MDL Approach for Multiple Low Observable Track Initiation

Huimin Chen, University of New Orleans
T K Kirubarajan, McMaster University
Yaakov Bar-Shalom, University of Connecticut - Storrs
K R Pattipati, University of Connecticut - Storrs
I. INTRODUCTION

A substantial threat comes from theater ballistic missiles (TBM). This has stimulated extensive research work in the area of ballistic missile defense (BMD). The first step is to detect the incoming missiles, then track their trajectories and eventually direct interceptors or other defense apparatus to counter them. The acquisition of a missile consists of missile track detection and its initial state (position and velocity) estimation at a certain time instant. During its exo-atmospheric flight, the predominant force acting on the missile is Earth’s gravity, which yields a deterministic target trajectory given by the laws of Kepler [5]. The estimation of the exo-atmospheric trajectory is equivalent to the estimation of the target state at some reference time $t_p$, i.e., it amounts to the estimation of a fixed dimensional parameter vector. The measurements are obtained by a ground-based radar searching for the missile in a predetermined “cueing region.” The radar is a phased array monopulse radar and the cueing region is informed by early warning satellites. The radar uses amplitude comparison monopulse processing to obtain the angle measurements. The technical challenges in this estimation problem consist of short time period available to acquire the missile tracks, presence of spurious measurements due to the low detection threshold made necessary by the desire to extend the acquisition range of the targets (which have unknown signal to noise ratio (SNR), but are known to be weak—we will be searching for targets with a certain low SNR), and most of all, unknown number of targets. Other issues such as the error in positioning the cueing region are not considered. We assume that the targets to be detected are in the cueing region throughout all radar scans. The situation where targets’ presence during the entire batch of scans is not guaranteed can be handled as in [9] via a variable length sliding window. The problem can be formally written as multiple composite hypothesis testing where the number of candidate hypotheses is yet to be determined. From a statistical standpoint, it is a model selection problem where each model represents a different number of tracks to explain the observed data.

Previous work [23] proposed the track initiation algorithm of a ballistic missile using line of sight (LOS) measurements. The monopulse radar model developed in [6] has been widely used in the benchmark tracking problems in [12, 22, 19] with SNR of at least 10 dB. Few results are available for the estimation of missile trajectories under lower SNR. We use a monopulse radar to measure the full position of the incoming target and want to extend the missile acquisition range, which requires acquisition capability of low SNR targets. A more recent paper [21] developed a track initiation algorithm for monopulse radar to work under average SNR as
low as 4 dB. The algorithm used in [21] was the maximum likelihood estimator with probabilistic data association (ML-PDA), which does not have to make a decision as to which measurement is target-originated. However, the above works are applicable only for acquisition of a single target. The acquisition of multiple targets in the presence of false alarms and missed detections is the topic of this paper.

The need for model selection arises when we want to determine the number of tracks using the ML-PDA based on the available data. There exist several criteria for model selection, namely, the Akaike information criterion (AIC) [1], the Bayes information criterion (BIC) [18], stochastic complexity (SC) [14], generalized maximum likelihood (ML) rule [11], the information-theoretic measure of complexity (ICOMP) [7] and minimum description length (MDL) [15]. AIC and BIC are the most commonly used methods which penalize a complicated model by some terms depending on the number of unknown parameters being determined from the data. However, they are not invariant under reparameterization. The only approach invariant under reparameterization is the MDL criterion, which compares different model classes by the likelihood function augmented with the model complexity [16]. The idea behind MDL is the notion that knowledge (some regularity property in the data) and data redundancy are interrelated. There must be some kind of redundancy in the data because without it, every point in the data will be unique. In such a case, there is no regularity to be learned or extracted from the data. In other words, the more we compress the data by extracting redundancy from it, the more we learn about the regularity underlying the data. The model with minimum redundancy would be the best choice. With the MDL criterion, we can gain some insights on the performance limit of the ML-PDA for initiating multiple tracks.

The problem of testing multiple composite hypotheses is formulated in Section II. Several one-dimensional examples to illustrate the performance limit of the ML-PDA estimator using the MDL criterion are presented in Section III. The application of the MDL criterion to the missile track acquisition is presented in Section IV with the simulation results based on scenarios with up to two targets under various SNRs. Section V summarizes the results.

II. MULTIPLE COMPOSITE HYPOTHESIS TESTING AND MDL CRITERION

The multiple composite hypothesis testing can be formulated as follows. Assume we have \( K \) different models for hypothesis testing, namely, \( H_1, H_2, \ldots, H_K \). These models exist throughout the data sequence \( Z = \{z_1, \ldots, z_N\} \) of length \( N \). Under hypothesis \( H_i \), we have the conditional probability density function (pdf) \( p(\mathbf{Z} | \mathbf{x}_i, H_i) \) with unknown parameter vector \( \mathbf{x}_i \) of dimension \( n_i \). Hence we may equate a model with a parametric pdf \( p(\mathbf{Z} | \mathbf{x}_i, H_i) \). Here the conditioning on \( H_i \) cannot be omitted since different models may have different functional forms of the conditional pdf. Assuming the priors \( P(H_i) \) (\( i = 1, \ldots, K \)) are equal, we want to choose model \( j \) for which \( p(\mathbf{Z} | H_j) \) is maximum. Since the parameter \( \mathbf{x}_i \) is unknown, we assign a prior pdf \( p(\mathbf{x}_i) \) for each hypothesis \( H_i \). The likelihood function of \( H_i \) is

\[
p(\mathbf{Z} | H_i) = \int p(\mathbf{Z} | \mathbf{x}_i, H_i) p(\mathbf{x}_i) d\mathbf{x}_i.
\]

Due to the difficulty in performing the integration and the lack of prior knowledge, this approach is not very practical. An alternative is to use the ML estimate \( \hat{\mathbf{x}}_i \) for the unknown \( \mathbf{x}_i \) and choose \( H_j \) for which \( p(\mathbf{Z} | \hat{\mathbf{x}}_j, H_j) \) is maximum, i.e.,

\[
j = \arg \max_j p(\mathbf{Z} | \hat{\mathbf{x}}_j, H_j)
\]

where

\[
\hat{\mathbf{x}}_i = \arg \max_{\mathbf{x}_i} p(\mathbf{Z} | \mathbf{x}_i, H_i), \quad i = 1, \ldots, K.
\]

This is the generalized likelihood ratio test (GLRT) for multiple composite hypothesis testing. However, the above decision rule tends to choose a more sophisticated model as the number of data points increases because the modeling error decreases as we add more unknown parameters to the model. To compare models with different parameters (as well as different functional forms), we have to penalize a sophisticated model which overfits the limited amount of data.

From an information theoretic point of view, we prefer to choose a model that has MDL in explaining the data. In other words, the data are encoded with the help of a particular model and the model itself has to be encoded too. The MDL criterion chooses the model with the minimum overall code length defined below. In [15], the model complexity is measured by the normalized maximum likelihood (NML), i.e.,

\[
\hat{p}(\mathbf{Z} | H_i) = \frac{p(\mathbf{Z} | \hat{\mathbf{x}}_i, H_i)}{\int_{\hat{\mathbf{x}}(\mathbf{Z}) \in \Omega_i} p(\mathbf{Z} | \hat{\mathbf{x}}_i, H_i) d\mathbf{Z}}
\]

where \( \mathbf{Z} \) is a data sequence of length \( N \) that yields the maximum likelihood estimate (MLE) \( \hat{\mathbf{x}}_i \) and \( \Omega_i \) is a subset of the parameter space for \( H_i \) that makes the integral finite.\(^1\) It is shown in [15] that under some mild conditions, the code length of the NML is

\[
- \ln \hat{p}(\mathbf{Z} | H_i) = - \ln p(\mathbf{Z} | \hat{\mathbf{x}}_i, H_i) + \frac{n_i}{2} \ln \left( \frac{N}{2\pi} \right)
+ \ln \int_{\Omega_i} |I(\mathbf{x}_i)|^{1/2} d\mathbf{x}_i + o(1).
\]

\(^1\)The integral is finite if \( \Omega_i \) is an open bounded set and its closure contains no singular point of the FIM.

CHEN ET AL.: MDL APPROACH FOR MULTIPLE LOW OBSERVABLE TRACK INITIATION 863
The MDL criterion is an invariant measure under reparameterization, i.e., invariant to any one-to-one mapping of the parameter space. However, in the Bayesian approach, only Jeffreys’ prior is invariant under reparameterization. In some simple cases, this prior may not exist.

III. ONE-DIMENSIONAL ILLUSTRATIVE EXAMPLES

In this section, we present several one-dimensional examples to show the MDL model selection used for track initiation. The ML-PDA is used to obtain an initial estimate of the track. The MDL approach is also compared with the GLRT for hypothesis testing. From this simple problem, some insights on the performance limits of the ML-PDA in track initiation are interpreted in terms of the sharpness of the test.

Example 1: Consider an estimation problem under a false alarm environment. We have \( N \) scans and at scan \( i \) we have \( m_i \) measurements namely \( z_{i1}, \ldots, z_{im_i} \). The overall measurement set is denoted as \( Z = \{ z_1, \ldots, z_N \} \). For each scan \( i \), we have as follows.

- \( H_0 \): The number of false measurements follows Poisson distribution with the expect number of false measurements equal to \( \lambda \) per unit volume. All false measurements are independent identically distributed (IID) \( U(-A,A) \), where \( U(-A,A) \) is the uniform distribution in \((-A,A)\).

- \( H_1 \): False measurements as in \( H_0 \) plus a target-originated measurement with probability \( P_D \) in each scan (independent from scan to scan) following \( N(\mu, \sigma^2) \) with unknown mean \( \mu \) and known variance \( \sigma^2 \).

This is a detection problem of a Gaussian signal with unknown mean observed together with uniformly distributed noises. We can view the Gaussian signal as a static object measured by a sensor. The track initiation problem consists of detecting this Gaussian signal and estimating its mean. Although it is a simple one-dimensional problem, the methodology also applies to the more complicated situation that is presented in the next section. Under \( H_0 \) the likelihood function is

\[
p(Z \mid H_0) = \prod_{i=1}^{N} p(z_i \mid H_0) = \prod_{i=1}^{N} \left( \frac{1}{2\lambda} \right)^{m_i} \exp \left\{ -\frac{Z_i^2}{2\lambda} \right\} . \quad (9)
\]

Under \( H_1 \), due to the measurement origin uncertainty, one has to use a certain data association technique combined with the parameter estimation. One approach is to formulate the ML-PDA. The likelihood function, using ML-PDA, is [4]

\[
p(Z \mid \mu, H_1) = \prod_{i=1}^{N} p(z_i \mid \mu, H_1) \quad (10)
\]

where

\[
p(z_i \mid \mu, H_1) = \frac{1}{C_i} \left[ (1 - P_D)P(m_i) \left( \frac{1}{2\lambda} \right)^{m_i} \right.
\]

\[
+ P_D P(m_i - 1) \left( \frac{1}{2\lambda} \right)^{m_i-1} \frac{1}{m_i} \sum_{j=1}^{m_i} P(m_j) \exp \left\{ \frac{(z_{ij} - \mu)^2}{2\lambda} \right\} \left. \right] \quad (11)
\]

\( P(m_i) \) is the Poisson distribution given by

\[
P(m_i) = \frac{e^{-2\lambda} (2\lambda)^m_i}{m_i!} \quad (12)
\]

and \( C_i \) is a normalizing constant (the probability of having \( m_i \) measurements) given by

\[
C_i = (1 - P_D)P(m_i) + P_D P(m_i - 1) \quad m_i > 0 . \quad (13)
\]

Note that the first term in the right hand side (RHS) of (8) is the log-likelihood function. The second term in the RHS of (8) is a penalty for the number of unknown parameters used to fit the data. The third term in the RHS of (8) is a penalty for the functional complexity measured over the subset \( \Omega_i \). In general, it is difficult to compute the third penalty term. In the AIC [1] the penalty term \( n_i \) is used while in the BIC [18] the penalty term \( n_i/2\log N \) is used. They all ignore the differences in model complexity in different functional forms. In [11], the penalty as a function of the FIM is used rather than as a function of the number of parameters, which again can only choose the correct model in an asymptotic sense.

The MDL criterion is an invariant measure under reparameterization [15]. It has the advantage of not favoring a particularly complex model, and is suitable for the multiple composite hypothesis testing problem considered here.
Notice that \( C_i = (1 - P_D)P(0) \) when \( m_i = 0 \). The MLE \( \hat{\mu} \) is
\[
\hat{\mu} = \underset{\mu}{\arg \max} \ p(Z \mid \mu, H_1).
\] (14)

The log-likelihood ratio is [4]
\[
\ln \left( \frac{p(Z \mid \mu, H_1)}{p(Z \mid H_0)} \right) = \sum_{i=1}^{N} \left[ \ln \left( 1 - P_D \frac{P_0}{\lambda (2\pi\sigma)} \sum_{j=1}^{m_i} \exp \left\{ -\frac{(z_{ij} - \mu)^2}{2\sigma^2} \right\} \right) - \ln c_i \right] \]
where \( c_i = \begin{cases} 1 & m_i = 0 \\ (1 - P_D) + \frac{m_i}{2A\lambda}P_D & m_i > 0. \end{cases} \) (15)

The maximization of (15) that yields the estimate \( \hat{\mu} \) has to be done numerically. This is discussed in detail later.

Applying the MDL criterion, we have the test statistic
\[
T(Z) = \ln \left( \frac{p(Z \mid \hat{\mu}, H_1)}{p(Z \mid H_0)} \right) - \frac{1}{2} \ln \left( \frac{S_N}{2\pi} \right) - \ln \int_{-A}^{A} |I(\mu)|^{1/2} d\mu
\] (17)
where
\[
S_N = \sum_{i=1}^{N} m_i.
\] (18)

Thus we choose \( H_1 \) when \( T(Z) > 0 \).

In simulation, we let \( A = 10 \), \( \mu = 0 \), \( \sigma = 1 \), and change \( P_D \) and \( \lambda \) to increase the difficulty in estimating \( \mu \). No gating is used here. For the difficult missile acquisition problem to be presented later, gating significantly reduces the computational load while giving similar ML-PDA performance. In traditional hypothesis testing we assume that the generalized log-likelihood ratio (GLLR)
\[
\ln \left( \frac{p(Z \mid \hat{\mu}, H_1)}{p(Z \mid H_0)} \right)
\] is asymptotically normal under both \( H_1 \) and \( H_0 \). We are interested in its mean and variance in both cases. We fix \( P_D = 0.7 \) and increase the number of scans \( N \) from 30 to 120.

For each case, 5000 Monte Carlo runs are used to estimate the mean and variance of the GLLR. Table I lists the mean and variance of the GLLR under \( H_1 \) and \( H_0 \) and the variance of the MLE \( \hat{\mu} \) under \( H_1 \) for \( \lambda = 0.10 \). Table II lists the mean and variance of the GLLR for \( \lambda = 0.15 \). From Tables I and II several comments are worth noting. First, the mean of \( \hat{\mu} \) (not listed) is close to \( \mu \) due to the asymptotic unbiasedness of the MLE while the variance of \( \hat{\mu} \) decreases as \( N \) increases. This indicates the effectiveness of the ML-PDA. Second, the difference of the means under \( H_1 \) and \( H_0 \) becomes larger as \( N \) increases. The most important observation is that the mean of the GLLR becomes positive under \( H_0 \) when \( \lambda \) increases. This indicates that \( H_1 \) is preferred in the sense of GLLR even though the data is generated from \( H_0 \).

Intuitively, we can see that the more sophisticated model is likely to fit the data better. Traditional hypothesis testing does not care about whether the GLLR under \( H_0 \) is positive or negative. Based on the means and variances under \( H_0 \) and \( H_1 \), two Gaussian densities are used to approximate the GLLRs for the two hypotheses and a threshold is chosen to obtain a certain power of the test. The criticisms to this GLRT technique are: 1) it cannot deal with multiple composite hypothesis testing; 2) no optimality is guaranteed for the test statistic under finite scans; 3) the Gaussian approximation of the GLLR under \( H_0 \) is hard to obtain.

To test whether the approximate MDL criterion can reduce the bias in favoring \( H_1 \), we add the penalty term to the GLLR as in (17) and estimate the target acquisition probability \( \hat{P}(H_1 \mid H_1) \) as well as false acquisition probability \( \hat{P}(H_1 \mid H_0) \) with the same parameter settings as above. To simplify the problem, we let
\[
\int_{-A}^{A} |I(\mu)|^{1/2} d\mu = \frac{2A}{\sigma}
\] (19)
without considering the information reduction factor (IRF) [21] due to false alarms. In this case, \( H_1 \) is penalized more than it should be since it has more flexibility to explain the data in a probabilistic data association (PDA) fashion.

We fix the detection probability and change the false alarm rate for various \( N \). The results are listed in Table III. We can see that \( H_1 \) is much more likely to be accepted even though the truth is \( H_0 \) when the expected number of false alarms per unit volume is greater than 0.1.3 In all cases, the target acquisition probability is close to 1 indicating that the ML-PDA is quite effective when the target is present. However, when \( \lambda > 0.15 \), the false acquisition probability does not decrease as \( N \) increases. This means that the ML-PDA approach explains the data well in the MDL sense even though they are purely uniformly distributed within the surveillance volume. The test of the two models becomes less sharp as \( \lambda \) increases.

**Example 2:** Now we assume the number of Gaussian signals is unknown. The number of false measurements is still Poisson with the expected number \( \lambda \) per unit volume in each scan. They are uniformly distributed in \((-A,A)\). When there is a target, its detection probability is \( P_D \). We assume \( P_D \) is known and is the same for all targets. All the target-originated measurements follow normal distribution with unknown mean and known variance.

---

3Notice that the size of a resolution cell is \( \sqrt{2}\sigma \) if the false measurement is uniformly distributed within the cell. In this example, the probability of false alarm in a resolution cell is \( P_{FA} = \sqrt{2}\sigma\lambda = 0.346 \) when \( \lambda = 0.1 \). This amounts to extremely high false alarm rates in a real application.
estimated target (within with high probability to be associated with the maximum) first, and then remove those measurements "strongest" target (i.e., the target that yields (17) to be the ML-PDA approach described as above. When hypothesis testing of one target versus no target using the hypothesis testing. Let \(A = 10, \sigma = 1 \) be fixed. We consider three typical scenarios, namely, no target \((H_0)\), one target \((H_1; \mu = 4)\), and two targets \((H_2; \mu_1 = 4, \mu_2 = -3)\). We begin with the hypothesis testing of one target versus no target using the ML-PDA approach described as above. When more than one target is present, we simply extract the "strongest" target (i.e., the target that yields (17) to be maximum) first, and then remove those measurements with high probability to be associated with the estimated target (within \(\pm 1.96\sigma \) region of the estimate in this case) and use the rest of the measurements to do the same hypothesis testing. The procedure is repeated until no target can be extracted out from the measurements (i.e., the test statistic (17) is less than 0). Notice that the measurements originated from other targets are considered as false alarms (which clearly do not follow uniform distribution) when no target has been extracted out initially. We still apply the MDL criterion in the multistage testing procedure and obtain the correct model selection probabilities among three scenarios under various \(\lambda \) and \(N\).

The results are listed in Table IV for comparison. We can see that the performance of the MDL criterion is sensitive to \(\lambda\). Comparing with Table III, the probability of choosing the correct model when more than one target is present is quite close to that of the one target case. Table V shows the probability of choosing the correct model under different \(P_D\) and \(\lambda\). Comparing with Table IV we can see that the performance of the MDL criterion is not very

<table>
<thead>
<tr>
<th>(N)</th>
<th>30</th>
<th>40</th>
<th>50</th>
<th>60</th>
<th>70</th>
<th>80</th>
<th>90</th>
<th>100</th>
<th>110</th>
<th>120</th>
</tr>
</thead>
<tbody>
<tr>
<td>mean(GLLR (</td>
<td>H_0 ))</td>
<td>-6.04</td>
<td>-8.94</td>
<td>-11.64</td>
<td>-14.74</td>
<td>-17.74</td>
<td>-20.74</td>
<td>-23.84</td>
<td>-26.64</td>
<td>-29.94</td>
</tr>
<tr>
<td>std(GLLR (</td>
<td>H_0 ))</td>
<td>4.8</td>
<td>5.5</td>
<td>6.2</td>
<td>6.8</td>
<td>7.3</td>
<td>7.6</td>
<td>8.2</td>
<td>8.6</td>
<td>9.1</td>
</tr>
<tr>
<td>mean(GLLR (</td>
<td>H_1 ))</td>
<td>18.1</td>
<td>23.8</td>
<td>29.7</td>
<td>35.6</td>
<td>41.2</td>
<td>47.1</td>
<td>52.8</td>
<td>58.8</td>
<td>64.6</td>
</tr>
<tr>
<td>std(GLLR (</td>
<td>H_1 ))</td>
<td>4.9</td>
<td>5.8</td>
<td>6.3</td>
<td>6.9</td>
<td>7.6</td>
<td>8.2</td>
<td>8.5</td>
<td>9.0</td>
<td>9.6</td>
</tr>
<tr>
<td>std((\hat{\mu}</td>
<td>H_1 ))</td>
<td>0.36</td>
<td>0.30</td>
<td>0.26</td>
<td>0.24</td>
<td>0.22</td>
<td>0.20</td>
<td>0.19</td>
<td>0.18</td>
<td>0.17</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>(N)</th>
<th>30</th>
<th>40</th>
<th>50</th>
<th>60</th>
<th>70</th>
<th>80</th>
<th>90</th>
<th>100</th>
<th>110</th>
<th>120</th>
</tr>
</thead>
<tbody>
<tr>
<td>mean(GLLR (</td>
<td>H_0 ))</td>
<td>4.2</td>
<td>4.9</td>
<td>5.3</td>
<td>5.8</td>
<td>6.3</td>
<td>6.8</td>
<td>7.2</td>
<td>7.4</td>
<td>7.7</td>
</tr>
<tr>
<td>std(GLLR (</td>
<td>H_0 ))</td>
<td>4.3</td>
<td>4.8</td>
<td>5.4</td>
<td>5.8</td>
<td>6.4</td>
<td>6.8</td>
<td>7.2</td>
<td>7.7</td>
<td>8.0</td>
</tr>
<tr>
<td>mean(GLLR (</td>
<td>H_1 ))</td>
<td>18.2</td>
<td>24.1</td>
<td>30.1</td>
<td>35.8</td>
<td>41.9</td>
<td>47.8</td>
<td>53.4</td>
<td>59.7</td>
<td>65.3</td>
</tr>
<tr>
<td>std(GLLR (</td>
<td>H_1 ))</td>
<td>3.9</td>
<td>4.5</td>
<td>5.1</td>
<td>5.5</td>
<td>6.0</td>
<td>6.4</td>
<td>6.9</td>
<td>7.2</td>
<td>7.6</td>
</tr>
<tr>
<td>std((\hat{\mu}</td>
<td>H_1 ))</td>
<td>0.47</td>
<td>0.39</td>
<td>0.33</td>
<td>0.29</td>
<td>0.27</td>
<td>0.25</td>
<td>0.23</td>
<td>0.22</td>
<td>0.21</td>
</tr>
</tbody>
</table>

| \(\lambda\) | \(P(H_1 | H_1)\) | \(P(H_1 | H_0)\) |
|----------|----------------|----------------|
| 0.09     | 1.0 1.0 1.0 1.0 | 0.007 0.003 0.001 0.000 0.000 0.000 0.000 0.000 0.000 0.000 |
| 0.11     | 0.988 0.996 0.999 1.0 1.0 1.0 1.0 1.0 1.0 1.0 |
| 0.13     | 0.985 0.993 0.997 0.999 1.0 1.0 1.0 1.0 1.0 1.0 |
| 0.15     | 0.981 0.992 0.995 0.998 0.999 1.0 1.0 1.0 1.0 1.0 |

\(\sigma^2\): Clearly, we need to test multiple composite hypotheses. In this case, the model selection criterion either from the Bayes point of view or the MDL approach has to be considered.

We assume that the Gaussian signals have enough separations so that any of the two signals can be individually estimated without resorting to their joint likelihood function. Thus there is no need to model the target-originated measurements as a Gaussian mixture in the hypothesis testing. Let \(A = 10, \sigma = 1 \) be fixed. We consider three typical scenarios, namely, no target \((H_0)\), one target \((H_1; \mu = 4)\), and two targets \((H_2; \mu_1 = 4, \mu_2 = -3)\). We begin with the hypothesis testing of one target versus no target using the ML-PDA approach described as above. When more than one target is present, we simply extract the "strongest" target (i.e., the target that yields (17) to be maximum) first, and then remove those measurements with high probability to be associated with the estimated target (within \(\pm 1.96\sigma \) region of the estimate in this case) and use the rest of the measurements to do the same hypothesis testing. The procedure is repeated until no target can be extracted out from the measurements (i.e., the test statistic (17) is less than 0). Notice that the measurements originated from other targets are considered as false alarms (which clearly do not follow uniform distribution) when no target has been extracted out initially. We still apply the MDL criterion in the multistage testing procedure and obtain the correct model selection probabilities among three scenarios under various \(\lambda\) and \(N\).

The results are listed in Table IV for comparison. We can see that the performance of the MDL criterion is sensitive to \(\lambda\). Comparing with Table III, the probability of choosing the correct model when more than one target is present is quite close to that of the one target case. Table V shows the probability of choosing the correct model under different \(P_D\) and \(\lambda\). Comparing with Table IV we can see that the performance of the MDL criterion is not very
sensitive to \( P_D \) when \( \lambda < 0.10 \). We conclude that 1) the iterative MDL approach used in this one-dimensional example is effective for multiple composite hypothesis testing when the targets have enough separation, and 2) the MDL criterion yields large false acquisition probability when \( \lambda > 0.10 \). As \( \lambda \) increases, even the simple versus composite alternative hypothesis testing (two model comparison) becomes less sharp as can be seen in Table III. Note that the upper limit of the expected number of false alarms in the gate is around 0.6. Another interesting phenomenon is that \( \hat{P}(H_2 | H_1) > \hat{P}(H_1 | H_2) \) for any fixed \( \lambda \) and \( N \). This is due to the following: when \( P_D \) is around 0.5 to 0.7, there can be enough false measurements aligned such that a false target is declared.

The example reveals that the region of acceptable performance of the ML-PDA for model selection is limited by \( \lambda < 0.10 \), which, as indicated before, corresponds to an extremely high \( P_{FA} \). If we think each model represented by the NML density function and the sharpness of the hypothesis testing measured by the distances between these NML density functions, we may conclude that the test becomes less sharp when \( \lambda \) increases. Note that the MDL criterion only indicates that the candidate models tend to be indistinguishable as \( \lambda \) increases. It does not provide any information on how to choose a good set of models that yields a sharper test. For the track initiation problem of higher dimensional parameter space, the same methodology applies but the NML density functions are not readily computable and certain approximation techniques are required.

**Example 3:** Now we consider the same one-dimensional target acquisition problem with amplitude information. At each scan, if a measurement is a false alarm, the pdf of its associated amplitude \( a \) follows a Swerling I model given by

\[
p_0(a) = \frac{1}{P_{FA}} a \exp\left(-\frac{a^2}{2}\right) \quad a > \tau
\]

where

\[
P_{FA} = \int_{\tau}^{\infty} a \exp\left(-\frac{a^2}{2}\right) da
\]

and \( \tau \) is a suitable threshold used to declare a detection. For acquisition of low SNR targets, \( \tau \) should be set fairly low to get a reasonable target detection probability. The density of the amplitude is a truncated Rayleigh distribution at the threshold \( \tau \). If the SNR is \( d \) (assuming known), the pdf of the amplitude from a target-originated measurement is

\[
p_1(a) = \frac{1}{P_D} \frac{a}{1+d} \exp\left(-\frac{a^2}{2(1+d)}\right) \quad a > \tau
\]
TABLE VI
Estimated Probability of Choosing Correct Model Among Three Different Scenarios with Amplitude Information Under Various Target SNR and \( N, P_D = 0.7, 5000 \) Runs

<table>
<thead>
<tr>
<th>( N )</th>
<th>30</th>
<th>40</th>
<th>50</th>
<th>60</th>
<th>70</th>
<th>80</th>
<th>90</th>
<th>100</th>
<th>110</th>
<th>120</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \lambda = 0.11 )</td>
<td>0.90</td>
<td>0.91</td>
<td>0.94</td>
<td>0.95</td>
<td>0.96</td>
<td>0.96</td>
<td>0.97</td>
<td>0.98</td>
<td>0.99</td>
<td>0.99</td>
</tr>
<tr>
<td>( \lambda = 0.13 )</td>
<td>0.93</td>
<td>0.94</td>
<td>0.95</td>
<td>0.97</td>
<td>0.97</td>
<td>0.98</td>
<td>0.98</td>
<td>0.99</td>
<td>0.99</td>
<td>0.99</td>
</tr>
<tr>
<td>( \lambda = 0.14 )</td>
<td>0.99</td>
<td>0.99</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>( \lambda = 0.13 )</td>
<td>0.81</td>
<td>0.83</td>
<td>0.85</td>
<td>0.87</td>
<td>0.89</td>
<td>0.90</td>
<td>0.91</td>
<td>0.91</td>
<td>0.92</td>
<td>0.93</td>
</tr>
<tr>
<td>( \lambda = 0.14 )</td>
<td>0.87</td>
<td>0.88</td>
<td>0.90</td>
<td>0.91</td>
<td>0.92</td>
<td>0.93</td>
<td>0.94</td>
<td>0.94</td>
<td>0.95</td>
<td>0.95</td>
</tr>
<tr>
<td>( \lambda = 0.14 )</td>
<td>0.98</td>
<td>0.99</td>
<td>0.99</td>
<td>0.99</td>
<td>0.99</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>( \lambda = 0.14 )</td>
<td>0.57</td>
<td>0.58</td>
<td>0.58</td>
<td>0.60</td>
<td>0.61</td>
<td>0.62</td>
<td>0.63</td>
<td>0.64</td>
<td>0.64</td>
<td>0.65</td>
</tr>
<tr>
<td>( \lambda = 0.14 )</td>
<td>0.70</td>
<td>0.70</td>
<td>0.71</td>
<td>0.72</td>
<td>0.73</td>
<td>0.74</td>
<td>0.74</td>
<td>0.75</td>
<td>0.75</td>
<td>0.75</td>
</tr>
<tr>
<td>( \lambda = 0.14 )</td>
<td>0.88</td>
<td>0.89</td>
<td>0.89</td>
<td>0.90</td>
<td>0.91</td>
<td>0.92</td>
<td>0.92</td>
<td>0.93</td>
<td>0.93</td>
<td>0.94</td>
</tr>
</tbody>
</table>

where

\[
P_D = \int_\tau^\infty \frac{a}{1 + d} \exp \left( -\frac{a^2}{2(1 + d)} \right) da. \quad (23)
\]

For a 9 dB target we have \( P_D = 0.7 \) and \( P_{\lambda} = 0.044 \) when \( \tau = 2.5 \). The measurement set is augmented with the amplitude associated to each point measurement. We still denote the total number of scans as \( N \). At scan \( i \), we have \( m_i \) measurements with the corresponding amplitude denoted as \( a_i = \{ a_{i1}, \ldots, a_{im_i} \} \). Similarly, the overall amplitude set is denoted as \( A = \{ a_i, \ldots, a_N \} \). We define the amplitude likelihood ratio \( \rho_{ij} \) as

\[
\rho_{ij} = \frac{p_i(a_{ij})}{p_0(a_{ij})} \quad (24)
\]

when an amplitude \( a_{ij} \) is observed (at scan \( i \) of measurement \( z_{ij} \)). With the amplitude information, the log-likelihood ratio in (15) becomes [4]

\[
\ln \left( \frac{\tilde{p}(Z_{ij} | \tilde{\mu}, H_i)}{\tilde{p}(Z_{ij} | \tilde{\mu}, H_0)} \right) = \sum_{i=1}^N \ln \left( 1 - P_D + \frac{P_D}{\lambda(\sqrt{2\pi}\sigma)} \sum_{j=1}^{m_i} \rho_{ij} \exp \left\{ -\frac{(z_{ij} - \tilde{\mu})^2}{2\sigma^2} \right\} \right) - \ln c_j \quad (25)
\]

where the MLE \( \tilde{\mu} \) is obtained by maximizing the new likelihood ratio given in (25). The amplitude provides additional information to resolve the measurement origin uncertainty [4].

We apply the same MDL criterion to model selection for measurements with the amplitude information and expect some improvement in performance measured in terms of the maximal tolerable \( \lambda \). We consider three typical scenarios as presented above and change the target SNR from 2 dB down to 0 dB. In each case, the amplitude information is used in the ML-PDA, \( P_D = 0.7 \) is fixed and \( N \) varies from 30 to 120. The results are listed in Table VI. Comparing with Table IV we can see that the amplitude information helps to improve the performance even for target SNR as low as 1 dB. However, the improvement in terms of tolerating heavy false alarm density is not significant (the maximum \( \lambda \) for which the performance is still good increases from 0.1 to 0.13; these correspond to SNR = 3 dB and 1 dB, respectively). Thus, we conclude that the MDL approach for model selection works well for \( \lambda \leq 0.13 \) by using the amplitude information.

To summarize the results from the above one-dimensional illustrative examples, we find that the performance limit of the ML-PDA for track initiation is bounded by a value of \( \lambda \) beyond which the candidate models become less distinguishable in the MDL sense. The MDL criterion works well when \( \lambda \leq 0.1 \) (or equivalently, the expected number of false alarms in the 3\( \sigma \) validation gate is up to 0.6) without using the amplitude information and it can achieve similar performance for \( \lambda \leq 0.13 \) (or equivalently, the expected number of false alarms in the 3\( \sigma \) validation gate is up to 0.8) with the amplitude information.

IV. APPLICATION: ACQUISITION OF BALLISTIC MISSILES

The main focus of the ballistic missile acquisition is to initialize the missile trajectories using the measurements obtained by a surface-based radar for a short period of time before the missile reentry phase. A track initiation algorithm has to determine how many targets are in the cueing region (model selection problem). If there are targets, the initial state of each target has to be estimated. In the following subsections, we present the target and radar model and the ML-PDA algorithm for missile trajectory parameter estimation. The detailed missile trajectory, radar model, and the monopulse processing technique can be found in [8, 21] and Appendices B–G.
A. Coordinate Systems and Target Trajectory

The motion of a missile above the atmosphere is governed by Kepler’s laws. The Earth-centered inertial (ECI) coordinate system is used for the state propagation of the targets. This coordinate system has its origin at the center of the Earth, with the positive x axis pointing along the vernal equinox and the positive z axis to the north pole. The radar and target locations are better interpreted in the Earth-centered Earth-fixed (ECEF) coordinate system. This coordinate system has its origin at the center of the Earth, with the positive x axis passing through the prime meridian at the equator and the positive z axis passing through the north pole. Let \( x_k = [r_k \_ r_k] \) be the 6-dimensional target state vector at time \( t_k \), where the vectors \( r_k = [\xi(t_k) \, \eta(t_k) \, \zeta(t_k)] \) and \( \_r_k = [\dot{\xi}(t_k) \, \dot{\eta}(t_k) \, \dot{\zeta}(t_k)] \) are the position and velocity of the target in ECI coordinates, respectively. Given the initial state \( x_0 \) of the target at time \( t_0 \), the state at \( t_k \) is

\[
x_k = [r_k(x_0,t_0,t_k) \_ r_k(x_0,t_0,t_k)]
\]

where \( r_k(x) \) and \( \_r_k(x) \) are deterministic functions given in [8].

Another coordinate system is located at the radar. This coordinate has its origin at the center of the Earth, with the positive x axis pointing toward north and positive z axis passing through the prime meridian. The radar performs a volumetric search over the cueing region based on prior information. It is assumed that all the dwells in a scan are made at (practically) the same time. Denote by \( m_i \) the number of measurements in the \( i \)-th scan obtained at time \( t_i \). The set is written as

\[
Z(i) = \{z_{ik}\}_{k=1}^{m_i}.
\]

The cumulative set of measurements during the entire period is

\[
Z = \{Z(i)\}_{i=1}^{N}.
\]

The following assumptions about the statistical characteristics of the measurements are made.

1) The measurements at two different scans are conditionally independent, i.e.,

\[
p(Z(i), Z(j) \mid x) = p(Z(i) \mid x)p(Z(j) \mid x).
\]

2) A measurement that originated from a target at a particular scan is given by

\[
r_{ik} = r_0 + (n + \frac{1}{2})\Delta_r,
\]

if \( r_0 + n\Delta_r < r_i(x_i) < r_0 + (n + 1)\Delta_r \).
\[
\beta_{ik} = \beta_i(x_i) + w_{ik}^\beta
\]
\[
\epsilon_{ik} = \epsilon_i(x_i) + w_{ik}^\epsilon
\]
\[
w_{ik}^\beta \sim \mathcal{N}(0,(\sigma_{ik}^\beta)^2)
\]
\[
w_{ik}^\epsilon \sim \mathcal{N}(0,(\sigma_{ik}^\epsilon)^2)
\]

where \(r_0\) is the reference range; \(\Delta_i\) is the length of a range cell, \(n\) is the index of the range bins (range cells), and the bearing and elevation measurement noise standard deviations \(\sigma_{ik}^\beta\) and \(\sigma_{ik}^\epsilon\) are obtained as in [8].

3) In the case of a false alarm, the bearing and elevation measurements are Gaussian distributed around the commanded bearing and elevation (beam boresight), respectively. That is, if a measurement \(z_{ik}\) is a false alarm, we have (if SNR in range bin \(n\) exceeds the detection threshold)
\[
r_{ik} = r_0 + (n + \frac{1}{2})\Delta_i
\]
\[
\beta_{ik} \sim \mathcal{N}(b_{ik},(\sigma_{ik}^\beta)^2)
\]
\[
\epsilon_{ik} \sim \mathcal{N}(e_{ik},(\sigma_{ik}^\epsilon)^2)
\]

where \(b_{ik}\) and \(e_{ik}\) are the bearing and elevation of the beam boresight for the \(k\)th measurement in the \(i\)th radar scan, and \(\sigma_{ik}^\beta\) and \(\sigma_{ik}^\epsilon\) are the standard deviations which can be obtained for various detection thresholds.\(^4\)

4) The number of false measurements in a volume \(V_g\) (measurement validation gate around the hypothesized trajectory) at each scan follows a Poisson probability mass function (pmf) with known expected number of false measurements \(\lambda\) per unit volume. That is, the probability that there are \(m_i\) false measurements in a volume \(V_g\) is given by
\[
\mu(m_i) = e^{-\lambda V_g}(\lambda V_g)^{m_i}/m_i!
\]

where the parameter \(\lambda\) is the ratio of the false alarm probability in a resolution cell to the cell volume.

If there are \(m_i\) measurements in the \(i\)th scan, we assume at most one measurement originates from the target. Hence the mutually exclusive and exhaustive events giving rise to these measurements are
\[
\chi_i(i) = \begin{cases} 
\text{all measurements are false} & k = 0 \\
\text{measurement } z_{ik} \text{ is from the target} & k = 1, \ldots, m_i.
\end{cases}
\]

This assumption may not hold exactly since a target may have multiple detections from several adjacent beams in one scan. In this case, only the detection from the main beam is considered as target-originated measurement and the rest of them are taken as false alarms. Denote by \(p_0(\hat{R}_{ij}), p_0(r_{ij}), p_0(\beta_{ij}),\) and \(p_0(\epsilon_{ij})\) the pdfs of SNNR, slant range, bearing, and elevation of the false measurement \(z_{ij}\) with the detection threshold \(\tau\), respectively. We have [8]
\[
p_0(\hat{R}_{ij}) = \frac{1}{PFA} p_0(\bar{R}_{ij}), \quad \bar{R}_{ij} > \tau.
\]

Denote by \(p_1(\hat{R}_{ij} | x_i), p_1(r_{ij} | x_i), p_1(\beta_{ij} | x_i),\) and \(p_1(\epsilon_{ij} | x_i)\) the pdfs of SNNR, slant range, bearing, and elevation from the target-originated measurement \(\hat{R}_{ij}\) at state \(x_i\) (the full position at reference time \(t_i\)) with the detection threshold \(\tau\), respectively. We have [8]
\[
p_1(\hat{R}_{ij} | x_i) = \frac{1}{PD} p(\hat{R}_{ij} | x_i), \quad \bar{R}_{ij} > \tau.
\]

Under \(H_0\), the pdf of the measurements \(Z(i)\) corresponding to the hypothesis that all measurements are false is given by
\[
p_0(Z(i) | x_0(i)) = \prod_{j=1}^{m_i} p_0(\hat{R}_{ij}) p_0(r_{ij}) p_0(\beta_{ij}) p_0(\epsilon_{ij}).
\]

Under \(H_1\), the pdf of the measurements \(Z(i)\) if the \(k\)th measurement originates from the target is given by
\[
p_1(Z(i) | x_i, \chi_k(i)) = p_1(\hat{R}_{ik} | x_i) p_1(r_{ik} | x_i) p_1(\beta_{ik} | x_i) p_1(\epsilon_{ik} | x_i)
\]
\[
\cdot \prod_{j=1,j \neq k}^{m_i} p_0(\hat{R}_{ij}) p_0(r_{ij}) p_0(\beta_{ij}) p_0(\epsilon_{ij})
\]

where \(x_i\) is the target position at time \(t_i\). From (42), (43) the amplitude likelihood ratio is [8]
\[
p_{ik} = \frac{p_1(\hat{R}_{ik} | x_i)}{p_0(\hat{R}_{ik})} = \zeta_i(1 + \nu_i R_{ik} e^{\nu_i \bar{R}_{ik}})
\]

and the likelihood ratio of the position component between a target-originated and a false measurement is
\[
\frac{p_1(r_{ik} | x_i) p_1(\beta_{ik} | x_i) p_1(\epsilon_{ik} | x_i)}{p_0(r_{ik}) p_0(\beta_{ik}) p_0(\epsilon_{ik})} = \frac{\sigma_i^\beta \sigma_i^\epsilon}{\sigma_{ik}^\beta \sigma_{ik}^\epsilon} \exp\{\psi(z_{ik}, x_i)\}
\]

where, with \(\alpha_i\Sigma_i^4\) denoting the average SNR (multiplied by 2),\(^5\)
\[
\zeta_i = \frac{16 PFA}{P_D (4 + \alpha_i \Sigma_i^4)^2}
\]

\(^4\)This is the consequence of the “centrality tendency” of the false measurements, discussed in [21].

\(^5\)The SNR is the ratio of the total target power to the total noise power in the I and Q radar receiver channel, with each of them normalized to unity. Thus \(\text{SNR} = \alpha_i \Sigma_i^4 / 2\) and \(\alpha_i \Sigma_i^4\) is the normalized signal power (average target return) as it passes through the antenna pattern. For more details, see [6, 21].
The log-likelihood ratio is given by

\[ \psi(c_{ik}, x_i) = \frac{[\beta_k - \beta_i(x_i)]^2}{2(\sigma_{ik}^2)} - \frac{[\epsilon_{ik} - \epsilon_i(x_i)]^2}{2(\sigma_{ij}^2)} + \frac{(\beta_k - b_{ik})^2}{2(\sigma_{ik}^2)} + \frac{(\epsilon_{ik} - e_{ik})^2}{2(\sigma_{ij}^2)} \]

and the radiation pattern \( \Sigma_i \) is given by (89) in Appendix D. Assuming the prior probabilities of each measurement to be target-originated are equal, the likelihood ratio of one target present versus no target is, using the PDA approach, given by

\[ \Phi(Z(i), x_i) = \frac{P_D}{XV_s} \sum_{k=1}^{m_i} \exp(\psi(z_{ik}, x_i)) + (1 - P_D) \]

where \( C_i \) is a normalizing constant. Using the conditional independence of the measurements among different scans, the likelihood ratio of the entire measurement set can be written as

\[ \Lambda(Z, x) = \prod_{i=1}^{N} \Phi(Z(i), x_i). \]

The log-likelihood ratio is given by

\[
\ln(\Lambda(Z, x)) = \sum_{i=1}^{N} \ln \left\{ \frac{P_D}{XV_s} \sum_{k=1}^{m_i} \exp(\psi(z_{ik}, x_i)) + (1 - P_D) \right\} - \sum_{i=1}^{N} \ln c_i
\]

where \( c_i \) is a constant given by

\[
c_i = \begin{cases} 
0 & m_i = 0 \\
(1 - P_D) + \frac{m_i}{XV_s} P_D & m_i > 0.
\end{cases}
\]

Note that (53) includes all the measurements without any decision as which are target-originated. This is the essence of the ML-PDA approach.

In (53) the second summation involving \( c_i \) can be omitted when finding the MLE. The maximization of the log-likelihood ratio was accomplished in [21] using the quasi-Newton method following a preliminary grid search. Here, instead of evaluating the log-likelihood ratio with a grid search, we use a less expensive grid search based on the number of nonempty gates. This will speed up finding the initial point before using the quasi-Newton method.

C. Acceptance of the Estimate

We first consider the hypothesis test of one target versus no target:

\( H_0 \): \( \hat{x} \) does not correspond to a valid track.

We use the log-likelihood ratio \( \ln(\Lambda(Z, x)) \) with the penalty terms given by the approximate MDL criterion as the test statistic for the hypothesis test of \( H_1 \) against \( H_0 \). Denote \( T(Z) \) as the test statistic. We have

\[
T(Z) = \ln(\Lambda(Z, \hat{x})) - \frac{k}{2} \ln \left( \frac{S_N}{2\pi} \right) - \ln \int_{\Omega} |f(x)|^{1/2} dx
\]

where \( S_N \) is the number of measurements given by

\[
S_N = \sum_{i=1}^{N} m_i
\]

and \( k \) is the dimension of the MLE \( \hat{x} \) (in this case \( k = 6 \)) which is

\[
\hat{x} = \arg \max_{x} \ln(\Lambda(Z, x)).
\]

The parameter space \( \Omega \) covers the cueing region and all possible velocity vectors of the missile. Without considering the IRF, the integral \( \int_{\Omega} |f(x)|^{1/2} dx \) is obtained numerically and can be computed off-line.

We choose \( H_1 \) when \( T(Z) > 0 \). Alternatively, we can use \( \ln(\Lambda(Z, \hat{x})) \) as the test statistic and compare it with a prespecified threshold for certain power based on the Gaussian approximation under \( H_1 \). The numerical approach for obtaining the mean and variance of the test statistic under \( H_1 \) can be found in [21]. We compare the MDL with the GLRT in simulation.

If a track is accepted, the MLE \( \hat{x} \) is used as the state estimate with the corresponding covariance given by approximate Cramer-Rao lower bound (CRLB) which is presented in the next subsection.

The estimator uses all the measurements \( Z \) to extract one track at a time. The measurements having the association probability greater than 0.5 to the presumed track at each scan are deleted after the track has been initiated. If a measurement at a specific scan is deleted, all the measurements from its adjacent beams (in the same range bin) are also deleted (if there is any) to compensate for the effect of multiple detections. It is very important to remove all measurements from the main beam and the adjacent beams after a track is declared. The number of false tracks decreases when using this procedure especially when the SNR is low. The rest of the measurements are used to extract another target using the same ML-PDA estimator with MDL model selection criterion. The procedure continues until the test statistic \( T(Z) \) is less than 0.

D. Cramer-Rao Lower Bound of Estimator in Clutter

The CRLB is a lower limit on the variance that can be achieved by an unbiased estimator. The MLE can be said to be efficient if its sample variance is statistically commensurate with the CRLB. For the
MLE $\hat{x}$ of the state $x$, we have
\[ \mathcal{E}[x - \hat{x}] (x - \hat{x})' \geq J^{-1} \]
where $J$ is the FIM. If the noises of target-terminated measurements are Gaussian, the FIM can be written as
\[ J = \sum_{i=1}^{N} \frac{1}{(\sigma')^2} \left( \frac{1}{(\sigma')^2} \left[ \nabla_{x} g_i(x) [\nabla_{x} g_i(x)]' \right] + \frac{1}{(\sigma')^2} \left[ \nabla_{x} g_i(x) [\nabla_{x} g_i(x)]' \right] \right) \]
where $g_i(.)$ is the IRF for the $i$th scan that accounts for the loss of information due to the presence of false alarms and the less-than-unity probability of detection [21]. However, the noise of the range measurement is not Gaussian. We use a moment matching technique and let
\[ \sigma' = \sqrt{\frac{(2\Delta)^2}{12}} = 51.96. \]

To simplify the calculation, we also assume that the IRF is a constant across different scans. Only those scans with at least one measurement in the gate are considered in obtaining the FIM. Thus we can write the FIM as
\[ J = q(P_d, \lambda, \rho, \alpha^4) \sum_{i=1}^{N} \left( \frac{1}{(\sigma')^2} \left[ \nabla_{x} g_i(x) [\nabla_{x} g_i(x)]' \right] + \frac{1}{(\sigma')^2} \left[ \nabla_{x} g_i(x) [\nabla_{x} g_i(x)]' \right] \right) \]
\[ \quad + \frac{1}{(\sigma')^2} \left[ \nabla_{x} g_i(x) [\nabla_{x} g_i(x)]' \right] + \frac{1}{(\sigma')^2} \left[ \nabla_{x} g_i(x) [\nabla_{x} g_i(x)]' \right] \right) \]
\[ \quad \right) \]

The standard deviations of the bearing and elevation measurements depend on the target position within the radar beam. Assuming the target is uniformly distributed within the rectangular region covered by one radar beam, the average measurement noise for a target is $\sigma' = \sigma'' = 0.0067$ rad when the detection threshold is 3.25. If the detection probability per scan is $P_D = P_{\text{Dave}} = 0.464$, we have the average target return $\alpha \Sigma^4 = 5.76$. The false alarm probability is obtained by averaging the false detections in eight adjacent beams when the target is uniformly distributed within the rectangle area of the main beam (the one pointing to the target). Using the integral approach developed in [21], the IRF is $q_i(.) = 0.346$ for a single target with Poisson distributed false alarms. This IRF $q_i(.)$ is averaged over 100 randomly generated target positions uniformly distributed within the rectangular region of the main beam. It approximately quantifies the estimation accuracy. For multiple target scenarios, we assume the approximate CRLB for a single target is still valid for quantifying the efficiency of the estimator. For Swerling III targets with SNR = 6 dB at boresight and various detection thresholds, we list the detection probability, false alarm probability and the information reduction factor averaged over the above rectangular region in Table VII.

### Table VII

<table>
<thead>
<tr>
<th>Threshold $\tau$</th>
<th>$P_D$ at Boresight</th>
<th>$P_{\text{Dave}}$</th>
<th>Information Reduction $q$</th>
</tr>
</thead>
<tbody>
<tr>
<td>3.15</td>
<td>0.112</td>
<td>0.586</td>
<td>0.483</td>
</tr>
<tr>
<td>3.30</td>
<td>0.115</td>
<td>0.593</td>
<td>0.489</td>
</tr>
<tr>
<td>3.45</td>
<td>0.119</td>
<td>0.599</td>
<td>0.496</td>
</tr>
<tr>
<td>3.60</td>
<td>0.122</td>
<td>0.605</td>
<td>0.503</td>
</tr>
<tr>
<td>3.75</td>
<td>0.126</td>
<td>0.611</td>
<td>0.509</td>
</tr>
<tr>
<td>3.90</td>
<td>0.130</td>
<td>0.617</td>
<td>0.516</td>
</tr>
<tr>
<td>4.05</td>
<td>0.134</td>
<td>0.624</td>
<td>0.523</td>
</tr>
<tr>
<td>4.20</td>
<td>0.138</td>
<td>0.630</td>
<td>0.530</td>
</tr>
<tr>
<td>4.35</td>
<td>0.143</td>
<td>0.636</td>
<td>0.537</td>
</tr>
<tr>
<td>4.50</td>
<td>0.147</td>
<td>0.643</td>
<td>0.544</td>
</tr>
<tr>
<td>4.65</td>
<td>0.151</td>
<td>0.649</td>
<td>0.551</td>
</tr>
<tr>
<td>4.80</td>
<td>0.156</td>
<td>0.656</td>
<td>0.559</td>
</tr>
<tr>
<td>4.95</td>
<td>0.161</td>
<td>0.662</td>
<td>0.566</td>
</tr>
</tbody>
</table>

### Table VIII

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Description</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>$N$</td>
<td>radiating elements per side</td>
<td>55</td>
</tr>
<tr>
<td>$P_t$</td>
<td>transmitter power (MW)</td>
<td>1</td>
</tr>
<tr>
<td>$G_r$</td>
<td>antenna gain $\frac{N^2}{\tau}$</td>
<td>(2.55 \times 10^{-23})</td>
</tr>
<tr>
<td>$\lambda$</td>
<td>nominal wavelength (m)</td>
<td>0.075</td>
</tr>
<tr>
<td>$F_t$</td>
<td>transmitter propagation factor</td>
<td>1</td>
</tr>
<tr>
<td>$F_r$</td>
<td>receiver propagation factor</td>
<td>1</td>
</tr>
<tr>
<td>$L_{\text{tot}}$</td>
<td>total system losses</td>
<td>144.5</td>
</tr>
<tr>
<td>$k_B$</td>
<td>Boltzmann constant (J/K)</td>
<td>1.38 \times 10^{-23}</td>
</tr>
<tr>
<td>$T_0$</td>
<td>reference temperature (K)</td>
<td>290</td>
</tr>
<tr>
<td>$F_n$</td>
<td>receiver noise figure</td>
<td>2</td>
</tr>
<tr>
<td>$R_{\text{all}}$</td>
<td>range for STC (km)</td>
<td>25</td>
</tr>
<tr>
<td>$s_{0}$</td>
<td>broadside squint angle (deg)</td>
<td>0.9</td>
</tr>
<tr>
<td>$b_{0}$</td>
<td>bearing broadside angle (deg)</td>
<td>0</td>
</tr>
<tr>
<td>$e_{0}$</td>
<td>elevation broadside angle (deg)</td>
<td>15</td>
</tr>
<tr>
<td>$\tau_{e}$</td>
<td>pulse length ((\mu)s)</td>
<td>675</td>
</tr>
</tbody>
</table>

E. Simulation Results

In this section we consider a two-target scenario to illustrate the operation of the ML-PDA estimator under various SNRs. The scenario contains two Swerling III targets with the same average radar cross section (RCS). We assume that the two targets enter the radar acquisition region at \(t = 0\) s with the initial state $r_0^1 = [720.0 \ 6050.0 \ 2180.0]'$ km, $r_0^2 = [718.0 \ 6052.0 \ 2179.5]'$ km, and $r_0^1 = [-0.045 \ -1.531 \ -0.694]'$ km/s, and $r_0^2 = [718.0 \ 6052.0 \ 2179.5]'$ km, $r_0^2 = [-0.045 \ -1.541 \ -0.695]'$ km/s; the radar position is $O_{\text{radar}} = [800 \ 6000 \ 2000]'$ km, in ECEF coordinates. The radar parameters are given in Table VIII. At this point the slant range of the radar to target 1 is 204.1 km and the slant range of radar to target 2 is 203.2 km.
It is assumed that the radar cueing region for initial radar beam pointing is available from prior information. The radar cueing volume is approximately (30 km)
3 and the radar beam packing is 7 by 7 rectangular with each dwell consisting 330 bins of length 90 m. The average output SNR at boresight for each target is 6 dB to 8 dB for the cases considered. Notice that the average SNR in a resolution cell is 4.4 dB when the target SNR at boresight is 6 dB. The detection threshold for acquisition of 6 dB targets is \( \tau = 2.87 \), yielding an average detection probability 0.52. The spatial density of false alarms in the validation gate is dominated by the target-due extraneous detections in the eight adjacent beams when the target is uniformly distributed within the rectangular region centered at the main beam.6 With the parameter settings specified as above we have \( \lambda_{\gamma} = 1.2 \). The radar scan rate is 10 Hz. The total number of scans \( N \) varies from 40 to 60. The scenarios are denoted as \( H_0 \): no target is present; \( H_1 \): only target 1 is present; \( H_2 \): both target 1 and target 2 are present. We are interested in the probability of choosing the correct model under each case. The target SNR at boresight varies from 6 dB to 8 dB. In each case, the detection threshold is chosen to maintain \( \lambda_{\gamma} \leq 1.2 \).

The results from 100 Monte Carlo runs for three scenarios \( (H_0, H_1, H_2) \) are listed in Table IX. When two tracks are initiated, the track to truth association is based on the geometric distance between the true and estimated state. If the distance is greater than three times the measurement standard deviation, the track is declared as false. If we want the probability of choosing the correct model to be greater than 0.95 for

\[ \lambda_{\gamma} = 2 \] and standard deviation \( \sigma_\lambda = 0.032 \), where \( C^0_i \) is the indicator function that the algorithm chooses the correct model in the \( i \)-th run. Assuming \( \Delta \) to be normal, we have more than 99% confidence that the MDL outperforms the GLRT in this case. For details, see [3, pp. 81–82].

![Table IX](image1)

<table>
<thead>
<tr>
<th>N</th>
<th>SNR = 6 dB</th>
<th>SNR = 7 dB</th>
<th>SNR = 8 dB</th>
</tr>
</thead>
<tbody>
<tr>
<td>40</td>
<td>( \tilde{P}(H_0 \mid H_0) )</td>
<td>( \tilde{P}(H_1 \mid H_1) )</td>
<td>( \tilde{P}(H_2 \mid H_2) )</td>
</tr>
<tr>
<td>50</td>
<td>0.96</td>
<td>0.92</td>
<td>0.98</td>
</tr>
<tr>
<td>60</td>
<td>0.98</td>
<td>0.96</td>
<td>0.99</td>
</tr>
</tbody>
</table>

![Table X](image2)

<table>
<thead>
<tr>
<th>N</th>
<th>SNR = 6 dB</th>
<th>SNR = 7 dB</th>
<th>SNR = 8 dB</th>
</tr>
</thead>
<tbody>
<tr>
<td>40</td>
<td>( \tilde{P}(H_0 \mid H_0) )</td>
<td>( \tilde{P}(H_1 \mid H_1) )</td>
<td>( \tilde{P}(H_2 \mid H_2) )</td>
</tr>
<tr>
<td>50</td>
<td>0.93</td>
<td>0.92</td>
<td>0.85</td>
</tr>
<tr>
<td>60</td>
<td>0.98</td>
<td>0.94</td>
<td>0.88</td>
</tr>
</tbody>
</table>

These unavoidable extra detections cannot be combined via clustering with the main beam detection because this leads to a significant degradation of the resulting single measurement. This is due to the fact that the measurements in the adjacent beams (if a target detection occurs there) have a strong tendency of being attracted toward the center of the corresponding adjacent beam, i.e., they are biased. Thus such detections/measurements are treated as false, i.e., lumped with the false alarms.
The collection of points created by varying the parameters of the model gives rise to a hyper-surface in which “similar” distributions are mapped to “nearby” points. The infinitesimal distance (squared) between points separated by the infinitesimal parameter difference \( d\mathbf{x} \) is given by

\[
dx^2 = \sum_{i,j=1}^{n} g_{ij}(\mathbf{x}) d(x_i) d(x_j)
\]

where \( g_{ij}(\mathbf{x}) \) is the Riemannian metric tensor and \( (x_i) \) is the \( i \)th component of the vector \( \mathbf{x} \). The subscript of \( \mathbf{x} \) for model index is dropped for simplicity. If we use Fisher information as the natural metric on a manifold of distributions \([10]\), we have the matrix with elements \( g_{ij} \) as

\[
[g_{ij}] = \lim_{N \to \infty} \frac{1}{N} E(\nabla_x \ln p(\mathbf{Z} | \mathbf{x}) | \nabla_x \ln p(\mathbf{Z} | \mathbf{x})') = I(\mathbf{x})
\]

where \( N \) is the number of data points in \( \mathbf{Z} \). The infinitesimal volume on the manifold parameterized by \( \mathbf{x} \) is \( dV = |I(\mathbf{x})|^{1/2} d\mathbf{x} \) where \( d\mathbf{x} = \prod_{i=1}^{n} d(x_i) \). The Riemannian volume of the manifold is obtained by integrating \( dV \) over the parameter space \( \Omega \):

\[
V_M = \int_{\Omega} dV = \int_{\Omega} |I(\mathbf{x})|^{1/2} d\mathbf{x}.
\]

In other words, the third term in the RHS of (8) penalizes those models that occupy a large volume in the space of probability distributions.

In fact, \( V_M \) is related to the number of “distinguishable” probability distributions indexed by the model \([13]\). Here two probability distributions are considered indistinguishable if the probability that one is mistaken for the other is close to 1 even in the presence of infinite amount of data. It can be shown \([2]\) that the volume of the parameter space which contains indistinguishable distributions is

\[
V_\epsilon(\mathbf{x}) = \frac{\epsilon(2\pi)^{n/2}}{|I(\mathbf{x})|^{1/2}}
\]

V. CONCLUSIONS

In this paper the track initiation problem was formulated as multiple composite hypothesis testing using ML-PDA. The number of tracks was determined based on the MDL criterion. We first reviewed some well-known approaches for statistical model selection and the advantage of using the MDL criterion. Then several one-dimensional examples were used to illustrate the MDL criterion for multiple composite hypothesis testing. Finally, we applied the MDL approach for the detection and initiation of tracks of incoming tactical ballistic missiles in the exo-atmospheric phase using a surface-based electronically scanned array (ESA) radar. The targets were characterized by low SNR, which lead to low detection probability and high false alarm rate.

A batch of radar scans were processed to detect the presence of up to two targets. The ML-PDA estimator was used to initiate the tracks assuming the target trajectories follow a deterministic state propagation. The approximate MDL criterion was used to determine the number of valid tracks in a surveillance region. The detector/estimator was shown to be efficient even at 4.4 dB average SNR within the beam, i.e., in a resolution cell.

APPENDIX A. GEOMETRIC INTERPRETATION OF MDL

Rissanen’s MDL principle was originally interpreted in a two-part code, where the first has all the useful information in the data while the rest has none (see \([15–17]\)). A geometric interpretation of the MDL was later proposed by Myung \([13]\). We feel that the MDL principle can be understood elegantly and intuitively from a geometric as well as the underlying Bayesian perspective. Assume that a parametric model family of probability distributions forms a Riemannian manifold embedded in the space of all probability distributions. Every distribution is a point in this space. The collection of points created

\[\begin{array}{|c|c|c|c|c|c|c|}
\hline
SNR & Parameter of Target 1 & \delta & \sigma_{CRLB} & Parameter of Target 2 & \delta & \sigma_{CRLB} \\
\hline
6 dB & \zeta (m) & 698.8 & 755.6 & \zeta (m) & 720.2 & 754.4 \\
6 dB & \eta (m) & 1012.2 & 860.9 & \eta (m) & 965.3 & 862.2 \\
6 dB & \zeta (m) & 398.2 & 371.7 & \zeta (m) & 408.4 & 381.6 \\
6 dB & \xi (m/s) & 214.3 & 233.6 & \xi (m/s) & 219.6 & 233.3 \\
6 dB & \hat{\eta} (m/s) & 297.8 & 265.5 & \hat{\eta} (m/s) & 317.1 & 265.9 \\
6 dB & \hat{\xi} (m/s) & 138.4 & 114.3 & \hat{\xi} (m/s) & 130.6 & 117.4 \\
\hline
8 dB & \xi (m) & 628.3 & 648.9 & \xi (m) & 650.6 & 647.9 \\
8 dB & \eta (m) & 864.6 & 739.3 & \eta (m) & 901.5 & 740.4 \\
8 dB & \zeta (m) & 310.4 & 319.2 & \zeta (m) & 346.3 & 327.7 \\
8 dB & \xi (m/s) & 209.7 & 200.6 & \xi (m/s) & 210.5 & 200.3 \\
8 dB & \hat{\eta} (m/s) & 267.8 & 228.0 & \hat{\eta} (m/s) & 268.1 & 228.3 \\
8 dB & \hat{\xi} (m/s) & 88.7 & 98.1 & \hat{\xi} (m/s) & 96.9 & 100.8 \\
\hline
\end{array}\]
Thus the negative log of distinguishable distributions decreases as the dimension $n$ of the parameter vector $x$ increases. Let $M$ be the number of all distinguishable distributions in the parameter space. We have

$$M = \int \frac{1}{V(x)} \, dx = \frac{\int |I(x)|^{1/2} \, dx}{\epsilon(2\pi)^{n/2}} = \frac{V_M}{\epsilon(2\pi)^{n/2}}.$$  

(66)

For a data sequence of length $N$, the number of distinguishable distributions is

$$M(N) = \left( \frac{N}{2\pi} \right) n/2 \frac{V_M}{\epsilon}.$$  

(67)

If we measure model complexity in terms of $\ln[\epsilon M(N)]$, we have

$$\ln[\epsilon M(N)] = \frac{n}{2} \ln \left( \frac{N}{2\pi} \right) + \ln \int |I(x)|^{1/2} \, dx$$  

(68)

which coincides with the penalty terms in the MDL. Thus the MDL criterion measures the number of distinguishable distributions based on the observed data. To see this, we take a Bayesian perspective on the second term in the RHS of (8). Let $f$ be the functional relationship between the parameter $x$ and the observed data $Z$. Using Bayes rule, the probability that the data originated from $f$ can be written as

$$P(f \mid Z) = \frac{P(f)}{P(Z)} \int f(Z \mid x) p(x) \, dx.$$  

(69)

where $p(x)$ is the prior density of the parameter $x$. If we assume the prior probabilities $P(f)$ and $P(Z)$ are noninformative among the model and data spaces, we are only interested in the integral given by

$$m(Z) = \int f(Z \mid x) p(x) \, dx.$$  

(70)

Lacking prior knowledge of the model, the prior density should be chosen to weight all distinguishable distributions in the family $f$ equally. We let $p(x) = 1/V_M$. For large sample sizes, the likelihood function $f(Z \mid \hat{x})$ localizes to a multivariate normal distribution centered at the MLE given by

$$\hat{x} = \arg\max_x f(Z \mid x).$$  

(71)

Thus the negative log of $m(Z)$ is

$$-\ln m(Z) = -\ln f(Z \mid \hat{x}) + \ln \left( \frac{V_M}{C_M} \right) + O \left( \frac{1}{N} \right)$$  

(72)

where

$$C_M = \left( \frac{2\pi}{N} \right)^{n/2} h(\hat{x})$$  

(73)

and $h(\hat{x})$ is a data dependent factor that goes to 1 for $N \to \infty$ if the data originate following the

function $f$. Here $C_M$ measures the number of distinguishable distributions within the family $f$ that lie close to the true data generating mechanism [2].

Another way to interpret the MDL is by considering $f(Z \mid \hat{x})V_M/C_M$ as the NML and the code length for $m(Z)$ is approximately

$$\text{MDL} = -\ln m(Z) \approx -\ln \left( \frac{f(Z \mid \hat{x})}{V_M/C_M} \right).$$  

(74)

The geometric meaning of the complexity penalty of the MDL now becomes clear: models having distinguishable distributions which occupy a relatively large volume distant from the truth are penalized while a relatively large fraction of the distinguishable distributions lying close to the truth are preferred. From this point of view, a better model is simply the one with more distinguishable distributions close to the truth but less distinguishable distributions in all.

APPENDIX B. ALGORITHM FOR STATE PROPAGATION

Presented below is an algorithm that propagates the state of an object in a ballistic trajectory around the Earth. Let $x(t) = [x(t) \ y(t) \ z(t)]$ be the unknown state to be computed at time $t$, given the state $x_0 = x(t_0) = [x_0, y_0, z_0]$ at the time $t_0$. The underlying theoretical concepts and the derivation of the equations used in this algorithm can be found in [5]. The gravitational parameter $\mu = 3.986012 \times 10^5$ km$^3$/s$^2$ and the convergence check parameter $TOL = 10^{-10}$ are used.

Step 1

$$r_0 := ||r_0||, \quad v_0 := ||v_0||, \quad q_0 := \frac{1}{\mu}r_0 v_0$$

$$a_0 := \frac{2}{r_0} - \frac{v_0^2}{\mu}, \quad p_0 := \frac{1 - a_0 r_0}{\sqrt{\mu}}$$

Step 2

$$\alpha := \frac{a_0(t - t_0)}{\sqrt{\mu}}, \quad \beta := a_0 \alpha^2$$

Step 3

$$c := \frac{1 - \cos(\sqrt{\beta})}{\beta}, \quad s := \sqrt{\beta - \sin(\sqrt{\beta})}$$

Step 4

$$\tau := p_0 \alpha^3 s + q_0 \alpha^2 c + \frac{r_0}{\sqrt{\mu}} \alpha$$

$$\frac{d\tau}{d\alpha} := p_0 \alpha^2 c + q_0 \alpha (1 - s_\beta) + \frac{r_0}{\sqrt{\mu}}$$

$$\alpha := \alpha + \left[ \frac{d\tau}{d\alpha} \right]^{-1} [t(t_0) - \tau]$$

Step 5

if $|(t(t_0) - \tau) > TOL$

goto STEP 3
\( f := 1 - \frac{\alpha^2 c}{r_0} \) \hspace{1cm} \( g := (t - t_0) - \frac{\alpha^3 s}{\sqrt{r}} \)

\( \mathbf{r}(t) := f \mathbf{r}_0 + g \mathbf{r}_0, \) \hspace{1cm} \( r := \| \mathbf{r}(t) \| \)

**Step 7**

\( f := \left( \frac{\sqrt{r}}{r_0} \right) (s\beta - 1) \left( \frac{\alpha}{r} \right) \)

\( g := 1 - \frac{\alpha^2 c}{r} \)

\( \dot{\mathbf{r}}(t) := \dot{f} \mathbf{r}_0 + \dot{g} \mathbf{r}_0 \)

The above steps yield the required state \( x(t)' = [\mathbf{r}(t)' \mathbf{\dot{r}}(t)'] \) at time \( t \). In order to predict the covariance of the position \( \mathbf{r}(t) \) from \( t_0 \) to \( t \), we need to compute the \( 6 \times 6 \) matrix \( \nabla_x \mathbf{r}(t) \). The computation of this matrix involves the following additional steps.

**Step 8**

\[ \nabla_x \mathbf{r}_0 := \left[ \begin{array}{c} \mathbf{r}_0 \\ 0 \end{array} \right], \quad \nabla_x v_0 := \left[ \begin{array}{c} 0 \\ \mathbf{v}_0 \end{array} \right] \]

\[ \nabla_x q_0 := \left[ \begin{array}{c} \mathbf{r}_0 \\ \mu \end{array} \right], \quad \nabla_x a_0 := (\nabla_x \mathbf{r}_0) \left( \frac{-2}{r_0^2} \right) + (\nabla_x \mathbf{v}_0) \left( \frac{-2v_0}{\mu} \right) \]

\[ \nabla_x p_0 := (\nabla_x \mathbf{r}_0) \left( \frac{-a_0}{\mu^2} \right) + (\nabla_x \mathbf{a}_0) \left( \frac{-p_0}{\mu^2} \right) \]

**Step 9**

\[ \frac{ds}{d\beta} = \frac{c - 3s}{2\beta}, \quad \frac{dc}{d\beta} = \frac{1 - s - 3c}{2\beta} \]

**Step 10**

\[ b_1 := (\nabla_x q_0)(-\alpha^2 c) + (\nabla_x p_0)(-\alpha^3 s) + (\nabla_x r_0) \left( -\frac{\alpha}{\sqrt{r}} \right) \]

\[ b_2 := (\nabla_x a_0)(-\alpha^2) \]

\[ A := \left[ \begin{array}{ccc} 3p_0\alpha - 2q_0\alpha + \frac{r_0}{\sqrt{r}} & p_0\alpha^3 \frac{ds}{d\beta} + q_0\alpha^2 \frac{dc}{d\beta} \\ 2a_0\alpha & 1 \end{array} \right] \]

\[ l(\nabla_x \alpha) (\nabla_x \beta) := [b_1, b_2] A^{-1} \]

**Step 11**

\[ \nabla_x f := \left( \nabla_x f_0 \right) \left( \frac{\alpha c}{r_0} \right) - (\nabla_x \alpha)(2c) - (\nabla_x \beta) \left( \frac{\alpha}{\sqrt{r}} \right) \]

\[ \nabla_x g := \left( \nabla_x \alpha \right)(3\alpha) - (\nabla_x \beta) \left( \frac{dx}{d\beta} \right) \left( \frac{-\alpha^2}{\sqrt{r}} \right) \]

**Step 12**

\[ \nabla_x \mathbf{r}(t) = \left[ \begin{array}{c} fI_3 \\ gI_3 \end{array} \right] + (\nabla_x f) \mathbf{r}_0' + (\nabla_x g) \mathbf{r}_0' \]

**APPENDIX C. CONVERSION OF COORDINATE SYSTEMS [5]**

Given the position \( \mathbf{r} = [\xi \eta \zeta]' \) in the ECEF frame, the latitude \( \varphi \), longitude \( \lambda \) and altitude \( h \) (which are the three components of \( \mathbf{r}_{geo} \)) are determined by

\[ \varphi = 0 \]

\[ \varphi_{old} = \varphi \]

\[ D_{\varphi} = R_e [1 - c \sin^2 \varphi_{old}]^{\frac{1}{2}} \]

\[ \varphi = \atan \left( \frac{\zeta + D_{\varphi} \sin \varphi_{old}}{\sqrt{\xi^2 + \eta^2}} \right) \]

until \( |\varphi - \varphi_{old}| < TOL \)

\[ \lambda = \atan \left( \frac{\eta}{\zeta} \right) \]

\[ h = \frac{\zeta}{\sin \varphi} - D_{\varphi}(1 - \epsilon_e^2) \]

where \( R_e = 6378.137 \text{ km} \) and \( \epsilon_e = 0.0818191 \) are the equatorial radius and eccentricity of the Earth, respectively. \( TOL \) is the error tolerance (e.g., \( 10^{-10} \)) and the convergence occurs normally within 10 iterations. The origin of the local Cartesian frame \( \mathbf{O} \) is given by

\[ \mathbf{O} = [0 \ D_{\varphi} \epsilon_e \sin \varphi \cos \psi \ D_{\varphi}(\epsilon_e^2 \sin^2 \varphi - 1)]' \]

and the rotation matrix is given by

\[ A = \left[ \begin{array}{ccc} -\sin \lambda & \cos \lambda & 0 \\ -\sin \varphi \cos \lambda & -\sin \varphi \sin \lambda & \cos \varphi \\ \cos \varphi \cos \lambda & \cos \varphi \sin \lambda & \sin \varphi \end{array} \right] \]

The position \( \mathbf{r} \) in the local Cartesian frame is given by

\[ \mathbf{r}_{loc} = A \mathbf{r} + \mathbf{O} \]

**APPENDIX D. RADAR MODEL [22]**

In the development of the problem a monopulse phased array radar with uniform illumination across the array is emulated [6]. Each radar dwell consists of one phase/frequency discrete coded pulse. The radar beam is quasi-circular with the 3 dB beamwidth, \( \theta_{bw} \), increasing as the beam is steered off the broadside direction.

The basic radar equation for the received power from a target at range \( R \) using a single pulse is given by

\[ P_r = \frac{P_G \sigma_0 \lambda^2 F_I^2 F_r^2 \sigma_0}{(4\pi)^3 L_{tot}} \left( \frac{G_{stc}(R)}{R^4} \right) \]

where the sensitivity time control (STC) gain is defined as

\[ G_{stc}(R) = \begin{cases} 1 & R \geq R_{stc} \\ \left( \frac{R}{R_{stc}} \right)^3 & R < R_{stc} \end{cases} \]

and \( R_{stc} = 25 \text{ km} \). The SNR\(^8\) of the output of the matched filter is given by

\(^8\text{In fact, the receiver observes SNR when a target is present.}\)
The meaning of all the variables at the right hand side (RHS) of (80) with their values used in simulations are listed in Table VIII.

For a radar with given power $P_t$, the SNR at the output of the matched filter can be increased by extending the length of the radar pulse. Pulse compression is used in order to achieve good range resolution with a long pulse. In discrete-coded pulse compression, the pulsewidth $\tau_p$ is composed of $N_\ell$ subpulses of width $\tau_s$ where each subpulse is coded with frequency and/or phase. The range resolution of the output of the matched filter can be increased where the pdf of the RCS is given by

$$p(\sigma_0 | \sigma_{\text{ave}}) = \frac{4\sigma_0}{\sigma_{\text{ave}}} \exp\left(-\frac{2\sigma_0}{\sigma_{\text{ave}}}\right)$$

with the average RCS $\sigma_{\text{ave}}$ varying among the scenarios. Assuming a long range target with RCS of 0.15 m$^2$ and the radar parameters in Table VIII, the output SNR for the compressed pulse is given by

$$\mathcal{R}_0 = \frac{\tau_p \sigma_0 \Omega_0}{R^4}$$

(83)

where

$$\Omega_0 = \frac{PG_G r^2 F_i^2 F_r^2}{(4\pi)^2 L_{\text{tot}} L_0 T_0 F_n} = 257.94 \text{ dB}. \quad (84)$$

Thus when $\tau_s = 675 \mu$s, the output SNR is 6 dB for a target with RCS of 0.15 m$^2$ and at the range of 200 km.

The range is computed as the range bin value that is closest to the true range. The discrete coding results in range measurements with errors that are uniformly distributed within the range bin. The phase array radar consists of $N^2$ individual elements with cosine illumination. The broadside of the array is directed at the bearing angle of $b_0$ and elevation angle of $e_0$. The received sum voltage $S_k$ at time $t_k$ is described by the in-phase and quadrature phase voltages. In the case of a noncoherent detection, the noisy in-phase and quadrature components of the sum voltage normalized to the receiver noise are

$$S_{ik} = \Gamma_k (\Sigma_k)^2 \cos \phi_k + n_{i,Q}$$

(85)

$$S_{Qk} = \Gamma_k (\Sigma_k)^2 \sin \phi_k + n_{Q,Q}$$

(86)

where $n_{i,Q}$ and $n_{Q,Q}$ are zero-mean unity-variance normal random variables. The normalized amplitude of the received signal $\Gamma_k$ can be expressed as

$$\Gamma_k = \frac{\kappa_0}{R^4} \sqrt{\sigma_0 G_{\text{sec}}(R)}$$

(87)

where

$$\kappa_0 = \sqrt{\frac{2PG_G r^2 F_i^2 F_r^2}{(4\pi)^2 L_{\text{tot}} L_0 T_0 F_n}} \quad (88)$$

is a constant. For the radar parameters given in Table VIII, we have $\kappa_0 = 114.60 \text{ dB}$. The resultant two dimensional normalized radiation pattern $\Sigma_k$ due to simultaneous lobing is given by

$$\Sigma_k = \Phi_k(sq_e, sq_b) + \Phi_k(-sq_e, sq_b) + \Phi_k(sq_e, -sq_b) + \Phi_k(-sq_e, -sq_b)$$

(89)

where

$$\Phi_k(x, y) = V_k(\epsilon(x_k), e_k + x) V_k(\beta(x_k), b_k + y)$$

(90)

and $sq_b$, $sq_e$ are the squint angles in bearing and elevation, which are given by

$$sq_b = sqb_0 / \cos(b_k - b_0)$$

(91)

$$sq_e = sqe_0 / \cos(e_k - e_0)$$

(92)

The angles $b_k$ and $e_k$ denotes the bearing and elevation of the radar pointing direction (boresight); $\beta(x_k)$ and $\epsilon(x_k)$ are the true target bearing and elevation at time $t_k$. The normalized radiation patterns in bearing $V_k(\beta(x_k), b_k)$ and elevation $V_k(\epsilon(x_k), e_k)$ are given by

$$V_k(\beta(x_k), b_k) = \frac{\pi}{4} \sin(N b_k) \sqrt{\frac{\sin(b_k + \frac{\pi}{2})}{b_k + \frac{\pi}{2}}} \sqrt{\frac{\sin(b_k - \frac{\pi}{2})}{b_k - \frac{\pi}{2}}}$$

(93)

$$V_k(\epsilon(x_k), e_k) = \frac{\pi}{4} \sin(N e_k) \sqrt{\frac{\sin(e_k + \frac{\pi}{2})}{e_k + \frac{\pi}{2}}} \sqrt{\frac{\sin(e_k - \frac{\pi}{2})}{e_k - \frac{\pi}{2}}}$$

(94)

where

$$b_1 = \frac{\pi}{4} \sin(\beta(x_k) - b_0)$$

(96)

$$b_2 = \frac{\pi}{2} \sin(\beta(x_k) - b_0) - \sin(b_k - b_0)$$

(97)

$$e_1 = \frac{\pi}{4} \sin(\epsilon(x_k) - e_0)$$

(98)

$$e_2 = \frac{\pi}{2} \sin(\epsilon(x_k) - e_0) - \sin(e_k - e_0)$$

(99)

The in-phase and quadrature difference voltages for bearing and elevation are given by

$$D_{ik} = \Gamma_k \Sigma_k \Omega_k^b \cos \phi_k + n_{i,bl}$$

(100)

$$D_{Qik} = \Gamma_k \Sigma_k \Omega_k^b \sin \phi_k + n_{i,Ql}$$

(101)

$$D_{ik} = \Gamma_k \Sigma_k \Omega_k^Q \cos \phi_k + n_{Q,bl}$$

(102)

$$D_{Qik} = \Gamma_k \Sigma_k \Omega_k^Q \sin \phi_k + n_{Q,Ql}$$

(103)
where \( n_{bl}, n_{bo}, n_{cl} \) and \( n_{e0} \) are zero-mean unity-variance normal random variables and

\[
\Omega^b_k = \Phi_k(s_q, s_q) + \Phi_k(-s_q, s_q) - \Phi_k(s_q, -s_q)
\]

\[
\Omega^e_k = \Phi_k(s_q, s_q) - \Phi_k(-s_q, s_q) + \Phi_k(s_q, -s_q)
\]

(104)

(105)

APPENDIX E. PROBABILITY DENSITY FUNCTION OF SNR

For a Swerling III fluctuating target with average RCS \( \sigma_{ave} \) at range \( R \), the conditional pdf is

\[
p(\Gamma_k | \mathbf{x}_k) = \frac{8\Gamma^3_k}{\alpha_k^2} \exp\left(-\frac{2\Gamma^2_k}{\alpha_k}\right)
\]

(106)

where

\[
\alpha_k = \frac{k^2 G_{ae}(R)}{R^4} \sigma_{ave}.
\]

The average amplitude is given by

\[
\Gamma_{ave} = \frac{3}{4} \sqrt{\frac{\pi}{2} \alpha_k}.
\]

(108)

The conditional densities of the in-phase and quadrature components of the sum voltage are

\[
p(S_{ik} | \Gamma_k, \mathbf{x}_k) = N(\Gamma_k \Sigma^2_k \cos \phi_k, 1)
\]

(109)

\[
p(S_{Qk} | \Gamma_k, \mathbf{x}_k) = N(\Gamma_k \Sigma^2_k \sin \phi_k, 1).
\]

(110)

Let the SNNR be

\[
\mathbb{R}_k = \frac{1}{2}(S_{ik}^2 + S_{Qk}^2)
\]

(111)

which represents the target amplitude information at the receiver. The pdf of \( \mathbb{R}_k \) in the presence of a target is

\[
p(\mathbb{R}_k | \Gamma_k, \mathbf{x}_k) = \exp\left(-\frac{2\mathbb{R}_k + \Gamma^2_k \Sigma^4_k}{2}\right) I_0(\Gamma_k \Sigma^2_k \sqrt{2\mathbb{R}_k})
\]

(112)

where \( I_0(\cdot) \) is the modified Bessel function of the first kind of order zero. According to [21], the conditional pdf of SNNR is given by

\[
p(\mathbb{R}_k | \mathbf{x}_k) = \int_0^\infty p(\mathbb{R}_k | \Gamma_k, \mathbf{x}_k)p(\Gamma_k | \mathbf{x}_k)d\Gamma_k
\]

\[
= \frac{16}{(4 + \alpha_k \Sigma^4_k)^2} \left(1 + \frac{\alpha_k \Sigma^4_k}{4 + \alpha_k \Sigma^4_k} \mathbb{R}_k\right)
\]

\[
\times \exp\left(-\frac{4}{4 + \alpha_k \Sigma^4_k} \mathbb{R}_k\right).
\]

(113)

The expected value of \( \mathbb{R}_k \) given the target state \( \mathbf{x}_k \) is \( 1 + \alpha_k \Sigma^4_k / 2 \). Given the detection threshold \( \tau \), the probability of target detection by one radar beam is

\[
P_D = \int_{\tau}^{\infty} p(\mathbb{R}_k | \mathbf{x}_k)d\mathbb{R}_k
\]

(114)

The pdf of the amplitude when the measurement is false (due to noise only) is

\[
p_0(\mathbb{R}_k) = \exp(-\mathbb{R}_k).
\]

(115)

APPENDIX F. MONOPULSE PROCESSING [6]

In a typical monopulse system, the angle of arrival (AOA) with respect to the antenna boresight is approximated by

\[
\theta \approx \frac{\theta_{bw}}{R_m} \eta
\]

(116)

where \( \theta_{bw} \) is the 3 dB beamwidth of the antenna pattern; \( \eta \), the monopulse ratio, contains the angle information for bearing \( (\eta^b = \Omega^b_k / \Sigma_k) \) and elevation \( (\eta^e = \Omega^e_k / \Sigma_k) \) at time \( t_k \), which can be estimated independently. Denote the in-phase and quadrature parts of the monopulse ratio as

\[
y^b_{ik} = \frac{D^b_{ik} S^b_{ik} + D^b_{Qk} S^b_{Qk}}{S^b_{ik} + S^b_{Qk}}
\]

(117)

\[
y^e_{Qk} = \frac{D^b_{Qk} S^b_{ik} - D^b_{Qk} S^b_{Qk}}{S^b_{ik} + S^b_{Qk}}
\]

(118)

\[
y^e_{Qk} = \frac{D^e_{ik} S^e_{ik} + D^e_{ik} S^e_{Qk}}{S^e_{ik} + S^e_{Qk}}
\]

(119)

\[
y^e_{Qk} = \frac{D^e_{ik} S^e_{ik} - D^e_{ik} S^e_{Qk}}{S^e_{ik} + S^e_{Qk}}.
\]

(120)

Typically, the angle estimates of bearing and elevation are obtained only with the in-phase components given by

\[
\theta^b_k = b_k + \frac{1}{k_m^b} y^b_{ik}
\]

(121)

\[
\theta^e_k = e_k + \frac{1}{k_m^e} y^e_{ik}
\]

(122)

where

\[k_m^b = 45.1 \cos(b_k - b_0)\]

(123)

\[k_m^e = 45.1 \cos(e_k - e_0)\]

(124)

are obtained from the average monopulse error slope with the radar parameters given in Table VIII by linear approximation. The above angle estimates are valid for \( \mathbb{R}_k > 3 \text{ dB} \) within 0.74° off boresight with the standard deviations given below

\[
\sigma^b_k = \frac{1}{k_m^b \sqrt{2 \mathbb{R}_k}} \left(1 + \frac{1}{2 \mathbb{R}_k}\right) \sqrt{1 + \frac{(y^b_{ik})^2}{2 \mathbb{R}_k + 1}}
\]

(125)

\[
\sigma^e_k = \frac{1}{k_m^e \sqrt{2 \mathbb{R}_k}} \left(1 + \frac{1}{2 \mathbb{R}_k}\right) \sqrt{1 + \frac{(y^e_{ik})^2}{2 \mathbb{R}_k + 1}}
\]

(126)
APPENDIX G. RADAR OPERATION

In a typical missile acquisition scenario the radar performs a volumetric search over the region using a number of dwells. The dwells are arranged into a rectangular lattice of rows and columns. The beam packing configurations are similar as presented in [21]. Assuming each dwell covers 3 dB beamwidth ($\theta_{bw} = 1.6^\circ$) in both bearing and elevation, the number of dwells required to cover the 30 km$^3$ cureing region in width without uncovered area is 49 with 1.23$^\circ$ angle change between the adjacent beams. The number of range bins to cover the region in depth is 330 when $\Delta_r = 90$ m.

Denote the number of resolution cells as $n$ (in this case we have $n = 49 \times 330 = 16170$ “pancake” shaped cells). The receiver uses monopulse processing to obtain the measurement of target position for each resolution cell. Given the detection threshold $\tau$ for each cell, the probability of false alarm is

$$P_{FA} = e^{-\tau}.$$  \hspace{1cm} (127)

We consider the detection of a target (range = 200 km, SNR = 6 dB at boresight) in one scan as following: if the bearing and elevation between the target and the radar boresight are within 0.74$^\circ$ and the measured SNNR exceeds detection threshold, the measurement obtained using monopulse processing is taken as a valid measurement. Notice that the adjacent radar beams may have multiple detections even when the target is near the main beam. The measurements obtained using monopulse processing are consistent only when the true target position is within 0.74$^\circ$ off boresight. Clearly, the probability of detection depends on the true target position.

Denote $P_{D_{\min}}$ as the minimum detection probability of a target inside the cueing region. With the radar parameters given in Table VIII and the rectangular beam packing, the position having minimum detection probability is off boresight by 0.62$^\circ$ in both bearing and elevation. In this case, we have $P_{D_{\min}} = 0.297$ when $\tau = 3.25$ for target SNR = 6 dB at boresight.

Comparing with the detection probability when the target is exactly at the radar boresight ($P_{D_{\max}} = 0.568$, ignoring the possible detections from adjacent beams), the difference between $P_{D_{\min}}$ and $P_{D_{\max}}$ is around 0.3. Assuming the target position is uniformly distributed within the rectangular region covered by the main beam with the center at radar boresight, we have the average detection probability $P_{D_{\text{ave}}} = 0.464$.

REFERENCES

A new look at the statistical model identification.

A geometric formulation of Occam’s razor for inference of parametric distributions.

Estimation with Applications to Tracking and Navigation: Theory, Algorithms and Software.

Multisensor Multitarget Tracking: Principles and Techniques.

Fundamentals of Astrodynamics.

Monopulse processing for tracking unresolved targets.

On the information-based measure of covariance complexity and its application to the evaluation of multivariate linear models.

Multiple low observable ballistic missile track initiation using an ESA radar.

An adaptive ML/PDA estimator for imaging sensors with heavy clutter.

Geometrical Foundations of Asymptotic Inference.


IMMPDAF for radar management and tracking benchmark with ECM.

Counting probability distributions: Differential geometry and model selection.

Stochastic complexity and the MDL principle.

‘Fisher information and stochastic complexity.

Hypothesis selection and testing by the MDL principle.

Strong optimality of the normalized ML models as universal codes and information in data.
IEEE Transactions on Information Theory, 47 (July 2001), 1712–1717.


Huimin Chen received the B.E. and M.E. degrees from Department of Automation, Tsinghua University, Beijing, China, in 1996 and 1998, respectively, and the Ph.D. degree from the Department of Electrical and Computer Engineering, University of Connecticut, Storrs, in 2002, all in electrical engineering.

He was a research post-doc at Physics and Astronomy Department, University of California, Los Angeles, and a visiting researcher with the Electrical and Computer Engineering Department, Carnegie Mellon University from July 2002. He has been an assistant professor with the Department of Electrical Engineering, University of New Orleans since Jan. 2003. His research interests are in signal processing, estimation theory, target detection and target tracking.

Yaakov Bar-Shalom (S’63—M’66—SM’80—F’84) was born on May 11, 1941. He received the B.S. and M.S. degrees from the Technion, Israel Institute of Technology, in 1963 and 1967 and the Ph.D. degree from Princeton University, Princeton, NJ, in 1970, all in electrical engineering.

From 1970 to 1976 he was with Systems Control, Inc., Palo Alto, CA. Currently he is Board of Trustees Distinguished Professor in the Dept. of Electrical and Computer Engineering and director of the ESP Lab (Estimation and Signal Processing) at the University of Connecticut. His research interests are in estimation theory and stochastic adaptive control and he has published over 280 papers in these areas. In view of the causality principle between the given name of a person (in this case, “(he) will track,” in the modern version of the original language of the Bible) and the profession of this person, his interests have focused on tracking.


He was program chairman of the 1982 American Control Conference, general chairman of the 1985 ACC, and cochairman of the 1989 IEEE International Conference on Control and Applications. During 1983–1987 he served as chairman of the Conference Activities Board of the IEEE Control Systems Society and during 1987–1989 was a member of the Board of Governors of the IEEE CSS. Currently he is a member of the Board of Directors of the International Society of Information Fusion and served as Y2K and Y2K2 President. In 1987 he received the IEEE CSS distinguished Member Award. Since 1995 he is a distinguished lecturer of the IEEE AESS. He is corecipient of the M. Barry Carlton Awards for the best paper in the IEEE Transactions on Aerospace and Electronic Systems in 1995 and 2000, and received the 1998 University of Connecticut AAUP Excellence Award for Research.
Krishna R. Pattipati (S’77—M’80—SM’91—F’95) received the B.Tech degree in electrical engineering with highest honors from the Indian Institute of Technology, Kharagpur, in 1975, and the M.S. and Ph.D. degrees in systems engineering from the University of Connecticut, Storrs, in 1977 and 1980, respectively.

From 1980–1986 he was employed by ALPHATECH, Inc., Burlington, MA, where he supervised and performed research on fault-tolerant computer architecture modeling, human decision modeling, multi-target tracking, queuing networks, systems testing and diagnosis, and large-scale mixed-integer optimization. Since 1986, he has been with the University of Connecticut, where he is a professor of Electrical and Systems Engineering.

Dr. Pattipati has published over 190 articles, primarily in the application of systems theory and optimization (continuous and discrete) techniques to large-scale systems. He has served as a consultant to ALPHATECH, Inc., IBM Research and Development, and Qualtech Systems, Inc. He was selected by the IEEE Systems, Man, and Cybernetics Society as the Outstanding Younger Engineer of 1984, and received the Centennial Key to the Future award. He was elected a Fellow of the IEEE in 1995 for his contributions to discrete-optimization algorithms for large-scale systems and team decision making. He won the best technical paper awards at the 1985, 1990, and 1994 IEEE AUTOTEST Conferences, and at the 1997 Command and Control Conference. Dr. Pattipati is currently Editor of the IEEE Transactions on SMC—Cybernetics (Part B), and is the Vice-President for Technical Activities of the IEE SMC Society.

Thiagalingam Kirubarajan (S’95—M’98) was born in Sri Lanka in 1969. He received the B.A. and M.A. degrees in electrical and information engineering from Cambridge University, England, in 1991 and 1993, respectively, and the M.S. and Ph.D. degrees in electrical engineering from the University of Connecticut, Storrs, in 1995 and 1998, respectively.

While in England, he worked for the Central Electricity Research Laboratories, Leatherhead, Surrey, as a research assistant. From 1991 to 1993 he was a lecturer in Electrical and Computer Engineering at the University of Peradeniya, Sri Lanka. Between 1993 and 1998 he was a research/teaching assistant at the University of Connecticut.

Dr. Kirubarajan’s research interests are in estimation, target tracking, signal/image processing, signal detection and fault diagnosis. He has published about 50 articles in these research areas. He has also served as a consultant to a number of companies, including Motorola Corporation, Northrop-Grumman Corporation, Pacific-Sierra Research Corporation, Qualtech Systems, Inc., and Orincon Corporation. He has worked on the development a number of commercially available engineering software programs, including BEARDAT for target localization from bearing and frequency measurements in clutter, FUSEDAT for fusion of possibly heterogeneous multisensor data for tracking, and MTI* for large-scale ground target tracking. He has also worked with Qualtech Systems, Inc., to develop advanced fault diagnosis techniques for the TEAMS-RT fault diagnosis engine. Currently, Dr. Kirubarajan is serving as an associate editor for the IEEE Transactions on Systems, Man and Cybernetics (Part B). From 1998 to 2001 he was assistant research professor in Electrical Engineering and the associate director of the Estimation and Signal Processing Research Laboratory at the University of Connecticut. Presently he is assistant professor of ELE at McMaster University, Hamilton, Ontario, Canada.