Proceedings of the
44th IEEE Conference on Decision and Control, and
the European Control Conference 2005
Seville, Spain, December 12-15, 2005

Distribution-Free Mode-Estimators for a Class of Discrete-Time Jump-Linear Systems

Daniel Choukroun and Jason L. Speyer

***Abstract*— This paper is concerned with the development of recursive distribution-free mode-estimators for a class of discrete-time jump-linear systems. The cornerstone of the proposed filters consists of an algebraic manipulation of the dynamics equation of the continuous state. This equation turns out to be linear with respect to the mode vector, and, under the assumption of perfect state information, provides a linear observation equation for the mode. Appending this equation to the known linear dynamics equation of the mode yields a linear non-Gaussian state-space model. A first mode-estimator is then derived using standard Linear Filtering results. A second filter is developed as an application of a general discrete-time filter, which approximates *the* continuous-time optimal non-linear filter (the conditional mean estimator for continuous time) for small sample times. The second filter is preferred from a performance point of view. Model order reduction is applied in order to avoid singularity issues in the filters implementations. The second filter is envisioned as a useful tool in the analysis and design of dual controllers for this type of hybrid systems.**

***Index Terms*— Finite-alphabet homogenous Markov chain, jump-linear system, optimal non-linear filtering**

I. INTRODUCTION

A very popular way of modeling dynamical systems in switching environments is by means of hybrid systems, where some state variables are continuous and some are discrete. The discrete variables characterize the environment in which the continuous variables evolve. Typically, the mode, that is the set of the discrete states, switches among a finite number of values and the switch may be deterministic [1] or stochastic [2]. For the stochastic case, assuming that the continuous state is known, finite-dimensional optimal non-linear mode-estimators can be developed; that is, algorithms which compute the conditional expectation of the mode. In a continuous-time setting (see [3, Chap. 9] and [4]), these algorithms consist of non-linear stochastic differential equations. In order to avoid the drawback of numerically integrating them, a recursive optimal non-linear filter can be developed in a discrete-time setting [5]. It appears, however, that this algorithm requires a complete knowledge of the probability distribution of the noise in the continuous state dynamics equation.

In this work we present two types of *recursive distribution-free* mode-estimators for a class of jump-linear systems. A novel contribution in this work consists of the modeling of the continuous state process equation. That equation turns out to be linear with respect to the mode. As a result, a linear non-Gaussian state-space model for the mode is developed. Using this linear structure, two filters are developed. The first filter is obtained by applying standard results of Linear Least-Squares theory. The second filter is a straightforward application of an algorithm, which, for fairly general discrete-time systems, is an approximation for small sample times of the optimal non-linear filter (a brief summary is provided in the Appendix). The linearity of the model equations and the fact that the mode is a probability vector, which components add to one, leads to singularity issues in various covariance matrices. This issue is dealt with via model order reduction.

The remainder of this paper is organized as follows. Section II includes the mathematical statement of the problem. Then, the linear state-space model for the mode is derived in Section III. The best linear filter is developed in Section IV. Section V presents the suboptimal non-linear filter. A discussion is proposed in Section VI. The results on the model order reduction are the topics of Section VII. Finally conclusions are drawn in the last section.

This work was sponsored by NASA-Ames Research Center under Grant NAG2-1484:4 and by AFOSR under Grant F49620-01-1-0361

The authors are with the Department of Mechanical and Aerospace Engineering at UCLA, Los Angeles, CA-90095, USA dach@ucla.edu, speyer@seas.ucla.edu

II. STATEMENT OF THE PROBLEM

Consider the discrete-time dynamical system

$$\mathbf{x}_{k+1} = A(y_k)\mathbf{x}_k + B(y_k)\mathbf{u}_k + \mathbf{w}_k \qquad k \in \mathbb{N} \tag{1}$$

where $\mathbf{x}_k \in \mathbb{R}^n$ is a known continuous state vector and $\{y_k : k \in \mathbb{N}\}$ is an unknown scalar finite-state homogenous discrete-time Markov chain with state space $S_y = \{\gamma_1, \gamma_2, \ldots, \gamma_\nu\}$ and transition matrix M; that is,

$$M[i,j] \triangleq Pr(y_{k+1} = \gamma_i \mid y_k = \gamma_j) \tag{2}$$

for all $\gamma_i, \gamma_j \in S_y$. The disturbance $\mathbf{w}_k$ is assumed to be a zero-mean white-noise sequence with known covariance matrix W_k, and is assumed to be independent from $\mathbf{x}_k$ and y_k. The control vector $\mathbf{u}_k \in \mathbb{R}^m$ is assumed to be a vector of known inputs that are function of the available information. It can be shown [6, p. 17] that the Markov chain $\{y_k,\ k \in \mathbb{N}\}$ defined on a given probability space $\{\Omega, \mathcal{F}, Pr\}$ can be equivalently replaced by the vector Markov chain $\{\mathbf{y}_k,\ k \in \mathbb{N}\}$ defined over the same probability space, where $\mathbf{y}_k$ is the random vector with components the characteristic random variables associated with each of the ν elementary events of $\Omega = \{\omega_i\}_{i=1}^{\nu}$; that is,

$$\begin{aligned} \mathbf{y}_k : \Omega = \{\omega_i\}_{i=1}^{\nu} &\rightarrow S_{\mathbf{y}} = \{\mathbf{e}_1, \mathbf{e}_2, \ldots, \mathbf{e}_\nu\} \\ \omega &\mapsto \mathbf{y}_k^T(\omega) = \left[\delta(\mathbf{y}_k, \mathbf{e}_1)\, \delta(\mathbf{y}_k, \mathbf{e}_2) \ldots \delta(\mathbf{y}_k, \mathbf{e}_\nu)\right] \end{aligned} \tag{3}$$

where $\mathbf{e}_i$, for $i = 1, 2, \ldots, \nu$ are the standard unit vectors in $\mathbb{R}^\nu$, and $\delta(\cdot,\cdot)$ represents the Kronecker delta. The dynamics of the Markov chain $\{\mathbf{y}_k \in \mathbb{R}^\nu, k \in \mathbb{N}\}$ is governed by the following linear $\nu \times 1$ vector difference equation

$$\mathbf{y}_{k+1} = M\mathbf{y}_k + \mathbf{v}_k \tag{4}$$

where $\{\mathbf{v}_k\}$ is a zero-mean white-noise sequence that is orthogonal to $\mathbf{y}_k$, and independent from $\mathcal{Y}^{k-1}$ and from $\mathcal{X}^k$. Moreover, $\mathbf{v}_k$ and $\mathbf{w}_k$ are assumed to be independent. It stems from the definition of $\mathbf{y}_k$ that, for $k \in \mathbb{N}$,

$$E\{\mathbf{y}_k\} = \left[\ldots Pr\{\mathbf{y}_k = \mathbf{e}_i\} \ldots\right]^T \triangleq \mathbf{p}_k \tag{5}$$

where $\mathbf{p}_k \in \mathbb{R}^\nu$ can be recursively computed from the Chapman-Kolmogorov equation of the MArkov chain $\mathbf{y}_k$:

$$\mathbf{p}_{k+1} = M\mathbf{p}_k \tag{6}$$

starting at $\mathbf{p}_0$. We are seeking for two types of estimator. The first one, denoted by $\overline{\mathbf{y}}_{k/k}$, is the unbiased linear estimator of $\mathbf{y}_k$ that minimizes the mean-square error subject to the constraints (1) and (4). The second one, denoted by $\widehat{\mathbf{y}}_{k/k}$, is an approximation of the conditional mean of $\mathbf{y}_k$ given $\mathcal{X}^k$, the past history of $\mathbf{x}_k$. That approximation is to the first order in Δt, where Δt denotes the discretization time increment that is associated with the difference equation (1). Notice that the estimate $\widehat{\mathbf{y}}_{k/k}$ approximates a vector of a posteriori probabilities; that is,

$$\begin{aligned}\widehat{\mathbf{y}}^T_{k/k} &= E\{\mathbf{y}_k^T|\mathcal{X}^k\} + \mathcal{O}(\Delta t)\\ &= \left[\ldots Pr\{\mathbf{y}_k = \mathbf{e}_i|\mathcal{X}^k\} \ldots\right] + \mathcal{O}(\Delta t)\end{aligned} \tag{7}$$

The following lemma will be used in the sequel of this work. For a proof, see e.g. [6, p. 19].

Lemma 1 : Let $\mathbf{u}$, $\widehat{\mathbf{u}}$, $\widetilde{\mathbf{u}}$, and $\mathcal{X}$, denote, respectively, a random vector defined as in Eq. (3), its conditional expectation given $\mathcal{X}$, the associated estimation error, and a collection of conditioning random variables. Then the following identity can be shown:

$$\mathrm{cov}\{\widetilde{\mathbf{u}}|\mathcal{X}\} = \mathrm{diag}\{\widehat{\mathbf{u}}\} - \widehat{\mathbf{u}}\widehat{\mathbf{u}}^T \tag{8}$$

III. LINEAR STATE-SPACE MODEL FOR $\mathbf{y}_k$

A. *Reformulation of the state-Space equation* (1)

The governing equations for the dynamics of the discrete-time jump-linear system $\{\mathbf{x}_k, \mathbf{y}_k\}$ described in the previous section can be written as follows:

$$\mathbf{x}_{k+1} = C(\mathbf{x}_k, \mathbf{u}_k)\,\mathbf{y}_k + \mathbf{w}_k \tag{9}$$

$$\mathbf{y}_{k+1} = M\mathbf{y}_k + \mathbf{v}_k \tag{10}$$

where $C(\mathbf{x}_k, \mathbf{u}_k)$ is an $n \times \nu$ matrix defined as

$$C(\mathbf{x}_k, \mathbf{u}_k) \triangleq \mathcal{A}\left(I_\nu \otimes \begin{bmatrix}\mathbf{x}_k\\ \mathbf{u}_k\end{bmatrix}\right) \tag{11}$$

In Eq. (11), $\otimes$ denotes the Kronecker product, I_ν denotes the $\nu \times \nu$ identity matrix, and $\mathcal{A}$ is the $n \times (n+m)\nu$ matrix expressed as

$$\mathcal{A} \triangleq \left[\ldots [A_i\, B_i] \ldots\right] \qquad i = 1, 2, \ldots, \nu \tag{12}$$

where the matrices $A_i \in \mathbb{R}^n$ and $B_i \in \mathbb{R}^m$ denote, respectively, the matrices $A(\gamma_i)$ and $B(\gamma_i)$, for $i = 1, 2, \ldots, \nu$. The novelty in this state-space model is in Eq. (9), which is equivalent to Eq. (1), but is re-written as a linear equation with respect to (w.r.t) the discrete state vector $\mathbf{y}_k$. Moreover, since we assume full knowledge of $\mathbf{x}_k$, Eq. (9) can be considered a linear observation equation for $\mathbf{y}_k$, where $\mathbf{x}_{k+1}$ is the observation, $C(\mathbf{x}_k, \mathbf{u}_k)$ is the observation matrix, and $\mathbf{w}_k$ is the observation noise. Together with the linear process equation (13), this observation equation yields a linear state-space model for $\mathbf{y}_k$.

B. *Development of Eq.* (9)

the matrices $A(\mathbf{y}_k)$ and $B(\mathbf{y}_k)$ can be rewritten as

$$A(\mathbf{y}_k) = \sum_{i=1}^{\nu} A_i\ \delta(\mathbf{y}_k, \mathbf{e}_i) \tag{13}$$

$$B(\mathbf{y}_k) = \sum_{i=1}^{\nu} B_i\ \delta(\mathbf{y}_k, \mathbf{e}_i) \tag{14}$$

Using Eqs. (13) and (14) in the Right-hand-Side (RHS) of Eq. (1), without the noise $\mathbf{w}_k$, yields

$$\begin{aligned}
&A(\mathbf{y}_k)\mathbf{x}_k + B(\mathbf{y}_k)\mathbf{u}_k =\\
&= \left(\sum_{i=1}^{\nu} A_i\delta(\mathbf{y}_k, \mathbf{e}_i)\right)\mathbf{x}_k + \left(\sum_{i=1}^{\nu} B_i\delta(\mathbf{y}_k, \mathbf{e}_i)\right)\mathbf{u}_k\\
&= \sum_{i=1}^{\nu} (A_i\mathbf{x}_k + B_i\mathbf{u}_k)\,\delta(\mathbf{y}_k, \mathbf{e}_i)\\
&= \left[\ldots [A_i\mathbf{x}_k + B_i\mathbf{u}_k] \ldots\right]\begin{bmatrix}\vdots\\ \delta(\mathbf{y}_k, \mathbf{e}_i)\\ \vdots\end{bmatrix}\\
&= \left[\ldots [A_i\, B_i]\begin{bmatrix}\mathbf{x}_k\\ \mathbf{u}_k\end{bmatrix} \ldots\right]\begin{bmatrix}\vdots\\ \delta(\mathbf{y}_k, \mathbf{e}_i)\\ \vdots\end{bmatrix}\\
&= \left[\ldots [A_i\, B_i] \ldots\right]\begin{bmatrix}\begin{bmatrix}\mathbf{x}_k\\ \mathbf{u}_k\end{bmatrix} & 0 & \ldots\\ 0 & \begin{bmatrix}\mathbf{x}_k\\ \mathbf{u}_k\end{bmatrix} & \ldots\\ \vdots & \vdots & \ddots\end{bmatrix}\begin{bmatrix}\vdots\\ \delta(\mathbf{y}_k, \mathbf{e}_i)\\ \vdots\end{bmatrix}\\
&= \mathcal{A}\left(I_\nu \otimes \begin{bmatrix}\mathbf{x}_k\\ \mathbf{u}_k\end{bmatrix}\right)\mathbf{y}_k
\end{aligned} \tag{15}$$

where $\otimes$ denotes the Kronecker product, I_ν is the identity matrix in $\mathbb{R}^\nu$, and the $n \times (n+m)\nu$ matrix $\mathcal{A}$ is defined from Eq. (15). Finally, defining the $n \times \nu$ matrix $C(\mathbf{x}_k, \mathbf{u}_k)$ as follows:

$$C(\mathbf{x}_k, \mathbf{u}_k) \triangleq \mathcal{A}\left(I_\nu \otimes \begin{bmatrix}\mathbf{x}_k\\ \mathbf{u}_k\end{bmatrix}\right) \tag{16}$$

and using Eqs. (15) and (16) in Eq. (1) yields Eq. (9).

IV. BEST LINEAR UNBIASED ESTIMATION

We are interested in finding $\overline{\mathbf{y}}_{k/k}$, the linear approximation of $\mathbf{y}_k$, which minimizes the mean squared error, $E\{\|\mathbf{y}_k - \overline{\mathbf{y}}_{k/k}\|^2\}$ subject to the linear state-space model equations (9) and (10), and to the unbiasedness condition. The solution to that problem, sometimes called the Best Linear Unbiased Estimator (BLUE) [7, p. 121], can be derived using standard techniques based, for instance, on the Orthogonality Principle [8, p. 202]. For such a derivation the following properties are central. First, the noise vector $\mathbf{w}_k$ is both independent from $\mathbf{x}_k$ and $\mathbf{y}_k$, and is zero-mean, which makes it orthogonal to $\mathbf{x}_k$ and $\mathbf{y}_k$. Second, the observation $\mathbf{x}_k$ and the control $\mathbf{u}_k$ are known quantities at time t_{k+1}, and this makes the matrix $C(\mathbf{x}_k, \mathbf{u}_k)$ a known quantity. The algorithm is next summarized without proof.

1) Initialization equations:

$$\widehat{\mathbf{y}}_{0/0} = \mathbf{p}_0 \tag{17a}$$
$$\overline{P}_{0/0} = \operatorname{diag}\{\mathbf{p}_0\} - \mathbf{p}_0\, \mathbf{p}_0^T \tag{17b}$$

2) Smoothing stage equations:

$$\widetilde{\mathbf{x}}_{k+1/k} = \mathbf{x}_{k+1} - C_k \widehat{\mathbf{y}}_{k/k} \tag{18a}$$
$$\overline{S}_{k+1} = C_k \overline{P}_{k/k} C_k^T + W_k \tag{18b}$$
$$\overline{K}_{k+1} = \overline{P}_{k/k} C_k^T S_{k+1}^{-1} \tag{18c}$$
$$\widehat{\mathbf{y}}_{k/k+1} = \widehat{\mathbf{y}}_{k/k} + \overline{K}_{k+1} \widetilde{\mathbf{x}}_{k+1/k} \tag{18d}$$
$$\overline{P}_{k/k+1} = \left(I_\nu - \overline{K}_{k+1} C_k\right) \overline{P}_{k/k} \left(I_\nu - \overline{K}_{k+1} C_k\right)^T + \overline{K}_{k+1} W_k \overline{K}_{k+1}^T \tag{18e}$$

3) Time-propagation stage equations:

$$\widehat{\mathbf{y}}_{k+1/k+1} = M \widehat{\mathbf{y}}_{k/k+1} \tag{19a}$$
$$V_k = \operatorname{diag}\{\mathbf{p}_{k+1}\} - M \operatorname{diag}\{\mathbf{p}_k\} M^T \tag{19b}$$
$$\overline{P}_{k+1/k+1} = M \overline{P}_{k/k+1} M^T + V_k \tag{19c}$$

where C_k, $\overline{P}_{k/k}$, and V_k denote, respectively, the matrices $C(\mathbf{x}_k, \mathbf{u}_k)$, $\operatorname{cov}\{\mathbf{y}_k - \overline{\mathbf{y}}_{k/k}\}$, and $\operatorname{cov}\{\mathbf{v}_k\}$. The expressions for $\overline{P}_{0/0}$ and V_k are easily obtained by using Lemma 1. The denomination of *Smoothing stage* instead of the classical *Filtering stage* is due to the fact that the observation, $\mathbf{x}_{k+1}$, is one-step delayed.

V. APPROXIMATE NON-LINEAR FILTERING

A. Algorithm Summary

The suboptimal non-linear filter equations result from applying the general algorithm presented in the Appendix. That algorithm recursively computes $\widehat{\mathbf{y}}_{k/k}$ and $P_{k/k}$, such that

$$\widehat{\mathbf{y}}_{k/k} = E\{\mathbf{y}_k | \mathcal{X}^k\} + \mathcal{O}(\Delta t) \tag{20}$$
$$P_{k/k} = \operatorname{cov}\{\widetilde{\mathbf{y}}_{k/k} | \mathcal{X}^k\} = \operatorname{cov}\{\mathbf{y}_k | \mathcal{X}^k\} + \mathcal{O}(\Delta t^2) \tag{21}$$

where $\widetilde{\mathbf{y}}$ denotes the estimation error $\mathbf{y} - \widehat{\mathbf{y}}$. The initialization equations are identical to Eqs. (17a) and (17b).

$$\widehat{\mathbf{y}}_{0/0} = \mathbf{p}_0 \tag{22a}$$
$$P_{0/0} = \operatorname{diag}\{\mathbf{p}_0\} - \mathbf{p}_0\, \mathbf{p}_0^T \tag{22b}$$

a) Smoothing stage equations:

$$\widetilde{\mathbf{x}}_{k+1/k} = \mathbf{x}_{k+1} - C_k\, \widehat{\mathbf{y}}_{k/k} \tag{23a}$$
$$S_{k+1} = C_k P_{k/k} C_k^T + W_k \tag{23b}$$
$$K_{k+1} = P_{k/k} C_k^T S_{k+1}^{-1} \tag{23c}$$
$$\boldsymbol{\delta}\widehat{\mathbf{y}}_k = K_{k+1}\, \widetilde{\mathbf{x}}_{k+1/k} \tag{23d}$$
$$\widehat{\mathbf{y}}_{k/k+1} = \widehat{\mathbf{y}}_{k/k} + \boldsymbol{\delta}\widehat{\mathbf{y}}_k \tag{23e}$$
$$\Delta P_k = \operatorname{diag}\{\boldsymbol{\delta}\widehat{\mathbf{y}}_k\} - \boldsymbol{\delta}\widehat{\mathbf{y}}_k\, \boldsymbol{\delta}\widehat{\mathbf{y}}_k^T - \boldsymbol{\delta}\widehat{\mathbf{y}}_k\, \widehat{\mathbf{y}}_{k/k}^T - \widehat{\mathbf{y}}_{k/k}\, \boldsymbol{\delta}\widehat{\mathbf{y}}_k^T \tag{23f}$$
$$P_{k/k+1} = \left(I_\nu - K_{k+1} C_k\right) \left(P_{k/k} + \Delta P_k\right) \left(I_\nu - K_{k+1} C_k\right)^T + K_{k+1} W_k K_{k+1}^T \tag{23g}$$

b) Time-propagation stage equations:

$$\widehat{\mathbf{y}}_{k+1/k+1} = M\, \widehat{\mathbf{y}}_{k/k+1} \tag{24a}$$
$$V_{k/k+1} = \operatorname{diag}\{\widehat{\mathbf{y}}_{k+1/k+1}\} - M \operatorname{diag}\{\widehat{\mathbf{y}}_{k/k+1}\} M^T \tag{24b}$$
$$P_{k+1/k+1} = M P_{k/k+1} M^T + V_{k/k+1} \tag{24c}$$

where $V_{k/k+1}$ denotes $\operatorname{cov}\{\mathbf{v}_k | \mathcal{X}^{k+1}\}$.

B. Algorithm Development

1) Smoothing stage: Using Eq. (20), the vector $\widehat{\mathbf{f}}_{k/k}$ is defined as follows:

$$E\{C_k \mathbf{y}_k | \mathcal{X}^k\} = C_k \widehat{\mathbf{y}}_{k/k} + \mathcal{O}(\Delta t) = \widehat{\mathbf{f}}_{k/k} + \mathcal{O}(\Delta t) \tag{25}$$

where C_k can be taken out of the expectation since it is function of $\mathbf{x}_k$. At time t_{k+1}, the residual $\widetilde{\mathbf{x}}_{k+1/k}$ is computed as

$$\widetilde{\mathbf{x}}_{k+1/k} = \mathbf{x}_{k+1} - \widehat{\mathbf{f}}_{k/k} \tag{26}$$

Using Eqs. (9) and (20) in Eq. (26), one can show that the conditional mean of $\widetilde{\mathbf{x}}_{k+1/k}$ given $\mathcal{X}^k$ is in $\mathcal{O}(\Delta t)$. An approximation of the conditional covariance matrix of $\widetilde{\mathbf{x}}_{k+1/k}$ given $\mathcal{X}^k$, to second order in Δt, can be computed using standard techniques:

$$\operatorname{cov}\{\widetilde{\mathbf{x}}_{k+1/k} | \mathcal{X}^k\} = E\{\widetilde{\mathbf{x}}_{k+1/k} \widetilde{\mathbf{x}}_{k+1/k}^T | \mathcal{X}^k\} + \mathcal{O}(\Delta t^2) = C_k P_{k/k} C_k^T + W_k + \mathcal{O}(\Delta t^2) \tag{27}$$

The cross-terms that are involved in the development of Eq. (27) cancel out since $\mathbf{w}_k$ is independent from $\widetilde{\mathbf{y}}_{k/k}$ and is zero-mean. The matrix S_{k+1} is defined as the first term on the RHS of Eq. (27). Let $\widetilde{\mathbf{f}}_{k/k}$ and $P_{\mathbf{y}\widetilde{\mathbf{f}}}(k/k)$ denote, respectively, the estimation error in $\widehat{\mathbf{f}}_{k/k}$, and the conditional cross-covariance matrix of $\mathbf{y}_k$ and $\widetilde{\mathbf{f}}_{k/k}$ given $\mathcal{X}^k$. Their expressions are obtained as follows:

$$\widetilde{\mathbf{f}}_{k/k} = C_k \widetilde{\mathbf{y}}_{k/k} \tag{28}$$
$$P_{\mathbf{y}\widetilde{\mathbf{f}}}(k/k) = P_{k/k} C_k^T + \mathcal{O}(\Delta t) \tag{29}$$

Using Eqs. (27) and (29), we define the gain matrix, K_{k+1}, as

$$K_{k+1} \triangleq P_{\mathbf{y}\widetilde{\mathbf{f}}}(k/k)\, S_{k+1}^{-1} \tag{30}$$

which proves Eq. (23c), and we compute a smoothed estimate of $\mathbf{y}_k$ given $\mathcal{X}^{k+1}$ through the following equation:

$$\widehat{\mathbf{y}}_{k/k+1} = \widehat{\mathbf{y}}_{k/k} + K_{k+1}\widetilde{\mathbf{x}}_{k+1/k} \tag{31}$$

which proves Eq. (23e). The proof of Eq. (23g) is as follows:

$$\begin{aligned} P_{k/k+1} &= \operatorname{cov}\{\widetilde{\mathbf{y}}_{k/k+1}|\mathcal{X}^{k+1}\} + \mathcal{O}(\Delta t^2) \\ &= (I_\nu - K_{k+1}C_k)\operatorname{cov}\{\widetilde{\mathbf{y}}_{k/k}|\mathcal{X}^{k+1}\}(I_\nu - K_{k+1}C_k)^T \\ &\quad + K_{k+1}W_k K_{k+1}^T + \mathcal{O}(\Delta t^2) \end{aligned} \tag{32}$$

Furthermore, the conditional covariance matrix on the Right-Hand-Side (RHS) of Eq. (32) can be computed using the following identity:

$$\operatorname{cov}\{\widetilde{\mathbf{y}}_{k/k}|\cdot\} = E\{\widetilde{\mathbf{y}}_{k/k}\widetilde{\mathbf{y}}_{k/k}^T|\cdot\} - E\{\widetilde{\mathbf{y}}_{k/k}|\cdot\}E\{\widetilde{\mathbf{y}}_{k/k}|\cdot\}^T \tag{33}$$

The conditional mean $E\{\widetilde{\mathbf{y}}_{k/k}|\mathcal{X}^{k+1}\}$ is expressed as

$$\begin{aligned} E\{\widetilde{\mathbf{y}}_{k/k}|\mathcal{X}^{k+1}\} &= \widehat{\mathbf{y}}_{k/k+1} - \widehat{\mathbf{y}}_{k/k} + \mathcal{O}(\Delta t) \\ &= \boldsymbol{\delta}\widehat{\mathbf{y}}_k + \mathcal{O}(\Delta t) \end{aligned} \tag{34}$$

The expression for the conditional second order moment on the RHS of Eq. (33) is developed as

$$\begin{aligned} &E\{\widetilde{\mathbf{y}}_{k/k}\widetilde{\mathbf{y}}_{k/k}^T|\mathcal{X}^{k+1}\} = \\ &= E\{\mathbf{y}_k\mathbf{y}_k^T|\mathcal{X}^{k+1}\} - \widehat{\mathbf{y}}_{k/k+1}\widehat{\mathbf{y}}_{k/k}^T - \widehat{\mathbf{y}}_{k/k}\widehat{\mathbf{y}}_{k/k+1}^T + \widehat{\mathbf{y}}_{k/k}\widehat{\mathbf{y}}_{k/k}^T \\ &= \left(\operatorname{diag}\{\widehat{\mathbf{y}}_{k/k}\} - \widehat{\mathbf{y}}_{k/k}\widehat{\mathbf{y}}_{k/k}^T\right) + \operatorname{diag}\{\boldsymbol{\delta}\widehat{\mathbf{y}}_k\} - \boldsymbol{\delta}\widehat{\mathbf{y}}_k\widehat{\mathbf{y}}_{k/k}^T - \widehat{\mathbf{y}}_{k/k}\boldsymbol{\delta}\widehat{\mathbf{y}}_k^T \\ &= P_{k/k} + \operatorname{diag}\{\boldsymbol{\delta}\widehat{\mathbf{y}}_k\} - \boldsymbol{\delta}\widehat{\mathbf{y}}_k\widehat{\mathbf{y}}_{k/k}^T - \widehat{\mathbf{y}}_{k/k}\boldsymbol{\delta}\widehat{\mathbf{y}}_k^T + \mathcal{O}(\Delta t^2) \end{aligned} \tag{35}$$

where the last equality stems from Lemma 1 and Eq. (21). Using Eqs. (33) to (35) in Eq. (32) yields Eqs. (23f) and (23g).

2) Time-propagation stage: Some preliminary results are needed.

a) Lemma 2 : The random vector $\mathbf{v}_k$ is independent from $\mathcal{X}^k$ and orthogonal to $\{\mathbf{x}_{k+1}\}$, and thus,

$$E\{\mathbf{v}_k|\mathcal{X}^{k+1}\} = 0 \qquad \forall k \in \mathbb{N} \tag{36}$$

Proof : It is known that $\mathbf{v}_k$ is orthogonal to $\mathbf{y}_k$, and is assumed independent from $\mathcal{W}^k \triangleq \{\mathbf{w}_i\}_{i=0}^k$. Furthermore, to ensure the Markov property of the chain $\{\mathbf{y}_k\}$, the vector $\mathbf{v}_k$ is necessarily independent from $\mathcal{X}^k$. As a result, using Eq. (9), $\mathbf{v}_k$ is orthogonal to $\mathbf{x}_{k+1}$, and, therefore, to $\mathcal{X}^{k+1}$.

b) Lemma 3 : The random vectors $\mathbf{y}_k$ and $\mathbf{v}_k$ are conditionally orthogonal given $\mathcal{X}^{k+1}$; that is,

$$E\{\mathbf{y}_k\mathbf{v}_k^T|\mathcal{X}^{k+1}\} = 0 \tag{37}$$

Proof:

$$\begin{aligned} E\{\mathbf{y}_k\mathbf{v}_k^T|\mathcal{X}^{k+1}\} &= E\{E\{\mathbf{y}_k\mathbf{v}_k^T\,|\,\mathbf{y}_k, \mathcal{X}^{k+1}\}\,|\,\mathcal{X}^{k+1}\} \\ &= E\{\mathbf{y}_k\,E\{\mathbf{v}_k^T\,|\,\mathbf{y}_k, \mathcal{X}^{k+1}\}\,|\,\mathcal{X}^{k+1}\} \\ &= 0 \end{aligned} \tag{38}$$

where we used Lemma 2 and the fact that $\mathbf{v}_k$ is orthogonal to $\mathbf{y}_k$ in the passage to the last line.

The propagation equation of the estimate $\widehat{\mathbf{y}}_{k/k}$ as given in Eq. (24a) is obtained by applying the conditional expectation on both sides of Eq. (4), by using Lemma 2, and by noting that here $D_k = 0$. The conditional covariance matrix $V_{k/k+1}$ of $\mathbf{v}_k$ given $\mathcal{X}^{k+1}$ is directly obtained by using Lemma 3 in its definition. An expression for $P_{k+1/k+1}$ can be developed as follows:

$$\begin{aligned} P_{k+1/k+1} &= E\{\widetilde{\mathbf{y}}_{k+1/k+1}\widetilde{\mathbf{y}}_{k+1/k+1}^T|\mathcal{X}^{k+1}\} + \mathcal{O}(\Delta t^2) \\ &= MP_{k/k+1}M^T + V_{k/k+1} + \mathcal{O}(\Delta t^2) \end{aligned} \tag{39}$$

where it can be shown, using Lemmas 2 and 3, that the cross-terms involving $\widetilde{\mathbf{y}}_{k/k+1}$ and $\mathbf{v}_k$ cancel out.

VI. DISCUSSION

The dynamics equation for the continuous state $\mathbf{x}_k$ was recast, through a special algebraic manipulation, as a linear equation in the discrete state $\mathbf{y}_k$ [Eq. (9)]. Adding the linear dynamics equation of $\mathbf{y}_k$ [Eq. (10)] to Eq. (9) results in a conventional linear state-space model for $\mathbf{y}_k$. Were $\mathbf{y}_k$, $\mathbf{v}_k$, and $\mathbf{w}_k$ continuous jointly Gaussian processes, we would have been in the standard linear *Conditionally Gaussian* case, and the optimal filter would be the standard Conditional Gaussian filter, which is an extension of the classical Kalman filter (KF) [9, Ch. 11]. The vectors $\mathbf{y}_k$ and $\mathbf{v}_k$ are however discrete-valued non-Gaussian random vectors, and $\mathbf{w}_k$ is not necessarily Gaussian, so that the given model differ from the above-mentioned standard case. Thanks to the linearity of the state-space equations for $\mathbf{y}_k$, a best linear unbiased recursive filter could be developed. It has the advantage of being simple to implement and analyze. The linearity of the $\mathbf{y}_k$-model was also utilized toward a straightforward application of the suboptimal non-linear filter presented in the Appendix. The difference between both filters were emphasized in the update stage of the second-order estimation error statistics. In particular, the gain computations in the second filter is highly coupled with the estimate computations. From a performance point of view, the second algorithm should be preferred since it yields an approximation of the conditional mean of $\mathbf{y}_k$ to first-order in Δt. Notice, using Lemma 1, that there is an alternative more efficient way of computing the P-matrices in the second filter; namely,

$$P = \operatorname{diag}\{\widehat{\mathbf{y}}\} - \widehat{\mathbf{y}}\,\widehat{\mathbf{y}}^T \tag{40}$$

It is interesting to emphasize that the proposed filters are distribution-free. In particular, the knowledge of the distribution of $\mathbf{w}_k$ is not required.

VII. MODEL ORDER REDUCTION

The ν-vector $\mathbf{y}_k$ is by definition a unity vector in the sense of the l_1-norm; that is;

$$\|\mathbf{y}_k\|_1 \triangleq \sum_{i=1}^{\nu} y_i(k) = 1 \tag{41}$$

By definition of the matrix M, we also know that the columns of M are probability vectors, which components add to one. As a result, the components of the noise vector $\mathbf{v}_k$ in Eq. (4) add to zero; that is, they satisfy to a linear constraint and are, therefore, perfectly correlated. This leads to a loss of one in the rank of the covariance matrix of $\mathbf{v}_k$. The same issue arises for the estimation errors in the proposed filters. Since the sum of the components of these errors is close to zero, the associated covariance matrices, $\overline{P}_{k/k}$ and $P_{k/k}$, are close to be singular. To avoid this, we propose a linear model reduction of the linear state-space equations (1) and (4) via the linear constraint (41). In the following, we only present the development of the reduced model. The applications to, respectively, Linear Least-Squares Filtering and to the approximate Non-Linear Filtering are straightforward.

A. Reduced-order state-space model

Let $\mathbf{y}_k^r$ denote the $(\nu-1)$-vector obtained by truncating the last component of $\mathbf{y}_k$; that is,

$$\mathbf{y}_k^r \triangleq \left[Pr\{\mathbf{y}_k = \mathbf{e}_i\}\right]_{i=1}^{\nu-1} \qquad \in \mathbb{R}^{\nu-1} \tag{42}$$

The process equation for $\mathbf{y}_k^r$ is given as the following $(\nu-1)$-vector equation:

$$\mathbf{y}_{k+1}^r = M^r \mathbf{y}_k^r + \mathbf{m}_\nu^r + \mathbf{v}_k^r \tag{43}$$

where

$$\mathbf{m}_\nu^r \triangleq \left[m_{i,\nu}\right]_{i=1}^{\nu-1} \qquad \in \mathbb{R}^{\nu-1} \tag{44}$$

$$\mathbf{v}_k^r \triangleq \left[v_i(k)\right]_{i=1}^{\nu-1} \qquad \in \mathbb{R}^{\nu-1} \tag{45}$$

$$M^r \triangleq M_{\nu-1} - \mathbf{m}_\nu^r\, \mathbf{1}^T \qquad \in \mathbb{R}^{\nu-1\times\nu-1} \tag{46}$$

and where $m_{i,\nu}$ denotes the element i,ν of the matrix M, $v_i(k)$ denotes the i^{th} component of the vector $\mathbf{v}_k$, the vector $\mathbf{1}$ has all its $\nu-1$ components equal to one, and the $\nu-1$ dimensional matrix $M_{\nu-1}$ is the principal submatrix of M obtained by extracting the $\nu-1$ first rows and columns. In Eq. (43), the vector $\mathbf{m}_\nu^r$ is a vector of known deterministic inputs. The stochastic properties of the reduced noise vector $\mathbf{v}_k^r$ can be directly deduced from those of the full vector $\mathbf{v}_k$. In particular, an expression for the covariance matrix of $\mathbf{v}_k^r$ is obtained by deleting the last row and column of the covariance matrix of $\mathbf{v}_k$.

The observation equation for the truncated vector $\mathbf{y}_k^r$ is given as the following $n\times 1$ vector equation:

$$\mathbf{z}_{k+1} = C_k^r \mathbf{y}_k^r + \mathbf{w}_k \tag{47}$$

where

$$\mathbf{z}_{k+1} \triangleq \mathbf{x}_{k+1} - \mathbf{c}_\nu(k) \qquad \in \mathbb{R}^n \tag{48}$$

$$C_k^r \triangleq \left[\mathbf{c}_i(k) - \mathbf{c}_\nu(k)\right]_{i=1}^{\nu-1} \qquad \in \mathbb{R}^{n\times\nu-1} \tag{49}$$

and $\mathbf{c}_j(k)$, for $j = 1, 2, \ldots, \nu$, denote the following $n\times 1$ vectors

$$\mathbf{c}_j(k) \triangleq A_j \mathbf{x}_k + B_j \mathbf{u}_k \tag{50}$$

B. Development of the reduced-order model

Consider the $\nu\times 1$ process equation for $\mathbf{y}_k$, which in expanded form is written as

$$\begin{bmatrix} y_0(k+1) \\ \vdots \\ y_\nu(k+1) \end{bmatrix} = \begin{bmatrix} m_{1,1} & \ldots & m_{1,\nu} \\ \vdots & & \vdots \\ m_{\nu,1} & \ldots & m_{\nu,\nu} \end{bmatrix} \begin{bmatrix} y_0(k) \\ \vdots \\ y_\nu(k) \end{bmatrix} + \begin{bmatrix} v_1(k) \\ \vdots \\ v_\nu(k) \end{bmatrix} \tag{51}$$

Then, substituting Eq. (41) in Eq. (51) and rearranging yields

$$\begin{aligned} &\begin{bmatrix} y_0(k+1) \\ \vdots \\ y_\nu(k+1) \end{bmatrix} = \\ &= \begin{bmatrix} m_{1,1}-m_{1,\nu} & \ldots & m_{1,\nu-1}-m_{1,\nu} & 0 \\ \vdots & & \vdots & \vdots \\ m_{\nu,1}-m_{\nu,\nu} & \ldots & m_{\nu,\nu-1}-m_{\nu,\nu} & 0 \end{bmatrix} \begin{bmatrix} y_0(k) \\ \vdots \\ y_\nu(k) \end{bmatrix} \\ &\quad + \begin{bmatrix} m_{1,\nu} \\ \vdots \\ m_{\nu,\nu} \end{bmatrix} + \begin{bmatrix} v_1(k) \\ \vdots \\ v_\nu(k) \end{bmatrix} \end{aligned} \tag{52}$$

and deleting the last equation from Eq. (52) yields the sought equation (43). The observation equation for $\mathbf{y}_k$ is developed as follows:

$$\begin{aligned} \mathbf{x}_{k+1} &= \left[\mathbf{c}_1(k) \ldots \mathbf{c}_{\nu-1}(k)\, \mathbf{c}_\nu(k)\right] \begin{bmatrix} y_0(k) \\ \vdots \\ y_{\nu-1}(k) \\ 1-\sum_{i=1}^{\nu-1} y_i(k) \end{bmatrix} + \mathbf{w}_k \\ &= \left[\ldots \mathbf{c}_i(k) - \mathbf{c}_\nu(k) \ldots\right] \begin{bmatrix} y_0(k) \\ \vdots \\ y_{\nu-1}(k) \end{bmatrix} + \mathbf{c}_\nu(k) + \mathbf{w}_k \\ &= C_k^r \mathbf{y}_k^r + \mathbf{c}_\nu(k) + \mathbf{w}_k \end{aligned} \tag{53}$$

which, using Eqs. (48) and (49) yields Eq. (47).

VIII. CONCLUSION

Two types of recursive distribution-free mode-estimators for a class of discrete-time jump-linear systems were developed in this work. Their derivation was straightforward thanks to the linear structure of the state space model for the mode. Model order truncation did away with the singularity issue without drawback. Performance-wise the second filter should be preferred since it better approximates the conditional expectation of the mode. Both algorithms have structures that allow for filter tuning in practical implementations. The envisioned work will be to implement the suboptimal non-linear filter in the analysis and design of a dual controller for this type of jump-linear systems.

APPENDIX

State-space model

Consider a joint vector random sequence $\{\mathbf{x}_k, \mathbf{y}_k\}$, $k \in \mathbb{N}$, where $\mathbf{x}_k \in \mathbb{R}^{n_x}$ is known and $\mathbf{y}_k \in \mathbb{R}^{n_y}$ is unknown. The dynamics of these sequences are governed by the following random difference equations

$$\mathbf{x}_{k+1} = \mathbf{f}_k(\mathcal{X}^k, \mathcal{Y}^k) + \mathbf{w}_k \tag{1}$$

$$\mathbf{y}_{k+1} = \mathbf{g}_k(\mathcal{X}^k, \mathcal{Y}^k) + \mathbf{v}_k \tag{2}$$

In Eqs. (1) and (2), $\mathbf{f}_k$ and $\mathbf{g}_k$ are mappings into $\mathbb{R}^{n_x}$ and $\mathbb{R}^{n_y}$, respectively, that satisfy to the conditions of existence and uniqueness of a solution, $\mathcal{X}^k$ and $\mathcal{Y}^k$ denote the past histories of $\mathbf{x}$ and $\mathbf{y}$, respectively, up to time t_k; that is, $\mathcal{X}^k = \{\mathbf{x}_0, \mathbf{x}_1, \ldots, \mathbf{x}_k\}$ and $\mathcal{Y}^k = \{\mathbf{y}_0, \mathbf{y}_1, \ldots, \mathbf{y}_k\}$, the sequence $\{\mathbf{w}_k, \mathbf{v}_k\} \in \mathbb{R}^{n_x} \times \mathbb{R}^{n_y}$ denotes a zero-mean white noise sequence with a known cross-covariance matrix given as

$$\operatorname{cov}\left\{\begin{bmatrix}\mathbf{w}_k \\ \mathbf{v}_k\end{bmatrix}\right\} = \begin{bmatrix} W_k & D_k^T \\ D_k & V_k \end{bmatrix} \tag{3}$$

The vector $\mathbf{w}_k$ is assumed to be independent from $\{\mathcal{X}^k, \mathcal{Y}^k\}$, $k \in \mathbb{N}$ and the vector $\mathbf{v}_k$ is assumed to be independent from $\mathcal{X}^k$ and orthogonal to $\mathcal{Y}^k$, $k \in \mathbb{N}$. The initial vector $\mathbf{x}_0$ is known as well as the mean and the covariance matrix of the initial vector $\mathbf{y}_0$; that is, $E\{\mathbf{y}_0\}$ and $P_{\mathbf{y}_0}$. Consider the optimal filtering problem that consists in solving for the mean-square estimate of $\mathbf{y}_k$ given $\mathcal{X}^k$, which is known to be the conditional expectation of $\mathbf{y}_k$ given $\mathcal{X}^k$. A general suboptimal algorithm for sequential computation of $\widehat{\mathbf{y}}_{k/k}$ is described in the following. The approximation is to first order in Δt, where Δt denotes the underlying discretization incremental time associated with Eq. (1). Let $\widehat{A}_{k/l}$ and $\widetilde{A}_{k/l}$ denote, respectively, the estimate and the estimation error at t_k given $\mathcal{X}^l$.

Summary of the filtering equations

1) Initialization stage equation:

$$\widehat{\mathbf{y}}_{0/0} = E\{\mathbf{y}_0\} \tag{4}$$

$$P_{0/0} = P_{\mathbf{y}_0} \tag{5}$$

2) Smoothing stage equations:
Assuming that the estimates $\widehat{\mathbf{y}}_{k/k}$ and $\widehat{\mathbf{f}}_{k/k}$ can be computed, a new observation, $\mathbf{x}_{k+1}$, is acquired at time t_{k+1} and processed in order to yield a smoothed estimate of $\mathbf{y}_k$ at t_{k+1}, denoted by $\widehat{\mathbf{y}}_{k/k+1}$. The smoothing stage equations are:

$$\widetilde{\mathbf{x}}_{k+1/k} = \mathbf{x}_{k+1} - \widehat{\mathbf{f}}_{k/k} \tag{6}$$

$$P_{\widetilde{\mathbf{f}}_{k/k}} = \operatorname{cov}\{\widetilde{\mathbf{f}}_{k/k} | \mathcal{X}^k\} \tag{7}$$

$$P_{\widetilde{\mathbf{x}}_{k+1/k}} = P_{\widetilde{\mathbf{f}}_{k/k}} + W_k \tag{8}$$

$$P_{\mathbf{y}\widetilde{\mathbf{f}}_{k/k}} = \operatorname{cov}\{\mathbf{y}_k, \widetilde{\mathbf{f}}_{k/k} | \mathcal{X}^k\} \tag{9}$$

$$\widehat{\mathbf{y}}_{k/k+1} = \widehat{\mathbf{y}}_{k/k} + P_{\mathbf{y}\widetilde{\mathbf{f}}_{k/k}} P^{-1}_{\widetilde{\mathbf{x}}_{k+1/k}} \widetilde{\mathbf{x}}_{k+1/k} \tag{10}$$

3) Time-Propagation stage equations:
Assuming that the estimate $\widehat{\mathbf{g}}_{k/k+1}$ can be computed, the estimate $\widehat{\mathbf{y}}_{k+1/k+1}$ is computed at time t_{k+1} using the cross-correlation between $\mathbf{w}_k$ and $\mathbf{v}_k$. The time-propagation equation is:

$$\widehat{\mathbf{y}}_{k+1/k+1} = \widehat{\mathbf{g}}_{k/k+1} + D_k P^{-1}_{\widetilde{\mathbf{x}}_{k+1/k}} \widetilde{\mathbf{x}}_{k+1/k} \tag{11}$$

The approximation stems from the measurement update Eq. (10). The second term on the RHS only represents the orthogonal projection of $\mathbf{y}_k$ onto the linear manifold generated by $\widetilde{\mathbf{x}}_{k+1/k}$, rather the orthogonal projection of $\mathbf{y}_k$ onto the Hilbert space generated by any bounded function of the residual.

References

[1] Liberzon, D. and Morse, A.S. “Basic Problems in Design and Stability of Switched Systems,” *IEEE Control Systems Letters*, Oct. 1999, pp. 59–70.

[2] Wonham, W.M., Random Differential Equations in Control Theory, In A. Bharucha-Reid, editor, *Probabilistic Methods in Applied Mathematics*, Volume 2, Academic Press, 1970, pp. 192–194.

[3] Liptser, R.S., and Shiryayev, A.N., *Statistics of Random Processes I: General Theory*, Springer-Verlag New-York Inc., 1977.

[4] Wonham, W.M., *Some Applications of Staaochastic Differential Equations to Optimal Non-Linear Filtering*, Tech-Rept. 64-3, Feb. 1964, RIAS, Baltimore.

[5] Liptser, R.S., *Lecture Notes in Stochastic Processes*, Lect. 8, Tel-Aviv University, Israel.

[6] Elliot, R.J., Aggoun, L., and Moore, J.B., *Hidden Markov Models, Estimation and Control*, Springer-Verlag New-York Inc., 1995

[7] Mendel, J.M., *Lessons in Estimation Theory for Signal Processing, Communications, and Control*, Prentice Hall PTR, Englewood Cliffs, N.J., 1995.

[8] Jazwinski, A.H., *Stochastic Processes and Filtering Theory*, New York: Academic, 1970.

[9] Liptser, R.S., and Shiryayev, A.N., *Statistics of Random Processes II: Applications*, Springer-Verlag New-York Inc., 1977.