Information Theoretic Enumeration and Tracking of Multiple Sources

Ahmad Khodayari-Rostamabad, *Student Member, IEEE*, and Shahrokh Valaee, *Senior Member, IEEE*

***Abstract*—The problem of multiple target tracking using a passive direction-finding system is addressed when the number of targets is not known *a priori*. A new method is proposed that is suitable for systems operating in low signal-to-noise ratio and high clutter. Such conditions cause unpredictable variations of stochastic characteristics of noise and signals (especially for wideband frequency ones) and create ambiguity in the output of direction-finding algorithms. In this paper, we use the predictive description length (PDL) technique, which is an information theoretic approach, and by suitable modeling, we minimize the predictive codelength for statistical data description of position measurements. The PDL-dynamic programming (DP) method is also presented, which employs the DP algorithm to reduce the computational load of the PDL technique. The concept of tracking time-varying number of targets makes PDL-DP a suitable technique for target tracking in practical systems.**

***Index Terms*—Enumeration, localization, multiple target tracking, predictive description length (PDL).**

I. INTRODUCTION

DIRECTION finding and target tracking have been the focus of active research for the last four decades. Tracking involves estimating trajectories and predicting the location of sources. Most target tracking algorithms use an estimator—as a preprocessor—to acquire raw estimates of target movement attributes such as location, velocity, and acceleration. The estimator output at each time instant is a set of entities treated as the candidates for the true attributes of the targets. A tracker is then employed to draw a temporal relationship among the candidates and to select appropriate trajectories.

In this paper, we focus on the problem of enumeration and tracking of multiple moving sources using an information theoretic approach. The information theoretic criteria, such as the *Akaike information criterion* (AIC), the *minimum description length* (MDL), and the *predictive description length* (PDL) [1] have been applied to signal enumeration and direction-of-arrival (DOA) estimation in array signal processing, [2]–[6].

Here, we derive a target tracking algorithm using the PDL principle. We consider an environment obscured by a high density of clutter and noise. In such an environment, the SNR might be well below the ambiguity threshold, resulting in a high probability of error in estimating the number of signals and their motion attributes. The description length is defined as the length of an encoded stream of data based on a prescribed model with some unknown parameters. Using the description length as a means of modeling resides on the algorithmic definition of complexity [1]. Due to this principle, the best model for the description of observed data is the one with the smallest codelength. The PDL information theoretic method is consistent and—due to its recursive structure—suits well to online applications and time varying environments.

The proposed PDL algorithm comprises two distinct parts: one part describes the estimate of the source DOAs and the other part describes the spurious observations induced by clutter and noise. We use Kalman filtering to model the kinematics of sources and formulate the multiple target tracking as a multidimensional minimization problem. In the proposed approach, the PDL cost is determined for a number of tentative models and then the true model is estimated by minimizing the description length over all tentative models. We also formulate the PDL algorithm using the dynamic programming approach to lower the computational cost. We call this method *predictive description length with dynamic programming* (PDL-DP) and show that it can also resolve a time-varying number of targets.

A. Related Literature

Several algorithms, such as *joint probabilistic data association* (JPDA), *multiple hypotheses tracking* (MHT), and *joint maximum likelihood* (ML) have been proposed for target tracking, (see [7]–[9] for excellent reviews). Several other methods exist for joint DOA estimation, data association, and tracking [10], [11]. To solve the multitarget data association, the Viterbi algorithm, dynamic programming (DP), and hidden Markov models have been proposed [12]–[14].

Recent technological developments in computational capabilities encourage the use of approaches like MHT and DP, to obtain near optimal data association in *ill-posed* multitarget tracking applications encountered with decision ambiguities. In MHT, all tentative observation-to-track association hypotheses are propagated in time, and the decision is delayed with the hope that future data can resolve the uncertainty. This results in exponential growth in the number of hypotheses or tentative track sets and, hence, in computational load versus time, [7]. The PDL-DP approach, proposed in this paper, also considers nearly all data association possibilities, but in a dynamic programming framework which is more practical.

The MDL principle is used in target tracking applications in [15]–[17]. A pioneer work was [15], where two-term MDL criterion is used for track file registration (or data fusion) in a

Manuscript received October 15, 2003; revised February 1, 2006. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Feng Zhao.

A. Khodayari-Rostamabad is with the J. Sazandeghi Engineering Research Institute (JERI), Tehran, Iran, and Tarbiat Modarres University, Tehran, Iran (e-mail: akhodayari@ieee.org).

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

Digital Object Identifier 10.1109/TSP.2007.893977

system of two platforms (sensors). It is assumed that the tracks reported by each platform consist of a set of sequences of independent measurements (state vectors) with the Gaussian estimation error and known associated error covariance matrices. Then, using these data sets, the MDL criterion is employed to find the number of targets. In [16], the MDL criterion with a modified penalty term (in place of the classical second term of MDL) has been used for multitarget enumeration. The technique employs an arbitrary constant factor in the penalty term to prevent over-parameterization. However, the constant factor has been determined in an *ad hoc* manner to obtain an unbiased estimate of the number of targets. In [17], a three-term MDL criterion [18] is used for detection and initiation of tracks using a batch of data. The MDL cost is used as the test statistics to detect multiple targets one-by-one in a multiple composite hypothesis testing framework.

The PDL principle relates to the Bayesian inference. Both techniques are based on probabilistic model selection. In other words, the PDL principle shares the view with the Bayesian inference that uncertainty can be modelled with some unknown probability distribution. However, unlike the Bayesian approach, the PDL does not require *a priori* knowledge of the parameter distribution. Indeed, the PDL does not necessitate that a "true" model should even exist [6]. Instead, the PDL simply encodes the data using a probabilistic predictor. If the predictor uses a probability model that closely reflects the intrinsic properties of the observed data, the compression gain is maximal. The PDL principle uses the length of such encoded data to select the "best-fit" model [1].

II. Problem Formulation

Assume an array of p sensors receiving the wavefronts of $q_k < p$ point sources, where k is the time index. We have assumed that the number of sources can vary in time. The observation interval is decomposed into consecutive snapshots, with each snapshot containing a segment of duration T_o of the observed data. At the end of each snapshot, an array signal processing algorithm—so-called *preprocessor*—is applied to ascertain candidates for the DOA of each source. It is assumed that the preprocessor operates under low signal-to-noise ratios (SNR), short window size, and high clutter. Due to these assumptions, the output of the array processor contains rough estimates of the angular location of signals along with some spurious angles.

Assume that the state of each source is evolved with the following linear recursion:

$$\mathbf{x}_u(k+1) = \mathbf{F}\mathbf{x}_u(k) + \mathbf{v}_u(k) \tag{1}$$

$$\mathbf{v}_u(k) = \Gamma v_u(k) \tag{2}$$

with $\mathbf{x}_u(k)$, the state of the uth source, defined as $\mathbf{x}_u(k) = [\theta_u(k), \dot{\theta}_u(k), \ddot{\theta}_u(k)]^T$, where $\theta_u(k)$ is the angular location, $\dot{\theta}_u(k)$ is the velocity, $\ddot{\theta}_u(k)$ is the acceleration, the superscript T denotes transposition, and $\mathbf{F}$ is the state transition matrix given by

$$\mathbf{F} = \begin{bmatrix} 1 & T_o & \frac{1}{2}T_o^2 \\ 0 & 1 & T_o \\ 0 & 0 & 1 \end{bmatrix} \tag{3}$$

$\mathbf{v}_u(k)$ is the state space (process) noise, $v_u(k)$ is a zero-mean Gaussian white stochastic process with unknown variance $\sigma_{v_u}^2$, and $\Gamma = [(1/2)T_o^2, T_o, 1]^T$. The correlation matrix of the noise vector is given by

$$E\left[\mathbf{v}_u(k)\mathbf{v}_l^T(j)\right] = \mathbf{Q}_u \delta_{ul}\delta_{kj} \tag{4}$$

for $u = 1, \ldots, q_k$ and $l = 1, \ldots, q_j$, where q_k is the number of sources at the kth snapshot, δ_{mn} is the Kronecker delta function, $\mathbf{Q}_u = \sigma_{v_u}^2 \Gamma\Gamma^T$, and $\sigma_{v_u}^2$ is the variance of process noise. The kinematic of sources is formulated as the piecewise constant Wiener process acceleration model.

The output of the DOA estimator for each source is represented by

$$z_u(k) = \mathbf{h}^T \mathbf{x}_u(k) + w_u(k), \quad u = 1, \ldots, q_k \tag{5}$$

where $\mathbf{h} = [1\ 0\ 0]^T$ is a vector associating the state of the source to its DOA, and $w_u(k)$—the observation noise—is a zero-mean Gaussian stochastic process with the cross correlation

$$E\left[w_u(k)w_l^T(j)\right] = \sigma_{w_u}^2 \delta_{kj}\delta_{ul} \tag{6}$$

where $\sigma_{w_u}^2$ is an unknown scalar representing the variance of the DOA estimation error.

The set of measurements at the output of preprocessor, up to time k, is represented by

$$\mathbf{Z}^k = \{\mathbf{z}(i)\}_{i=1}^k \tag{7}$$

where $\mathbf{z}(i)$ is the vector of DOA estimates at the ith snapshot. Note that $\mathbf{z}(i)$ is a collection of two types of estimates: the estimates associated to the source DOAs, denoted by $\mathbf{z}_s(i)$, and the spurious estimates resulted from noise and clutter, $\mathbf{z}_c(i)$.

The source DOAs estimate $\mathbf{z}_s(i)$ can be written as

$$\mathbf{z}_s(i) = \mathbf{A}_i^{(q_i)}\mathbf{z}(i) \tag{8}$$

where $\mathbf{A}_i^{(q_i)}$, a $q_i \times n_i$ unknown association matrix, is defined as

$$\left[\mathbf{A}_i^{(q_i)}\right]_{(l,m)} = \begin{cases} 1, & \text{if } z_m(i) \text{ is associated with source } l \\ 0, & \text{otherwise} \end{cases} \tag{9}$$

for $m = 1, \ldots, n_i$ and $l = 1, \ldots, q_i$, where n_i is the number of DOAs detected by the preprocessor at the ith snapshot.

A *track* for a source u is a smoothed version of a collection of DOA estimates associated to the uth target, such that the ith element in the set is selected from $\mathbf{z}(i)$. We represent the track for source u up to time k by

$$\begin{aligned} \mathbf{z}_s^{k,u} &= \{z_{s_u}(1), \ldots, z_{s_u}(k)\} \\ &= \{\mathbf{a}_i^u \mathbf{z}(i); \quad i = 1, \ldots, k\} \end{aligned} \tag{10}$$

where $\mathbf{a}_i^u$ is the uth row of $\mathbf{A}_i^{(q_i)}$, and $z_{s_u}(i)$ is an observation associated to the uth target. We indicate the number of possible associations at time step i by $J_i^{(q_i)}$. We assume that there is only one entity with value "1" in each row of $\mathbf{A}_i^{(q_i)}$, and the rest of the entities are zero. Similarly, there is at most one entity with value "1" in each column, with the rest of the entities equal to zero.

Assuming q_i sources $(q_i \leq n_i, i = 1, \ldots, k)$ at the ith snapshot, the total number of possible tracks is

$$J(k; \mathbf{q}) = J_1^{(q_1)} J_2^{(q_2)} \ldots J_k^{(q_k)} = \prod_{i=1}^{k} J_i^{(q_i)} \tag{11}$$

where $\mathbf{q} = [q_1, \ldots, q_k]^T$, and

$$J_i^{(q_i)} = \begin{cases} \dfrac{n_1!}{q_1!(n_1 - q_1)!}, & \text{for } i = 1 \\ \dfrac{n_i!}{(n_i - q_i)!}, & \text{for } i = 2, \ldots, k. \end{cases} \tag{12}$$

A tracker should be able to allocate an appropriate trajectory to each detected source among all tentative tracks. As seen, the number of tracks grows exponentially in time. We will show in Section IV that dynamic programming can be used to implement the multitarget trajectory estimator.

III. Predictive Description Length

We use the PDL [1] as an information theoretic criterion for target tracking. The PDL for a stream of data is defined as the smallest number of bits required to encode the data using a preselected generating model. The presumed model can be parameterized with some unknown parameters. The true model is then estimated by minimizing the PDL cost over all tentative models.

For a stream of data, indicated by $x_i, i = 1, \ldots, k$, the PDL cost is defined as

$$\mathrm{PDL}(k) \triangleq -\sum_{i=1}^{k} \log g(x_i \,|\, \hat{\psi}_{i-1}) \tag{13}$$

where $g(x_i \,|\, \psi)$ is the parameterized probability density function (the prospective generating model) and $\hat{\psi}_{i-1}$ is the causal *maximum likelihood* (ML) estimate of the unknown parameter vector ψ based on data history (observations up to time $i-1$). It has been shown in [1] that PDL generalizes the Shannon information to cases where the inherent data generating model is unknown. In the present section, we formulate the PDL criterion for our target tracking problem.

Let $\mathbf{Z}_s^k$ and $\mathbf{Z}_c^k$ be, respectively, the set of all estimates for the source DOAs and the spurious estimates over the interval of length k, that is

$$\mathbf{Z}_s^k \triangleq \{\mathbf{z}_s(i)\}_{i=1}^k \tag{14}$$

$$\mathbf{Z}_c^k \triangleq \{\mathbf{z}_c(i)\}_{i=1}^k. \tag{15}$$

Here, we assume that q_k, the number of sources, is constant over the observation interval, and all sources have been detected by the preprocessor. In Section V, we will present the PDL algorithm for time-varying number of sources.

Since $\mathbf{z}_s(i)$ and $\mathbf{z}_c(i)$ are generated by two independent phenomena, we can assume that $\mathbf{z}_s(i)$ and $\mathbf{z}_c(i)$ are independent. Therefore, the total PDL is the summation of the description length of each of the vectors when encoded separately. We assume that, at most M targets can exist where $M \leq \min_{i=1,\ldots,k} n_i$. The PDL cost is calculated for all $0 \leq m \leq M$ and the smallest is selected as the description length; the corresponding m will indicate the estimated number of targets.

Assuming a model with m targets, the PDL terms for $\mathbf{Z}_s^k$ and $\mathbf{Z}_c^k$ are denoted, respectively, by $\mathrm{PDL}(\mathbf{Z}_s^k; \Psi_k^m)$ and $\mathrm{PDL}(\mathbf{Z}_c^k; \Psi_k^m)$, where Ψ_k^m indicates the parameter set for a model of order m. We will shortly discuss what the elements of Ψ_k^m can be. For the moment, we assume that all unknown parameters of model m are reflected in the vector Ψ_k^m. The independence of $\mathbf{Z}_s^k$ and $\mathbf{Z}_c^k$ results in

$$\mathrm{PDL}\left(\mathbf{Z}^k; \Psi_k^m\right) = \mathrm{PDL}\left(\mathbf{Z}_s^k; \Psi_k^m\right) + \mathrm{PDL}\left(\mathbf{Z}_c^k; \Psi_k^m\right). \tag{16}$$

Each term in this summation is computed separately and the results are added to get the total description length.

First, we calculate the description length for the spurious estimates. Using the definition of the PDL algorithm in (13), we get

$$\mathrm{PDL}\left(\mathbf{Z}_c^k; \Psi_k^m\right) = -\sum_{i=1}^{k} \log p\left(\mathbf{z}_c(i) \,\middle|\, \hat{\Psi}_{i-1}^m, \mathbf{Z}^{i-1}\right). \tag{17}$$

Generally, the spurious estimates $\mathbf{z}_c(i)$ are independent of $\hat{\Psi}_{i-1}^m$ and $\mathbf{Z}^{i-1}$. We model the spurious estimates by a random variable with uniform probability density function W_o^{-1}, where W_o is the angular extent of the region at which the sources can exist. Therefore, the PDL term for the spurious angles will be

$$\mathrm{PDL}\left(\mathbf{Z}_c^k; \Psi_k^m\right) = (N_k - mk) \log W_o \tag{18}$$

where $N_k \triangleq \sum_{i=1}^k n_i$.

The PDL term for the DOA estimates is

$$\begin{aligned} \mathrm{PDL}(\mathbf{Z}_s^k; \Psi_k^m) &= -\sum_{i=1}^{k} \log p\left(\mathbf{z}_s(i) \,\middle|\, \hat{\Psi}_{i-1}^m, \mathbf{Z}^{i-1}\right) \\ &= -\sum_{i=1}^{k} \log p\left(\mathbf{A}_i^{(m)} \mathbf{z}(i) \,\middle|\, \hat{\Psi}_{i-1}^m, \mathbf{Z}^{i-1}\right). \end{aligned} \tag{19}$$

Let $\mathbf{Z}_s^k$ be partitioned into m tracks $\{\mathbf{z}_s^{k,u}\}_{u=1}^m$. If we assume that the track of each source is independent of tracks of other sources, then

$$\mathrm{PDL}\left(\mathbf{Z}_s^k; \Psi_k^m\right) = \sum_{u=1}^{m} \mathrm{PDL}\left(\mathbf{z}_s^{k,u}; \Psi_k^m\right) \tag{20}$$

where $\mathrm{PDL}(\mathbf{z}_s^{k,u}; \Psi_k^m)$, the predictive description length of the uth track, is given by

$$\mathrm{PDL}\left(\mathbf{z}_s^{k,u}; \Psi_k^m\right) = -\sum_{i=1}^{k} \log p\left(\mathbf{a}_i^u \mathbf{z}(i) \,\middle|\, \hat{\Psi}_{i-1}^m, \mathbf{Z}^{i-1}\right). \tag{21}$$

We use the Kalman filter for state estimation in noisy measurements. In particular

$$\hat{\mathbf{x}}_u(i \,|\, i-1) = \mathbf{F} \hat{\mathbf{x}}_u(i-1 \,|\, i-1) \tag{22}$$

$$\hat{z}_{s_u}(i \,|\, i-1) = \mathbf{h}^T \hat{\mathbf{x}}_u(i \,|\, i-1) \tag{23}$$

$$\hat{\mathbf{x}}_u(i \,|\, i) = \hat{\mathbf{x}}_u(i \,|\, i-1) + \hat{\mathbf{K}}_u(i) \hat{\Lambda}_u(i) \tag{24}$$

and

$$\hat{\Lambda}_u(i) = z_{s_u}(i) - \hat{z}_{s_u}(i\,|\,i-1) \\ = \mathbf{a}_i^u \mathbf{z}(i) - \hat{z}_{s_u}(i\,|\,i-1) \tag{25}$$

where, for time step i and target u,, $\hat{\mathbf{x}}_u(i\,|\,i-1)$ is the predicted state vector, $\hat{\mathbf{x}}_u(i\,|\,i)$ is the updated state estimate, $\hat{\mathbf{K}}_u(i)$ is the Kalman filter gain, and $\hat{\Lambda}_u(i)$ is the filter innovation (see [7] and [9] for further details).

The innovation covariance, $\hat{C}_u(i)$, is

$$\hat{C}_u(i) = \mathrm{E}\left\{\hat{\Lambda}_u^2(i)\,|\,\mathbf{z}_s^{i-1,u}\right\} \\ = \mathbf{h}^T \hat{\mathbf{P}}_u(i\,|\,i-1)\mathbf{h} + \sigma_{w_u}^2 \tag{26}$$

where $\mathrm{E}\{\cdot\}$ is the expectation operator, and $\hat{\mathbf{P}}_u(\cdot\,|\,\cdot)$ is the covariance matrix of the state estimate updated by

$$\hat{\mathbf{P}}_u(i\,|\,i-1) = \mathbf{F}\hat{\mathbf{P}}_u(i-1\,|\,i-1)\mathbf{F}^T + \mathbf{Q}_u \tag{27}$$
$$\hat{\mathbf{K}}_u(i) = \hat{C}_u^{-1}(i)\hat{\mathbf{P}}_u(i\,|\,i-1)\mathbf{h} \tag{28}$$
$$\hat{\mathbf{P}}_u(i\,|\,i) = \hat{\mathbf{P}}_u(i\,|\,i-1) - \hat{C}_u(i)\hat{\mathbf{K}}_u(i)\hat{\mathbf{K}}_u^T(i). \tag{29}$$

Using a Gaussian model for the estimation error, we have

$$p\left(z_{s_u}(i)\,\middle|\,\hat{\Psi}_{i-1}^m, \mathbf{Z}_s^{i-1}\right) \sim \mathcal{N}\left(\hat{z}_{s_u}(i\,|\,i-1), \hat{C}_u(i)\right) \tag{30}$$

where $\mathcal{N}(\eta, \sigma^2)$ represents the Gaussian distribution with mean η and variance σ^2. With this modeling, the predictive description length of source DOAs is

$$\mathrm{PDL}\left(\mathbf{Z}_s^k; \Psi_k^m\right) = \frac{1}{2}\sum_{u=1}^{m}\sum_{i=1}^{k}\left(\log(2\pi\hat{C}_u(i)) + \frac{\hat{\Lambda}_u^2(i)}{\hat{C}_u(i)}\right) \tag{31}$$

where $\hat{\Lambda}_u(i)$ and $\hat{C}_u(i)$ are given in (25) and (26), respectively.

The estimated number of targets is found by minimizing the PDL cost over all models, that is

$$\hat{q}_k = \arg\min_m \left\{\frac{1}{2}\sum_{u=1}^{m}\sum_{i=1}^{k}\left(\log(2\pi\hat{C}_u(i)) + \frac{\hat{\Lambda}_u^2(i)}{\hat{C}_u(i)}\right) + (N_k - mk)\log W_o\right\}. \tag{32}$$

The PDL criterion in (32) depends on the parameter set Ψ_k^m, defined as

$$\Psi_k^m = \left\{m, W_o, \{\sigma_{w_u}^2\}_{u=1}^m, \{\sigma_{v_u}^2\}_{u=1}^m \left\{\mathbf{A}_i^{(m)}\right\}_{i=1}^k, \{\mathbf{x}_u(0\,|\,0)\}_{u=1}^m\right\} \tag{33}$$

where $\{\mathbf{x}_u(0\,|\,0)\}_{u=1}^m$ denotes the set of initial state vectors. For any fixed m, a parameter estimator should be used to get all other elements of Ψ_k^m. In this paper, the 1-point, 2-point or 3-point initialization methods [9] can be used to estimate $\{\mathbf{x}_u(0\,|\,0)\}_{u=1}^m$. We use a short window size of K_o snapshots to estimate $\{\sigma_{v_u}^2\}_{u=1}^m$ and $\{\sigma_{w_u}^2\}_{u=1}^m$. We have observed that the detector is not very sensitive to the estimated values of these parameters. Therefore, for simplicity, we find the PDL cost for all $\{\sigma_{v_u}^2\}_{u=1}^m$ and $\{\sigma_{w_u}^2\}_{u=1}^m$ selected on a grid of discrete values in a window of size K_o and choose the ones corresponding to the minimum PDL. These estimated values are then used in the subsequent snapshots. Another option is to use an interacting multiple model (IMM) filter instead of a single fixed Kalman filter (see [7]).

IV. Dynamic Programming

In this section, we present a DP formulation of the PDL algorithm. We first assume that the number of targets is constant over a window of k snapshots. In Section V, we will extend the results to time varying number of sources.

We apply the dynamic programming method based on the extended trellis approach [19] for multitarget association. Assume that the preprocessor has produced the set of prospective DOAs over a window of k snapshots. At each time i, with $1 < i \leq k$, assuming m targets over the window of k snapshots, we make $J_i^{(m)} = ((n_i!)/((n_i - m)!))$ new ordered extended states, and then apply the DP algorithm to the extended trellis. We denote the jth possible extended state at time step i by $\mathbf{y}_j^{(m)}(i)$, which is defined as

$$\mathbf{y}_j^{(m)}(i) = \mathbf{A}_{j,i}^{(m)}\mathbf{z}(i) \tag{34}$$

where $\mathbf{A}_{j,i}^{(m)}$ is an $m \times n_i$ association matrix and $j = 1, \ldots, J_i^{(m)}$. Note that $\mathbf{y}_j^{(m)}(i)$ is equivalent to $\mathbf{z}_s(i)$ introduced in the previous sections for a model with m targets. The PDL-DP algorithm for all $i = 1, \ldots, k$, can be written as

$$\mathcal{L}\left(\mathbf{y}_j^{(m)}(i)\right) = \min_{1 \leq l \leq J_{i-1}^{(m)}} \left\{\mathcal{L}\left(\mathbf{y}_l^{(m)}(i-1)\right) + f\left(\mathbf{y}_j^{(m)}(i), \mathbf{y}_l^{(m)}(i-1)\right)\right\} + (n_i - m)\log W_d \tag{35}$$

where $\mathcal{L}(\mathbf{y}_j^{(m)}(i))$ denotes the PDL cost for the state variable $\mathbf{y}_j^{(m)}(i)$, and

$$f\left(\mathbf{y}_j^{(m)}(i), \mathbf{y}_l^{(m)}(i-1)\right) = \frac{1}{2}\sum_{u=1}^{m}\left(\log(2\pi\hat{C}_u(i)) + \frac{\hat{\Lambda}_{u,j,l}^2(i)}{\hat{C}_u(i)}\right) \tag{36}$$

is the cost function (description length) of transition from state $\mathbf{y}_l^{(m)}(i-1)$ to state $\mathbf{y}_j^{(m)}(i)$. The PDL-DP algorithm (35) is performed for all $i = 1, \ldots, k$ and the PDL cost at time instant k for the mth model is obtained from

$$\mathrm{PDL}(\mathbf{Z}^k; \Psi_k^m) = \min_{1 \leq j \leq J_k^{(m)}} \mathcal{L}\left(\mathbf{y}_j^{(m)}(k)\right). \tag{37}$$

At time instant k, the best model is selected from

$$\hat{q}_k = \arg\min_m \mathrm{PDL}\left(\mathbf{Z}^k; \hat{\Psi}_k^m\right). \tag{38}$$

Assuming $\mathbf{y}_j^{(m)}(i) = [y_{1,j}(i), \ldots, y_{m,j}(i)]^T$, and for each $u = 1, \ldots, m$, the Kalman filter equations are

$$\hat{\mathbf{x}}_{u,l}(i\,|\,i-1) = \mathbf{F}\hat{\mathbf{x}}_{u,l}(i-1\,|\,i-1) \tag{39}$$
$$\hat{\Lambda}_{u,j,l}(i) = y_{u,j}(i) - \mathbf{h}^T\hat{\mathbf{x}}_{u,l}(i\,|\,i-1) \tag{40}$$
$$\hat{\mathbf{x}}_{u,j}(i\,|\,i) = \hat{\mathbf{x}}_{u,l}(i\,|\,i-1) + \hat{\mathbf{K}}_u(i)\hat{\Lambda}_{u,j,l}(i) \tag{41}$$

where $\hat{\mathbf{K}}_u(i)$ is found from (28), and $\hat{\mathbf{x}}_{u,j}(i\,|\,i)$ is the updated state estimate of target u (among m ordered ones) at time instant i and the association index j.

For each state $\mathbf{y}_j^{(m)}(i)$, and for each model of order m, the following parameters should be stored in memory: $\{\hat{\mathbf{x}}_{u,j}(i\,|\,i)\}_{u=1}^m$, $\{\hat{\mathbf{P}}_u(i\,|\,i)\}_{u=1}^m$, $\mathcal{L}(\mathbf{y}_j^{(m)}(i))$, and $\xi(m,j,i)$ where

$$\xi(m,j,i) \triangleq \arg \min_{1\le l\le J_{i-1}^{(m)}} \left\{ \mathcal{L}\left(\mathbf{y}_l^{(m)}(i-1)\right) + f\left(\mathbf{y}_j^{(m)}(i), \mathbf{y}_l^{(m)}(i-1)\right) \right\} \quad (42)$$

represents the index of the previous node that corresponds to the smallest cost. Note that $\hat{C}_u(i)$ and $\hat{\mathbf{K}}_u(i)$ can be easily calculated from $\hat{\mathbf{P}}_u(i-1\,|\,i-1)$ using (26)–(28).

Based on the data up to and including time k, we can find the DOAs that correspond to the best model. To achieve this goal, we define $d_m(i)$ as the index of the state in trellis m that gives the smallest PDL at time i. Therefore, the index of the state for the minimum cost trellis at time k is given by

$$d_{\hat{q}_k}(k) = \arg \min_{1\le j\le J_k^{(\hat{q}_k)}} \left\{ \mathcal{L}\left(\mathbf{y}_j^{(\hat{q}_k)}(k)\right) \right\} \quad (43)$$

where $\hat{q}_k$ is estimated from (38). The values of $d_{\hat{q}_k}(i)$ for $i = 1,\ldots,k-1$ are found by back-tracing

$$d_{\hat{q}_k}(i) = \xi\left(\hat{q}_k, d_{\hat{q}_k}(i+1), i+1\right), \quad i = k-1,\ldots,1. \quad (44)$$

We will drop the subscript $\hat{q}_k$ in $d_{\hat{q}_k}(i)$ in next section, for convenience. Note also that some simple track maintenance rules can be used to avoid divergence from acceptable kinematic data model (and avoid fitting an unacceptable one); for example, simply limiting the maximum velocity and acceleration variation at each time step can create a smooth track for each target.

V. Tracking Time-Varying Number of Targets

In this section, we use an extended trellis method to detect time varying number of sources. In the previous section, we formed M trellises and used DP to compute the PDL cost on each trellis. Since we assumed that the number of targets was constant, there was no link connecting two states in two different trellises. In this section, we allow the number of targets to change in the observation window. Using the same approach as the previous section, we create M base trellises formed by collecting the DOAs obtained by the preprocessor. Each base trellis is identified by its corresponding parameter m that represents the number of targets (the dimensionality of each state vector in that trellis). By extending edges to connect states in two different base trellises, we indeed permit a variable number of the signals over the observation window.

Using the same state definition as (34), the PDL cost for state $\mathbf{y}_j^{(m)}(i)$ at time instant i, is given by

$$\mathcal{L}\left(\mathbf{y}_j^{(m)}(i)\right) = \min_{1\le n\le M} \min_{1\le l\le J_{i-1}^{(n)}} \left\{ \mathcal{L}\left(\mathbf{y}_l^{(n)}(i-1)\right) + f\left(\mathbf{y}_j^{(m)}(i), \mathbf{y}_l^{(n)}(i-1)\right) \right\} + (n_i - m)\log W_d \quad (45)$$

where $f(\mathbf{y}_j^{(m)}(i), \mathbf{y}_l^{(n)}(i-1))$ is the PDL cost for transition between states $\mathbf{y}_l^{(n)}(i-1)$ and $\mathbf{y}_j^{(m)}(i)$. We will show shortly how $f(\mathbf{y}_j^{(m)}(i), \mathbf{y}_l^{(n)}(i-1))$ can be calculated. The total PDL cost can be obtained as

$$\mathrm{PDL}(\mathbf{Z}^i;\Psi_i) = \min_m \left\{ \min_j \mathcal{L}\left(\mathbf{y}_j^{(m)}(i)\right) \right\}. \quad (46)$$

We define the following parameters:

$$\eta(m,j,i) \triangleq \arg \min_{1\le n\le M} \left\{ \min_{1\le l\le J_{i-1}^{(n)}} \left\{ \mathcal{L}\left(\mathbf{y}_l^{(n)}(i-1)\right) + f\left(\mathbf{y}_j^{(m)}(i), \mathbf{y}_l^{(n)}(i-1)\right) \right\} \right\} \quad (47)$$

$$\xi(m,j,i) \triangleq \arg \min_{1\le l\le J_{i-1}^{(\eta(m,j,i))}} \left\{ \mathcal{L}\left(\mathbf{y}_l^{(\eta(m,j,i))}(i-1)\right) + f\left(\mathbf{y}_j^{(m)}(i), \mathbf{y}_l^{(\eta(m,j,i))}(i-1)\right) \right\} \quad (48)$$

where $\eta(m,j,i)$ represents the best previous model order in the trellis, and $\xi(m,j,i)$ denotes the index of the best previous extended state in model $\eta(m,j,i)$. For each state $\mathbf{y}_j^{(m)}(i)$, we need to store the following parameters: $\{\hat{\mathbf{x}}_{u,j}^{(m)}(i\,|\,i)\}_{u=1}^m$, $\mathcal{L}(\mathbf{y}_j^{(m)}(i))$, $\xi(m,j,i)$, and $\eta(m,j,i)$. Assume that the PDL cost is calculated over the window $i = 1,\ldots,k$. The best model at time instant k is selected from

$$\hat{q}_k = \arg\min_m \left\{ \min_j \mathcal{L}\left(\mathbf{y}_j^{(m)}(k)\right) \right\} \quad (49)$$

$$\hat{d}(k) = \arg\min_j \mathcal{L}\left(\mathbf{y}_j^{(\hat{q}_k)}(k)\right). \quad (50)$$

The number of targets at time instants $i = 1,\ldots,k-1$ is obtained by back tracing

$$\hat{q}_i = \eta(\hat{q}_{i+1}, \hat{d}(i+1), i+1) \quad (51)$$

$$\hat{d}(i) = \xi(\hat{q}_{i+1}, \hat{d}(i+1), i+1). \quad (52)$$

In this section, we assume that the dynamic characteristics of tracks (i.e., σ_w^2 and σ_v^2) are identical. Therefore, the Kalman filter parameters only depend on time duration or time index. This assumption is only for the simplification of the notations. The proposed approach is, however, general and can be applied to tracks with unequal parameters.

To compute $f(\mathbf{y}_j^{(m)}(i), \mathbf{y}_l^{(n)}(i-1))$, first note that if $m = n$, the two states belong to the same base trellis and the PDL cost can be found from (36). We differentiate two other cases.

- **Case 1:** $m < n$ (deleting tracks): In this case, we will need to terminate $(n-m)$ tracks. In other words, we have to choose m elements of $\mathbf{y}_l^{(n)}(i-1)$ as the candidate DOAs, and apply the Kalman filtering on those elements. Let us represent the transition cost for this case by $f(\mathbf{y}_j^{(m)}(i), \mathbf{y}_l^{(n)}(i-1); m < n)$. Since, we have assumed that the states comprise all ordered estimated DOAs, it suffices to assume that

$$f\left(\mathbf{y}_j^{(m)}(i), \mathbf{y}_l^{(n)}(i-1); m < n\right) = f\left(\mathbf{y}_j^{(m)}(i), \tilde{\mathbf{y}}_l^{(m)}(i-1)\right) \quad (53)$$

where $\tilde{\mathbf{y}}_l^{(m)}(i-1)$ is the vector of the first m elements of $\mathbf{y}_l^{(n)}(i-1)$. Note that since all permutations of the DOA candidates are reflected in the extended states in the trellis, choosing the first m elements of $\mathbf{y}_l^{(n)}(i-1)$ suffices to guarantee that all mappings between $\mathbf{y}_l^{(n)}(i-1)$ and $\mathbf{y}_j^{(m)}(i)$ are considered.

- **Case 2:** $m > n$ (introducing new tracks): We use the spurious DOAs in step $(i-1)$ to initiate $(m-n)$ new tracks. Here, we choose a very simple method to initialize the tracks; more sophisticated track initialization schemes may also be suggested.

  Let us represent $\mathbf{y}_j^{(m)}(i) = [\tilde{\mathbf{y}}_j^{(n)}(i), \tilde{\mathbf{y}}_j^{(m-n)}(i)]^T$, where $\tilde{\mathbf{y}}_j^{(n)}(i)$ is the first n elements of $\mathbf{y}_j^{(m)}(i)$, and the rest of the elements are indicated by $\tilde{\mathbf{y}}_j^{(m-n)}(i)$. The transition cost for new tracks is estimated as follows:

$$\begin{aligned} f&\left(\mathbf{y}_j^{(m)}(i), \mathbf{y}_l^{(n)}(i-1); m > n\right) \\ &= f\left(\tilde{\mathbf{y}}_j^{(n)}(i), \mathbf{y}_l^{(n)}(i-1)\right) \\ &\quad + f\left(\tilde{\mathbf{y}}_j^{(m-n)}(i), \mathbf{z}_{c,l}^{(n_{i-1}-n)}(i-1)\right) \end{aligned} \tag{54}$$

  where $\mathbf{z}_{c,l}^{(n_{i-1}-n)}(i-1)$ represents the vector of spurious observations corresponding to $\mathbf{y}_l^{(n)}(i-1)$, and $f(\tilde{\mathbf{y}}_j^{(m-n)}(i), \mathbf{z}_{c,l}^{(n_{i-1}-n)}(i-1))$ is the PDL cost of the $(m-n)$ new tracks and is given by

$$\begin{aligned} f&\left(\tilde{\mathbf{y}}_j^{(m-n)}(i), \mathbf{z}_{c,l}^{(n_{i-1}-n)}(i-1)\right) \\ &= \min_e \left\{ \frac{1}{2} \sum_{u=1}^{m-n} \left(\log(2\pi \hat{C}_u(1)) + \frac{\hat{\Lambda}_{u,j,e}^2(i)}{\hat{C}_u(1)} \right) \right\} \end{aligned} \tag{55}$$

  where $\hat{\Lambda}_{u,j,e}^2(i)$ is found from

$$\hat{\Lambda}_{u,j,e}(i) = \tilde{y}_{u,j}(i) - \mathbf{h}^T \mathbf{F} [z_{c,e,u}(i-1), 0, 0]^T \tag{56}$$

  where $\tilde{y}_{u,j}(i)$ is the uth element of the state vector $\tilde{\mathbf{y}}_j^{(m-n)}(i)$, and $z_{c,e,u}(i-1)$ is the uth spurious DOA at time step $(i-1)$ corresponding to the association index e. Note that in (55), we have assumed that the newly generated tracks only depend on $\mathbf{y}_l^{(n)}(i-1)$. Therefore, we use $\hat{C}_u(1)$ that corresponds to a track of size 1.

VI. Computational Cost

In PDL-DP, the computational cost grows linearly with time. However, since the algorithm is based on dynamic programming, parallel processors can be efficiently used to implement the proposed technique. Note that the complexity of the PDL-DP technique is in the same order of the Viterbi algorithm, which has been successfully implemented in chipsets presently being used in various wireless terminals and cell phones.

Here, we find an approximate value for the computational cost of the PDL-DP algorithm. Computing the association matrix in (45) for each value of (m, j) needs about $4m$ operations,[1] which results in a total of $\sum_{m=1}^{M} 4mJ_i^{(m)}$ operations. Then, for each (n, l), we have to construct the association matrix (with $4n$ operations) and compute the transition function f (with about $6n$ operations). This results in a total of $\sum_{m=1}^{M}\{J_i^{(m)} \sum_{n=1}^{M}(4+6)nJ_{i-1}^{(n)}\}$ operations. Then, for each n, we have to find the minimum over $J_{i-1}^{(n)}$ values. Calculating the predicted state parameters using the best path needs about $12n$ operations. We must also find the minimum over the resulting M PDL values. We assume that finding the minimum over M values uses about M operations.[2] So, this part requires about $\sum_{m=1}^{M} J_i^{(m)}\{\sum_{n=1}^{M} J_{i-1}^{(n)} + \sum_{n=1}^{M} 12n + M\}$ operations. Therefore, the total computational cost from time step 1 up to k will be approximately

$$\begin{aligned} \text{COST} &= \sum_{i=2}^{k} \sum_{m=1}^{M} J_i^{(m)} \left\{ 4m + 10 \sum_{n=1}^{M} nJ_{i-1}^{(n)} + M \right. \\ &\qquad \left. + \sum_{n=1}^{M} \left(12n + J_{i-1}^{(n)}\right) \right\} \\ &= \sum_{i=2}^{k} \sum_{m=1}^{M} J_i^{(m)} \left\{ 4m + M + 6M(M+1) \right. \\ &\qquad \left. + \sum_{n=1}^{M} (1+10n) J_{i-1}^{(n)} \right\} \\ &\approx 10 \sum_{i=2}^{k} \sum_{m=1}^{M} \sum_{n=1}^{M} nJ_i^{(m)} J_{i-1}^{(n)}. \end{aligned} \tag{57}$$

VII. Afterthoughts

In this section, we present some afterthoughts that should be considered when the PDL-DP algorithm is employed in practice.

First, in the proposed technique, for each (m, j, i) in (45), only one best history with the lowest PDL is saved. This is a limitation of the dynamic programming method. One method to deal with this problem is to sort the PDL cost in (45), and store the L lowest description lengths, instead of just the minimum one. So, at any time k, when making a decision, we will have L best tracks that may be used by the postprocessing (and data fusion) subsystem to make a better final decision.

Second, such as any other target tracking system, the performance of the PDL-DP method depends on the number of targets, the number of spurious observations, the closeness of the targets, and the estimates of the tracking parameters. For instance, when the number of targets increases, data association might be more ambiguous. This is because the preprocessor, which provides the observation data to the tracker, has a lower performance for larger number of targets. This can result in model fitting error, especially in the case of high clutter and low SNR. Therefore, like any target tracking algorithm, the PDL-DP tracker is expected to have a larger detection threshold as the number of targets increases. Using multiple sensors/trackers and employing data fusion may help by providing a higher dimensional observation space.

Third, because of limited resolution of the preprocessor for crossing or *closely-spaced tracks*, the observation will be biased

[1]By one operation, we mean one flop (floating point operation).

[2]We assume that each comparison uses one flop.

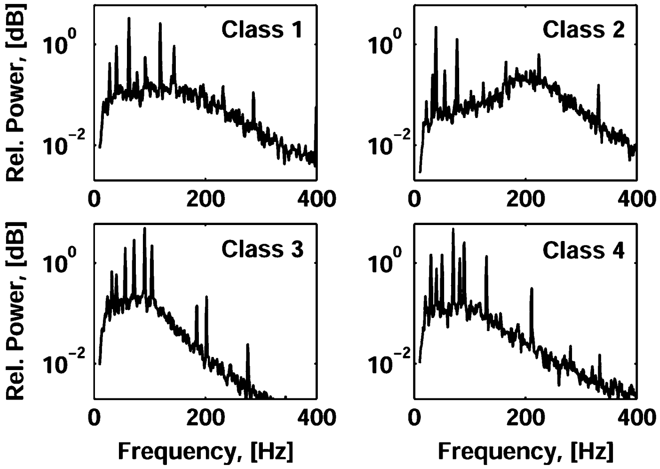

Fig. 1. Power spectral density of the four simulated wideband signals.

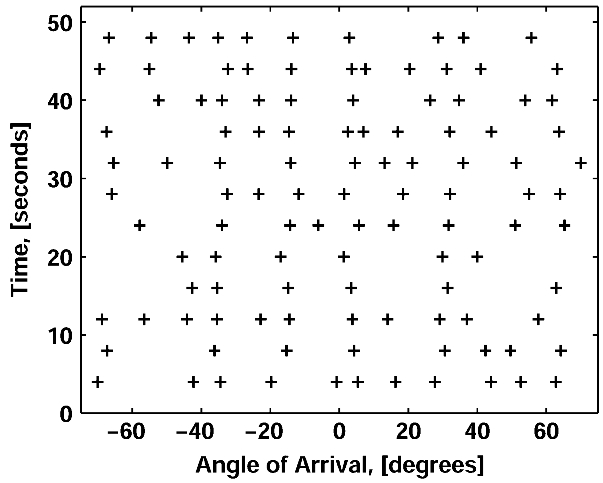

Fig. 2. Estimated DOAs using the method of [21].

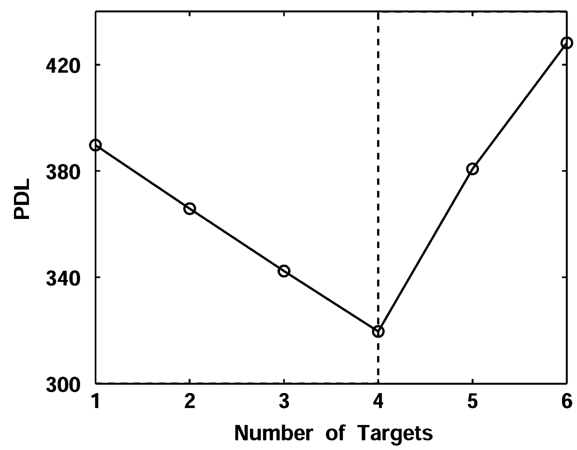

Fig. 3. Predictive description length versus the selected model.

and the innovation sequence will be correlated among those targets. Furthermore, in the crossing point, the number of detected DOAs is smaller than the number of targets. To overcome this limitation, one can modify the tracking system to recognize the approaching tracks and when the targets are in a certain distance to use the history of the tracks for target association. In this case, a known data model for crossing and closely-spaced sources can be included in the PDL framework. Due to space limitation, this approach is not discussed in the present paper.

VIII. Simulation Results

We have simulated a passive system with a uniform linear array of 18 omni-directional sensors to localize underwater vehicles. Practical measurements of acoustic signals emitted from large underwater vehicles show that they consist of a continuous spectrum and of a number of narrowband components. Here, a Gaussian auto-regressive (AR) model is used for continuous part of the spectrum, and a number of narrowband signals with random relative power, location and bandwidth are added. Fig. 1 shows the power spectrum of 4 wideband signals that are used in the simulations. We only process the frequency band of [100, 200] Hz; the sampling frequency is 2560 Hz. The distance between the sensors of the array is 3.75 m—half the wavelength of the maximum frequency of the processing bandwidth. A spatially and temporally white noise model is used. In each time step, a block of 4 s of data received by sensors of the array is processed. We assume $T_o = 4$ s, and $W_o = 150^\circ$.

In the first scenario, the DOA estimates of the wideband sources is obtained by [20] using the parameter estimation algorithm of [21]. We have assumed $\mathrm{SNR} = -10$ dB where the SNR is defined as the ratio of signal power to noise power at the processing bandwidth. Since the level of clutter is high and the SNR is low, the MDL and AIC enumeration methods [2] fail and cannot detect the proper number of signals. Therefore, we use [21] which only requires a detection threshold. At the final estimation stage of [21], no threshold is set and all peak points in the array spatial spectrum are taken as raw observations fed to the tracking system.

Fig. 2 shows the raw DOA estimates. In this figure, the raw estimated DOAs are shown by "+." We look for unknown but constant number of targets in a window of 12 time steps. Fig. 3 illustrates the PDL cost as a function of assumed number of targets. Note that the minimum is achieved for four targets. Therefore, PDL can enumerate the true number of signals.

Fig. 4 shows the estimated tracks for four sources in the PDL-DP algorithm. The dashed lines in the figure indicate true tracks (slowly moving) and the solid lines indicate the estimated ones. Note that the estimated DOAs do not necessarily coincide with the values suggested by the preprocessor. It is seen that the number of targets and their trajectories are estimated reasonably well.

As the second scenario, we study a case with a time-varying number of sources. Here, we have used the incoherent signal processing with spatial smoothing for wideband DOA estimation [22]. The number of subarrays is 3. At the final estimation stage of [21] no threshold is set and all peak points in the array spatial spectrum are taken as raw observations fed to the tracking system. The processing frequency bandwidth is divided into 30 segments with uniform spacing. Figs. 5 and 6 show the PDL-DP tracking results using the extended trellis method. Because of the low resolution of the DOA estimator, when the two center tracks intersect, one of them dominates and the DOA estimator fails to recognize the two tracks. Such phenomenon seems to be a common problem in passive wideband localization techniques. However, the proposed approach for tracking a varying

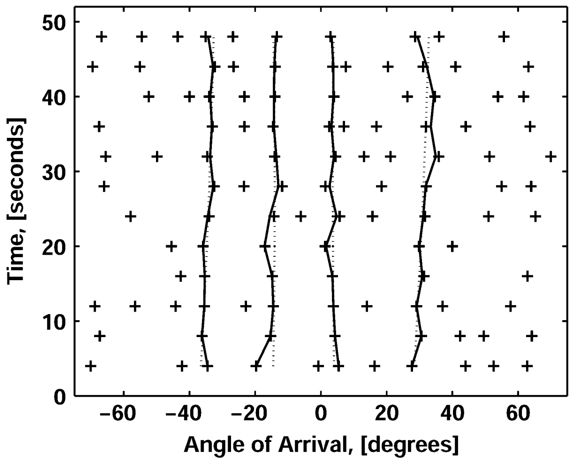

Fig. 4. Associated tracks for the four sources.

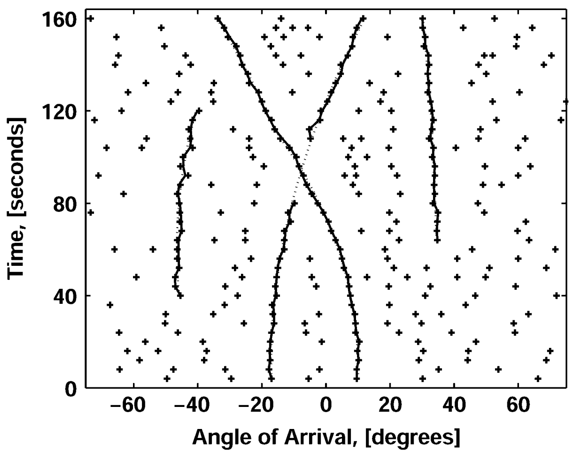

Fig. 5. Resolved tracks for four moving sources.

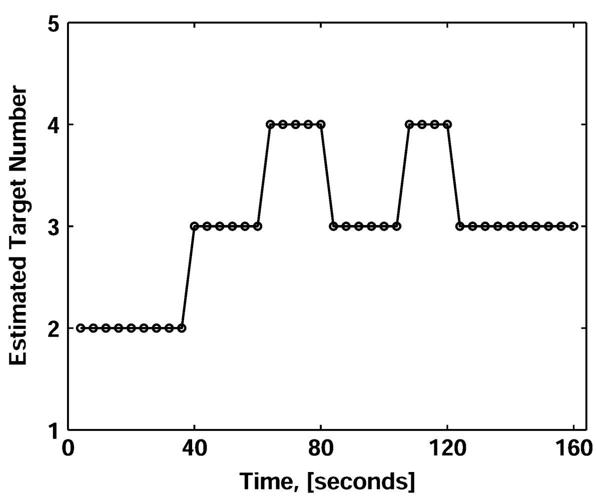

Fig. 6. Estimated number of targets as a function of time using the extended trellis method.

number of targets allows us to detect the two center tracks when the tracks separate. In such cases, a postprocessing stage in the tracking system might be helpful. The post processor should take into account the history of the tracks to resolve crossing targets.

In the third example, we use a uniform linear array with eight sensors, two moving narrowband signal sources, and iterative sample covariance matrix estimation with the smoothing factor $\alpha = 0.95$ [5]. The number of snapshots in each block of $T_o = 1$ s of array data is 20. We study the performance for four different SNR values $\{-12.5, -10, -7.5, -5\}$ dB. We assume that $\sigma_w^2 = \{2, 1.75, 1.5, 1\}$ respectively for each SNR value. The Monte-Carlo simulation includes 100 tracking runs, where each data window (fed to the tracker) includes nine time samples, (i.e., a total of 9 s each window). Therefore, for each SNR, the DOAs are estimated 900 times in total. In each data window, the two sources are initially located at $-16°$ and $8°$. Then, they move with constant velocity of $0.5°$ per second, and at the end of each observation window, they reach $-12°$ and $12°$. For DOA estimation, the range of $[-45°, 45°]$ is uniformly searched (so $W_o = 90°$). The performance of the PDL-DP tracker is compared to the performance of the PDL [5], MDL, and AIC criteria (see [2]). For the PDL method of [5], the alternating projection method is used for maximum likelihood DOA estimation [23]. The input to the PDL-DP tracker consists of DOAs reported by [21]. The results of enumeration for $M = 3$ are reported in Table I. Note that PDL-DP outperforms earlier methods.

TABLE I
COMPARISON OF THE PERFORMANCE OF THE PDL-DP, PDL [5], MDL, AND AIC [2] ENUMERATORS FOR FOUR DIFFERENT SNR VALUES. THERE ARE TWO MOVING SOURCES (I.E., A NONSTATIONARY DATA MODEL)

		SNR			
Alg.	m	-12.5dB	-10dB	-7.5dB	-5dB
PDL-DP	1	50	8	2	0
	2	49	91	98	99
	3	1	1	0	1
PDL ([5])	1	62	31	5	0
	2	32	62	87	93
	3	6	7	8	7
MDL ([2])	1	97	88	50	5
	2	0	10	48	93
	3	3	2	2	2
AIC ([2])	1	67	40	9	1
	2	13	40	71	80
	3	20	20	20	19

In the last example, the PDL-DP is compared to the 2-D assignment tracking algorithm (2-DA) [8] (the auction algorithm). The variance of the observation noise changes in the range $\sigma_w^2 = \{1, 3, 5, 7\}$, and the tracking results are analyzed using Monte-Carlo simulations with 200 runs. The number of clutter points is modeled as a Poisson process with average rate of 2 (indeed it is approximately Poisson because it is forced to have at least one clutter at each snapshot), and the clutter points are approximately distributed with a uniform pdf in the field-of-view, $W_o = 150°$. The clutter points are separated from other observations (and also from each other) by at least $4°$. There are two targets separated by $12°$, moving with constant velocity of $0.5°$ per second. $T_o = 1$ and $P_D = 1$. The number of time samples is 7, and the targets are initially located at -5 and $7°$. One-point initialization is used. The percentage of correct detection is shown in Fig. 7 for each estimator. Note that the PDL-DP has a higher detection percentage than 2-DA, especially for larger values of the observation noise variance.

IX. CONCLUSION

In this paper, the predictive description length was used to estimate the number of sources in a target tracking system. The

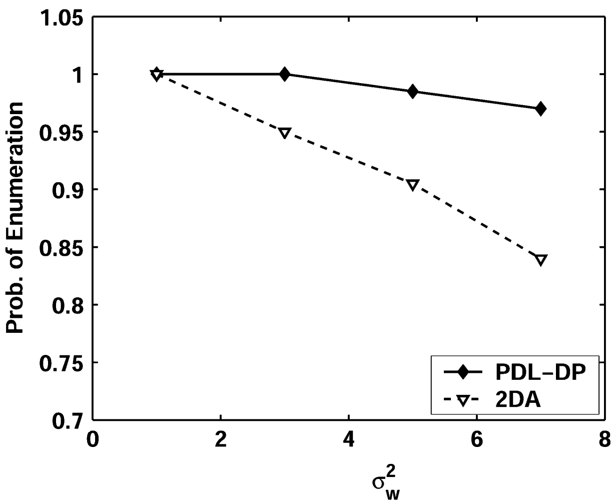

Fig. 7. Probability of correct enumeration versus observation noise variance for two tracking methods: 2-DA [8] and PDL-DP.

PDL algorithm is applied to the set of raw DOA estimates to detect the true number of DOAs and select the appropriate DOAs. The prospective DOAs are decomposed into two sets representing the actual DOAs and the spurious DOAs induced from clutter and noise. The PDL cost is computed for each set separately and the results are added to calculate the total description length. The PDL cost is computed for all tentative models and the smallest is chosen as the best model. We have devised a dynamic programming formulation of the algorithm.

REFERENCES

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

[2] M. Wax and T. Kailath, "Detection of signals by information theoretic criteria," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 33, no. 4, pp. 387–392, Apr. 1985.

[3] A. Barron, J. Rissanen, and B. Yu, "The minimum description length principle in coding and modeling," *IEEE Trans. Inf. Theory*, vol. 44, no. 6, pp. 2743–2760, Oct. 1998.

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

[5] S. Valaee and P. Kabal, "An information theoretic approach to source enumeration in array signal processing," *IEEE Trans. Signal Process.*, vol. 52, no. 5, pp. 1171–1178, May 2004.

[6] P. Grunwald, I. J. Myung, and M. Pitt, Eds., *Advances in Minimum Description Length: Theory and Applications*. Cambridge, MA: MIT Press, 2005.

[7] S. S. Blackman and R. Popoli, *Design and Analysis of Modern Tracking Systems*. Norwood, MA: Artech House, 1999.

[8] Y. Bar-Shalom and W. D. Blair, Eds., *Multitarget-Multisensor Tracking: Applications and Advances*. Norwood, MA: Artech House, 2000, vol. III.

[9] Y. Bar-Shalom, X. R. Li, and T. Kirubarajan, *Estimation with Applications to Tracking and Navigation: Theory, Algorithms and Software*. New York: Wiley, 2001.

[10] C. R. Rao, C. R. Sastry, and B. Zhou, "Tracking the direction of arrival of multiple moving targets," *IEEE Trans. Signal Process.*, vol. 42, no. 5, pp. 1133–1144, May 1994.

[11] Y. Zhou, P. C. Yip, and H. Leung, "Tracking the direction of arrival of multiple targets by passive arrays: Algorithm," *IEEE Trans. Signal Process.*, vol. 47, no. 10, pp. 2655–2666, Oct. 1999.

[12] J. Arnold, S. Shaw, and H. Pasternack, "Efficient target tracking using dynamic programming," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 29, no. 1, pp. 44–56, Jan. 1993.

[13] J. P. Le Cadre and O. Thrémois, "Bearings-only tracking for maneuvering sources," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 34, no. 1, pp. 179–193, Jan. 1998.

[14] L. A. Johnston and V. Krishnamurthy, "Performance analysis of a dynamic programming track before detect algorithm," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 38, no. 1, pp. 228–242, Jan. 2002.

[15] R. J. Kenefic, "Local and remote track file registration using minimum description length," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 29, no. 3, pp. 651–655, Jul. 1993.

[16] R. A. Iltis and K. L. Anderson, "A consistent estimation criterion for multisensor bearings-only tracking," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 32, no. 1, pp. 108–120, Jan. 1996.

[17] H. Chen, Y. Bar-Shalom, K. R. Pattipati, and T. Kirubarajan, "MDL approach for multiple low observable track initiation," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 39, no. 3, pp. 862–882, Jul. 2003.

[18] J. Rissanen, "Fisher information and stochastic complexity," *IEEE Trans. Inf. Theory*, vol. 42, no. 1, pp. 40–47, Jan. 1996.

[19] J. K. Wolf, A. M. Viterbi, and G. S. Dixon, "Finding the best set of K paths through a trellis with applications to multitarget tracking," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 25, no. 2, pp. 287–296, Mar. 1989.

[20] M. Allam and A. Moghaddamjoo, "Two-dimensional DFT projection for wideband direction of arrival estimation," *IEEE Trans. Signal Process.*, vol. 43, no. 7, pp. 1728–1732, Jul. 1995.

[21] J. Choi, I. Song, and H. M. Kim, "On estimating the direction of arrival when the number of signal sources is unknown," *Signal Process.*, vol. 34, no. 2, pp. 193–205, Nov. 1993.

[22] H. Wang and M. Kaveh, "Coherent signal-subspace processing for the detection and estimation of angles of arrival of multiple wide-band sources," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. ASSP-33, no. 4, pp. 823–831, Aug. 1985.

[23] I. Ziskind and M. Wax, "Maximum likelihood localization of multiple sources by alternating projection," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 36, no. 10, pp. 1553–1560, Oct. 1988.

Ahmad Khodayari-Rostamabad (S'04) received the B.Sc. degree from the Iran University of Science and Technology, Tehran, and the M.Sc. degree from Tarbiat Modarres University, Tehran, Iran, both in electrical engineering.

From 1995 to 2004, he was with the J. Sazandeghi Engineering Research Institute (JERI), Tehran, where he was involved in acoustic signal processing. From 2000 to 2003, he was also with Tarbiat Modarres University, Tehran, Iran.

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, Montréal, QC, Canada, all in electrical engineering.

From 1994 to 1995, he was a Research Associate at INRS Telecom, University of Quebec, Montréal. From 1996 to 2001, he was an Assistant Professor in the Department of Electrical Engineering, Tarbiat Modares University, Tehran, and in the Department of Electrical Engineering, Sharif University of Technology, Tehran. During this period, he was also a Consultant to the Iran Telecommunications Research Center. Since September 2001, he has been an Associate Professor in the Edward S. Rogers, Sr. Department of Electrical and Computer Engineering, University of Toronto, Toronto, ON, Canada, where he holds the Nortel Institute Junior Chair of Communication Networks. His current research interests are in wireless ad hoc, sensor, and cellular networks.

Dr. Valaee is the Co-Chair for the Wireless Communications Symposium of IEEE GLOBECOM 2006; the Editor for the *IEEE Wireless Communications Magazine* Special Issue on Toward Seamless Internetworking of Wireless LAN and Cellular Networks; and the Editor of a Special Issue of the Wiley *Journal on Wireless Communications and Mobile Computing* on Radio Link and Transport Protocol Engineering for Future Generation Wireless Mobile Data Networks.