psi_m) \right\|} \qquad (36)$$

then

$$\hat{\theta}^k_{i-1,m} = \arg \min_{\psi_m} \operatorname{tr}\left[\mathbf{b}^H(\psi_m)\bar{\mathbf{R}}_{i-1}\mathbf{b}(\psi_m) \right]. \qquad (37)$$

The mth subproblem is solved by fixing $\boldsymbol{\psi}^k_{(m)}$ and varying ψ_m to get the smallest cost. This procedure is continued for all parameters. The alternating projection algorithm converges to a local minimum [19].

Since the information theoretic criteria are sensitive to the estimate of the parameter vector, asymptotic bias in the DOA estimation degrades the performance of the detector. The alternating projection algorithm might converge to a local minimum [19]. If this method is applied to a cost function that is determined by the data collected in a batch of snapshots, such as in the MDL algorithm, a local minimum is found. If this local minimum is significantly far from the global minimum, the performance of the detector is degraded. In an iterative method of estimation such as PDL, since the location of the local minimum changes with every new sample of data, it is possible for the estimator to move out of the local minimum with the sample correlation matrix updates. For all the trials in our simulations, the recursive estimator converged to the true DO's within a few sample times.

The computational complexity of the PDL algorithm grows with the size of the observation window. The computational cost

TABLE I
NUMBER OF UNCORRELATED SOURCES DETECTED BY THE MDL, (38), PDL, (30), AND G&T [9] ALGORITHMS FOR 100 INDEPENDENT RUNS

detection method	k	SNR (dB)				
		−10	−5	0	5	10
MDL	1	92	10	0	0	0
	2	8	86	98	100	100
	3	0	4	2	0	0
PDL	1	89	8	0	0	0
	2	11	92	100	100	100
	3	0	0	0	0	0
G&T	1	82	100	100	36	0
	2	8	0	0	64	100
	3	10	0	0	0	0

of the technique is due to the eigenvalue decomposition of the sample correlation matrix at each snapshot. It is important to note that the PDL technique does not require the actual eigenvalues and eigenvectors of the sample correlation matrix. Indeed, approximate values for the eigenvalues and eigenvectors suffice to apply the algorithm. Using subspace tracking algorithms [21] substantially reduces the computational complexity of the PDL algorithm. Since the main objective of this paper is to introduce the PDL method, we will not discuss further the computational cost reduction techniques.

V. SIMULATION RESULTS

Consider a linear array of eight sensors exposed to two planar narrowband wavefronts arriving from $10°$ and $20°$. The spacing between two consecutive sensors is half the wave length. In the first example, we study uncorrelated sources. The total number of snapshots is 100. To compute the sample correlation matrix, we use the recursion (19) with $\alpha = 0.9$. We compute the MDL [3], G&T [9], and PDL criteria and use them to detect the number of signals. We run 100 independent trials for each SNR $= -10, -5, 0, 5, 10$ dB and find the number of times that MDL, G&T, and PDL enumerate the signals. The SNR is defined as the ratio of the variance of signal to the variance of noise. We have assumed that all signals have the same power. In the following experiments, we have set the signal variance to one and altered the variance of noise accordingly. For the MDL algorithm, we have used the method suggested in [3], that is

$$
\begin{aligned}
\mathrm{MDL}_k(N) &= N \log \zeta\left(\bar{\mathbf{R}}^k_{s_N}\right) \\
&+ N(p-k) \log\left(\frac{1}{p-k}\operatorname{tr}\bar{\mathbf{R}}^k_{n_N}\right) + \frac{k(k+1)}{2}\log N.
\end{aligned} \qquad (38)
$$

Table I shows the number of estimated sources using the MDL, G&T, and PDL algorithms. In this table and in the sequel, we use (38) for the MDL and (30) for the PDL algorithms. As shown in the table, the PDL algorithm outperforms both MDL and G&T for small SNR. In this example, for SNR $= -10$ dB, G&T randomly selects the number of signals, and the reported results do not precisely reflect the resolution capability of this algorithm.

In Table II, we show the detected number of sources in a coherent experiment. In this example, both sources emit the same signal. Note that all techniques can detect coherent signals when the operating SNR is high. The superiority of PDL is again evident in the table.

TABLE II
NUMBER OF COHERENT SOURCES DETECTED BY THE MDL, (38), PDL, (30),
AND G&T [9] ALGORITHMS FOR 100 INDEPENDENT RUNS

detection method	k	SNR (dB)				
		−10	−5	0	5	10
MDL	1	58	1	0	0	0
	2	40	88	93	93	95
	3	2	11	7	7	5
PDL	1	57	0	0	0	0
	2	43	100	100	100	100
	3	0	0	0	0	0
G&T	1	68	95	10	0	0
	2	11	5	90	100	100
	3	21	0	0	0	0

Fig. 2.   Estimated number of sources as a function of the observation window. The source at $10°$ is added at the 41st snapshot.

Fig. 1.   Probability of resolution for the PDL and MDL algorithms using an array of eight sensors exposed to two uncorrelated far-field signals. The probability of resolution is obtained for two cases: using the estimated DOAs for the model of order 2 and the true DOAs used for that model.

Fig. 3.   Difference between the PDL terms of a model of order k and the corresponding terms of the model of order $k - 1$. The source at $10°$ is added at the 41st snapshot.

The probability of resolution is defined as the ratio of the number of trials at which the true number of sources has been resolved to the total number of trials. In the second example, we assume two sources located at $10°$ and $15°$. We compute the probability of resolution for this example using 100 independent trials. In this example, since the two sources are close, G&T cannot resolve the number of signals. The probability of resolution for this example has been illustrated in Fig. 1 as a function of SNR. In this figure, we have studied two cases of interest. In the first setting, we have used the alternating projection algorithm to estimate the DOAs. The estimated DOAs are then used in the MDL and PDL enumerators. In the second setting, we have used the true value of the DOAs in the enumerator. In this case, the estimation and the detection problems are separated; the DOA uncertainty of the estimator does not affect the performance of the enumerator, and a fair comparison of the MDL and PDL algorithms is obtained. As shown in Fig. 1, in both cases, PDL outperforms MDL.

The recursive structure of the PDL algorithm can be very useful in nonstationary environment. We assume that the source at $10°$ is absent at the beginning and is added at the 41st snapshot. The SNR for this experiment is 10 dB. PDL is evaluated on a variable length window, where the size of the window
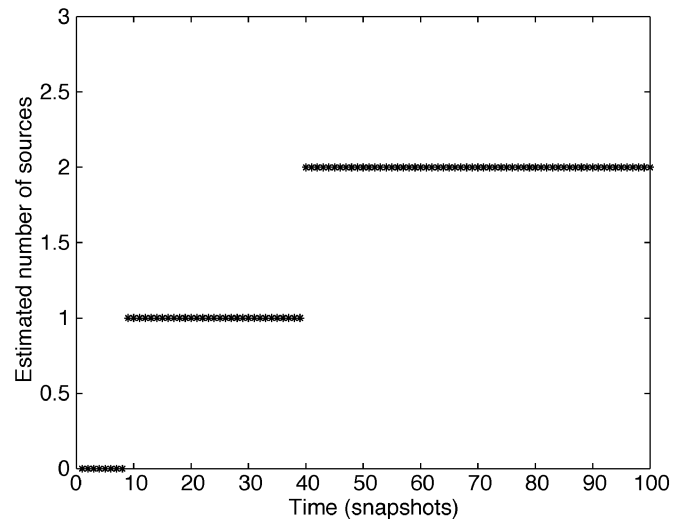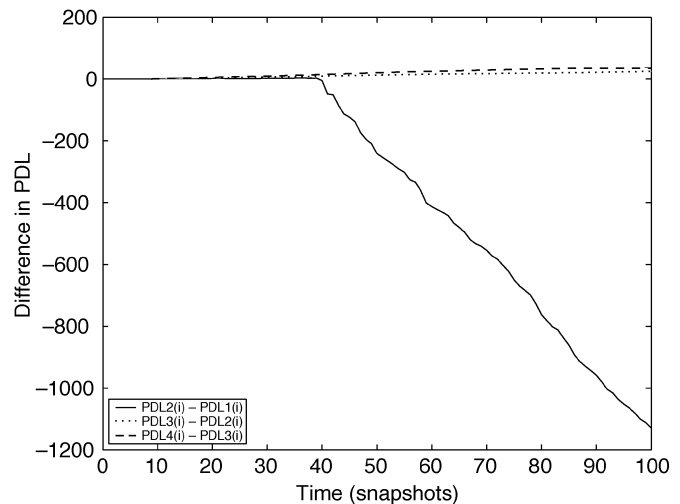
grows with the number of observations. The number of estimated sources has been illustrated in Fig. 2 as a function of time. In this example, PDL can resolve the true number of signals and can locate the change point.

For the same example, Fig. 3 illustrates the difference between the PDL terms of a model of order k and the corresponding terms of the model of order $k - 1$. Note that there exists a large difference between the PDL terms of the models of order 2 and 1 and that the difference between the two terms has a knee-point in the vicinity of the 40th snapshot. The sudden change in the rate suggests that a source appears at the 41st snapshot. This property of PDL can be used to locate the time instant at which the underlying model changes substantially.

We have also studied the case at which both sources are present at the time origin and the source at $10°$ vanishes at the 40th time instant. The result obtained by the PDL enumerator, with SNR $= 10$ dB, is illustrated in Fig. 4. Notice that PDL cannot resolve the true number of sources. PDL fails to detect
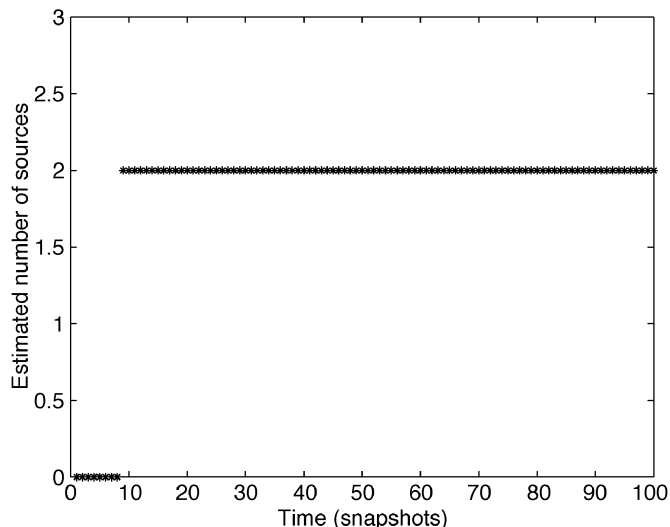
Fig. 4. Estimated number of sources as a function of the observation window. The source at $10°$ disappears at the 41st snapshot.
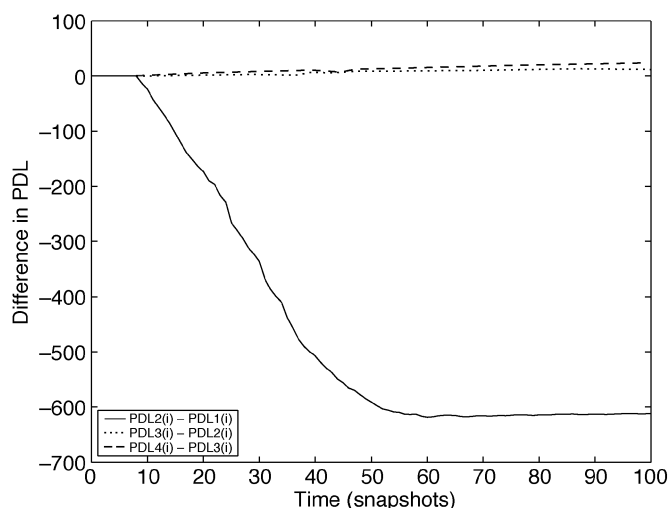

Fig. 5. Difference between the PDL terms of a model of order k and the corresponding terms of the model of order $k-1$. The source at $10°$ disappears at the 41st snapshot.

TABLE III
NUMBER OF NONCOHERENT SOURCES DETECTED BY THE MDL, PDL, AND G&T ALGORITHMS FOR 100 INDEPENDENT RUNS. THE SOURCE AT $20°$ MOVES TOWARD THE OTHER SOURCE IN A CONSTANT SPEED

detection method	k	SNR (dB)				
		−10	−5	0	5	10
MDL	1	52	26	2	0	0
	2	14	29	43	49	45
	3	34	45	55	51	55
PDL	1	76	8	0	0	0
	2	24	92	100	100	100
	3	0	0	0	0	0
G&T	1	97	100	100	37	1
	2	2	0	0	63	99
	3	1	0	0	0	0

moves toward the other source in a constant speed. We have assumed that the speed of movement is $0.05°$ per snapshot interval, and the total number of snapshots is 100. Therefore, at the end of the observation window, the sources are located at $10°$ and $15°$. As seen in the table, MDL cannot resolve the sources. Since MDL assumes that the sources are stationary over the window of observation, the technique cannot work properly in nonstationary environment. G&T only uses the rank of the correlation matrix; it does not incorporate the location of the sources in the detection algorithm. Therefore, for large SNR, G&T can resolve the sources. Unlike MDL, PDL has a recursive structure, and therefore, it can handle the variation of the model in a more effective way and results in a better performance.

We should note that there exist various techniques for target tracking of maneuvering sources; see, for instance, [24] and [25]. The authors in these two papers are looking for methods to estimate the number of signals for their far-field DOA tracking algorithms. Similar to the technique we have proposed in this paper, they also consider a nonstationary array correlation matrix (corresponding to a batch period) caused by a rapidly moving target.

VI. CONCLUSION

In this paper, we have introduced a new information theoretic method for signal enumeration. Our approach is based on the predictive description length. The PDL is the length of a predictive code of data. The proposed technique exploits the specific structure of the correlation matrix by decomposing it into two components in the signal and noise subspaces. Each component is encoded separately, and the results are added to obtain the total codelength. This process is performed for all candidate models, and the one with the smallest codelength is selected as the best model. It has been shown that the proposed technique enumerates coherent as well as noncoherent sources. To reduce the computational complexity of the ML estimator of the parameter vector, we have used the alternating projection method. The simulation results show that the performance of the PDL algorithm is better than that of the MDL [3] and G&T [9] methods.

the correct number of sources since it has a cumulative structure. Indeed, the aggregated PDL over $1, \ldots, 40$ becomes a dominant factor, and the PDL components over the window $41, \ldots, 100$ cannot nullify the effect of having two sources in $1, \ldots, 40$. To resolve this problem, we study the variation of the PDL cost for different models. The difference between the PDL criteria for the models $k = 1, \ldots, 4$ has been illustrated in Fig. 5. Note that the slope of the curve $\text{PDL}_2(t) - \text{PDL}_1(t)$ changes in the vicinity of the 40th snapshot. Due to the windowing effect of the sample correlation matrix with $\alpha = 0.9$, the location of the change point is slightly shifted. For smaller values of α, the change point will be detected faster. The change in the slope indicates that the underlying model may no longer be valid. Therefore, the PDL cost should be reset and should be re-evaluated over the new observation window. This approach allows a change on the number of sources be detected during the window of observation.

Table III illustrates the resolution capability of the MDL, PDL, and G&T algorithms for a case at which the source at $20°$

REFERENCES

[1] P. M. Djurić, "Selection of signal and system models by Bayesian predictive densities," Ph.D. dissertation, Univ. Rhode Island, Kingston, RI, 1990.

[2] M. Wax and T. Kailath, "Detection of signals by information theoretic criteria," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-33, pp. 387–392, Apr. 1985.

[3] M. Wax and I. Ziskind, "Detection of the number of coherent signals by the MDL principle," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 37, pp. 1190–1196, Aug. 1989.

[4] L. C. Zhao, P. R. Krishnaiah, and Z. D. Bai, "On detection of number of signals in presence of white noise," *J. Multivariate Anal.*, vol. 20, pp. 1–25, 1986.

[5] Q. T. Zhang, K. M. Wong, P. C. Yip, and J. P. Reilly, "Statistical analysis of the performance of information theoretic criteria in the detection of the number of signals in array processing," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 37, pp. 1557–1567, Oct. 1989.

[6] Q. Wu and D. R. Fuhrmann, "A parametric method for determining the number of signals in narrow-band direction finding," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 39, pp. 1848–1857, Aug. 1991.

[7] T. E. Shrimpton and S. Schell, "Source enumeration using a signal-selective information theoretic criterion," in *Proc. Military Commun. Conf.*, vol. 3, 1997, pp. 1092–1097.

[8] K. C. Tan, M. Zhang, and M. Wax, "Signal enumeration and direction-of-arrival estimation for coherent signals in correlated noise with unknown parameters," in *Proc. Int. Conf. Signal Process.*, vol. 1, 1998, pp. 339–342.

[9] P. J. Green and D. P. Taylor, "Dynamic signal enumeration algorithm for smart antennas," *IEEE Trans. Signal Processing*, vol. 50, pp. 1307–1314, June 2002.

[10] H. Akaike, "A new look at the statistical model identification," *IEEE Trans. Automat. Contr.*, vol. AC-19, pp. 716–723, 1974.

[11] J. Rissanen, "Modeling by shortest data description," *Automatica*, vol. 14, pp. 465–471, 1979.

[12] ——, "Stochastic complexity and modeling," *Ann. Statist.*, vol. 14, pp. 1080–1100, 1986.

[13] A. Barron, J. Rissanen, and B. Bin Yu, "The minimum description length principle in coding and modeling," *IEEE Trans. Inform. Theory*, vol. 44, pp. 2743–2760, Oct. 1998.

[14] M. H. Hansen and B. Yu, "Model selection and the principle of minimum description length," *J. Amer. Stat. Assoc.*, vol. 96, no. 454, pp. 746–774, June 2001.

[15] G. Schwarz, "Estimating the dimension of a model," *Ann. Statist.*, vol. 6, pp. 461–464, 1978.

[16] R. Schmidt, "A signal subspace approach to multiple emitter location and spectral estimation," Ph.D. dissertation, Stanford Univ., Stanford, CA, 1981.

[17] A. Di, "Multiple source location—a matrix decomposition approach," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-33, pp. 1086–1091, Oct. 1985.

[18] J. Rissanen, *Stochastic Complexity in Statistical Inquiry*. Singapore: World Scientific, 1989.

[19] I. Ziskind and M. Wax, "Maximum likelihood localization of multiple sources by alternating projection," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 36, pp. 1553–1560, Oct. 1988.

[20] J. Rissanen, "Fisher information and stochastic complexity," *IEEE Trans. Inform. Theory*, vol. 42, pp. 40–47, Jan. 1996.

[21] B. Champagne, "Adaptive eigendecomposition of data covariance matrices based on first-order perturbations," *IEEE Trans. Signal Processing*, vol. 42, pp. 2758–2770, Oct. 1994.

[22] T. W. Anderson, "Asymptotic theory for principal component analysis," *Ann. Math. Stat.*, vol. 34, pp. 122–148, 1963.

[23] E. Fishler and H. Messer, "On the use of order statistics for improved detection of signals by the MDL criterion," *IEEE Trans. Signal Processing*, vol. 48, pp. 2242–2247, Aug. 2000.

[24] Y. Zhou, P. C. Yip, and H. Leung, "Tracking the direction-of-arrival of multiple moving targets by passive arrays: algorithm," *IEEE Trans. Signal Processing*, vol. 47, pp. 2655–2666, Oct. 1999.

[25] M. Orton and W. Fitzgerald, "A Baysian approach to tracking multiple targets using sensor arrays and particle filters," *IEEE Trans. Signal Processing*, vol. 50, pp. 216–223, Feb. 2002.

Shahrokh Valaee (S'88–M'00–SM'02) was born in Tabriz, Iran. He received the B.Sc. and M.Sc. degrees from Tehran University, Tehran, Iran, and the Ph.D. degree from McGill University, Montreal, QC, Canada, all in electrical engineering.

From 1994 to 1995, he was a Research Associate at INRS Telecom, University of Quebec, Montreal. From 1996 to 2001, he was an Assistant Professor with the Department of Electrical Engineering, Tarbiat Modares University, Tehran, and an Adjunct Assistant Professor with the Department of Electrical Engineering, Sharif University of Technology, Tehran. Since September 2001, he has been an Associate Professor with the Edward S. Rogers, Sr. Department of Electrical and Computer Engineering, University of Toronto, Toronto, ON, Canada, and holds the Nortel Institute Junior Chair of Communication Networks. His current research interests are in wireless *ad hoc* and cellular networks and quality-of-service (QoS) provisioning in high-speed networks.
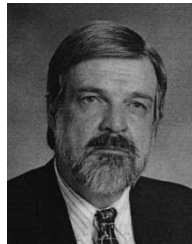
Peter Kabal received the Ph.D. degree in electrical engineering from the University of Toronto, Toronto, ON, Canada, in 1975.

He is a Professor of electrical and computer engineering at McGill University, Montreal, QC, Canada, and holds the NSERC/Nortel Industrial Research Chair. His current research interests focus on digital signal processing as applied to speech and audio processing, adaptive filtering, and data transmission.