

Linear Optimal State Estimation in Systems with Independent Mode Transitions

Daniel Sigalov, Tomer Michaeli and Yaakov Oshman
Technion—Israel Institute of Technology

Abstract—A generalized state space representation of a dynamical system with random modes is presented. The dynamics equation includes the effect of the state’s linear minimum mean squared error (LMMSE) optimal estimate, representing the behavior of a closed loop control system featuring a state estimator. The measurement equation is allowed to depend on past LMMSE estimate of the state, which can be used to represent the fact that measurements are obtained from a validation window centered at the predicted measurement position and not from the entire surveillance region. The matrices comprising the system’s mode constitute an independent stochastic process. It is shown that the proposed formulation generalizes several important problems considered in the past, and allows a unified modeling of new ones. The LMMSE optimal filter is derived for the considered general problem and is shown to reduce, in some special cases, to some well known classical algorithms. The new concept, as well as the derived algorithm, are demonstrated for the problem of target tracking in clutter, and are shown to attain performance that is competitive to that of several popular nonlinear methods.

I. INTRODUCTION

State estimation in dynamical systems with randomly switching coefficients is an important problem in a variety of applications. Perhaps the most natural examples are maneuvering target tracking, and fault detection and isolation (FDI) methods featured, e.g., in aerospace navigation systems. Standard state space modeling presumes that the dynamics of the continuously-valued random state and, possibly, its measurement equation, are controlled by an evolving mode that takes discrete values, which is the well known concept of hybrid systems [1]. In problems involving uncertain, or intermittent observations [2]–[8], the mode affects the matrices of the measurement equation. In maneuvering target tracking applications [9]–[11], the mode usually affects the matrices of the dynamics equation.

We consider a state space representation of dynamical systems with switching coefficients that is more general than standard state space modeling. First, we allow the state and measurement equations to depend on functions of past measurements. Specifically, the state evolution dynamics depends, in addition to the previous state and process noise, on some latest estimate of the state vector. This modeling may represent the behavior of a closed loop control system featuring a state estimator. The measurement equation is

D. Sigalov is with the Interdepartmental Program for Applied Mathematics. dansigal@tx.technion.ac.il.

T. Michaeli is with the Department of Electrical Engineering. tomermic@tx.technion.ac.il.

Y. Oshman is with the Department of Aerospace Engineering. yaakov.oshman@technion.ac.il.

allowed to depend, in addition to the current state and measurement noise, on some current state estimate. This can be used to represent the fact that observations are not taken in the entire space, but in a small validation window set around the predicted state of the target. Further, we broaden the class of problems that may be treated under the hybrid systems framework by allowing the mode to take values in a continuous, rather than in some finite, domain. Finally, the random (matrix) coefficients may change not only in their values, but also in their dimensions, provided that the dimensions of the corresponding state vectors are properly chosen.

It is well known [9] that, even for the simplest case of independently switching modes, the optimal estimate of the state cannot be obtained without resorting to exhaustive enumeration. Therefore, significant efforts have been dedicated to developing suboptimal approaches. In the sequel we consider optimality within a narrower family of linear filters. Specifically, we derive a linear optimal algorithm that may be conveniently implemented in a recursive form, thus eliminating the need for unbounded memory. We also show that the filter reduces to previously reported results when the parameters of the underlying problem are adjusted appropriately.

To illustrate the applicability of the proposed framework we show how it can be used to model the problem of tracking a target in clutter, and solve it optimally in the linear MMSE sense. The derived filter is shown to attain performance comparable to that of several popular nonlinear approaches.

The remainder of the paper is organized as follows. In Section II we define and discuss the proposed dynamics and measurement models. Related contributions are surveyed in Section III. The linear optimal recursive state estimation algorithm is developed in Section IV. An application of the approach to target tracking in clutter is presented in Section V. A representative numerical example is given in Section VI. Concluding remarks are made in Section VII.

II. SYSTEM MODEL

We consider the following dynamical system:

$$x_{k+1} = A_k x_k + C_k \hat{x}_k + B_k w_k \quad (1a)$$

$$y_k = H_k x_k + F_k \hat{x}_{k-1} + G_k v_k, \quad (1b)$$

where $x_k \in \mathbb{R}^n$ and $y_k \in \mathbb{R}^m$ are, respectively, the state and measurement vectors at time k . The process and measurement noise sequences, $\{w_k\}$ and $\{v_k\}$, respectively, are zero mean, unit variance and white, and x_0 is a random vector

with mean \bar{x}_0 and covariance P_0 . The vectors $\hat{x}_k \in \mathbb{R}^n$ and $\hat{x}_{k-1} \in \mathbb{R}^n$ are linear optimal, in the Minimum Mean-Squared Error (MMSE) sense, estimates of x_k and x_{k-1} using the measurement histories \mathcal{Y}_k and \mathcal{Y}_{k-1} , respectively, where, $\mathcal{Y}_k \triangleq \{y_1, \dots, y_k\}$. Finally, $A_k, B_k, C_k, H_k, G_k, F_k$ are matrices of appropriate dimensions.

The mode of the system, $\mathcal{M}_k \triangleq \{A_k, B_k, C_k, H_k, G_k, F_k\}$, constitutes an independent random process such that the distribution at time k , $\mathbb{P}(\mathcal{M}_k)$, is known. The random quantities $\{w_k\}$, $\{v_k\}$, $\{\mathcal{M}_k\}$, and x_0 are assumed to be mutually independent.

Affecting the state at time $k+1$, the term $C_k \hat{x}_k$ represents the effect of a linear closed loop control input. Because the controller does not possess perfect knowledge of the state due to the process and measurement noises, a state estimator is used to generate an estimate of the state using the available measurements; this estimate is then used to generate the control signal $C_k \hat{x}_k$, where the gain C_k may be obtained as a solution of an appropriate optimal control problem.

Affecting the measurement at time k , the term $F_k \hat{x}_{k-1}$ can be used to represent the fact that observations are not taken in the entire (feasible) space, but, rather, in a small (admissible) validation window, set about the predicted state.

We seek to obtain a linear optimal estimate \hat{x}_{k+1} using the measurements $\mathcal{Y}_{k+1} \triangleq \{y_1, \dots, y_{k+1}\}$. It will be shown in the sequel that this linear optimal estimate conveniently possesses the following, recursive form:

$$\hat{x}_{k+1} = L_k \hat{x}_k + K_k y_{k+1}. \quad (2)$$

Namely, the desired linear optimal solution will be shown to fuse the (extrapolated) previous estimate with the most recently acquired measurement vector, effectively eliminating the need to store the entire measurement sequence.

The system (1) may be viewed as a generalization of the popular Jump Markov Linear System (JMLS) model, which is a special case of the hybrid systems paradigm, in which continuous uncertainties associated with the system state are accompanied by discrete ones representing the system mode. Specifically, in the present case, the system mode does not have to assume values in a discrete domain. Moreover, as opposed to the classical multiple model based treatment of hybrid systems, the present formulation allows evolution not only of the entries of the matrices constituting the mode, but also of their dimensions. This distinction provides a convenient basis for treating problems that, to the best of the authors' knowledge, have not been previously considered, as will be discussed in detail in Section V.

III. RELATED WORK AND MODEL GENERALITY

In the sequel we show how several estimation problems, that were considered and solved in the past, may be cast as special cases of the general formulation (1) addressed herein.

Nahi [2] considered the problem of state estimation with uncertain observations:

$$y_k = \gamma_k H_{\text{nom}} x_k + G_{\text{nom}} v_k, \quad (3)$$

where H_{nom} and G_{nom} are some known deterministic matrices representing sensor geometry and measurement noise covariances, respectively, and $\{\gamma_k\}$ is a sequence of independent Bernoulli random variables. Nahi's formulation (3) is obtained from (1b) by substituting $H_k = \gamma_k H_{\text{nom}}$, $G_k = G_{\text{nom}}$, and $F_k = 0$.

A generalization of Nahi's work has been recently proposed in [8], where multiplicative faults, $\{\gamma_k\}$, were accompanied by additive ones:

$$y_k = \gamma_k H_{\text{nom}} x_k + \delta_k G_{\text{nom}} v_k, \quad (4)$$

with $\{\delta_k\}$ being a sequence of random variables taking values in $\{1, C\}$, and $C > 1$ is some known constant. For the case of independently distributed $\{\gamma_k, \delta_k\}$ a linear optimal recursive algorithm has been developed. This formulation may be obtained from (1b) by setting $H_k = \gamma_k H_{\text{nom}}$, $G_k = \delta_k G_{\text{nom}}$, and $F_k = 0$.

In [12] Costa considered the problem of state estimation with a random mode variable taking values in some discrete domain, and derived a linear optimal estimation scheme. This modeling follows in a direct manner from the formulation proposed herein by setting $C_k = 0$ and $F_k = 0$. It is important to note that, unless the mode matrices A_k and B_k are deterministic, the solution proposed in [12] requires state augmentation, as opposed to the solution proposed here.

It should be noted that both [8] and [12] also consider, in addition to independently evolving modes, Markov mode dynamics, thus deviating from the focus of this paper. When the Markov property is degenerated to independently evolving modes, both contributions follow as special cases of the present work as discussed above.

Several additional contributions related to the considered problem include [3], in which a generalized version of Nahi's problem was addressed by allowing correlated fault indicators; [4], that allowed correlations between subsequent fault variables; [13] that proposed a linear optimal estimator for the Static Multiple Model (SMM) problem [14]; and [11], that considered linear optimal estimation in systems with bounded and unbounded numbers of actuator faults. Nonlinear suboptimal solutions for related problems were proposed in [9]–[11], [15] and references therein.

In passing, we note that all the contributions mentioned above consider modes taking values in some discrete (and finite) domain, but fail to address (except in a few simple cases) problems where a continuum of mode values should be assumed. Such cases include, for example, terrain-based target tracking applications, in which the measurement noise intensity may evolve in a continuous manner. In contradistinction, the model considered in this paper, as well as the algorithm developed in the sequel, do not make this restrictive assumption, and allow treatment of continuously as well as discretely varying modes, as long as $\{\mathcal{M}_k\}$ is an independent sequence.

IV. LINEAR OPTIMAL RECURSIVE ESTIMATION

It is well known [16], [17] that the estimate \hat{x}_{k+1} is optimal if and only if the following orthogonality conditions

are satisfied:

$$\mathbb{E}[(\hat{x}_{k+1} - x_{k+1})y_i^T] = 0, \quad i = 1, \dots, k+1. \quad (5)$$

Given a linear optimal estimate \hat{x}_k , we will search for matrices L_k and K_k such that \hat{x}_{k+1} in (2) satisfies (5). We separate orthogonality to past measurements, y_i , $i = 1, \dots, k$, from orthogonality to the latest observation, y_{k+1} .

A. Orthogonality w.r.t. Past Observations

Consider the conditions (5) for $i = 1, \dots, k$. Substituting x_{k+1} from (1a) as well as the conjectured form (2) yields

$$\begin{aligned} 0 &= \mathbb{E}[(A_k x_k + B_k w_k + C_k \hat{x}_k - L_k \hat{x}_k - K_k y_{k+1})y_i^T] \\ &= \mathbb{E}[(A_k x_k + C_k \hat{x}_k - L_k \hat{x}_k - K_k y_{k+1})y_i^T], \end{aligned} \quad (6)$$

which follows from the independence of B_k , w_k and y_i , $i = 1, \dots, k$. Substituting (1b) for y_{k+1} , (6) becomes

$$\begin{aligned} 0 &= \mathbb{E}[(A_k x_k + C_k \hat{x}_k - L_k \hat{x}_k \\ &\quad - K_k(H_{k+1}x_{k+1} + G_{k+1}v_{k+1} + F_{k+1}\hat{x}_k))y_i^T]. \end{aligned} \quad (7)$$

Utilizing the independence of G_{k+1} , v_{k+1} and y_i , $i = 1, \dots, k$, and substituting (1a) for x_{k+1} , (7) reads

$$\begin{aligned} 0 &= \mathbb{E}[(A_k x_k + C_k \hat{x}_k - L_k \hat{x}_k - (K_k H_{k+1} A_k x_k \\ &\quad + K_k H_{k+1} C_k \hat{x}_k + K_k F_{k+1} \hat{x}_k))y_i^T] \\ &= \mathbb{E}[(A_k - K_k H_{k+1} A_k)x_k y_i^T \\ &\quad + (C_k - K_k H_{k+1} C_k - L_k - K_k F_{k+1})\hat{x}_k y_i^T], \end{aligned} \quad (8)$$

where we have also utilized the independence of B_k , w_k and y_i , $i = 1, \dots, k$. Note that $x_k y_i^T$ and $\hat{x}_k y_i^T$ are deterministic functions of $\{w_j, A_j, B_j, C_j\}_{j=1}^{k-1}$, $\{v_j, H_j, G_j, F_j\}_{j=1}^k$, and x_0 . Bearing in mind that the mode sequence is independent, we conclude that for $i = 1, \dots, k$, the random matrices $x_k y_i^T$ and $\hat{x}_k y_i^T$ are independent of $\{A_k, H_{k+1}\}$ and $\{C_k, F_{k+1}, H_{k+1}\}$, respectively.

By assumption, \hat{x}_k is a linear optimal estimate of x_k , satisfying the orthogonality conditions w.r.t. y_i , $i = 1, \dots, k$. In particular,

$$\mathbb{E}[\hat{x}_k y_i^T] = \mathbb{E}[x_k y_i^T], \quad i = 1, \dots, k, \quad (9)$$

which is essentially the same relation as (5), satisfied by x_k and \hat{x}_k . Combining (9) with the above independence property, (8) reads

$$\begin{aligned} &\mathbb{E}[(A_k + C_k - L_k - K_k(H_{k+1}(A_k + C_k) + F_{k+1})) \\ &\quad \times \mathbb{E}[x_k y_i^T] = 0. \end{aligned} \quad (10)$$

Finally, utilizing the independence of H_{k+1} and A_k , C_k , we find that orthogonality with respect to the measurements y_i , $i = 1, \dots, k$ is guaranteed by requiring that the matrices L_k and K_k satisfy the following relation:

$$L_k = (I - K_k \mathbb{E}[H_{k+1}]) \mathbb{E}[A_k + C_k] - K_k \mathbb{E}[F_{k+1}]. \quad (11)$$

B. Orthogonality w.r.t. Latest Observation

First, we define the following symmetric matrices:

$$S_k \triangleq \mathbb{E}[x_k x_k^T] \quad (12)$$

$$U_k \triangleq \mathbb{E}[\hat{x}_k \hat{x}_k^T] = \mathbb{E}[\hat{x}_k x_k^T], \quad (13)$$

where the RHS of (13) follows from the orthogonality principle satisfied by x_k . Substituting (2) and (1b) in the last orthogonality condition of (5) we have:

$$\begin{aligned} &\mathbb{E}[(\hat{x}_{k+1} - x_{k+1})y_{k+1}^T] \\ &= L_k \mathbb{E}[\hat{x}_k y_{k+1}^T] + K_k \mathbb{E}[y_{k+1} y_{k+1}^T] \\ &\quad - \mathbb{E}[x_{k+1}(H_{k+1}x_{k+1} + G_{k+1}v_{k+1} + F_{k+1}\hat{x}_k)^T] \\ &= L_k \mathbb{E}[\hat{x}_k y_{k+1}^T] + K_k \mathbb{E}[y_{k+1} y_{k+1}^T] \\ &\quad - S_{k+1} \mathbb{E}[H_{k+1}^T] - \mathbb{E}[x_{k+1} \hat{x}_k^T] \mathbb{E}[F_{k+1}^T], \end{aligned} \quad (14)$$

where the last transition follows from the independence of x_{k+1} and $G_{k+1}v_{k+1}$ as well as the independence of F_{k+1} and $x_{k+1} \hat{x}_k^T$. We compute next each of the terms $\mathbb{E}[\hat{x}_k y_{k+1}^T]$, $\mathbb{E}[x_{k+1} \hat{x}_k^T]$, and $\mathbb{E}[y_{k+1} y_{k+1}^T]$ separately.

Using (1b), the independence of \hat{x}_k and $G_{k+1}v_{k+1}$, the independence of $\{\hat{x}_k, x_{k+1}\}$ and H_{k+1} , and the independence of \hat{x}_k and F_{k+1} , we have:

$$\begin{aligned} &\mathbb{E}[\hat{x}_k y_{k+1}^T] \\ &= \mathbb{E}[\hat{x}_k (H_{k+1}x_{k+1} + G_{k+1}v_{k+1} + F_{k+1}\hat{x}_k)^T] \\ &= \mathbb{E}[\hat{x}_k (H_{k+1}x_{k+1} + F_{k+1}\hat{x}_k)^T] \\ &= \mathbb{E}[\hat{x}_k x_{k+1}^T] \mathbb{E}[H_{k+1}^T] + U_k \mathbb{E}[F_{k+1}^T] \\ &= \mathbb{E}[\hat{x}_k (A_k x_k + B_k w_k + C_k \hat{x}_k)^T] \mathbb{E}[H_{k+1}^T] + U_k \mathbb{E}[F_{k+1}^T] \\ &= U_k (\mathbb{E}[A_k^T + C_k^T]) \mathbb{E}[H_{k+1}^T] + U_k \mathbb{E}[F_{k+1}^T], \end{aligned} \quad (15)$$

where the fourth transition follows from using (1a), and the last one from the independence of $\{B_k, w_k\}$ and \hat{x}_k , the independence of $\{A_k, C_k\}$ and $\{x_k, \hat{x}_k\}$, and the definition (13).

Similar arguments yield:

$$\mathbb{E}[x_{k+1} \hat{x}_k^T] = \mathbb{E}[A_k + C_k] U_k. \quad (16)$$

Finally, due to the independence of $H_{k+1}x_{k+1}$ and $G_{k+1}v_{k+1}$,

$$\begin{aligned} &\mathbb{E}[y_{k+1} y_{k+1}^T] = \mathbb{E}[(H_{k+1}x_{k+1} + G_{k+1}v_{k+1} + F_{k+1}\hat{x}_k) \\ &\quad \times (H_{k+1}x_{k+1} + G_{k+1}v_{k+1} + F_{k+1}\hat{x}_k)^T] \\ &= \mathbb{E}[H_{k+1}x_{k+1}x_{k+1}^T H_{k+1}^T] + \mathbb{E}[G_{k+1}v_{k+1}v_{k+1}^T G_{k+1}^T] \\ &\quad + \mathbb{E}[F_{k+1}\hat{x}_k \hat{x}_k^T F_{k+1}^T] + \mathbb{E}[H_{k+1}] \mathbb{E}[x_{k+1} \hat{x}_k^T] \mathbb{E}[F_{k+1}^T] \\ &\quad + \mathbb{E}[F_{k+1}] \mathbb{E}[\hat{x}_k x_{k+1}^T] \mathbb{E}[H_{k+1}^T], \end{aligned}$$

which, using (16) becomes

$$\begin{aligned} &\mathbb{E}[y_{k+1} y_{k+1}^T] = \mathbb{E}[H_{k+1}x_{k+1}x_{k+1}^T H_{k+1}^T] \\ &\quad + \mathbb{E}[G_{k+1}v_{k+1}v_{k+1}^T G_{k+1}^T] + \mathbb{E}[F_{k+1}\hat{x}_k \hat{x}_k^T F_{k+1}^T] \\ &\quad + \mathbb{E}[H_{k+1}] \mathbb{E}[A_k + C_k] U_k \mathbb{E}[F_{k+1}^T] \\ &\quad + \mathbb{E}[F_{k+1}] U_k \mathbb{E}[A_k^T + C_k^T] \mathbb{E}[H_{k+1}^T]. \end{aligned} \quad (17)$$

It remains to compute $\mathbb{E}[H_{k+1}x_{k+1}x_{k+1}^T H_{k+1}^T]$, $\mathbb{E}[G_{k+1}v_{k+1}v_{k+1}^T G_{k+1}^T]$, and $\mathbb{E}[F_{k+1}\hat{x}_k \hat{x}_k^T F_{k+1}^T]$. Consider

the first term. From the smoothing property of the conditional expectation,

$$\begin{aligned} & \mathbb{E}[H_{k+1}x_{k+1}x_{k+1}^T H_{k+1}^T] \\ &= \mathbb{E}[\mathbb{E}[H_{k+1}x_{k+1}x_{k+1}^T H_{k+1}^T \mid H_{k+1}]] \\ &= \mathbb{E}[H_{k+1}\mathbb{E}[x_{k+1}x_{k+1}^T \mid H_{k+1}] H_{k+1}^T] \\ &= \mathbb{E}[H_{k+1}S_{k+1}H_{k+1}^T], \end{aligned} \quad (18)$$

where the conditioning on H_{k+1} was omitted due to the independence of x_{k+1} and H_{k+1} . Bearing in mind that $\mathbb{E}[v_{k+1}v_{k+1}^T] = I$, and $\mathbb{E}[\hat{x}_k\hat{x}_k^T] = U_k$, the other two terms are computed in a similar manner and result in the following:

$$\mathbb{E}[G_{k+1}v_{k+1}v_{k+1}^T G_{k+1}^T] = \mathbb{E}[G_{k+1}G_{k+1}^T] \quad (19)$$

$$\mathbb{E}[F_{k+1}\hat{x}_k\hat{x}_k^T F_{k+1}^T] = \mathbb{E}[F_{k+1}U_k F_{k+1}^T]. \quad (20)$$

For future reference, we also note that

$$\mathbb{E}[A_k x_k x_k^T A_k^T] = \mathbb{E}[A_k S_k A_k^T] \quad (21)$$

$$\mathbb{E}[A_k x_k \hat{x}_k^T C_k^T] = \mathbb{E}[A_k U_k C_k^T] \quad (22)$$

$$\mathbb{E}[C_k \hat{x}_k \hat{x}_k^T C_k^T] = \mathbb{E}[C_k U_k C_k^T] \quad (23)$$

$$\mathbb{E}[B_k w_k w_k^T B_k^T] = \mathbb{E}[B_k B_k^T], \quad (24)$$

where (24) follows from the fact that $\{w_k\}$ is a unit-variance sequence. Using (18)-(20) in (17) and substituting (15)-(17) in (14) yields the second relation between L_k and K_k :

$$\begin{aligned} 0 &= L_k U_k (\mathbb{E}[A_k^T + C_k^T] \mathbb{E}[H_{k+1}^T] + \mathbb{E}[F_{k+1}^T]) \\ &+ K_k \left(\mathbb{E}[H_{k+1} S_{k+1} H_{k+1}^T] \right. \\ &+ \mathbb{E}[H_{k+1}] \mathbb{E}[A_k + C_k] U_k \mathbb{E}[F_{k+1}^T] \\ &+ \mathbb{E}[G_{k+1} G_{k+1}^T] \\ &+ \mathbb{E}[F_{k+1}] U_k \mathbb{E}[A_k^T + C_k^T] \mathbb{E}[H_{k+1}^T] \\ &+ \left. \mathbb{E}[F_{k+1} U_k F_{k+1}^T] \right) \\ &- S_{k+1} \mathbb{E}[H_{k+1}^T] - \mathbb{E}[A_k + C_k] U_k \mathbb{E}[F_{k+1}^T]. \end{aligned} \quad (25)$$

Using (11) and canceling out identical terms yields

$$\begin{aligned} K_k &= (S_{k+1} - V_{k+1}) \mathbb{E}[H_{k+1}^T] \\ &\times \left(\mathbb{E}[H_{k+1} S_{k+1} H_{k+1}^T] - \mathbb{E}[H_{k+1}] V_{k+1} \mathbb{E}[H_{k+1}^T] \right. \\ &+ \mathbb{E}[F_{k+1} U_k F_{k+1}^T] - \mathbb{E}[F_{k+1}] U_k \mathbb{E}[F_{k+1}^T] \\ &+ \left. \mathbb{E}[G_{k+1} G_{k+1}^T] \right)^{-1}, \end{aligned} \quad (26)$$

where

$$V_{k+1} \triangleq \mathbb{E}[A_k + C_k] U_k \mathbb{E}[A_k^T + C_k^T]. \quad (27)$$

Next we provide recursive mechanisms for the computation of S_{k+1} and U_{k+1} .

C. Computation of Second Order Moments Matrices

Consider first the second-order moment matrix of x_{k+1} . We have

$$\begin{aligned} S_{k+1} &= \mathbb{E}[x_{k+1}x_{k+1}^T] \\ &= \mathbb{E}[(A_k x_k + B_k w_k + C_k \hat{x}_k)(A_k x_k + B_k w_k + C_k \hat{x}_k)^T] \\ &= \mathbb{E}[A_k S_k A_k^T] + \mathbb{E}[A_k U_k C_k^T] + \mathbb{E}[C_k U_k A_k^T] \\ &+ \mathbb{E}[C_k U_k C_k^T] + \mathbb{E}[B_k B_k^T], \end{aligned} \quad (28)$$

where we utilized the independence of $B_k w_k$ and $\{A_k x_k, C_k \hat{x}_k\}$ and (21)-(24).

Next consider U_{k+1} . Using (1b) and (2) we have:

$$\begin{aligned} U_{k+1} &= \mathbb{E}[\hat{x}_{k+1}x_{k+1}^T] \\ &= \mathbb{E}[(L_k \hat{x}_k + K_k y_{k+1})x_{k+1}^T] \\ &= L_k \mathbb{E}[\hat{x}_k x_{k+1}^T] + K_k \mathbb{E}[(H_{k+1}x_{k+1} + F_{k+1}\hat{x}_k)x_{k+1}^T] \\ &= (L_k + K_k \mathbb{E}[F_{k+1}]) \mathbb{E}[\hat{x}_k x_{k+1}^T] + K_k \mathbb{E}[H_{k+1}] S_{k+1}, \end{aligned} \quad (29)$$

where the fifth transition follows from the independence of $\{G_{k+1}v_{k+1}, H_{k+1}\}$ and x_{k+1} , and the independence of $\hat{x}_k x_{k+1}^T$ and F_{k+1} . Using (16) and taking into account the fact that U_k is symmetric yields

$$\begin{aligned} U_{k+1} &= (L_k + K_k \mathbb{E}[F_{k+1}]) U_k \mathbb{E}[A_k^T + C_k^T] \\ &+ K_k \mathbb{E}[H_{k+1}] S_{k+1}. \end{aligned} \quad (30)$$

D. Filter Summary

The complete algorithm comprises the following steps:

- a) *Initialization*: $\hat{x}_0 = \bar{x}_0$, $S_0 = P_0 + \bar{x}_0 \bar{x}_0^T$, $U_0 = \bar{x}_0 \bar{x}_0^T$.
- b) *Recursion*: For $k = 1, 2, \dots$ perform the main routine summarized in Alg. 1.

Algorithm 1

Input: $y_{k+1}, \hat{x}_k, S_k, U_k$

- 1: Using the known distribution of \mathcal{M}_k compute $\mathbb{E}[A_k]$, $\mathbb{E}[B_k B_k^T]$, $\mathbb{E}[C_k]$, $\mathbb{E}[A_k S_k A_k^T]$, $\mathbb{E}[A_k U_k C_k^T]$, and $\mathbb{E}[C_k U_k C_k^T]$.
- 2: Compute V_{k+1} and S_{k+1} using Eqs. (27) and (28).
- 3: Using the known distribution of \mathcal{M}_{k+1} compute $\mathbb{E}[H_{k+1}]$, $\mathbb{E}[G_{k+1} G_{k+1}^T]$, $\mathbb{E}[F_{k+1}]$, $\mathbb{E}[F_{k+1} U_k F_{k+1}^T]$, and $\mathbb{E}[H_{k+1} S_{k+1} H_{k+1}^T]$.
- 4: Compute K_k using Eq. (26).
- 5: Compute L_k using Eq. (11).
- 6: Compute U_{k+1} using Eq. (30).
- 7: Compute \hat{x}_{k+1} using Eq. (2).

Output: $\hat{x}_{k+1}, S_{k+1}, U_{k+1}$

E. Special Cases

The derived algorithm is a general, linear MMSE solution to the estimation problem posed in Section II. We next consider several special cases it reduces to, when the general problem is degenerated appropriately.

1) *Standard Kalman Filter*: A standard Kalman Filter (KF) solution is obtained for the case where $\{\mathcal{M}_k\}$ is a deterministic sequence with $C_k = 0$, $F_k = 0$. In this case, (11) and (26) become KF matrices when the time and measurement updates are combined in a single update:

$$L_k = A_k - K_k H_{k+1} A_k, \quad (31)$$

$$K_k = P_{k+1} H_{k+1}^T (H_{k+1} P_{k+1} H_{k+1}^T + G_{k+1} G_{k+1}^T)^{-1}. \quad (32)$$

Here $P_{k+1} \triangleq S_{k+1} - V_{k+1}$ is the KF estimation error covariance, since, in the considered case, $V_{k+1} = A_k U_k A_k^T$, which is the second-order moment of the estimate of x_k using $\{y_1, \dots, y_{k-1}\}$ (the one-step ahead predicted estimate of x_k).

2) *Generalized Kalman Filter*: By letting the mode sequence $\{\mathcal{M}_k\}$ be deterministic we may generalize the standard KF to incorporate the additional terms of (1):

$$L_k = (I - K_k H_{k+1}) (A_k + C_k) - K_k F_{k+1}, \quad (33)$$

$$K_k = P_{k+1} H_{k+1}^T (H_{k+1} P_{k+1} H_{k+1}^T + G_{k+1} G_{k+1}^T)^{-1}. \quad (34)$$

Similarly to the previous case, $P_{k+1} \triangleq S_{k+1} - V_{k+1}$ is the KF estimation error covariance. Note that the filter compensates for the deterministic, (non-informative in this case) term $F_k \hat{x}_{k-1}$ in the measurement equation by subtracting $K_k F_{k+1}$ from the nominal L_k . The resulting filter operates in essentially the same manner as the standard KF.

3) *Estimation with Uncertain Observations*: Considered in [2], the case of linear optimal estimation with uncertain observations follows by setting $A_k = A$, $B_k = B$, $C_k = 0$, $F_k = 0$, $G_k = G_{\text{nom}}$, and $H_k = \gamma_k H_{\text{nom}}$. In this case, (11) and (26) become

$$L_k = A - K_k \mathbb{E}[\gamma_{k+1}] H_{\text{nom}} A, \quad (35)$$

$$K_k = \mathbb{E}[\gamma_{k+1}] (S_{k+1} - V_{k+1}) H_{\text{nom}}^T \\ \times \left(\mathbb{E}[\gamma_{k+1}] H_{\text{nom}} S_{k+1} H_{\text{nom}}^T - (\mathbb{E}[\gamma_{k+1}])^2 H_{\text{nom}} V_{k+1} H_{\text{nom}}^T + G_{\text{nom}} G_{\text{nom}}^T \right)^{-1}, \quad (36)$$

which coincide with Nahi's solution [2]. Similarly, the linear optimal estimator for random fault variables considered in [8] follows by setting $A_k = A$, $B_k = B$, $C_k = 0$, $F_k = 0$, $G_k = \delta_k G_{\text{nom}}$, and $H_k = \gamma_k H_{\text{nom}}$.

4) *Estimation with Actuator Faults*: Considered in [18], the problem of tracking with actuator faults follows by setting $A_k \in \{A^1, \dots, A^r\}$, $B_k \in \{B^1, \dots, B^r\}$, such that $\mathbb{P}\{\{A_k, B_k\} = \{A^i, B^i\}\} = p_i$, $C_k = 0$, $F_k = 0$, $G_k = G_{\text{nom}}$, $H_k = H_{\text{nom}}$. In this case, (11) and (26) read

$$L_k = \mathbb{E}[A_k] - K_k H_{\text{nom}} \mathbb{E}[A_k], \quad (37)$$

$$K_k = (V_{k+1} - S_{k+1}) H_{\text{nom}}^T \\ \times (H_{k+1} (S_{k+1} - V_{k+1}) H_{\text{nom}}^T + G_{\text{nom}} G_{\text{nom}}^T)^{-1}. \quad (38)$$

V. APPLICATION: TARGET TRACKING IN CLUTTER

A. System and Clutter Models

Consider a single target obeying a linear dynamical model. The evolution of the state is obtained from (1a) by setting $A_k = A$, $B_k = B$, and $C_k = 0$ which results in

$$x_{k+1} = A x_k + B w_k. \quad (39)$$

Here A and B are deterministic matrices representing the state dynamics and the square root of the process noise covariance, respectively.

At time k the target state is observed via the following linear measurement equation:

$$y_{k,\text{true}} = H_{\text{nom}} x_k + G_{\text{nom}} v_{k,\text{true}}. \quad (40)$$

Here, $y_{k,\text{true}}$ and $v_{k,\text{true}}$ represent the true measurement of the target and the true measurement noise, respectively.

In addition to the actual measurement, $y_{k,\text{true}}$, a number of clutter measurements are obtained. These will be denoted as $y_{k,\text{cl}}$. Clutter measurements originate from false (or ghost) targets and do not carry any information about the target of interest. They are, however, indistinguishable from true detections. At each time, the clutter measurements are assumed to be independent of each other, of the clutter measurements at other times, and of the true state and observation. In addition, we assume that these measurements are uniformly distributed in space, which is a common assumptions in such applications [19].

Instead of scanning the entire surveillance region, the sensor initiates a validation window centered at the predicted target position, and the algorithm operates on measurements obtained within this window. Since the clutter is uniformly distributed in space, it is also uniformly distributed within the validation window. For simplicity of exposition, we assume that the true measurement is always present in the validation window. This assumption can be relaxed, as discussed in the sequel.

In this setting, the acquired measurement vector at time k , y_k , is a concatenation of N measurements, for some integer N , such that $N - 1$ of them originate from false targets, or clutter, and only one originates from the (single) true target. The false measurements are centered around the predicted target state (since this is the center of the validation window), as opposed to the true measurement which is generated using the true target state (40).

The described observation model follows from (1b) by defining the following probability distribution of the mode \mathcal{M}_k which comprises the matrices $\{H_k, G_k, F_k\}$, such that:

$$\{H_k, G_k, F_k\} = \left\{ \begin{array}{l} \left\{ \left(\begin{array}{c} H_{\text{nom}} \\ 0 \\ \vdots \\ 0 \end{array} \right), \text{diag} \left(\begin{array}{c} G_{\text{nom}} \\ G_{\text{cl}} \\ \vdots \\ G_{\text{cl}} \end{array} \right), \left(\begin{array}{c} 0 \\ H_{\text{nom}} A \\ \vdots \\ H_{\text{nom}} A \end{array} \right) \right\}, \text{ w.p. } p_1 \\ \left\{ \left(\begin{array}{c} 0 \\ H_{\text{nom}} \\ 0 \\ \vdots \\ 0 \end{array} \right), \text{diag} \left(\begin{array}{c} G_{\text{cl}} \\ G_{\text{nom}} \\ G_{\text{cl}} \\ \vdots \\ G_{\text{cl}} \end{array} \right), \left(\begin{array}{c} H_{\text{nom}} A \\ 0 \\ H_{\text{nom}} A \\ \vdots \\ H_{\text{nom}} A \end{array} \right) \right\}, \text{ w.p. } p_2 \\ \vdots \\ \left\{ \left(\begin{array}{c} 0 \\ \vdots \\ 0 \\ H_{\text{nom}} \end{array} \right), \text{diag} \left(\begin{array}{c} G_{\text{cl}} \\ \vdots \\ G_{\text{cl}} \\ G_{\text{nom}} \end{array} \right), \left(\begin{array}{c} H_{\text{nom}} A \\ \vdots \\ H_{\text{nom}} A \\ 0 \end{array} \right) \right\}, \text{ w.p. } p_N, \end{array} \right. \quad (41)$$

where the following notation was introduced:

$$\text{diag} \begin{pmatrix} A_1 \\ A_2 \\ \vdots \\ A_N \end{pmatrix} \triangleq \begin{pmatrix} A_1 & 0 & \dots & 0 \\ 0 & A_2 & 0 & 0 \\ \vdots & \ddots & \ddots & \vdots \\ 0 & \dots & 0 & A_N \end{pmatrix}. \quad (42)$$

Here, G_{cl} is the square-root of the clutter covariance matrix.

Consider, for example, the first realization of $\{H_k, G_k, F_k\}$ in (41). Namely,

$$\{H_k, G_k, F_k\} = \left\{ \begin{pmatrix} H_{\text{nom}} \\ 0 \\ \vdots \\ 0 \end{pmatrix}, \text{diag} \begin{pmatrix} G_{\text{nom}} \\ G_{\text{cl}} \\ \vdots \\ G_{\text{cl}} \end{pmatrix}, \begin{pmatrix} 0 \\ H_{\text{nom}}A \\ \vdots \\ H_{\text{nom}}A \end{pmatrix} \right\}. \quad (43)$$

This realization corresponds to the case where the first of the N acquired measurements is the true target measurement, $y_{k,\text{true}}$, generated according to (40). The other $N - 1$ measurements are clutter, each of which is generated according to the following model:

$$y_{k,\text{cl}} = H_{\text{nom}}A\hat{x}_{k-1} + G_{\text{cl}}v_{k,\text{cl}}. \quad (44)$$

Note that $H_{\text{nom}}A\hat{x}_{k-1}$ is the predicted measurement at time k , which is also the center of the validation window as set by the sensor. Thus, clutter measurements acquired at time k may be viewed as generated uniformly around this quantity. We make the distinction between the true and clutter measurement noises ($v_{k,\text{true}}$ and $v_{k,\text{cl}}$, respectively) to emphasize that the two types of observations are generated with different driving noises that are independent, in compliance with the modeling assumption of Section II, where v_k is a unit variance vector.

The fact that exactly one of the N observations is target-originated is reflected in (41) by the fact that exactly one of the blocks of F_k is set to H_{nom} with others being set to 0. Likewise, all but one block of F_k are taken to be $H_{\text{nom}}A$.

Captured by the matrix G_{cl} , the covariance of the clutter measurements is typically higher than the true measurement noise covariance (represented by G_{nom}). However, it is not known a-priori which of the concatenated measurements carries useful information. We assume that any ordering of the true and clutter observations in the validation window is equiprobable and, thus, set $p_i = \frac{1}{N}$, $i = 1, \dots, N$. Hence H_k , G_k , and F_k correspond to random permutations of N possible positions of the true measurement among the clutter measurements. Note that the overall number of validated measurements (i.e., those that are in the validation window), N , is assumed to be known, but may vary in time.

As previously stated, we assume herein, for simplicity of exposition, that the true measurement is always present in the validation window. This is not a very restrictive assumption, that may be relaxed without sacrificing the optimality of Alg. 1 in the linear MMSE sense. To account for the possibility that the true measurement does not fall in the validation window with $N > 0$, the following option needs to

be added to the set of possible realizations of $\{H_k, G_k, F_k\}$ in (41)

$$\left\{ \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix}, \text{diag} \begin{pmatrix} G_{\text{cl}} \\ G_{\text{cl}} \\ \vdots \\ G_{\text{cl}} \end{pmatrix}, \begin{pmatrix} H_{\text{nom}}A \\ H_{\text{nom}}A \\ \vdots \\ H_{\text{nom}}A \end{pmatrix} \right\}. \quad (45)$$

Note that if $N = 0$, namely, there are no measurements in the validation window, (2) becomes $\hat{x}_{k+1} = L_k\hat{x}_k$, which corresponds to a simple prediction (time update) without a consecutive measurement update.

B. Matrix Computations

Since the matrices in (39) are deterministic, we have

$$\mathbb{E}[A_k] = A_k, \mathbb{E}[B_k B_k^T] = B_k B_k^T, \mathbb{E}[C_k] = 0. \quad (46)$$

Similarly, $\mathbb{E}[A_k U_k C_k^T] = 0$, $\mathbb{E}[C_k U_k C_k^T] = 0$, and

$$\mathbb{E}[A_k S_k A_k^T] = A_k S_k A_k^T. \quad (47)$$

According to the probability distribution defined above

$$\mathbb{E}[H_{k+1}] = \frac{1}{N} \begin{pmatrix} H_{\text{nom}} \\ \vdots \\ H_{\text{nom}} \end{pmatrix}, \mathbb{E}[F_{k+1}] = \frac{N-1}{N} \begin{pmatrix} H_{\text{nom}}A \\ \vdots \\ H_{\text{nom}}A \end{pmatrix}. \quad (48)$$

Computation of the remaining expectations yields the following results:

$$\mathbb{E}[H_{k+1} S_{k+1} H_{k+1}^T] = \frac{1}{N} \text{diag} \begin{pmatrix} H_{\text{nom}} S_{k+1} H_{\text{nom}}^T \\ H_{\text{nom}} S_{k+1} H_{\text{nom}}^T \\ \vdots \\ H_{\text{nom}} S_{k+1} H_{\text{nom}}^T \end{pmatrix}, \quad (49)$$

$$\mathbb{E}[G_{k+1} G_{k+1}^T] = \frac{1}{N} \text{diag} \begin{pmatrix} G_{\text{nom}} G_{\text{nom}}^T + (N-1)G_{\text{cl}} G_{\text{cl}}^T \\ G_{\text{nom}} G_{\text{nom}}^T + (N-1)G_{\text{cl}} G_{\text{cl}}^T \\ \vdots \\ G_{\text{nom}} G_{\text{nom}}^T + (N-1)G_{\text{cl}} G_{\text{cl}}^T \end{pmatrix}, \quad (50)$$

$$\mathbb{E}[F_{k+1} U_k F_{k+1}^T] = \Xi \otimes (H_{\text{nom}} A U_k A^T H_{\text{nom}}^T), \quad (51)$$

where \otimes is the Kronecker product, and Ξ is defined by

$$\Xi = \begin{cases} 0, & N = 1 \\ \frac{1}{N} ((N-2)\mathbf{1}_N \mathbf{1}_N^T + I_N), & N > 1 \end{cases} \quad (52)$$

where $\mathbf{1}_N$ is an $N \times 1$ vector comprising all ones.

It remains to describe the computation of the clutter covariance matrix, $G_{\text{cl}} G_{\text{cl}}^T$. Following the uniform distribution assumption, the spatial distribution of the clutter measurements is uniform in the validation window. Thus, $G_{\text{cl}} G_{\text{cl}}^T$ is the covariance matrix of a random vector uniformly distributed in the validation window. For simplicity of the computation we assume that the window is rectangular, so that the covariance computation boils down to finding the variance of a scalar uniform random variable.

Together with $G_{\text{cl}} G_{\text{cl}}^T$, Eqs. (46)-(52) enable the computation of the coefficients L_k and K_k of the linear optimal filter, according to (11) and (26), respectively.

C. Algorithm Analysis

Lemma 1: The linear optimal algorithm for tracking a target in clutter is algebraically equivalent to a linear optimal estimator operating on a manipulated measurement sequence, where the measurement at each time is obtained by summing up all available observations at that time.

Proof: The lemma follows by showing that, for the considered case, the filter gain, K_k , given in (26), has the form $K_k = \mathbf{1}_N^T \otimes \Psi_k$, where Ψ_k is some (time-dependent) matrix. Then, since y_{k+1} is a concatenated vector of all the observations acquired at time $k+1$, the product $K_k y_{k+1}$ in (2) is nothing but a sum of the individual measurements constituting y_{k+1} premultiplied by Ψ_k . The details of the proof are omitted due to space constraints. ■

Corollary 1: The linear optimal estimator may be implemented using a generalized KF (defined by (33) and (34)) designed for the system defined by the matrices $A_k = A$, $B_k = B$, $C_k = 0$, $H_k = H_{\text{nom}}$, and $G_k = (G_{\text{nom}} G_{\text{nom}}^T + (N-1)G_{\text{cl}} G_{\text{cl}}^T)^{1/2}$.

Proof: The Corollary follows from Lemma 1 in a straightforward manner. The details of the proof are omitted due to space constraints. ■

D. Discussion

The result of Lemma 1 should not be a surprising one. First, recall that the filter implements the linear optimal estimator of the state, meaning that only linear manipulations on the measurements are allowed. Moreover, the classical Nearest Neighbor (NN) [20] and the Probabilistic Data Association (PDA) [19] filters may also be viewed as implementing a standard KF procedure on a (nonlinear) function of the observations acquired at the current step.

By definition, the NN algorithm applies a standard KF on the measurement nearest to the predicted estimate, ignoring all other measurements. In other words, the measurement used for update purpose is obtained by performing a (nonlinear) comparison of the distances between the predicted measurement and all observations from the current scan.

Utilizing all measurements from the latest scan, the PDA filter updates the predicted target state by assigning weights to these measurements in accordance with their distance from the predicted state. This is performed via a number of standard KF updates for every measurement from the current set, and then fusing these local estimates into a single one. However, since the KF is a linear operator, the operation of the PDA may be equivalently viewed as a single KF update applied to an effective measurement, which is a weighted average of all currently available observations. Obtaining this effective measurement is a nonlinear operation, since it requires calculation of the likelihoods of the measurements in the set.

Thus, all three algorithms have conceptually similar modes of operation – at each estimation cycle an effective measurement is computed, followed by an application of a standard KF update to this effective measurement.

VI. NUMERICAL STUDY

A. Simulation Details

In the following numerical study we consider a one-dimensional tracking scenario in which a target is represented via a two-dimensional state, comprising position and velocity information, $x_k = (p_k, v_k)^T$. Starting at $x_0 = (0, 0)^T$ with $P_0 = \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}$, the target is simulated for $N = 1000$ time units driven by the dynamical equation (39) with $A = \begin{pmatrix} 0.95 & 0.2 \\ 0 & 0.95 \end{pmatrix}$ and $B = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$. This particular choice for the transition matrix A keeps the target in close vicinity to the origin and prevents it from diverging. This is done for better visualization as well as for easier implementation – instead of generating clutter measurements in a large (potentially unbounded) one-dimensional space, we can restrict ourselves to a narrow neighborhood of interest about the origin, from which the target does not escape.

The true measurement generation matrices of (40) used in the simulation are $H_{\text{nom}} = [1 \ 0]$, $G_{\text{nom}} = 0.32$. For a (one-dimensional) window of length d the clutter variance of (44) is $G_{\text{cl}} G_{\text{cl}}^T = d^2/12$.

B. Compared Algorithms and Performance Measures

We compare the performance of the linear optimal filter to that of the NN filter [20] and the PDA filter [19].

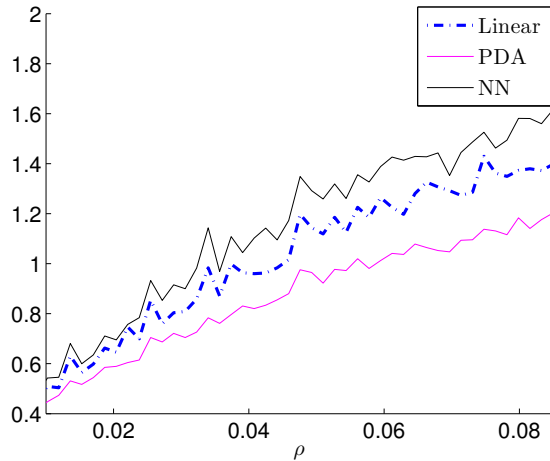
All the considered algorithms are designed to reduce the mean-squared estimation error in either a heuristic or an analytical manner. However, when dealing with tracking in clutter, using the MSE as the only performance measure may result in misleading conclusions. This is due to the fact that once the estimated state draws away from the true measurement, clutter measurements are more likely to be treated as the true ones, eventually resulting in target loss and rendering the algorithms' MSE meaninglessly large. Therefore, the MSE should be treated as a meaningful performance measure only as long as the target is not lost. We thus use the following two measures of performance to evaluate the performance of the algorithms.

The first performance measure is the time until the target is lost. The target is defined lost after the distance between the predicted position and the true state has deviated by more than five standard deviations of the (true) measurement noise for three time units. The second measure is the average squared error, calculated over the time interval until the first of the three algorithms has lost track. This makes the comparison fair, in the sense that none of the algorithms incorporates meaninglessly large errors corresponding to a lost target.

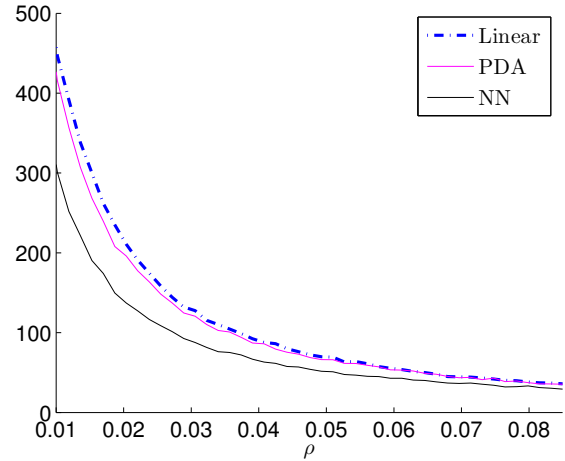
A good algorithm is expected to have long track loss times while maintaining low average squared errors.

C. Results and Discussion

We test the algorithms versus increasing clutter density. To this end we define ρ to be the average number of clutter measurements falling in an interval of one standard deviation of the (true) measurement noise. Averaged over 1000 independent Monte Carlo runs, the resulting track loss times and average squared position errors for the three tested



(a) Average Squared Position Errors



(b) Track Loss Times

Fig. 1: Average squared position estimation errors and track loss times vs clutter density.

filters are plotted in Figure 1. It is readily seen that the linear filter attains the longest track loss times while keeping the estimation errors at a reasonable compromise between the nonlinear PDA and NN filters.

VII. CONCLUSION

We have proposed a new general formulation of dynamical systems with independently switching parameter matrices. The dynamics and measurement equations are allowed to depend on previously obtained estimates of the state. The matrix coefficients constituting the system mode are allowed to take values in a continuous domain, thus significantly generalizing the popular hybrid systems paradigm. Moreover, the dimensions of these matrices are allowed to vary in time. Consequently, the proposed modeling framework generalizes the classical ones and allows formulation of several new problems within a single state space model.

A linear optimal recursive algorithm has been derived. The algorithm has been shown to reduce to previously reported methods when the parameters of the underlying problem are adjusted appropriately.

The utility of the new approach has been demonstrated by using it to derive a linear optimal filter for tracking a maneuvering target in clutter. The linear optimal algorithm has been numerically compared to the classical nonlinear NN and PDA filters, exhibiting competitive performance.

REFERENCES

- [1] E.-K. Boukas and Z.-K. Liu, *Deterministic and Stochastic Time-Delay Systems*. Boston: Birkhäuser, 2002.
- [2] N. Nahi, "Optimal recursive estimation with uncertain observation," *IEEE Trans. on Information Theory*, vol. IT-15, pp. 457–462, July 1969.
- [3] M. Hadidi and S. Schwartz, "Linear recursive state estimators under uncertain observations," *IEEE Trans. on Automatic Control*, vol. AC-24, pp. 944–948, December 1979.
- [4] Y. Sawaragi, T. Katayama, and S. Fujishige, "Adaptive estimation for a linear system with interrupted observation," *IEEE Trans. on Automatic Control*, vol. 18, no. 2, pp. 152–154, 1973.
- [5] A. Jaffer and S. Gupta, "Optimal sequential estimators of discrete processes with markov interrupted observations," *IEEE Trans. on Information Theory*, vol. IT-17, pp. 614–616, September 1971.
- [6] I. Rapoport and Y. Oshman, "Efficient fault tolerant estimation using the IMM methodology," *IEEE Trans. on Aerospace and Electronic Systems*, vol. 43, no. 2, pp. 492–508, 2007.
- [7] I. Rapoport and Y. Oshman, "A new estimation error lower bound for interruption indicators in systems with uncertain measurements," *IEEE Trans. on Information Theory*, vol. 50, no. 12, pp. 3375–3384, 2004.
- [8] D. Sigalov and Y. Oshman, "A new formulation of fault-tolerant estimation problems and some solutions," in *Proc. 26th Convention of Electrical and Electronic Engineers in Israel*, 2011.
- [9] G. Ackerson and K. Fu, "On state estimation in switching environments," *IEEE Trans. on Automatic Control*, vol. 15, no. 1, pp. 10–17, 1970.
- [10] H. Blom and Y. Bar-Shalom, "The interacting multiple model algorithm for systems with Markovian switching coefficients," *IEEE Trans. on Automatic Control*, vol. 33, no. 8, pp. 780–783, 1988.
- [11] D. Sigalov and Y. Oshman, "State estimation in hybrid systems with a bounded number of mode transitions," in *Proc. Fusion 2010, 13th International Conference on Information Fusion*, 2010.
- [12] O. Costa, "Linear minimum mean square error estimation for discrete-time Markovian jump linear systems," *IEEE Trans. on Automatic Control*, vol. 39, no. 8, pp. 1685–1689, 1994.
- [13] N. Nahi and E. Knobbe, "Optimal linear recursive estimation with uncertain system parameters," *IEEE Trans. on Automatic Control*, vol. 21, no. 2, pp. 263–266, 1976.
- [14] D. Magill, "Optimal adaptive estimation of sampled stochastic processes," *IEEE Trans. on Automatic Control*, vol. 10, no. 4, pp. 434–439, 1965.
- [15] A. Jaffer and S. Gupta, "On Estimation of Discrete Processes Under Multiplicative and Additive Noise Conditions," *Information Sciences*, vol. 3, no. 3, pp. 267–276, 1971.
- [16] A. Jazwinski, *Stochastic Processes and Filtering Theory*. Academic Press, 1970.
- [17] B. Anderson, J. Moore, and J. Barratt, *Optimal Filtering*. Prentice-Hall Englewood Cliffs, NJ, 1979.
- [18] D. Sigalov and Y. Oshman, "Linear optimal state estimation in hybrid dynamical systems with actuator faults," in *Proc. Fusion 2011, 14th International Conference on Information Fusion*, 2011.
- [19] Y. Bar-Shalom and E. Tse, "Tracking in a cluttered environment with probabilistic data association," *Automatica*, vol. 11, no. 5, pp. 451–460, 1975.
- [20] Y. Bar-Shalom and X. Li, *Multitarget-Multisensor Tracking: Principles and Techniques*. Storrs, CT: YBS Publishing, 1995.