

Epistemic Decision Theory Applied to Multiple-Target Tracking

T. K. Moon, *Member, IEEE*, S. E. Budge, W. C. Stirling, *Member, IEEE*, and J. B. Thompson

Abstract—A decision philosophy that seeks the avoidance of error by trading off belief of truth and value of information is applied to the problem of recognizing tracks from multiple targets (MTT). A successful MTT methodology should be robust in that its performance degrades gracefully as the conditions of the collection become less favorable to optimal operation. By stressing the avoidance, rather than the explicit minimization, of error, we obtain a decision rule for trajectory-data association that does not require the resolution of all conflicting hypotheses when the database does not contain sufficient information to do so reliably. This rule, coupled with a set-valued Kalman filter for trajectory estimation, results in a methodology that does not attempt to extract more information from the database than it contains.

I. INTRODUCTION

MULTIPLE Target Tracking (MTT) is a time-varying joint decision and estimation problem consisting of a rule to make decisions concerning the association of sensor outputs and target vehicle trajectories; and an estimator to incorporate the target dynamics model, the sensor model, and the collection geometry to calculate the target trajectory based upon the associated sensor outputs. As such, it provides a paradigm for other time-varying decision problems. The solution to the MTT problem requires the extraction from the data of as much information as possible about the number of targets and the trajectory of each. It is desirable to distinguish reliably all targets of interest from background noise, and to associate accurately each target with the available data.

A reliable MTT procedure must be able to deal with a wide range of scenarios. Targets may be diverging, merging, and maneuvering, background clutter may be present, and there may be an unknown (and changing) number of targets. A sufficient condition for bounded estimation error covariance for single-target tracking is observability [1, page 242]. For the multiple target tracking problem, however, joint observability may not be enough to ensure that the target-data association problem can be solved reliably in the sense of achieving uniformly small decision error probabilities and bounded estimation errors. A theoretical analysis yielding sufficient conditions for acceptable performance has not been developed; the difficulty of such analysis being due, most

likely, to the coupled nature of the joint decision-estimation problem.

Many MTT methods reported in the literature appear to perform satisfactorily, however, under favorable circumstances. For example, they work well if the targets are sufficiently separated spatially, if maneuvers can be adequately modeled, and if the trajectories can be initialized reliably. But the success of any track-association methodology whose design implicitly assumes robust collection circumstances is problematic. With many collection scenarios, there may simply not be sufficient reliable data to guarantee the desired performance. Thus, care must be taken to ensure that the analysis methodology does not attempt to extract more information from the database than it contains.

A standard approach to the MTT problems is to optimize the decision rule and the estimation rule separately, then attempt to merge them into a global solution [2]–[6]. In [7], an attempt is made to unify the two components of the problem by casting it as a systems identification problem. These approaches all invoke classical decision rules, however, such as maximum likelihood and Bayesian methods, to choose the track-data association that meets an optimality criterion. Also, they usually employ a point-valued estimator to calculate the optimal trajectory estimate for each track. They are designed to resolve all conflicts between possible trajectory-data association decisions, and to provide trajectory estimates that are (at least theoretically, when viewed in isolation from the decision problem) the very best possible.

Trajectory-data associations must be made even at the risk of a possibly large probability of error. Conventional measures of trajectory estimation error such as covariance analysis, while perhaps useful for assessing precision, cannot be used reliably to assess the accuracy of the decision-estimation solution because they do not properly account for association decision errors. Although such methods may apparently work well under ideal circumstances, their performance may not degrade gracefully when the conditions of the collection become less favorable to the performance of the algorithms.

We present an approach to the multiple target tracking problem that is based largely upon an epistemology developed by Levi [8], whose contributions have led to the development of a shift in the philosophy of decision making. This approach stresses the avoidance of error, rather than the explicit minimization of error. In other words, if insufficient information is available to make a unique “best” decision, we will not attempt to do so; rather, we will eliminate as many hypotheses as possible, but not insist at all but one be removed

Manuscript received September 5, 1992; revised May 3, 1993. This work was supported in part by ESL, Inc., Sunnyvale, CA.

T. K. Moon and S. E. Budge are with the Electrical Engineering Department, Utah State University, Logan, UT.

W. C. Stirling is with the Electrical and Computer Engineering Department, Brigham Young University, Provo, UT.

J. B. Thompson is with the Naval Ocean Systems Center, San Diego, CA. IEEE Log Number 9214537.

from consideration. To implement such a philosophy, we must be able to deal with unresolved conflict. Consequently, we must search for a new approach to the trajectory-data association problem, and a new technique to address the trajectory estimation issue.

The result of this search is a trajectory-data association decision rule expressed in terms of two probabilities: one governing the informational value of the association hypothesis, and one governing the subjective belief, or credal, probability of the association hypothesis [9]. With this methodology, a criterion of *serious possibility* is defined, and all trajectory-data associations that are seriously possible are retained; it is not necessary to resolve all conflicting hypotheses before processing more data.

The trajectory estimation problem is addressed by introducing a *set-valued* estimator, rather than the conventional point-valued estimator, to describe the evolution of the vehicle trajectory. The set-valued estimator is based upon the set-valued Kalman filter [10], which computes a convex set of trajectories, all with equal claims to validity, given the observations. If a trajectory is completely observable, the radius of the convex set decreases to zero as the quantity of data increases, and the set-valued estimate asymptotically becomes point-valued. Under less favorable circumstances, the radius of the set remains finite and may even grow, thus providing a more comprehensive characterization of the trajectory than does a single point estimate.

The output of the set-valued multiple target tracking methodology is a family of convex sets (in position-velocity space) of trajectories, which we shall refer to as *tracks*, or *track sets*. As more data are obtained, observable track sets tend to decrease in radius. This provides a useful method of assessing the credibility of a given track-data association hypothesis. Additionally, the decision rule permits the furcation of track sets into separate tracks as well as the merging of separate tracks into a single track. The resulting track estimation/track association algorithm has several useful features, including the incorporation of the phases of track initialization, track confirmation, track spawning, track merging, and track deletion into a single, unified methodology. In a companion paper [11] we extend the methods of this paper to target classification in the multiple-target environment.

II. METHODOLOGY

A. Decision Theory

We summarize some of the key features of convex Bayes decision theory [8], [9] used in this analysis. For an inquiry under investigation, suppose there are finitely many hypotheses that may be considered. Let U denote this set of possible answers and assume that exactly one element of U is correct and that all elements of U are consistent with our present state of knowledge. Using Levi's terminology, U is said to be an *ultimate partition*, and a *potential answer* is the collection of hypotheses remaining after we have rejected all members of a subset of U . Each element of a potential answer is said to be a *serious possibility*. A potential answer is degenerate

if no elements of U are rejected, and we may not reject all members of U . If all but one element of U is rejected, the surviving hypothesis is the strongest potential answer. When more than one element of U survives rejection, we remain in suspense between the rival serious possibilities.

Epistemic utility is a probability that is composed of a convex combination of two probabilities, one measuring the importance of acquiring new information, the other measuring the importance of avoiding error. For any $g \subset U$, we define the utility of accepting g in the interest of avoiding error as $\mathcal{T}(g; \ell) = 1$ if $\ell = \text{true}$, and $\mathcal{T}(g; \ell) = 0$ if $\ell = \text{false}$. In addition to the cost of error, we also apportion a unit of informational value to each hypothesis $h_i \in U$ by assigning to elements of U non-negative real values such that their sum is unity. If we enumerate $U = \{h_1, h_2, \dots, h_n\}$, and let $\mathcal{M}(h_j) \geq 0$ denote the value assigned to h_j , then $\sum_{j=1}^n \mathcal{M}(h_j) = 1$, and for any set $g \subset U$, we define

$$\mathcal{M}(g) = \sum_{h_j \in g} \mathcal{M}(h_j) \quad (1)$$

as the informational value of rejecting g . The function $\mathcal{M}(\cdot)$ thus defined is a probability, termed an *information-determining* probability, and is intended to regulate the evaluation of information regardless of its truth-value. The utility of accepting g in the interest of acquiring new information regardless of its truth-value is, then, $\mathcal{C}(g) = 1 - \mathcal{M}(g)$.

We may address the conflict that exists between the goals of avoiding error and acquiring information by defining an *epistemic utility function* for acquiring error-free knowledge (that is, making a decision) as the convex combination $u(g; \ell) = \alpha \mathcal{T}(g; \ell) + (1 - \alpha) \mathcal{C}(g)$. The quantity α represents the relative importance attached to avoiding error versus acquiring new information. We must restrict $\frac{1}{2} \leq \alpha \leq 1$ to ensure that no erroneous answer is preferred to any correct answer. Since all utility functions that are related by a positive linear transformation are equivalent, we may simplify this utility function by defining $u^\alpha(g; \ell) = \frac{1}{\alpha} u(g; \ell) - \frac{1-\alpha}{\alpha}$. The resulting utility function for accepting g in the interest of both avoiding error and acquiring new knowledge is

$$u^\alpha(g; \ell) = \begin{cases} 1 - b\mathcal{M}(g) & \text{if } \ell = t \\ -b\mathcal{M}(g) & \text{if } \ell = f \end{cases} \quad (2)$$

where $b = \frac{1-\alpha}{\alpha}$ is the *coefficient of boldness*. This coefficient is constrained to lie in the interval $(0, 1]$. The closer b is to unity, the less caution is exercised that error will be introduced (increased boldness in accepting hypotheses); the closer b is to zero, the lower the risk of error (decreased boldness).

We must also establish a probabilistic measure of belief for the hypotheses that are available. *Credal* probability is probability formed on the basis of subjective judgment, represents the likelihood that an option is true and is independent of any informational value or demand that might be associated with the option. Whereas the information-determining probability is used to determine the utility of error-free knowledge, credal probability may be viewed as *expectation-determining* probability.

For a given ultimate partition U , let $\mathcal{Q}(g)$ denote the credal probability assignment to any element $g \in U$. For $g \in U$, the expected utility is

$$E_{\mathcal{Q}}u_a(g; \ell) = [1 - b\mathcal{M}(g)]\mathcal{Q}(g) - b\mathcal{M}(g)[1 - \mathcal{Q}(g)] \\ = \mathcal{Q}(g) - b\mathcal{M}(g) \quad (3)$$

where $E_{\mathcal{Q}}(\cdot)$ is mathematical expectation. This expected utility represents a tradeoff between the desire to acquire new knowledge and the desire to avoid error. The choice of b establishes a threshold at which the demand for knowledge renders the risk of error worthwhile.

We may adopt any set of hypotheses in the Boolean algebra generated by the elements of the ultimate partition, U . This expands our possibilities; we are not constrained to select only the elementary hypotheses, h_i , but may choose any subset of them. This decision philosophy may be summarized as follows:

Levi's Rule of Expected Utility [8, page 53]: Given a finite ultimate partition U , an information-determining probability function \mathcal{M} defined over the Boolean algebra of elements of U , an expectation-determining probability function \mathcal{Q} defined over the same algebra, and an index of boldness b , the agent should reject all and only those elements of $h_i \in U$ satisfying $\mathcal{Q}(h_i) < b\mathcal{M}(h_i)$.

B. Set-Valued Estimation

The notational convention used in this paper is to use boldface, lowercase symbols to denote random vectors, boldface, uppercase symbols to denote sets of random vectors, under-barred, italic symbols to denote sample values, and uppercase, under-barred italic symbols to denote sets of sample values. Matrices will be denoted by uppercase, non-under-barred italic symbols. We shall use the notation $\mathcal{N}(\underline{x}, P)$ to denote the Gaussian density with mean \underline{x} and covariance matrix P .

Consider a linear stochastic system of the form

$$\mathbf{x}_t = F_t \mathbf{x}_{t-1} + G_t \mathbf{u}_t \quad (4)$$

$$\mathbf{z}_t = H_t \mathbf{x}_t + \mathbf{v}_t \quad (5)$$

for $t = 1, 2, \dots$, where F_t is $n \times n$, G_t is $n \times p$, H_t is $r \times n$, $\{\mathbf{u}_t\}$ and $\{\mathbf{v}_t\}$ are p - and r -dimensional vector Gaussian (zero-mean) white noise processes with positive-definite covariance matrices Q_t and R_t , respectively.

The conventional Kalman filtering solution to this problem assumes that we know the prior distribution for \mathbf{x}_0 ; that is, that we know the mean and covariance matrix for this Gaussian random variable. The resulting estimate is *point-valued*, since the output of the estimator is a point vector in the state space. Now let us relax the assumptions regarding the prior distribution, and assume that we only know that the mean value of \mathbf{x}_0 lies in a convex set of the form $\underline{X}_0 = \{\underline{x} \in \mathbb{R}^n : (\underline{x} - \underline{c}_0)^T [K_0 K_0^T]^{-1} (\underline{x} - \underline{c}_0) \leq 1\}$, where K_0 is a non-singular matrix and \underline{c}_0 is a known vector in state space. There is a continuum of possible Kalman filters associated with this problem, one for each possible initial condition \underline{X}_0 , and it is not tractable to implement a Kalman filter for each of them and thereby propagate the convex region as observations are made.

The *set-valued* Kalman filter, however, provides a mechanism for propagating this set of estimates. We quote the following theorem [10, Theorem 3]:

Theorem 1 Consider the system given by (4) and (5), Let $K_{0|0}$ be an invertible $n \times n$ matrix, and let $\underline{c}_{0|0}$ be an n -vector. Let $\{\underline{z}_1, \underline{z}_2, \dots\}$ be observations. Suppose the initial state vector \mathbf{x}_0 is a member of the set of random vectors

$$\mathbf{X}_{0|0} = \{\mathbf{x} \sim \mathcal{N}(\underline{x}, P_{0|0}) : \underline{x} \in \underline{X}_{0|0}\} \quad (6)$$

where

$$\underline{X}_{0|0} = \{\underline{x} \in \mathbb{R}^n : (\underline{x} - \underline{c}_{0|0})^T [K_{0|0} K_{0|0}^T]^{-1} (\underline{x} - \underline{c}_{0|0}) \leq 1\} \quad (7)$$

and $P_{0|0}$ is a positive-definite matrix.

The set-valued Kalman filter is as follows:

Prediction Step:

$$\underline{X}_{t|t-1} = \left\{ \underline{x} \in \mathbb{R}^n : (\underline{x} - \underline{c}_{t|t-1})^T [K_{t|t-1} K_{t|t-1}^T]^{-1} (\underline{x} - \underline{c}_{t|t-1}) \leq 1 \right\} \quad (8)$$

where

$$\underline{c}_{t|t-1} = F_{t-1} \underline{c}_{t-1|t-1} \quad (9)$$

$$P_{t|t-1} = F_{t-1} P_{t-1|t-1} F_{t-1}^T + G_{t-1} Q_{t-1} G_{t-1}^T \quad (10)$$

$$K_{t|t-1} = F_{t-1} K_{t-1|t-1} \quad (11)$$

Filter Step:

$$\underline{X}_{t|t} = \left\{ \underline{x} \in \mathbb{R}^n : (\underline{x} - \underline{c}_{t|t})^T [K_{t|t} K_{t|t}^T]^{-1} (\underline{x} - \underline{c}_{t|t}) \leq 1 \right\} \quad (12)$$

where

$$\underline{c}_{t|t} = \underline{c}_{t|t-1} + W_t [\underline{z}_t - H_t \underline{c}_{t|t-1}] \quad (13)$$

$$P_{t|t} = [I - W_t H_t] P_{t|t-1} \quad (14)$$

$$K_{t|t} = [I - W_t H_t] K_{t|t-1} \quad (15)$$

with $W_t = P_{t|t-1} H_t^T [H_t P_{t|t-1} H_t^T + R_t]^{-1}$, the Kalman gain.

III. MULTIPLE TARGET TRACKING

We assume that each track is characterized by a linear stochastic dynamics model of the form

$$\mathbf{x}_{jt} = F_{j,t-1} \mathbf{x}_{j,t-1} + G_{j,t-1} \mathbf{u}_{j,t-1} \quad (16)$$

where $j = 1, 2, \dots, \mathcal{T}_t$; there are \mathcal{T}_t active tracks at time t (\mathcal{T}_t is unknown). The number of tracks is allowed to vary since tracks may be initiated or terminated at any time. We assume that all tracks lie in the same state space.

In the interest of brevity, we shall restrict attention to the outputs of a single sensor, and assume that this output may be characterized by a linear stochastic model of the form

$$\mathbf{z}_{it} = H_{it} \mathbf{x}_{jt} + \mathbf{v}_{it}, \quad i = 1, \dots, s_t \quad (17)$$

where s_t is the number of observations vectors at time t and \mathbf{z}_{it} is an r_{it} -dimensional random vector. Each observation

vector, therefore, lies in an r_{it} dimensional space, termed the it -th data space, which corresponds to the column space of H_{it} . Let \mathcal{Z}_{it} denote the it -th data space. We assume that each data space is a subset of the state space ($r_{it} \leq n$). We do not, however, require all observations to lie in the same dimensional subspace of the state space, and we permit the dimensionality of the data spaces to be time-varying.

Before data are collected, we characterize the target environment with one set-valued track defined by an initial credal matrix K_0 , an initial centroid state \underline{c}_0 , and a prior covariance matrix Π_0 . The initial state-vector set is

$$\mathbf{X}_{0|0} = \left\{ \mathbf{x} \sim \mathcal{N}(\underline{x}, P_{0|0}) : \underline{x} \in \underline{X}_{0|0} \right\} \quad (18)$$

where

$$\underline{X}_{0|0} = \left\{ \underline{x} \in \mathbb{R}^n : (\underline{x} - \underline{c}_{0|0})^T [S_{0|0}]^{-1} (\underline{x} - \underline{c}_{0|0}) \leq 1 \right\} \quad (19)$$

$P_{0|0} = \Pi_0$ is a positive-definite matrix, and $S_{0|0} = K_0 K_0^T$. For sample times $t = 1, 2, \dots$, we observe s_t observations $\underline{Z}_t = \{\underline{z}_{1t}, \dots, \underline{z}_{s_t t}\}$. Let us suppose, at time t , that we have τ_{t-1} sets of predicted (from time $t-1$) random-variables of the form

$$\mathbf{X}_{t|t-1}^j = \left\{ \mathbf{x} \sim \mathcal{N}(\underline{x}, P_{t|t-1}^j) : \underline{x} \in \underline{X}_{t|t-1}^j \right\} \quad (20)$$

where

$$\underline{X}_{t|t-1}^j = \left\{ \underline{x} \in \mathbb{R}^n : (\underline{x} - \underline{c}_{t|t-1}^j)^T [S_{t|t-1}^j]^{-1} (\underline{x} - \underline{c}_{t|t-1}^j) \leq 1 \right\} \quad (21)$$

where $S_{t|t-1}^j = [K_{t|t-1}^j] [K_{t|t-1}^j]^T$, with

$$\underline{c}_{t|t-1}^j = F_{t-1} \underline{c}_{t-1|t-1}^j \quad (22)$$

$$P_{t|t-1}^j = F_{t-1} P_{t-1|t-1}^j F_{t-1}^T + G_{t-1} Q_{t-1} G_{t-1}^T \quad (23)$$

$$K_{t|t-1}^j = F_{t-1} K_{t-1|t-1}^j \quad (24)$$

for $j = 1, \dots, \tau_{t-1}$. We shall use the same F_t , G_t , and Q_t matrices for each track, thus it will not be necessary to index these matrices with their track identifications.

A. The Ultimate Partition

At each time, t , there exist s_t observation sample vectors, $\{\underline{z}_{it}\}_{i=1}^{s_t}$ of the random variables $\{\mathbf{z}_{it}\}_{i=1}^{s_t}$, and τ_{t-1} predicted track sets, $\{\underline{X}_{t|t-1}^j\}_{j=1}^{\tau_{t-1}}$. We wish to make decisions regarding the association of each \underline{z}_{it} with each track set $\underline{X}_{t|t-1}^j$. We desire to apply Levi's decisionmaking methodology to this problem and will, therefore, adopt the strategy of accepting all track/data associations that cannot be rejected on the basis of Levi's rule of epistemic utility. We first define the ultimate partition, then we may specify the information-value determining probability, or \mathcal{M} -function, along with the credal probability, or \mathcal{Q} -function, and apply Levi's rule.

We must define an ultimate partition for each sample vector, resulting in a set of s_t ultimate partitions of the form $U_{it} = \{h_{it1}, \dots, h_{it\tau_{t-1}}, h_{it s_t}\}$, $i = 1, \dots, s_t$, where

$$h_{itj} = \begin{cases} \text{there exists } \underline{x} \in \underline{X}_{t|t-1}^j & j = 1, \dots, \tau_{t-1} \\ \text{that generated } \underline{z}_{it} & \\ \text{no track generated } \underline{z}_{it} & \text{otherwise} \end{cases} \quad (25)$$

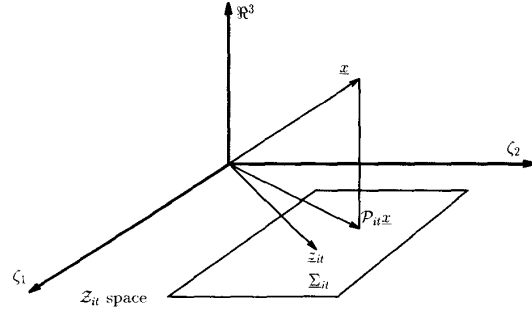


Fig. 1. Geometry of the \mathcal{M} -function for $n = 3$ showing the projection of the state onto the observation space \mathcal{Z}_{it} and the seriously possible region.

We shall say that track set $\underline{X}_{t|t-1}^j$ is associated with observation sample vector \underline{z}_{it} if we fail to reject the hypothesis h_{itj} . Each ultimate partition, U_{it} , has the property that exactly one element is true, although each is logically possible. According to Levi's theory of expected utility, we may reject only those members of the ultimate partition that are not seriously possible. We do not insist that the decision that one and only one element of U_{it} be chosen as the association decision.

B. Calculation of the \mathcal{M} -Function

The information-determining probability, or \mathcal{M} -function, is intended to measure the information value of rejecting an association, rather than the truth-value of an association. A measure of the information value of rejecting the association of a predicted track set $\underline{X}_{t|t-1}^j$ with an observation \underline{z}_{it} is the distance between sample values of the observation and the track set; if the distance is small there is little value in rejecting the association (in other words, there is great value in accepting it).

Observations are related to points in the state space via (17), where H_{it} is an $r_{it} \times n$ matrix (with $r_{it} \leq n$) of rank r_{it} . Since the dimensionality of the observations is generally different than the dimension of the state space, care must be taken to obtain a meaningful definition of distance between points in these spaces. For $\underline{x} \in \underline{X}_{t|t-1}^j \subset \mathbb{R}^n$ and $\underline{z}_{it} \in \mathbb{R}^{r_{it}}$, we define the generalized distance between them as

$$d(\underline{x}, \underline{z}_{it}) = \|H_{it}\underline{x} - \underline{z}_{it}\| \stackrel{\text{def}}{=} [(H_{it}\underline{x} - \underline{z}_{it})^T (H_{it} - \underline{z}_{it})]^{1/2}. \quad (26)$$

We desire to normalize this distance to permit its interpretation as a probability density function. This normalization is accomplished by defining a region of the data space \mathcal{Z}_{it} which we may assume contains all predicted observation values that may be feasibly associated with the given value for \underline{z}_{it} , and restricting attention only to this region, termed the seriously possible region for \underline{z}_{it} , which we shall denote as $\underline{\Sigma}_{it}$. An illustration of the geometry is given in Fig. 1.

Let \underline{z}_{it} denote a sample value of the observation random vector, \mathbf{z}_{it} . We may view \underline{z}_{it} as a sample from a normal random variable with distribution $\mathcal{N}(\underline{z}_{it}^*, R_{it})$ for some real vector \underline{z}_{it}^* . The mean, \underline{z}_{it}^* , is unknown, but we will assume it is an element of $\underline{\Sigma}_{it}$, the seriously possible region for \underline{z}_{it} . We shall assume that $\underline{\Sigma}_{it}$ is a convex set centered at \underline{z}_{it} of the form

$$\underline{\Sigma}_{it} = \left\{ \underline{z} \in \mathbb{R}^{r_{it}} : |\zeta_1 - z_{it1}| \leq \theta \rho_{i1}, \dots, |\zeta_{r_{it}} - z_{itr_{it}}| \leq \theta \rho_{ir_{it}} \right\} \quad (27)$$

where we assume that R_{it} is a diagonal matrix of the form $R_{it} = \text{diag}\{\rho_{i1}^2, \dots, \rho_{ir_{it}}^2\}$ and θ is a given constant. Note that the size of the seriously possible region is determined by the covariance of the observation \underline{z}_{it} .

Before we can use the distance function for the calculation of the information-determining probability function \mathcal{M} , we must characterize the set of all $\underline{x} \in \mathbb{R}^n$ that corresponds to the seriously possible region. To do this, it is convenient to employ a *projection operator* [12, page 105], to project the state space onto the space spanned by the columns of the matrix H_{it} ; that is, onto the data space \mathcal{Z}_{it} . Define

$$\mathcal{P}_{it} = H_{it}^T [H_{it} H_{it}^T]^{-1} H_{it} \quad (28)$$

and note that $\mathcal{P}_{it} = \mathcal{P}_{it}^T$. For any $\underline{x} \in \underline{X}_{t|t-1}^j$, its projection onto the data space \mathcal{Z}_{it} is given by $\underline{z}_{it} = \mathcal{P}_{it} \underline{x}$. The vector \underline{z}_{it} is termed the *minimum-norm* solution to the equation $\underline{\zeta} = H_{it} \underline{x}$.

The operator $\mathcal{P}_{it}^\perp = I - \mathcal{P}_{it}$, where \mathcal{P}_{it} is defined by (28), is a projection operator onto the subspace that is orthogonal to the column space of H_{it} . We may decompose \mathbb{R}^n into the direct sum of the column space of H_{it} and the subspace orthogonal to the column space of H_{it} . Let $\mathcal{P}_{it} \mathbb{R}^n$ and $\mathcal{P}_{it}^\perp \mathbb{R}^n$, respectively, denote these two orthogonal subspaces, and define

$$\Xi_{it} = \{ \underline{x} \in \mathbb{R}^n : H_{it} \underline{x} \in \Sigma_{it} \}. \quad (29)$$

If $\underline{x} \in \Xi_{it}$, then $|(H_{it} \underline{x})_k - z_{itk}| \leq \rho_{ik} \theta$, $k = 1, \dots, r_{it}$, where $(H_{it} \underline{x})_k$ is the k th element of $H_{it} \underline{x}$. But $\underline{x} = (\mathcal{P}_{it} + \mathcal{P}_{it}^\perp) \underline{x} = \mathcal{P}_{it} \underline{x} + \mathcal{P}_{it}^\perp \underline{x}$, thus $H_{it} \underline{x} = H_{it} \mathcal{P}_{it} \underline{x} + H_{it} \mathcal{P}_{it}^\perp \underline{x}$. Since $\mathcal{P}_{it}^\perp \underline{x}$ is orthogonal to the column space of H_{it} , $H_{it} \mathcal{P}_{it}^\perp \underline{x} \equiv 0$ and $H_{it} \underline{x} = H_{it} \mathcal{P}_{it} \underline{x}$. Then

$$|(H_{it} \underline{x})_k - z_{itk}| = |(H_{it} \mathcal{P}_{it} \underline{x})_k - z_{itk}|, \quad k = 1, \dots, r_{it} \quad (30)$$

and there is no constraint on vectors in \mathbb{R}^n that lie in the subspace orthogonal to the column space of H_{it} . We may, therefore, express Ξ_{it} as the direct sum of constrained vectors that lie in the column space and unconstrained vectors that lie in the space orthogonal to the column space; that is, $\Xi_{it} = \Xi_{it}^{H_{it}} \oplus \mathcal{P}_{it}^\perp \mathbb{R}^n$, where

$$\Xi_{it}^{H_{it}} = \{ \underline{x} \in \mathcal{P}_{it} \mathbb{R}^n : |(H_{it} \underline{x})_k - z_{itk}| \leq \rho_{ik} \theta, \quad k = 1, \dots, r_{it} \}. \quad (31)$$

Since we wish \mathcal{M} to be a probability, we require that the distance function, $d(\underline{x}, \underline{z}_{it})$, be normalized, thereby admitting the interpretation as an information-determining probability density function. We must normalize this function by the seriously possible region; that is, for fixed \underline{z}_{it} , let

$$m_{it}(\underline{x}) = \begin{cases} \frac{d(\underline{x}, \underline{z}_{it})}{\int_{\Xi_{it}} \|\underline{\zeta} - \underline{z}_{it}\| d\underline{\zeta}} = \frac{\|H_{it} \underline{x} - \underline{z}_{it}\|}{\int_{\Xi_{it}} \|\underline{\zeta} - \underline{z}_{it}\| d\underline{\zeta}} & \underline{x} \in \Xi_{it} \\ \text{undefined} & \underline{x} \notin \Xi_{it} \end{cases} \quad (32)$$

(If the dimension of the data space is constant and the noise covariance matrix is constant, the denominator can be precomputed.)

The function $m_{it}(\underline{x})$ may be viewed as a normalized distance between \underline{z}_{it} and $\underline{x} \in \underline{X}_{t|t-1}^j$, and is a measure of the information gained by rejecting the association of $\underline{x} \in \underline{X}_{t|t-1}^j$ and \underline{z}_{it} . Intuitively, the greater the distance between the predicted observation and the actual observation, the greater the value in rejecting the hypothesis that the predicted state estimate and the observation are associated.

Let $\underline{B}_{\underline{x}}$ be a ball with center at $\underline{x} \in \underline{X}_{t|t-1}^j$. The information value of rejecting the association of $\underline{B}_{\underline{x}}$ with \underline{z}_{it} is

$$\mathcal{M}_{it}(\underline{B}_{\underline{x}}) = \int_{\mathcal{P}_{it} \underline{B}_{\underline{x}} \cap \Xi_{it}} m_{it}(\underline{\xi}) d\mathcal{P}_{it} \underline{\xi}, \quad i = 1, \dots, s_t, \quad (33)$$

where $\mathcal{P}_{it} \underline{B}_{\underline{x}}$ is the projection of the ball $\underline{B}_{\underline{x}}$ onto the data space \mathcal{Z}_{it} . The vector $\mathcal{P}_{it} \underline{x}$ is the projection of \underline{x} onto the same data space. We emphasize that the informational-value determining probability places all of the probability mass in the data space. This result is appropriate, since there is no way of assessing the informational value of rejecting track/data associations by means of components of the state that lie in the subspace orthogonal to the column space of H_{it} .

For $\underline{x} \in \underline{X}_{t|t-1}^j$ and $\underline{x}' \in \underline{X}_{t|t-1}^j$ with $\underline{x} \neq \underline{x}'$, let the diameter of the balls $\underline{B}_{\underline{x}}$ and $\underline{B}_{\underline{x}'}$ become arbitrarily small. Then the condition $\mathcal{M}_{it}(\underline{B}_{\underline{x}}) < \mathcal{M}_{it}(\underline{B}_{\underline{x}'})$ indicates that the information value of *accepting* the association of \underline{x} with \underline{z}_{it} is greater than the value of accepting the association of \underline{x}' and \underline{z}_{it} .

C. Calculation of the Q-Functions

The credal probability, or Q-function, is the probability that a given track-data association is correct. This probability is a function of the statistical descriptions of the target dynamics and of the observation errors, and is characterized by the density of the predicted track, projected onto the observation space. For each $j = 1, \dots, \tau_{t-1}$, the set of predicted random vectors is given by

$$\mathbf{X}_{t|t-1}^j = \{ \mathbf{x} \sim \mathcal{N}(\underline{x}, P_{t|t-1}^j) : \underline{x} \in \underline{X}_{t|t-1}^j \} \quad (34)$$

where

$$\underline{X}_{t|t-1}^{ij} = \{ \underline{x} \in \mathbb{R}^n : (\underline{x} - \underline{c}_{t|t-1}^{ij})^T [S_{t|t-1}^{ij}]^{-1} (\underline{x} - \underline{c}_{t|t-1}^{ij}) \leq 1 \} \quad (35)$$

and $S_{t|t-1}^{ij} = K_{t|t-1}^{ij} [K_{t|t-1}^{ij}]^T$. Here the superscript i denotes the observation vector and the superscript j denotes the track number.

The probability distributions of $\{ \mathbf{x} \in \mathbf{X}_{t|t-1}^{ij} \}$ assume the form

$$p_{\mathbf{X}_{t|t-1}^{ij}}(\underline{\xi}; \underline{x}) = \mathcal{N}(\underline{x}, P_{t|t-1}^j), \quad \underline{x} \in \underline{X}_{t|t-1}^j \quad (36)$$

so that the density is

$$p_{\mathbf{X}_{t|t-1}^{ij}}(\underline{\xi}; \underline{x}) = \frac{1}{(2\pi)^{\frac{n}{2}} |P_{t|t-1}^{ij}|^{\frac{1}{2}}} e^{-\frac{1}{2}(\underline{\xi} - \underline{x})^T [P_{t|t-1}^{ij}]^{-1} (\underline{\xi} - \underline{x})} \quad \underline{x} \in \underline{X}_{t|t-1}^j \quad (37)$$

By projecting this density onto the space of observations, a measure of the truth value of the association of the track $\underline{X}_{t|t-1}$ with the observation z_{it} may be obtained. Let

$$p_{\mathbf{x}_{t|t-1}}^{ij}(z_{it}, \underline{x}) = \frac{1}{(2\pi)^{\frac{r_{it}}{2}} |P_{t|t-1}^H|^{\frac{1}{2}}} e^{-\frac{1}{2}(z-H\underline{x})^T [P_{t|t-1}^H]^{-1} (z-H\underline{x})} \quad (38)$$

where $P_{t|t-1}^H = H P_{t|t-1}^{ij} H^T$ projects the covariance matrix onto the observation space. We may view this family of densities as credal probability densities, and interpret them as characterizing the belief that the above association is correct.

Let $\underline{B}_{\underline{x}}$ be a ball with center at \underline{x} . The credal probability that the true state lies in $\underline{B}_{\underline{x}}$ conditioned on the observational value z_{it} is

$$Q_{ijt}(\underline{B}_{\underline{x}}; \underline{x}) = \int_{\underline{B}_{\underline{x}}} p_{\mathbf{x}_{t|t-1}}^{ij}(z_{it}; \underline{x}) d\underline{x}. \quad (39)$$

For $\underline{x} \in \underline{X}_{t|t-1}^j$ and $\underline{x}' \in \underline{X}_{t|t-1}^j$ with $\underline{x} \neq \underline{x}'$, the condition $Q_{ijt}(\underline{B}_{\underline{x}}; \underline{x}) > Q_{ijt}(\underline{B}_{\underline{x}'}; \underline{x}')$ indicates that the association of x with z_{it} is more credible, or believable, than the association of x' and z_{it} .

D. The Association Likelihood Ratio Test

The association decision problem is to determine whether or not the track $\underline{X}_{t|t-1}^j$ is associated with the observation z_{it} . We shall assume a conservative attitude, and say that the entire track set $\underline{X}_{t|t-1}^j$ is associated with z_{it} if any $\underline{x} \in \underline{X}_{t|t-1}^j$ is associated with z_{it} . While it is possible to refine the notion of association to determine a subset of $\underline{X}_{t|t-1}^j$ that is associated with z_{it} , we will not make this refinement in this development.

The decision rule may be formulated in terms of the information-determining probability density, $m_{it}(\underline{x})$ (where $\underline{x} \in \underline{X}_{t|t-1}^j$), and the family of subjective belief, or credal, probability density functions, $\{p_{\mathbf{x}_{t|t-1}}^{ij}(z; \underline{x}), \underline{x} \in \underline{X}_{t|t-1}^j\}$. We desire to apply Levi's rule of expected utility to this problem.

For $\underline{x} \in \underline{X}_{t|t-1}^j$, let $\underline{B}_{\underline{x}}$ be a ball with center at \underline{x} . Using (33) and (39), Levi's rule of expected utility indicates we may not reject the association of $\underline{B}_{\underline{x}}$ and z_{it} if

$$Q_{ijt}(\underline{B}_{\underline{x}}; \underline{x}) \geq b \mathcal{M}_{it}(\underline{B}_{\underline{x}}). \quad (40)$$

Now let the radius of $\underline{B}_{\underline{x}}$ go to zero, and define the function

$$q_{ijt}(\underline{x}) \stackrel{\text{def}}{=} p_{\mathbf{x}_{t|t-1}}^{ij}(z_{it}; \underline{x}). \quad (41)$$

Since the densities are continuous at \underline{x} , a necessary condition for (40) to hold for all balls $\underline{B}_{\underline{x}}$ is that

$$q_{ijt}(\underline{x}) \geq b m_{it}(\underline{x}). \quad (42)$$

If (42) holds for any $\underline{x} \in \underline{X}_{t|t-1}^j$, we may not reject the association of \underline{x} with z_{it} ; consequently, we may not reject the association of the track set $\underline{X}_{t|t-1}^j$ with z_{it} .

We emphasize that both $m_{it}(\cdot)$ and $q_{ijt}(\cdot)$, are independent of the values that \underline{x} assumes in the orthogonal subspace, since, in both expressions, \underline{x} appears premultiplied by H_{it} . Thus, both the informational-value determining probability

and the credal probability are completely determined by the component of \underline{x} that lies in the column space of H_{it} (the data space \mathcal{Z}_{it}).

Given the information-value determining probability density (32) and the credal probability density (38), we may form the Association Likelihood Ratio Test (ALRT):

Let $\underline{X}_{t|t-1}^j$ be a predicted track set, and let z_{it} be a sample value of the observation vector \mathbf{z}_{it} . We shall say that $\underline{X}_{t|t-1}^j$ and z_{it} are associated with boldness b if $q_{ijt}(\underline{x}) \geq b m_{it}(\underline{x})$ for some $\underline{x} \in \Xi_{it} \cap \underline{X}_{t|t-1}^j$. That is, there exists $\underline{x} \in \Xi_{it} \cap \underline{X}_{t|t-1}^j$ such that

$$p_{\mathbf{x}_{t|t-1}}^{ij}(z_{it}, \underline{x}) = \frac{1}{(2\pi)^{\frac{r}{2}} |P_{t|t-1}^H|^{\frac{1}{2}}} e^{-\frac{1}{2}(z-H\underline{x})^T [P_{t|t-1}^H]^{-1} (z-H\underline{x})} \geq b \frac{\|H_{it}\underline{x} - z_{it}\|}{\int_{\Xi_{it}} \|H_{it}\underline{\xi} - z_{it}\| d\underline{\xi}} \quad (43)$$

in which case the track set is filtered according to (44) as described in the next section. If $\underline{X}_{t|t-1}^j \cap \Xi_{it} = \emptyset$, then we deem z_{it} and $\underline{X}_{t|t-1}^j$ to be dissociated.

If $\underline{X}_{t|t-1}^j$ survives the ALRT for z_{it} , then the likelihood that $\underline{X}_{t|t-1}^j$ is associated with z_{it} is greater than the information value gained by rejecting the association. If the threshold $b = 1$, then the decision strategy is maximally bold in the sense that as many associations $(\underline{X}_{t|t-1}^j, z_{it})$ will be rejected as possible. As b approaches zero, the decision strategy is maximally cautious, and the likelihood diminishes that a correct association (indeed, any association) will be rejected. A brief discussion on some numerical aspects of the ALRT computations is provided in the appendix.

E. Track Filtering

When an observation z_{it} associates with a track set $\underline{X}_{t|t-1}^{ij}$ then we calculate the filtered set-valued estimate according to

$$\underline{X}_{t|t}^{ij} = \left\{ \underline{x} \in \mathfrak{R}^n : (\underline{x} - \underline{c}_{t|t}^{it})^T [S_{t|t}^{ij}]^{-1} (\underline{x} - \underline{c}_{t|t}^{it}) \leq 1 \right\} \quad i = 1, 2, \dots, s_t \quad (44)$$

where $S_{t|t}^{ij} = K_{t|t}^{it} [K_{t|t}^{it}]^T$ and

$$\underline{c}_{t|t}^{ij} = \underline{c}_{t|t-1}^i + W_{ijt} [z_{it} - H_{it} \underline{c}_{t|t-1}^j] \quad (45)$$

$$P_{t|t}^{ij} = [I - W_{ijt} H_{it}] P_{t|t-1}^j \quad (46)$$

$$K_{t|t}^{ij} = [I - W_{ijt} H_{it}] K_{t|t-1}^j \quad (47)$$

for $j = 1, 2, \dots, \tau_{t-1}$, with W_{ijt} the Kalman gain defined by $W_{ijt} = P_{t|t-1}^j H_{it}^T [H_{it} P_{t|t-1}^j H_{it}^T + R_{it}]^{-1}$.

IV. TRACK MAINTENANCE

The set $\underline{V}_{t|t}^i = \bigcup_{j=1}^{\tau_{t-1}} \underline{X}_{t|t}^{ij}$, $i = 1, \dots, s_t$, is composed of all track sets that survive the association test with z_{it} . If $\underline{V}_{t|t}^i = \underline{X}_{t|t}^{ij}$, for some j , then there exists only one predicted track set that associates with z_{it} . If $\underline{V}_{t|t}^i$ consists of more than one $\underline{X}_{t|t}^{ij}$, then there exists a *conflict of the first kind*, which means that more than one track set is associated with z_{it} . We

do not need to force a resolution of this conflict by choosing one "best" association. Indeed, doing so may result in an unacceptably large probability of error, which may adversely affect future decisions and track estimates. Fortunately, the application of the set-valued estimator permits us to avoid imposing such arbitrary and objectionable constraints on the problem. We may deal with this situation by combining into one track set all track sets that survive the association test with a given data vector.

Consider the collection $\{\underline{X}_{t|t}^{i j_1}, \dots, \underline{X}_{t|t}^{i j_{k_{it}}}\}$ of filtered track sets associated with data vector \underline{z}_{it} . If $k_{it} = 1$, there is a unique association with \underline{z}_{it} . If $k_{it} > 1$, the following *ad hoc* procedure may be used to combine these commonly-associated track sets into one track set.

- 1) *Centroid calculation.* The centroid of the new ellipsoidal set, denoted $\underline{c}_{t|t}^i$, is the average of the centroids; that is,

$$\underline{c}_{t|t}^i = \frac{1}{k_{it}} \sum_{\ell=1}^{k_{it}} \underline{c}_{t|t}^{i j_\ell}.$$

- 2) *Ellipsoid calculation.* The ellipsoid of the merged track set, denoted The credal matrix of the new ellipsoidal set, denoted $K_{t|t}^i$, is obtained as follows:

- (a) Compute the eigenvalues and eigenvectors, denoted, for each ellipsoidal matrix $K_{t|t}^{i j_k} \left[K_{t|t}^{i j_k} \right]^T$, by $\{\lambda_{\ell k}\}_{\ell=1}^n$ and $\{\underline{\mu}_{\ell k}\}_{\ell=1}^n$, respectively (we assume that the eigenvectors are of unit length). Define the vectors

$$\{\underline{a}_1, \dots, \underline{a}_n\} = \{\sqrt{\lambda_{1k}} \underline{\mu}_{1k} - \underline{c}_{t|t}^i, \dots, \sqrt{\lambda_{nk}} \underline{\mu}_{nk} - \underline{c}_{t|t}^i\}. \quad (48)$$

- (b) Let $\kappa_1 = \arg\{\max_{\ell}\{|\underline{a}_{\ell k}|\}\}$. Then \underline{a}_{κ_1} defines the semi-major axis of the new ellipsoid. Let $\eta_1 = \frac{\underline{a}_{\kappa_1}}{\|\underline{a}_{\kappa_1}\|}$, and $\phi_1 = \|\underline{a}_{\kappa_1}\|^2$. Then ϕ_1 and η_1 represent the largest eigenvalue and associated eigenvector of the new credal matrix $K_{t|t}^i$.

- (c) Compute the orthogonal complement of all other \underline{a}_{ℓ} 's with respect to η_1 . The resulting vectors then lie in the $n-1$ -dimensional space orthogonal to η_1 . We define the vectors $\underline{a}'_{\ell} = (I - \eta_1 \eta_1^T) \underline{a}_{\ell}$, $\ell \neq \kappa_1$.

- (d) Repeat Steps (b) and (c) above to find κ_2 , ϕ_2 , and η_2 , and continue until all eigenvalues and vectors of $K_{t|t}^i$ have been formed, thus defining the new ellipsoidal matrix $K_{t|t}^i \left[K_{t|t}^i \right]^T$. The credal matrix $K_{t|t}^i$ can then be obtained by a Cholesky factorization.

- 3) *Covariance calculation.* We assume that each credal set $\underline{X}_{t|t}^{i j_k}$ is governed by the same dynamics matrix F_t but the covariance matrices $P_{t|t}^{i j_k}$ may all be different. Thus, it is necessary to compute a new covariance matrix to associate with the new centroid and credal matrix. We shall define this matrix as the element of the set $\{P_{t|t}^{i k}\}_{k=1}^{k_{it}}$ with the largest Euclidean norm.

This track-merge method is reasonable, but no claims of optimality are made and it is by no means the only one that may be proposed. It suffers from requiring a full eigendecomposition for each of the track sets. Also, it can be shown that the merged track does not necessarily contain all parts of the

track sets that merge into it. Another merging method may be obtained by observing that since the track sets are usually initialized to have their axes parallel with the coordinate axes (that is, $S = K K^T$ is diagonal and the ellipses are unrotated), the eigendecomposition of track sets is initially trivial. If the system noise covariance of the track states is diagonal (a common assumption), then for many observation matrices the ellipse matrices remain diagonal; a merging algorithm which preserves the diagonal track ellipse matrices obviates the need for ever doing a complicated eigendecomposition. This may be accomplished by simply finding the largest extent of all of the ellipses in each coordinate direction. This approach, however, leads to generally larger merged tracks than the method above. Another variation suggested by an anonymous reviewer is to take into consideration the accountability of each track by weighting each track by the inverse of its covariance when computing the centroid. In the simulations below, however, we have used the track merge method described above.

If $\underline{V}_{t|t}^i = \emptyset$, then no predicted track sets associate with the observation vector \underline{z}_{it} , and \underline{z}_{it} is deemed to be an *unassociated observation*, which must correspond to either a new track or to a false alarm (a signal detection that does not correspond to a legitimate target). Since there is no warrant for deciding arbitrarily that \underline{z}_{it} is a false alarm, we may assume that \underline{z}_{it} represents the start of a new track. We then define a new track set associated with this observation. The new track set is formed by choosing its centroid to be any vector \underline{c} such that $H_{it} \underline{c} = \underline{z}_{it}$; the credal matrix may be any nonsingular matrix.

The set $\underline{Y}_{t|t}^j = \bigcup_{i=1}^{s_t} \underline{X}_{t|t}^{i j}$ is composed of all filtered track sets corresponding to the j th predicted track set that survive the association test with at least one data vector. If $\underline{Y}_{t|t}^j = \underline{X}_{t|t}^{i j}$ for some i , then exactly one data vector associates with $\underline{X}_{t|t-1}^j$. If $\underline{Y}_{t|t}^j$ consists of more than one filtered track set, then more than one data vector is associated with $\underline{X}_{t|t-1}^j$, resulting in a *conflict of the second kind*. This situation may correspond to the initiation of a new track (for example, one vehicle may split into multiple vehicles), or it may mean that multiple tracks are crossing, or it may correspond to parallel tracks. In all of these cases, it is undesirable to force a resolution of this conflict; it is far better to retain all tracks.

If $\underline{Y}_{t|t}^j = \emptyset$, then the set $\underline{X}_{t|t-1}^j$ is an *unassociated predicted track set* at time t ; either the track has terminated or there is a data drop-out due to a probability of detection less than 1. Since there is no warrant for deciding arbitrarily that the track has terminated, we must assume initially that $\underline{X}_{t|t-1}^j$ represents a track with no data attached at time t . We may deal with this situation by propagating the predicted track set to form the set

$$\underline{X}_{t+1|t-1}^j = \left\{ \underline{x} \in \mathbb{R}^n : \left(\underline{x} - \underline{c}_{t+1|t-1}^j \right)^T \left[S_{t+1|t-1}^j \right]^{-1} \left(\underline{x} - \underline{c}_{t+1|t-1}^j \right) \leq 1 \right\} \quad (49)$$

with $S_{t+1|t-1}^j = K_{t+1|t-1}^j \left[K_{t+1|t-1}^j \right]^T$ and

$$\underline{c}_{t+1|t-1}^j = F_t \underline{c}_{t|t-1}^j \quad (50)$$

$$P_{t+1|t-1}^j = F_t P_{t|t-1}^j F_t^T + G_t Q_t G_t^T \quad (51)$$

$$K_{t+1|t-1}^j = F_t K_{t|t-1}^j. \quad (52)$$

If, at time $t + 1$, no association is made with this track, we must decide whether to continue to propagate it or to terminate it. We must define a track termination decision rule to decide whether or not to reject this track as being valid. We do this by following the same philosophy as used to develop the ALRT. Let the ultimate partition for this track be $U_p = \{\text{retain}, \text{terminate}\}$. To invoke Levi's rule of expected utility, we must define the information-determining and credal probabilities for rejection, or *pruning* of the track set.

Let t' , be the time associated with the most recent association of this track with any previous observations ($t' \leq t$). The larger the value $t - t'$, the lower value it has as a legitimate track. It is reasonable to suppose that there exists some integer, N_p , such that $t - t' > N_p$ entails maximum informational value of termination. We may, therefore, define the information-value of terminating a track

$$\mathcal{M}_p = \begin{cases} \frac{t-t'}{N_p} & t-t' < N_p \\ 1 & t-t' \geq N_p \end{cases}. \quad (53)$$

The only rational basis for terminating a track is the possibility that it is not currently associated with a real signal source of interest. This may occur because of the following reasons: (a) the signal is too weak to be detected by the receiver at time t , (b) the source physically ceases to exist at time t , or (c) the track is bogus and was created by a false alarm, such as clutter, at a previous time. Reliable analysis of the probabilities of the first two reasons is problematic, and we will ignore them in this analysis. The third reason, however, is directly related to the probability of false alarm, which can often be estimated from knowledge of the characteristics of the receiver and the signal environment. Even if not precisely known, it may be possible to determine a largest seriously possible probability of false alarm, P_{FA}^* . Then $Q_p \leq P_{FA}^*$. For each such Q_p , the credal probability of track retention is then $1 - Q_p$.

In accordance with Levi's rule of expected utility, we may determine the retention or termination status of a given track with the *pruning likelihood ratio test* (PLRT):

An unassociated track is rejected if and only if $1 - Q_p < b_p \mathcal{M}_p$ for all $Q_p \leq P_{FA}^$, where $b_p \in [0, 1]$ is the coefficient of boldness associated with termination. Thus, a track is terminated if and only if $1 - P_{FA}^* < b_p \frac{t-t'}{N_p}$; that is, if $t - t' > \frac{N_p(1-P_{FA}^*)}{b_p}$.*

The structure of this rule permits the false-alarm probability to be a function of time, as might be the case for nonstationary collection environments. The coefficient of boldness b_p may also be made time-varying if desired, depending upon the criticalness of the decision to terminate a track. A high value of b_p entails boldness in terminating tracks, and a low value entails extreme reluctance to terminate tracks.

Example: To illustrate the concept of the ALRT methodology, consider the case of tracking targets constrained to planar motion [13] with intensity observations available. Let $\mathbf{x} = [x, y, \dot{x}, \dot{y}, i]^T$ denote the kinematic and intensity state

of a target in some convenient coordinate system ($n = 6$). The dynamics equation is

$$\mathbf{x}_t = \begin{bmatrix} 1 & 0 & \Delta_t & 0 & 0 & 0 \\ 0 & 1 & 0 & \Delta_t & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & \Delta_t \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix} \mathbf{x}_{t-1} + \begin{bmatrix} u_{x_t} \\ u_{y_t} \\ u_{\dot{x}_t} \\ u_{\dot{y}_t} \\ u_{i_t} \\ u_{i_t} \end{bmatrix} \quad (54)$$

where Δ_t is the sample interval, and we have set $G \equiv I$. Possible target maneuvers are assumed to be characterized by the process noise, \mathbf{u}_t , whose covariance is Q_t . We assume that angle-of-arrival data are available; further, we assume that the sensor is sufficiently far from the target that the linearized model is adequate. For convenience we also assume that the coordinate system is resolved along the azimuth and elevation angles and that the intensity is observable,

$$H_{it} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \end{bmatrix} \quad (55)$$

that is, $r_{it} \equiv 3$. Due to the intrinsic decoupling of the kinematic and intensity state variables, it will be convenient to consider the track sets as ellipsoidal cylinders, rather than as pure ellipsoids. Let $\underline{\mathbf{x}} = [x_1, x_2, x_3, x_4, x_5, x_6]^T \in \underline{X}_{t|t-1}^j$, $\underline{\mathbf{z}}_{it} = [z_{it1}, z_{it2}, z_{it3}]^T$, and suppose $R_{it} = \text{diag}\{\rho_1^2, \rho_2^2, \rho_3^2\}$. Then the seriously possible region is

$$\Sigma_{it} = \left\{ \begin{bmatrix} \zeta_1 \\ \zeta_2 \\ \zeta_3 \end{bmatrix} : |\zeta_1 - z_{it1}| \leq \theta \rho_1, \quad |\zeta_2 - z_{it2}| \leq \theta \rho_2 \right. \\ \left. |\zeta_3 - z_{it3}| \leq \theta \rho_3 \right\} \quad (56)$$

and the information-value determining probability density is

$$m_{it}(\underline{\mathbf{x}}) = \frac{\sqrt{(x_1 - z_{it1})^2 + (x_2 - z_{it2})^2 + (x_5 - z_{it3})^2}}{\int_{-\theta \rho_3}^{\theta \rho_3} \int_{-\theta \rho_2}^{\theta \rho_2} \int_{-\theta \rho_1}^{\theta \rho_1} \sqrt{\zeta_1^2 + \zeta_2^2 + \zeta_3^2} d\zeta_1 d\zeta_2 d\zeta_3}. \quad (57)$$

Each $\underline{\mathbf{x}} \in \underline{X}_{t|t-1}^j$ represents the mean value of a predicted conditional distribution. For each such $\underline{\mathbf{x}}$ there exists a filtered conditional density of the form (38); this can be applied in the ALRT (43) to determine if $\underline{X}_{t|t-1}^j$ and $\underline{\mathbf{z}}_{it}$ associate.

We wish to examine the decision boundaries and track sets as time evolves. Figure 2 illustrates the evolution of the azimuth and elevation components of the predicted and filtered track sets for a recursion from t to $t + 1$. The solid lines correspond to time t , and the dashed lines correspond to time $t + 1$ (one may think of a time axis that runs out of the paper, and the solid and dashed frames correspond to different points along this time axis). Since predicted track $\underline{X}_{t|t-1}^1$ is associated with $\underline{\mathbf{z}}_{it}$ via the ALRT decision rule, the filtered track set $\underline{X}_{t|t}^1$ is computed using (44) and used as the set-valued estimate of this track; it is denoted in the figure. Once a track has been associated, there is no need to retain the index i to identify it. We may simply refer to this track as $\underline{X}_{t|t}^1$; we also may drop the data index i for the filtered centroid, filtered covariance matrix, and filtered credal matrix in (45) through (47). We may obtain the predicted track set $\underline{X}_{t+1|t}^1$ by propagating track $\underline{X}_{t|t}^1$ via (21) through (24).

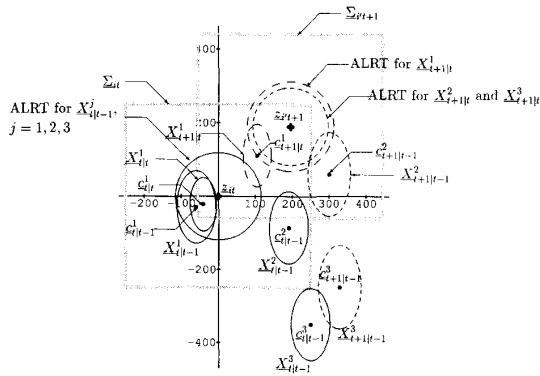
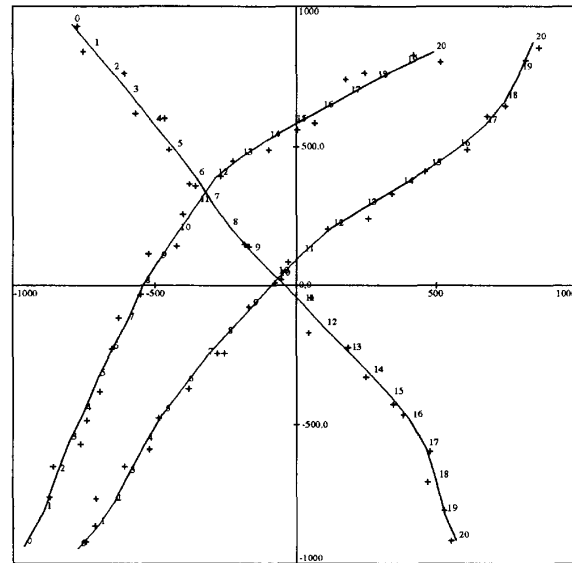


Fig. 2. Example of predicted and filtered track sets at times t and $t + 1$.

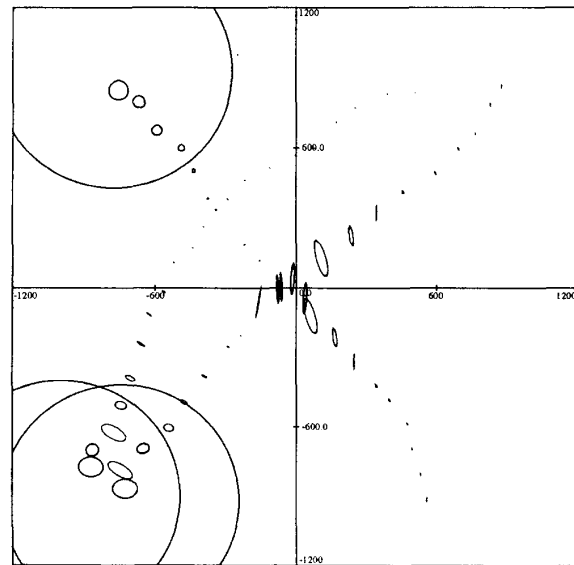
Tracks $X_{t|t-1}^2$ and $X_{t|t-1}^3$ did not associate with observation z_{it} , however, so the filtered track sets $X_{t|t}^{i2}$ and $X_{t|t}^{i3}$ cannot be used to propagate those tracks. If there are no other observations with which these tracks associate, we must regard them as dissociated at time t . They may be associated with future observations, however, so we need to propagate them to the next time, $t + 1$. We perform this propagation according to (49) for $j = 2, 3$.

At time $t + 1$, the observation $z_{i't+1}$ is obtained, and the ALRT decision region is calculated. For tracks that have been previously associated, this decision region is a function of the filtered covariance matrix, $P_{t+1|t+1}^j$; for tracks that were not previously associated, the decision region is a function of the two- (or more) step prediction covariance matrix, $P_{t+1|t-1}^j$. Hence, there are two decision boundaries indicated in this figure: the outer boundary, denoted by the coarser dashed lines, corresponds to the previously associated track, $X_{t+1|t}^1$, while the inner boundary, denoted by the finer dashed lines, corresponds to the previously dissociated tracks. We observe, from Fig. 2, that tracks $X_{t+1|t}^1$ and $X_{t+1|t-1}^2$ associate with $z_{i't+1}$, but track $X_{t+1|t-1}^3$ does not associate with this observation. Consequently, the filtered track sets $X_{t+1|t+1}^{i'1}$ and $X_{t+1|t+1}^{i'2}$ may be combined in accordance with the algorithm provided in Section IV and then predicted to time $t + 2$. The predicted track set $X_{t+1|t-1}^3$ will be tested for termination and, if it survives that test, will be used to predict track $X_{t+2|t-1}^3$. These predicted tracks may then be tested for association with observation $z_{i''t+2}$, and the process may be continued until all data are processed.

Figure 3(a) illustrates a family of three crossing trajectories. The tracks move generally from left to right at time increases; the lines correspond to the x and y position components and the $+$ symbols correspond to noise-corrupted observations. Figure 3(b) displays the filtered track sets corresponding to this simulation, with the ellipses representing the projections of the track sets onto position space. The initial predicted track set $X_{0|-1}$ includes the entire field of view, and is not shown. The three large elliptical regions correspond to the filtered track sets after the first set of observations have been processed. As time increases, the size of these track sets



(a)



(b)

Fig. 3. Three crossing tracks: (a) Simulated trajectories and observations. (b) Filtered track sets.

decreases rapidly; for the first few observations corresponding to Tracks 1 and 3, however, there are multiple associations, since the tracks are fairly close and the track sets are still fairly large. As more confidence is obtained in the associations, these tracks become uniquely associated, and the elliptical regions decrease rapidly in size, and will asymptotically become point tracks. At sample ten, Tracks 1 and 2 nearly intersect, and both

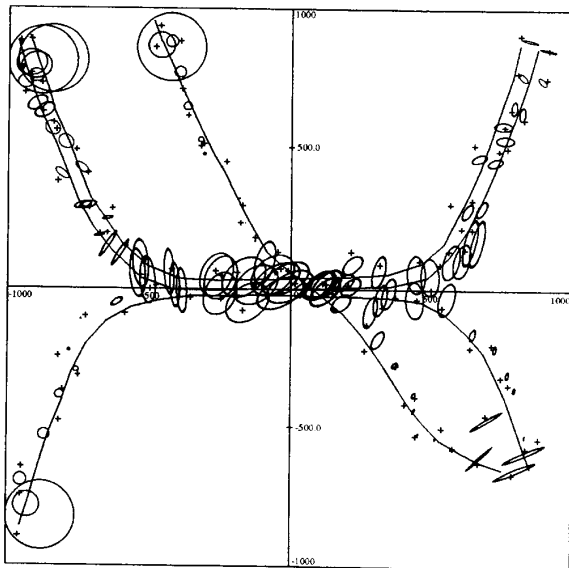
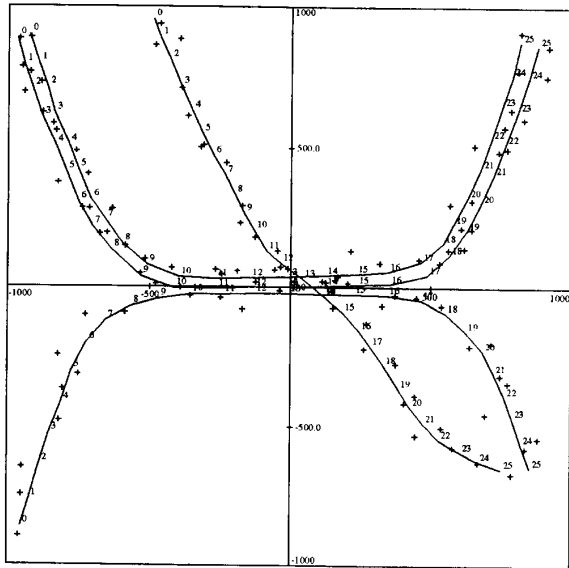


Fig. 4. Four closely spaced tracks: (a) Simulated trajectories and observations. (b) Filtered track sets.

tracks associate with the observations (conflicts of the first kind). These multiple associations persist for a few samples, but as the tracks diverge, the associations again become unique. Track 2 is unambiguously associated and estimated, as evidenced by the radius of the track set converging to zero, and the set-valued estimates asymptotically become point-valued.

A more complicated scenario is illustrated in Fig. 4, consisting of four trajectories with closely spaced track segments. Under such circumstances, the probability of misassociation will be extremely high. In regions of association conflict,

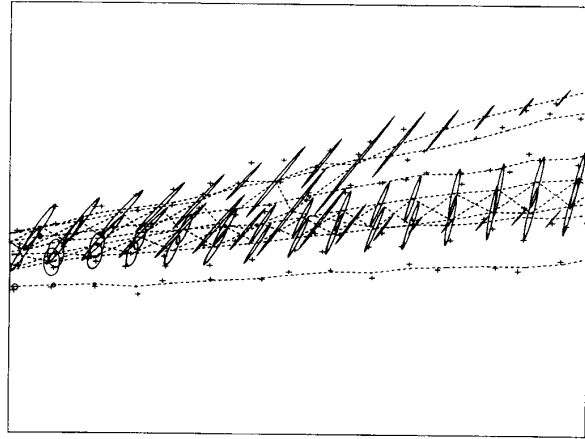


Fig. 5. Tracking with $b = 0.0$.

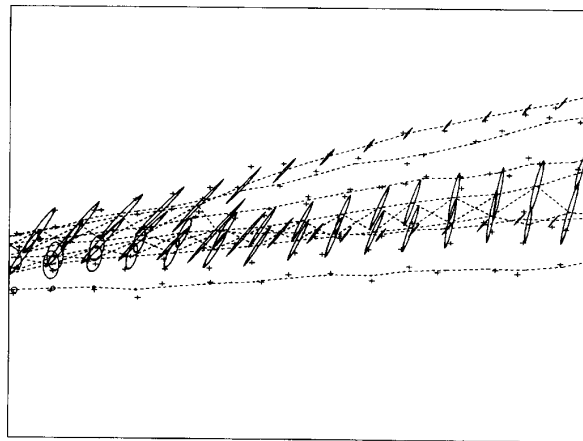


Fig. 6. Tracking with $b = 0.4$.

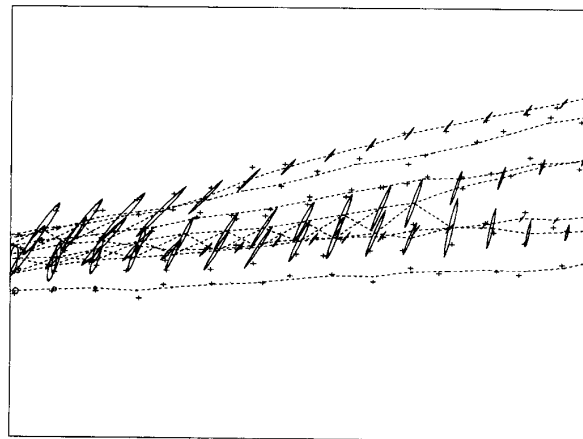


Fig. 7. Tracking with $b = 1$.

however, the ALRT methodology increases the size of the track sets to compensate for the highly ambiguous associations.

While it is impossible to distinguish exactly which track is associated with which data, it may be argued that this is as it should be, since there is insufficient information available to make unique associations with small probability of error. Thus, the set-valued estimates, though less precise than would be point-valued estimates in the ideal case, may ultimately be more useful, since they do not force unique decisions in an arbitrary and objectionable way.

To illustrate the effect of the boldness b on the tracking performance, we ran a simulation of seven diverging tracks. The tracks emerge from a single point and separate over about 50 time steps. This is a scenario that might be typical for a multiple-warhead vehicle. Each track has a different observed intensity which helps to distinguish between the targets, but is not a definitive aid because the intensities are similar. The results of the tracking algorithm are shown in Figs. 5 through 7 for $b = 0$, $b = 0.4$ and $b = 1$. In these figures, the + signs represent observations and the track estimates are shown using ellipses (which collapse to points). The estimated track lines are shown with dashed lines. Crossover from one apparent track to another is due to conflict of the first kind, which occurs when multiple observations associate with a single track. The higher b is, the less likelihood there is for such conflicts to occur. On the other hand, if b is too large, an association that should occur (as we would interpret the data) does not take place, and a new track is initiated. There are other factors that affect the performance of the tracking, namely the size of the seriously possible region (determined by θ) and the initial track size. Optimal settings of the boldness and other system parameters is a topic of ongoing investigation.

Simulations have also been run to test tracking in the presence of target clutter and dropout. Due to space limitations and the difficulty of interpreting the display on a black-and-white medium, the simulation results are not graphically displayed here. The results, however, are described briefly. In the case of data dropout (missed detection), an unassociated track propagates forward with the track ellipse growing over time. When the data reappears, the track ellipse begins shrinking as before. When clutter is present, new tracks are initiated for a short number of observations, then pruned out according to the pruning likelihood ratio test.

V. DISCUSSION

Many MTT algorithms, indeed many pattern recognition methodologies, are designed to optimize the performance of various components of the joint decision/estimation problem, regardless of the overall performance that results. The theme of this paper, however, runs somewhat counter to the dominant trends of estimation and decision to optimize performance. It is certainly important to extract all information possible from a given database, but in doing so, we may run the risk of attempting to squeeze more from the database than it contains. When this happens the results may lose credibility. The goal of the methodology presented in this paper is to develop a mechanism that adjusts the quality of the algorithm output to match the information content of the data. If the quality of the collection is degraded, then the preciseness of the decisions and the estimates must also be diminished.

Clearly needed to complete the investigation are comparisons of this method with other techniques (such as JPDA) and a more thorough evaluation of the effect of the boldness and other parameters. These investigations are being pursued.

APPENDIX—SOME NUMERICAL ISSUES FOR ELLIPSOIDAL SETS

Computation of the ALRT requires determination of $\underline{x} \in \Xi_{it} \cap \underline{X}_{t|t-1}^j$ such that (43) is satisfied. If \underline{z}_{it} is not in the projection of the track set $\underline{X}_{t|t-1}$ onto the observation space, then the closest distance from the observation \underline{z}_{it} to the track set $\underline{X}_{t|t-1}$ must be determined. The projection of the track ellipse onto the space of observations can be expressed as

$$\|L\underline{z}\| \leq 1 \quad (58)$$

where

$$L^T L = \left(H \begin{bmatrix} K_{t|t-1}^j \\ K_{t|t-1}^j \end{bmatrix} \begin{bmatrix} K_{t|t-1}^j \\ K_{t|t-1}^j \end{bmatrix}^T H^T \right)^{-1} \quad (59)$$

and the norm is defined as

$$\|\underline{z}\|^2 = \underline{z}^T \underline{z}. \quad (60)$$

For an observation \underline{z}_{it} outside the track set, the nearest point to the track set is that \underline{z} which satisfies

$$\min \|\underline{z} - \underline{z}_{it}\| \quad (61)$$

subject to

$$\|L\underline{z}\| = 1. \quad (62)$$

This is a constrained quadratic minimization problem of the type discussed in [14]. The normal equation can be readily written as

$$(I + \lambda L^T L)\underline{z} = \underline{z}_{it} \quad (63)$$

where λ is a Lagrange multiplier. It remains to determine λ , from which \underline{z} follows readily. Solving (63) for \underline{z} in terms of the unknown λ ,

$$\underline{z}(\lambda) = (I + \lambda L^T L)^{-1} \underline{z}_{it}. \quad (64)$$

To simplify this equation, it is convenient to compute the SVD decomposition of L as

$$L = UDV^T \quad (65)$$

where $D = \text{diag}[d_1, d_2, \dots, d_r]$ and U and V are orthogonal matrices. Using the orthogonality properties, it is straightforward to show that

$$\underline{z}(\lambda) = V(I + \lambda D^2)^{-1} \underline{e} \quad (66)$$

where $\underline{e} = V^T \underline{z}_{it}$ and

$$Lz(\lambda) = UD(I + \lambda D^2)^{-1} \underline{e} \quad (67)$$

from which the constraint $\|L\underline{z}\|^2 = 1$ becomes

$$\sum_{j=1}^r \frac{e_j^2 d_j^2}{(1 + \lambda e_j^2)^2} = 1. \quad (68)$$

Equation (68) is known [14] to have a unique solution for $\lambda > 0$, which by geometric arguments is the correct sign for λ

in this problem. Solution of this equation can be accomplished using a few Newton iterations. Substitution in (64) gives the desired value.

Note that if the track ellipses are aligned with the coordinate axes then the SVD decomposition becomes trivial.

REFERENCES

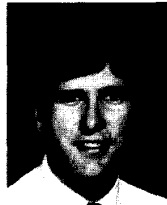
- [1] A. H. Jazwinski, *Stochastic Processes and Filtering Theory*. New York: Academic Press, 1970.
- [2] S. S. Blackman, *Multiple Target Tracking With Radar Applications*. Norwood, MA: Artech House, 1986.
- [3] Y. Bar-Shalom and T. E. Fortmann, *Tracking and Data Association*. San Diego, Ca: Academic Press, 1988.
- [4] Y. N. Chung, D. L. Gustafson, and E. Emre, "Extended solution to multiple maneuvering target tracking," *IEEE Trans. Aerospace and Electronic Sys.*, vol. 26, pp. 876-887, Sept. 1990.
- [5] A. K. Mahalanabis, B. Zhou, and N. K. Bose, "Improved multi-target tracking in clutter by PDA smoothing," *IEEE Trans. Aerospace and Electronic Sys.*, vol. 26, pp. 113-121, Jan. 1990.
- [6] C. K. Sword, M. Simaan, and E. W. Kamen, "Multiple target angle tracking using sensor array outputs," *IEEE Trans. Aerospace and Electronic Sys.*, vol. 26, pp. 367-373, Mar. 1990.
- [7] E. Emre and J. Seo, "A unifying approach to multitarget tracking," *IEEE Trans. Aerospace and Electronic Sys.*, vol. 25, pp. 520-528, July 1989.
- [8] I. Levi, *The Enterprise of Knowledge*. Cambridge, MA: MIT Press, 1980.
- [9] W. C. Stirling and D. R. Morrell, "Convex Bayes decision theory," *IEEE Trans. Syst., Man, Cybern.*, vol. 21, pp. 173-183, Jan./Feb. 1991.
- [10] D. R. Morrell and W. C. Stirling, "Set-valued filtering and smoothing," *IEEE Trans. Syst., Man, and Cybern.*, vol. 21, pp. 184-193, Jan./Feb. 1991.
- [11] T. K. Moon and S. E. Budge, "Classification using set-valued Kalman filtering and Levi's decision theory," *IEEE Trans. Syst., Man, Cybern.*, pp. 313-319, this issue.
- [12] G. Strang, *Introduction to Applied Mathematics*. Wellesley Cambridge, 1986.
- [13] W. C. Stirling, D. R. Morrell, G. R. Morell, and J. B. Thompson, "Convex Bayes decision theory and set-valued estimation: A new approach to multiple target tracking." Tech. Rep. TR-S107-91.1, Electrical and Computer Engineering Dept., Brigham Young Univ., Provo, UT, 1991.
- [14] W. Gander, "Least squares with a quadratic constraint," *Numerische Mathematik*, vol. 36, 1981.



Todd Moon (M'92) received the B.S. degree (summa cum laude) in electrical engineering and mathematics in 1988 and the M.S. degree in electrical engineering in 1988, both from Brigham Young University, and the Ph.D. degree in electrical engineering in 1991 from the University of Utah.

He is currently an Assistant Professor at Utah State University, where his research interests include decision theory, applications of wavelets, and speech processing. He has also worked as a consultant to industry in controls.

Dr. Moon is a member of Sigma Xi, Tau Beta Pi, and Phi Kappa Phi.



Scott Budge received the B.S., M.S., and Ph.D. degrees in electrical engineering from Brigham Young University in 1984, 1985, and 1990, respectively.

He joined the faculty at Utah State University in 1989, where he has been involved in research into image data compression algorithms for transmission systems and space-based observation platforms. His research interests are in the areas of data compression, image processing, adaptive systems, and pattern recognition.



Wynn C. Stirling (M'73) received the Honors B.A. degree (summa cum laude) in mathematics and the M.S. degree in electrical engineering from the University of Utah in 1969 and 1971, respectively, and the Ph.D. degree in electrical engineering from Stanford University, CA, in 1983.

From 1972 to 1975 he was with Rockwell International Corp., Anaheim, CA, and from 1975 to 1984 he was employed by ESL, Inc., Sunnyvale, CA. Since 1984 he has been with the Department of Electrical and Computer Engineering at Brigham

Young University, where he is an Associate Professor. His current research interests include decision and estimation theory, information theory, and stochastic processes.

Dr. Stirling is a member of Phi Beta Kappa and Tau Beta Pi.



Jeff Thompson received the M.S. degree in electrical and computer engineering from Brigham Young University in 1991.

He is with NCCOSC RDT&E (NRaD) in San Diego, CA. His current research continues with multiple-target tracking and now includes distributed data fusion and communications.