DEALING WITH UNCERTAINTY IN THE HYBRID WORLD*

Luís Pina and Miguel Ayala Botto
Department of Mechanical Engineering, IDMEC, Instituto Superior Técnico
Technical University of Lisbon, Portugal
luispina@dem.ist.utl.pt, ayalabotto@ist.utl.pt

Keywords: Hybrid systems, Hybrid Estimation, Interacting multiple-model estimation, Observability.

Abstract: This paper presents an efficient state estimation algorithm for hybrid systems based on a least-squares Interacting Multiple-Model setup. The proposed algorithm is shown to be computationally efficient when compared with the Moving Horizon Estimaton algorithm that is a brute force optimization algorithm for simultaneous discrete mode and continuous state estimation of a hybrid system. The main reason has to do with the fact that the proposed algorithm is able to disregard as many discrete mode sequence estimates as possible. This is done by rapidly computing good estimates, separating the constrained and unconstrained estimates, and using some auxiliary coefficients computed off-line. The success of this state estimation algorithm is shown for a fault detection problem of the benchmark AMIRA DTS200 three-tanks system experimental setup.

1 INTRODUCTION

In the last decade hybrid systems have become a major research topic in Control Engineering (Antsaklis, 2000). Hybrid systems are dynamical systems composed by both discrete valued and continuous valued states. The dynamics of a hybrid system is governed by a mode selector that determines, at each time instant, which discrete mode is active from endogenous and/or exogenous variables. The continuous state is then updated through a dynamic relation that is selected from a set of possible dynamics according to the value of the active discrete mode. In fact, the presence of physical components such as on/off switches or valves, gears or speed selectors, or behaviors dependent on if-then-else rules imply explicitly or implicitly the discrete/continuous interaction. This interaction can be found in many real world applications such as automotive control, urban and air traffic control, communications networks, embedded control systems, and in the control of complex industrial systems via the combination of classical continuous control laws with supervisory switching logic.

The hybrid nature has attracted the interest of mathematicians, control engineers and computer scientists, therefore leading to different modeling languages and paradigms that influenced the line of research on hybrid systems in several different ways. For instance, the computer science research community is more focused on systems whose variables take values in a finite set, so adopted the discrete events modeling formalism to model hybrid systems, using finite state machines, Petri nets, temporal logic, etc. On the other hand, the control systems community typically considers a continuous valued world, where time is continuously changing, thus considering a hybrid system as described by a differential (or difference) equation with some switching mechanism. Examples of such hybrid models include Piece-Wise Affine (PWA) (Sontag, 1981) and Mixed Logical Dynamical (MLD) (Bemporad and Morari, 1999) models. A PWA model is the most intuitive representation of a hybrid system since it provides a direct relation to linear systems while still capturing very complex dynamical behaviors. However, a MLD representation is most adequate to be used in optimization problems since it is able to embed both propositional logic statements (if-then-else rules) and operating constraints in a state linear dynamics equation by transforming them to mixed-integer linear inequalities. Despite these differences, PWA and MLD are equivalent models of hybrid systems in respect to well-posedness and boundness of input, state, output or auxiliary variables (Heemels et al., 2001). This fact allows to interchange analysis and synthesis tools between them.

*This work was supported by project PTDC/EME-CRO/69117/2006 co-sponsored by FEDER, Programa Operacional Ciência e Inovação 2010, Portugal.

Research on hybrid systems spans to a wide range of topics (and approaches), from modeling to stability analysis, reachability analysis and verification, study of the observability and controllability properties, methods of state estimation and fault detection, identification techniques, and control methodologies. Typically, hybrid tools rely on the solution of optimization problems. However, due to the different nature of the optimization variables involved (integer and continuous) the main source of complexity becomes the combinatorial (yet finite) number of possible switching sequences that have to be considered. A hybrid optimal solution thus requires solving mixed-integer non-convex optimization algorithms with NP-complete complexity (Torrisi and Bemporad, 2001).

Analysis and synthesis procedures for hybrid systems when disturbances are present either on the continuous dynamics or on the discrete mode of the hybrid system, is still an open research topic that has being tackled by several authors using distinct approaches. In the state estimation problem two distinct approaches are usually followed, the main difference being the knowledge of the active mode: some approaches consider only continuous state uncertainty with known discrete mode, while others assume that both the discrete mode and the continuous state are unknown. The combination of both uncertainties (state and mode) on the estimation process of a hybrid system presents a very difficult problem for which a global solution is not yet found. When the discrete mode is known in advance, the problem is greatly simplified and the state estimation methodologies for linear systems can be applied with very little modifications. For example in (Böker and Lunze, 2002) a bank of Kalman filters is used and in (Alessandri and Coletta, 2003) an LMI based algorithm computes the stabilizing gains for a set of Luenberger observers. If, on the other hand, the discrete mode must also be estimated the estimation problem becomes much more complex and every discrete mode sequence (*dms*) must be checked to choose the one that provides the best fit for the observed data. The continuous state estimates are then computed for the estimated *dms*. Several works address this problem, see (Balluchi et al., 2002) where a location observer is used to estimate the discrete mode and a Luenberger observer is then used to estimate the continuous state. In (Ferrari-Trecate et al., 2002) and (Pina and Botto, 2006) a Moving Horizon Estimation (MHE) scheme simultaneously estimates the discrete mode and the continuous state, differing in the fact that the latter can also estimate the input disturbances.

The derivation of the truly optimal filter for systems with switching parameters was first presented in (Athans and Chang, 1976). The objective was to perform simultaneous system identification and state estimation for linear systems but the derivation is quite general and is directly applicable to the hybrid state estimation problem. This method requires the consideration of all admissible *dms* starting from the initial time instant, being obviously unpractical since the number of *dms* grows exponentially in time, and so, suboptimal methods were developed. From the various possibilities, considering all the admissible *dms* of a given length is usually the preferred methodology. In view of this, suboptimal multiple model estimation schemes where then developed and applied for tracking maneuvering vehicles, as surveyed in (Mazor et al., 1998), and systems with Markovian switching coefficients, (Blom and Bar-Shalom, 1988), proving their efficiency for state estimation in multiple model systems. Multiple model estimation algorithms use a set of filters, one for each possible dynamic of the system. In this paper an efficient state estimation algorithm for stochastic hybrid systems, based on the Interacting Multiple-Model (IMM) estimation algorithm, is proposed. The method is applicable to most of the existing models of hybrid systems subject to disturbances with explicitly known probability density function, so being rather general. This estimation method will be further compared to the Moving Horizon Estimation (MHE) algorithm and tested in the benchmark AMIRA DTS200 three-tanks system experimental setup.

The paper is organized as follows. Section 2 provides a description of the considered PWA model and in section 3 the proposed Interacting Multiple-Model estimation algorithm is presented. Section 4 presents an experimental application of the proposed algorithms to the AMIRA DTS200 three-tanks system experimental setup. First the experimental setup is presented and modelled, including a full characterization of all uncertainties. Then the proposed algorithms are tested and their performance is compared. Finally, in section 5 some conclusions are drawn along with some possible future developments.

2 SYSTEM DESCRIPTION

The proposed estimation algorithm is developed for PWA systems which were introduced in (Sontag, 1981). The following stochastic PWA model will be considered:

$$x(k+1) = A_{i(k)}x(k)+B_{i(k)}u(k)+f_{i(k)}+L_{i(k)}w(k) \tag{1a}$$

$$y(k) = C_{i(k)}x(k)+D_{i(k)}u(k)+g_{i(k)}+v(k) \tag{1b}$$

$$\text{iff } \begin{bmatrix} x(k) \\ u(k) \\ w(k) \end{bmatrix} \in \Omega_{i(k)} \tag{1c}$$

where k is the discrete time, $x(k) \in \mathbb{X} \subset \mathbb{R}^{n_x}$ is the continuous state, $u(k) \in \mathbb{U} \subset \mathbb{R}^{n_u}$ is the input, $y(k) \in \mathbb{R}^{n_y}$ is the output, $i(k) \in I = \{1,\ldots,s\}$ is the discrete mode, and s is the total number of discrete modes. The matrices and vectors A_i, B_i, f_i, L_i, C_i, D_i, g_i depend on the discrete mode $i(k)$ and have appropriate dimensions. The input disturbance $w(k)$ and the measurement noise $v(k)$ are modelled as independent identically distributed random variables, belonging to the sets $\mathbb{W}_i$ and $\mathbb{V}_i$, with expected values $E\{w(k)\} = 0$, $E\{v(k)\} = 0$ and covariances Σ_{w_i} and Σ_{v_i}, respectively. These conditions are not restrictive at all since the zero mean can be imposed by summing a constant vector to the disturbances and compensated in the affine term of the system dynamics (1) and, the sets $\mathbb{W}_i$ and $\mathbb{V}_i$ can be considered large enough to contain all possible disturbances relevant for practical applications, for instance 99.99% of all admissible values. Notice that the input disturbance and measurement noise *pdf*s may depend on the actual mode of the system $i(k)$. The sets $\mathbb{W}_i$ and $\mathbb{V}_i$ are respectively defined for each mode $i(k)$ by:

$$H_{\mathbb{W}_{i(k)}}\, w(k) \leq h_{\mathbb{W}_{i(k)}} \quad , \quad \forall_{k\in\mathbb{N}_0} \tag{2}$$

$$H_{\mathbb{V}_{i(k)}}\, v(k) \leq h_{\mathbb{V}_{i(k)}} \quad , \quad \forall_{k\in\mathbb{N}_0} \tag{3}$$

The discrete mode $i(k)$ is a piecewise constant function of the state, input and input disturbance of the system whose value is defined by the regions Ω_i:

$$\Omega_i \quad : \quad S_i\, x(k) + R_i\, u(k) + Q_i\, w(k) \leq T_i \tag{4}$$

Some helpful notation regarding the time-compressed representation of (Kamen, 1992) for system (1) will now be introduced. The time-compressed representation of a system defines the dynamics of the system over a sequence of time instants in opposition to the single time step state-space representation. Consider the time interval $[k, k+T-1]$, the sequence of discrete modes over this interval is represented as $\mathbf{i}_T = \mathbf{i}_T(k) \triangleq \{i(k),\ldots,i(k+T-1)\}$. To simplify the notation, the time index k is removed from the discrete mode sequence (*dms*) whenever it is obvious from the other elements in the equations. In view of this, the output sequence over the same interval can be computed by:

$$Y_T(k) = \mathbf{C}_{\mathbf{i}_T}x(k)+\mathbf{D}_{\mathbf{i}_T}U_T(k)+\mathbf{g}_{\mathbf{i}_T}+\mathbf{L}_{\mathbf{i}_T}W_T(k)+V_T(k) \tag{5}$$

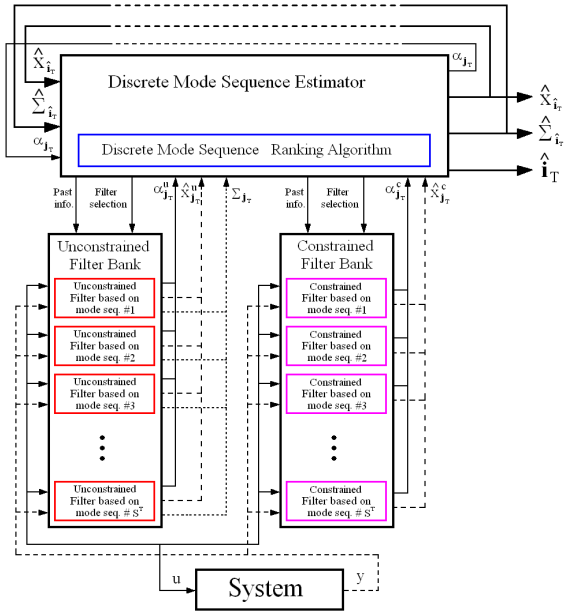

Figure 1: Interacting Multiple-Model Estimation Algorithm.

where the input, input disturbance and measurement noise sequences $U_T(k)$, $W_T(k)$ and $V_T(k)$ respectively are defined in the same way as the output sequence $Y_T(k) \triangleq [y(k)^{\mathrm{T}},\ldots,y(k+T-1)^{\mathrm{T}}]^{\mathrm{T}}$. The matrices and vectors $\mathbf{C}_{\mathbf{i}_T}$, $\mathbf{D}_{\mathbf{i}_T}$, $\mathbf{g}_{\mathbf{i}_T}$ and $\mathbf{L}_{\mathbf{i}_T}$ are computed from the system dynamics (1a-1b) according to what is presented in (Kamen, 1992). The same reasoning can be applied to the constraints $\Omega_{\mathbf{i}_T}$:

$$\Omega_{\mathbf{i}_T} : \quad \mathbf{S}_{\mathbf{i}_T}\, x(k) + \mathbf{R}_{\mathbf{i}_T}\, U_T(k) + \mathbf{Q}_{\mathbf{i}_T}\, W_T(k) \leq \mathbf{T}_{\mathbf{i}_T} \tag{6}$$

where the matrices $\mathbf{S}_{\mathbf{i}_T}$, $\mathbf{R}_{\mathbf{i}_T}$, $\mathbf{Q}_{\mathbf{i}_T}$ and $\mathbf{T}_{\mathbf{i}_T}$ can be computed from the system dynamics (1a) and partitions (4). The inequalities that define the disturbance and noise sets over a *dms* $\mathbf{i}_T$, $\mathbb{W}_{\mathbf{i}_T}$ and $\mathbb{V}_{\mathbf{i}_T}$ respectively, can also be easily found from equations (2) and (3):

$$\mathbf{H}_{\mathbb{W}_{\mathbf{i}_T}}\, W_T(k) \leq \mathbf{h}_{\mathbb{W}_{\mathbf{i}_T}} \tag{7}$$

$$\mathbf{H}_{\mathbb{V}_{\mathbf{i}_T}}\, V_T(k) \leq \mathbf{h}_{\mathbb{V}_{\mathbf{i}_T}} \tag{8}$$

3 INTERACTING MULTIPLE MODEL ESTIMATION

The proposed Interacting Multiple-Model (IMM) Estimation algorithm is composed of three parts; the Unconstrained Filter Bank (UFB), the Constrained Filter Bank (CFB) and, the Discrete Mode Sequence Estimator (DMSE). A schematic representation is presented in figure 1.

The estimation algorithm works as follows: first the continuous state estimates are computed in the UFB without considering the constraints. Then, the DMSE computes the squared errors of these estimates and ranks them. Finally, starting with the estimate with the lowest squared error, the estimates are recomputed in the CFB considering the presence of constraints. When the most accurate estimate is already a constrained estimate the whole process stops.

As the estimation is based on sequences of measurements $Y_T(k)$ and discrete modes $\mathbf{i}_T(k)$, two distinct time instants must be considered: the time instant at the beginning of the sequences, k, and the time instant at the end of these sequences, which is the present time instant $t = k+T-1$. The state estimates will be computed at time instant k, and can be propagated to the present time instant according to the estimated dynamics.

3.1 Unconstrained Filter Bank

The UFB computes the unconstrained state estimates. It is composed by a set of unconstrained least-squares filters, one for each possible *dms* $\mathbf{j}_T$:

$$\hat{x}^u_{\mathbf{j}_T}(k|t) = \hat{x}_{\mathbf{j}_T}(k|t-1) + \tag{9}$$

$$\mathbf{K}_{\mathbf{j}_T}(k|t-1)\left[\left(Y_T(k)-\mathbf{D}_{\mathbf{j}_T}U_T(k)-\mathbf{g}_{\mathbf{j}_T}\right)-\mathbf{C}_{\mathbf{j}_T}\hat{x}_{\mathbf{j}_T}(k|t-1)\right]$$

where $\hat{x}_{\mathbf{j}_T}(k|t-1)$ is the *a priori* continuous state estimate for mode sequence $\mathbf{j}_T$ using measurements up to time instant $t-1$. $\mathbf{K}_{\mathbf{j}_T}(k|t-1)$ is the filter gain:

$$\mathbf{K}_{\mathbf{j}_T}(k|t-1) = \left(\Sigma^{-1}_{x_{\mathbf{j}_T}}(k|t-1)+\mathbf{C}^{\mathrm{T}}_{\mathbf{j}_T}\Sigma^{-1}_{Y_{\mathbf{j}_T}}\mathbf{C}_{\mathbf{j}_T}\right)^{-1}\mathbf{C}^{\mathrm{T}}_{\mathbf{j}_T}\Sigma^{-1}_{Y_{\mathbf{j}_T}} \tag{10}$$

$$\Sigma_{Y_{\mathbf{j}_T}} = \left[\mathbf{L}_{\mathbf{j}_T}\ I_{T.n_y}\right]\begin{bmatrix}\Sigma_{W_{\mathbf{j}_T}} & 0\\ 0 & \Sigma_{V_{\mathbf{j}_T}}\end{bmatrix}\left[\mathbf{L}_{\mathbf{j}_T}\ I_{T.n_y}\right]^{\mathrm{T}} \tag{11}$$

The covariance of the obtained unconstrained estimate can also be computed:

$$\Sigma_{x_{\mathbf{j}_T}}(k|t) = \left(\Sigma^{-1}_{x_{\mathbf{j}_T}}(k|t-1)+\mathbf{C}^{\mathrm{T}}_{\mathbf{j}_T}\Sigma^{-1}_{Y_{\mathbf{j}_T}}\mathbf{C}_{\mathbf{j}_T}\right)^{-1} \tag{12}$$

This covariance matrix not only provides some insight on the accuracy of the continuous state estimate $\hat{x}^u_{\mathbf{j}_T}(k|t)$, but also defines the confidence on the past information at the subsequent time instant $\hat{x}_{\mathbf{j}_T}(k+1|t)$:

$$\Sigma_{x_{\mathbf{j}_T}}(k+1|t) = A_{j(k)}\Sigma_{x_{\mathbf{j}_T}}(k|t)A^{\mathrm{T}}_{j(k)} + L_{j(k)}\Sigma_{wj(k)}L^{\mathrm{T}}_{j(k)} \tag{13}$$

When computing the unconstrained state estimate, no *a priori* information may be available or one may be interested in discarding it, then $\Sigma^{-1}_{x_{\mathbf{j}_T}}(k|t-1)$ should be set to 0. The corresponding unconstrained state estimate is referred to as $\hat{x}^{u*}_{\mathbf{j}_T}(k|t)$.

3.2 Constrained Filter Bank

The CFB will recompute the state estimates but now considering the constraints (6), (7) and (8). The constrained least-squares filter is somehow more complicated. First the least-squares state vector must be augmented to incorporate both the input disturbance and measurement noise vectors, since there exist explicit constraints on these variables:

$$\begin{bmatrix} x_{\mathbf{j}_T}(k) \\ W_{\mathbf{j}_T}(k) \\ V_{\mathbf{j}_T}(k)\end{bmatrix} \tag{14}$$

Notice that by explicitly considering the input disturbance and measurement noise sequences, all the uncertainty is removed from the observation equation (5) and it becomes an equality constraint:

$$\mathbf{H_e}\ .\begin{bmatrix} x_{\mathbf{j}_T}(k) \\ W_{\mathbf{j}_T}(k) \\ V_{\mathbf{j}_T}(k)\end{bmatrix} = \mathbf{h_e} \quad \Leftrightarrow \tag{15}$$

$$\Leftrightarrow\ \left[\mathbf{C}_{\mathbf{j}_T}\ \mathbf{L}_{\mathbf{j}_T}\ I_{n_Y}\right].\begin{bmatrix} x_{\mathbf{j}_T}(k) \\ W_{\mathbf{j}_T}(k) \\ V_{\mathbf{j}_T}(k)\end{bmatrix} = \left[Y_T(k)-\mathbf{D}_{\mathbf{j}_T}U_T(k)-\mathbf{g}_{\mathbf{j}_T}\right]$$

The constraints of the *dms* (6) and the bounds on the input disturbance and measurement noise vectors defined by the sets $\mathbb{W}_{\mathbf{j}_T}$ and $\mathbb{V}_{\mathbf{j}_T}$ described by equations (7) and (8) compose the inequality constraints of the least-squares problem, according to:

$$\mathbf{H_i}\ .\begin{bmatrix} x_{\mathbf{j}_T}(k) \\ W_{\mathbf{j}_T}(k) \\ V_{\mathbf{j}_T}(k)\end{bmatrix} \leq \mathbf{h_i} \quad \Leftrightarrow \tag{16}$$

$$\Leftrightarrow \begin{bmatrix}\mathbf{S}_{\mathbf{j}_T} & \mathbf{Q}_{\mathbf{j}_T} & 0\\ 0 & \mathbf{H}_{\mathbb{W}_{\mathbf{j}_T}} & 0\\ 0 & 0 & \mathbf{H}_{\mathbb{V}_{\mathbf{j}_T}}\end{bmatrix}.\begin{bmatrix} x_{\mathbf{j}_T}(k) \\ W_{\mathbf{j}_T}(k) \\ V_{\mathbf{j}_T}(k)\end{bmatrix} \leq \begin{bmatrix}\mathbf{T}_{\mathbf{j}_T}-\mathbf{R}_{\mathbf{j}_T}U_T(k) \\ \mathbf{h}_{\mathbb{W}_{\mathbf{j}_T}} \\ \mathbf{h}_{\mathbb{V}_{\mathbf{j}_T}}\end{bmatrix}$$

Having defined the constraints matrices, the constrained least-squares filter corresponding to the mode sequence $\mathbf{j}_T$ is given by:

$$\begin{bmatrix} \hat{x}_{\mathbf{j}_T}(k|t) \\ \hat{W}_{\mathbf{j}_T}(k|t) \\ \hat{V}_{\mathbf{j}_T}(k|t)\end{bmatrix} = \begin{bmatrix} \hat{x}_{\mathbf{j}_T}(z,k|t-1) \\ \hat{W}_{\mathbf{j}_T}(k|t-1) \\ \hat{V}_{\mathbf{j}_T}(k|t-1)\end{bmatrix} + \mathbf{K}_{\mathbf{j}_T}(k|t)\left(\begin{bmatrix}\mathbf{h_e}\\ \mathbf{h_i}\end{bmatrix}-\begin{bmatrix}\mathbf{H_e}\\ \mathbf{H_i}\end{bmatrix}.\begin{bmatrix} \hat{x}_{\mathbf{j}_T}(k|t-1) \\ \hat{W}_{\mathbf{j}_T}(k|t-1) \\ \hat{V}_{\mathbf{j}_T}(k|t-1)\end{bmatrix}\right) \tag{17}$$

The constrained least-squares filter gain is defined as:

$$\mathbf{K}_{\mathbf{j}_T}(k|t) = \tag{18}$$

$$\left(\begin{bmatrix}\Sigma_{x_{\mathbf{j}_T}}(k|t-1) & 0 & 0\\ 0 & \Sigma_{W_{\mathbf{j}_T}} & 0\\ 0 & 0 & \Sigma_{V_{\mathbf{j}_T}}\end{bmatrix}^{-1}+\begin{bmatrix}\mathbf{H_e}\\ \mathbf{H_i}\end{bmatrix}^{\mathrm{T}}\mathbf{Z}_{\mathbf{j}_T}(k|t)\begin{bmatrix}\mathbf{H_e}\\ \mathbf{H_i}\end{bmatrix}\right)^{-1}\begin{bmatrix}\mathbf{H_e}\\ \mathbf{H_i}\end{bmatrix}^{\mathrm{T}}\mathbf{Z}_{\mathbf{j}_T}(k|t)$$

where $\Sigma_{x_{\mathbf{j}_T}}(k|t-1)$ is the covariance matrix associated with the *a priori* state estimate $\hat{x}_{\mathbf{j}_T}(k|t-1)$. $\mathbf{Z}_{\mathbf{j}_T}(k|t)$ is the diagonal matrix that defines the active constraints.

There are several methods, most of them iterative, for determining the matrix $\mathbf{Z}_{\mathbf{j}_T}(k|t)$, or equivalently the set of active constraints. Here, the active set method presented in (Fletcher, 1987) will be used.

As in the unconstrained case, *a priori* information may be discarded by setting $\Sigma^{-1}_{x_{\mathbf{j}_T}}(k|t-1)$ to 0. The corresponding constrained state estimate is referred to as $\hat{x}^{c*}_{\mathbf{j}_T}(k|t)$.

3.3 Discrete Mode Sequence Estimator

The DMSE deals with the estimation of the discrete mode sequence and, consequently, selects the filter which will provide the final continuous state estimate.

According to the least-squares philosophy, an approximation of the measured output sequence is computed for every possible *dms* and then, the one providing the smallest squared error should be selected as the least-squares estimate.

The *dms* estimate is then selected as the one that presents the lowest constrained squared error, $\alpha^c_{\mathbf{j}_T}$:

$$\hat{\mathbf{i}}_T(k|t) = \arg\min_{\mathbf{j}_T} \alpha^c_{\mathbf{j}_T}(k|t) \tag{19}$$

The squared error associated with the *dms* $\mathbf{j}_T$ is given by:

$$\begin{aligned}\alpha_{\mathbf{j}_T}(k|t) &= \left\|\hat{Y}^*_{\mathbf{j}_T}(k|t) - Y_T(k)\right\|^2_{\Sigma^{-1}_{Y_{\mathbf{j}_T}}} = \\ &= \left[\hat{Y}^*_{\mathbf{j}_T}(k|t) - Y_T(k)\right]^{\mathrm{T}} \Sigma^{-1}_{Y_{\mathbf{j}_T}} \left[\hat{Y}^*_{\mathbf{j}_T}(k|t) - Y_T(k)\right]\end{aligned} \tag{20}$$

where:

$$\hat{Y}^*_{\mathbf{j}_T}(k|t) = \mathbf{C}_{\mathbf{j}_T}\hat{x}^*_{\mathbf{j}_T}(k|t) + \mathbf{D}_{\mathbf{j}_T}U_T(k) + \mathbf{g}_{\mathbf{j}_T} \tag{21}$$

and $\hat{x}^*_{\mathbf{j}_T}(k|t)$ is the estimated state of the *dms* $\mathbf{j}_T$ when all past information is discarded, $\left(\Sigma^{-1}_{x_{\mathbf{j}_T}}(k|t-1) = 0\right)$.

The squared errors computed by equation and (20) are useful when comparing continuous state estimates from the same *dms*. However, when the covariance matrices are different, an additional factor, $\bar{\alpha}_{\mathbf{j}_T}$, must be considered to allow a meaningful comparison between squared errors. Recalling the relation between least-squares and the maximization of the Gaussian likelihood function (or its logarithm), the value of $\bar{\alpha}_{\mathbf{j}_T}$ should be defined as:

$$\bar{\alpha}_{\mathbf{j}_T} = -\frac{1}{2}\ln\left((2\pi)^{n_Y}\det\left(\Sigma_{Y_{\mathbf{j}_T}}\right)\right) \tag{22}$$

Equation (20) should be modified to:

$$\alpha_{\mathbf{j}_T}(k|t) = \bar{\alpha}_{\mathbf{j}_T} + \left\|\hat{Y}^*_{\mathbf{j}_T}(k|t) - Y_T(k)\right\|^2_{\Sigma^{-1}_{Y_{\mathbf{j}_T}}} \tag{23}$$

Equation (23) can be used to compute the squared errors of both the unconstrained estimates, $\alpha^u_{\mathbf{j}_T}(k|t)$, and the constrained estimates, $\alpha^c_{\mathbf{j}_T}(k|t)$, using $\hat{x}^{u*}_{\mathbf{j}}(k|t)$ and $\hat{x}^{c*}_{\mathbf{j}}(k|t)$, respectively.

3.4 Computational Issues

Concerning computational requirements, it is noticed that there can be as many as n_s^T *dms*, which becomes an extremely large number even for relatively small n_s and T. So, computationally demanding calculations should be preformed for the minimum number of *dms* possible.

Analyzing the required computations one concludes that $\hat{x}^{u*}_{\mathbf{j}_T}(k|t)$ can be determined by simple matrix sums and multiplications if the filter gain $\mathbf{K}_{\mathbf{j}_T}(k|t-1)$ is computed off-line, since there are no varying terms as can be seen in equation (9). The corresponding squared error $\alpha^u_{\mathbf{j}_T}(k|t)$, computed through equation (23), can also be determined using simple matrix sums and multiplications from $\hat{x}^{u*}_{\mathbf{j}_T}(k|t)$. The continuous state estimate $\hat{x}^u_{\mathbf{j}_T}(k|t)$ on the other hand, requires a matrix inversion to determine the corresponding filter gain using equation (10) since the matrix $\Sigma^{-1}_{x_{\mathbf{j}_T}}(k|t-1)$ is not known in advance.

The constrained estimates require much more complex computations in the solution of the inequality constrained least-squares problem. An iterative algorithm has to be preformed online, and involves one matrix inversion at each iteration which is computationally heavy. There is the possibility that the solution corresponding to the true *dms* is the same as the unconstrained solution and the iterative algorithm stops at the first iteration. In general, however, this will not be the case. So, the computation of constrained solutions should only be done in cases of absolute necessity. The squared error of the constrained estimates $\alpha^c_{\mathbf{j}_T}(k|t)$ can be determined using simple matrix sums and multiplications from $\hat{x}^{c*}_{\mathbf{j}}(k|t)$.

The proposed algorithm should take these knowledge into account and arrive at the final estimates in the most efficient way possible.

To avoid the computation of the constrained least-squares estimates from all discrete mode sequences, the following relation between the constrained and unconstrained squared errors for a given discrete mode sequence is used:

$$\alpha^u_{\mathbf{j}_T}(k|t) \leq \alpha^c_{\mathbf{j}_T}(k|t) \tag{24}$$

An efficient reduction on the number of constrained estimates that have to be computed can be achieved by computing all unconstrained estimates $\hat{x}^{u*}_{\mathbf{j}}(k|t)$ and the corresponding squared errors $\alpha^u_{\mathbf{j}_T}(k|t)$ and then, start replacing the unconstrained solutions with the

corresponding constrained ones, from the lower values of the squared error. Whenever the lowest squared error corresponds to a constrained solution, the algorithm stops since no further reduction of the squared error can be done. The discrete mode sequence and continuous state estimates are the ones corresponding to that lowest squared error.

This algorithmic procedure may provide a substantial reduction in the number of inequality constrained least-squares problems to be solved since the increase in the squared error should be small, or even zero, for the true *dms*. However, the unconstrained solutions of incorrect *dms* may have low squared errors, which rise substantially only when the respective constrained solutions are computed. An efficient procedure to detect these incorrect *dms* before computing the respective constrained estimates would reduce the computational requirements even more.

To further improve the algorithm, the following $\mathcal{B}$ matrix must be introduced. Each coefficient $\beta_{\mathbf{i}_T,\mathbf{j}_T}$ of the matrix $\mathcal{B}$ is defined as the maximum value of $\alpha^c_{\mathbf{i}_T}$ under which $\alpha^c_{\mathbf{i}_T}$ is always smaller than $\alpha^c_{\mathbf{j}_T}$, or in an even more restrictive way, under which $\mathbf{j}_T$ is never the estimated sequence. The coefficients $\beta_{\mathbf{i}_T,\mathbf{j}_T}$ can be computed off-line by the following optimization problem, which falls in the general class of Second-Order Cone Programs for which efficient solvers have already been developed, for instance, by (Alizadeh and Goldfarb, 2001):

$$\begin{aligned} \beta_{\mathbf{i}_T,\mathbf{j}_T} \;=\; & \min_{Y_T,U_T} \;\; \alpha^c_{\mathbf{i}_T}(Y_T,U_T) \\ \text{subject to :} & \\ & U_T \in \mathbb{U}^T \\ & \hat{\mathbf{i}}_T = \mathbf{j}_T \end{aligned} \tag{25}$$

By this definition of $\beta_{\mathbf{i}_T,\mathbf{j}_T}$, when the constrained solution of a *dms* $\mathbf{i}_T$ is computed, all *dms* $\mathbf{j}_T$ such that $\beta_{\mathbf{i}_T,\mathbf{j}_T}$ is greater than $\alpha^c_{\mathbf{i}_T}(k|t)$ can be discarded. This algorithmic procedure provides an even greater reduction on the number of constrained problems to be solved. Notice that this procedure does not even require the computation of the unconstrained solutions of the *dms* to be discarded.

Both previous modifications to the algorithm require the existence of one constrained solution to discard any other *dms*. Furthermore, the number of discarded *dms* depends on the quality of the constrained solution. In the following, some attention will be given to the recursiveness of the DMSE and the methodology to determine the *dms* that will most likely provide good constrained estimates.

At a given time instant $t+1$ the following quantities have been computed at the previous time instant: the discrete mode sequence estimate, $\hat{\mathbf{i}}_T(k|t)$, the squared errors (or lower bounds) of all *dms*, $\alpha^c_{\mathbf{j}_T}(k|t)$ and, the continuous state estimates $\hat{x}^{c*}_{\mathbf{j}_T}(k|t)$ and the values of the estimated input disturbances $\hat{W}_{\mathbf{j}_T}(k|t)$ for the *dms* whose squared errors have been computed, including the *dms* estimate. These quantities allow the computation of the *a priori* continuous state estimate corresponding to the discrete mode sequence estimate at the following time instant:

$$\begin{aligned} \hat{x}^*_{\mathbf{j}_T}(t+1|t) = & \Big(A_{j(t)} \ldots A_{j(k)}\Big)\hat{x}^*_{\mathbf{j}_T}(k|t) + \\ & \Big[A_{j(t)} \ldots A_{j(k+1)} B_{j(k)}, \ldots, B_{j(t)}\Big] U_T(k) + \\ & \Big[A_{j(t)} \ldots A_{j(k+1)} W_{j(k)}\,, \ldots,\, W_{j(t)}\Big] \hat{W}_{\mathbf{j}_T}(k|t) + \\ & \Big(A_{j(t)} \ldots A_{j(k+1)} f_{j(k)} + \ldots + f_{j(t)}\Big) \end{aligned} \tag{26}$$

This estimate can be used to obtain some insight on the likelihood of the discrete mode at the next time instant $j(t+1)$. The discrete modes $j(t+1)$ can be sorted by ascending values of:

$$\begin{aligned} & \gamma_{\mathbf{j}_T,j}(t+1|t) = \\ & \max\Big(S_j\,\hat{x}^*_{\mathbf{j}_T}(t+1|t) + R_j u(t+1) + Q_j \hat{w}(t+1|t) - T_j\Big) \end{aligned} \tag{27}$$

The value of $\hat{w}(t+1)$ should be set to $\mathrm{E}\{w_j\}$.

The discrete modes $j(t+1)$ that provide the lower values of $\gamma_{\mathbf{j}_T,j}(t+1|t)$ correspond the discrete mode sequences $\mathbf{j}_T = \big\{j(k+1), \ldots, j(t), j(t+1)\big\}$ at time instant $t+1$ most likely to succeed to $\mathbf{j}_T$ at time instant t.

Applying this methodology to the discrete mode sequence estimate at the previous time instant, $\hat{\mathbf{i}}_T(k|t)$, should provide *dms* with very low squared errors that discard most of the other candidate *dms*. The same reasoning should be applied to all other discrete mode sequences of the previous time instant that have not been discarded yet, starting from the ones that present lowest squared errors and then the ones with the lowest bounds.

4 EXPERIMENTAL APPLICATION

To demonstrate the applicability of the hybrid estimation algorithms, the laboratory setup of the DTS200 three-tanks system from AMIRA® (Amira, 2002) will be used to simulate different situations common in hybrid estimation. A photo of the three-tanks system is presented in figure 2 showing the different components of the experimental setup. The plant consists of three plexiglas cylinders or tanks, T_1, T_2 and

T_3 with similar cross section. These are connected in series with each other by cylindrical pipes with cross section S_n. Located at T_2 is the single so called nominal outflow valve V_0 which also has a circular cross section S_n. The outflowing liquid (colored distilled water) is collected in a reservoir, which supplies the pumps P_1 and P_2. Here the water circuit is closed. h_{max} denotes the highest possible liquid level in any of the tanks. In case the liquid level of T_1 or T_2 exceeds this limit the corresponding pump will be switched off automatically. Q_1 and Q_2 are the flow rates from pumps P_1 and P_2, respectively.

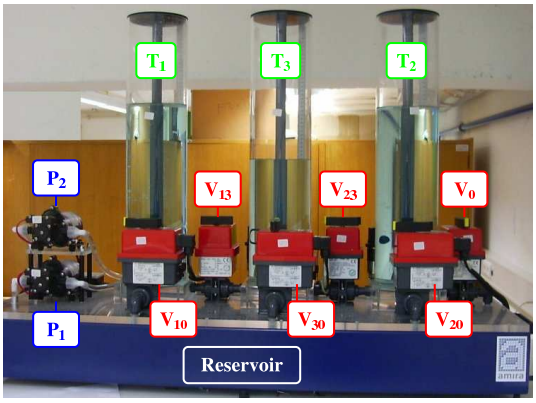

Figure 2: The three-tanks setup.

The pump flow rates Q_1 and Q_2 and the position of the valves V_{13}, V_{23}, V_0, V_{10}, V_{20}, V_{30}, denote the controllable variables, while the liquid levels of h_1, h_2 and h_3 are the output variables. The necessary level measurements are carried out by piezo-resistive difference pressure sensors. There are also potentiometric sensors that measure the position of each valve. The sensor signals are preprocessed to the interval $[0;1]$ and so need to be adjusted to $[0;h_{max}]$ for the water levels. For the remainder of this section the three-tanks system will be adapted so that more realistic hybrid estimation problems can be studied while simultaneously simplifying the presentation of results. The new model is present in figure 3 where the elements in grey are assumed to be nonexistent, the elements in green are fully operational and the elements in red may be subject to faults and will be used to model input disturbances.

Pump P_1 is considered to be a fully operational on/off valve. Valve V_{13} will have two nominal values "on" and "off", while Valve V_{10} will remain closed. Both these valves are subject to a possible fault resulting in an unmeasurable flow to cross them and described as an input disturbance. The water level sensor of tank 3 can also be subject to a fault. The Valve V_{30} is considered to be a fully operational "on/off" valve with no possible faults, while Valves V_{20}, V_{23} and V_0 will remain closed and so can be considered to be nonexistent.

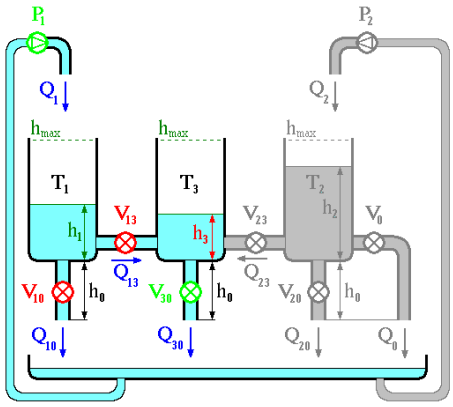

Figure 3: Final model of the three-tanks system.

The system can exhibit a large number of different dynamics, depending on the state of each discrete variable. The full hybrid model description of the system can be found in (Pina, 2007).

4.1 Estimation of the Fault in Valve $\mathbf{V_{10}}$

In this example, the estimation algorithm will have to estimate the discrete mode that indicates a fault on valve $\mathbf{V_{10}}$. As the analysis will focus on valve $\mathbf{V_{10}}$, the faults on valve $\mathbf{V_{13}}$ and sensor $\mathbf{h_3}$ will be considered nonexistent. A single test will be performed where various situations arise and are then analyzed separately. The system is excited according to the discrete variables presented in table 1. Various positions for the valve $\mathbf{V_{10}}$ are considered, corresponding to different intensities of the fault.

Table 1: Evolution of the discrete variables.

Time(s)	0-49	50-99	100-149	150-199	200-249	250-300
$\mathbf{V_{10}}$	"ok"	"faulty" *med*	"faulty" *max*	"faulty" *med*	"faulty" *max*	"ok"
$\mathbf{V_{13}}$	"ok"	"ok"	"ok"	"ok"	"ok"	"ok"
$\mathbf{h_3}$	"ok"	"ok"	"ok"	"ok"	"ok"	"ok"
$\mathbf{P_1}$	"on"	"on"	"on"	"on"	"on"	"on"
$\mathbf{V_{13}}$	"open"	"open"	"open"	"open"	"open"	"open"
$\mathbf{V_{30}}$	"open"	"open"	"open"	"open"	"open"	"open"

The measured outputs and the estimated water levels are presented in figure 4, where the influence of the intensity of the fault can be clearly seen.

The real (observed) and estimated values of the fault using the IMM algorithm are shown in figure 5. As the fault in valve $\mathbf{V_{10}}$ takes one time instant to be reflected in the water level measurements, only the value of $f_{V_{10}}(k-1|k)$ is relevant. Note that $f_{V_{10}}(k-1|k)$ is a discrete variable that takes value 1 when a leak occurs, and value 0 when there is no fault.

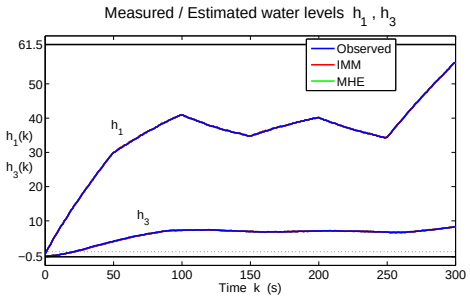

Figure 4: Water levels estimation using the IMM and MHE estimation algorithms.

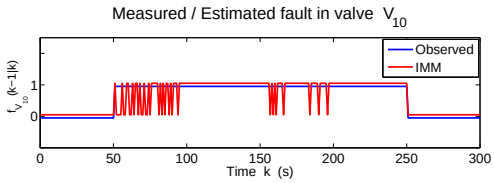

Figure 5: Estimation of the discrete mode sequence relative to the fault in valve $\mathbf{V_{10}}$.

The corresponding estimated continuous input disturbances by both algorithms are shown in figure 6. As the fault in valve $\mathbf{V_{10}}$ takes one time instant to be reflected in the water level measurements, only the value of $w_{V_{10}}(k-1|k)$ is estimated. The variable $w_{V_{10}}$ determines the leaking flow and is considered to be a uniformly distributed random variable defined in the interval $[-0.4;0.4]$ cm, with zero mean and variance $\frac{0.8^2}{12}$ cm^2 for all k, where 0.8 is the maximum water level change when the valve $\mathbf{V_{10}}$ is fully open.

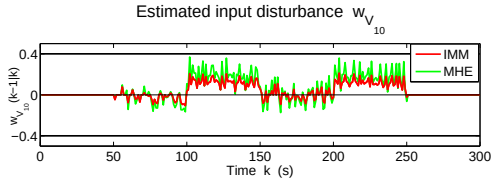

Figure 6: Estimation of the input disturbance $w_{V_{10}}(k-1|k)$ corresponding to the fault in valve $\mathbf{V_{10}}$.

The difference observed in both algorithms for the estimation of the disturbance $w_{V_{10}}(k-1|k)$ shows that the MHE algorithm is not able to weight the disturbance with any prior value so allowing it to change freely, which increases the variation of the input disturbance estimates.

The estimation results presented in figures 4 and 5 will now be analyzed independently for the 3 considered valve $\mathbf{V_{10}}$ fault intensities.

4.1.1 Case 1 - Fault Inactive

For time intervals $[0;50[$ s and $[250;300]$ s valve $\mathbf{V_{10}}$ remained closed and the fault is considered inactive. Despite being inactive, there is still a possibility of a wrong estimate reflected on the value of the discrete variable $f_{V_{10}}$. However, as shown in figure 7, the valve's true state was correctly estimated during these time periods.

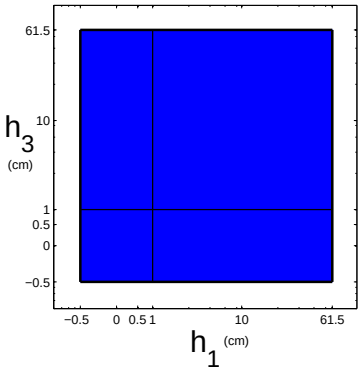

Figure 7: Map of probability of correct mode estimation, with 1s delay, when valve $\mathbf{V_{10}}$ is fully closed. (Red - probability of correct mode estimation 0, Blue - probability of correct mode estimation 1)

Figure 7 shows that if the valve $\mathbf{V_{10}}$ is closed there is no possibility of estimating a discrete mode sequence corresponding to an open valve condition. Thus the inactive fault is always correctly estimated.

4.1.2 Case 2 - Fault Active with Intermediate Intensity

The valve $\mathbf{V_{10}}$ has an intermediate open position during time intervals $[50;100[$ s and $[150;200[$ s allowing an unmeasured flow to cross it. In this case, a fully closed valve was estimated by the IMM algorithm in several time instants. These wrong estimates are understandable since the effect on the water level of tank 1 is not too drastic and can be mistaken by any other source of uncertainty, like measurement noise for instance. This difficulty in discerning whether the valve is slightly open or fully closed is patent in the map of probability of correct mode estimation shown in figure 8. It can also be concluded that the probability of an incorrect estimation of the valve's condition increases as the water level of tank 3 becomes lower.

The map of probability of correct mode estimation is not able to show the existing dependence between the probability of correctly determining the valve's condition and its real position. It is clear from figure

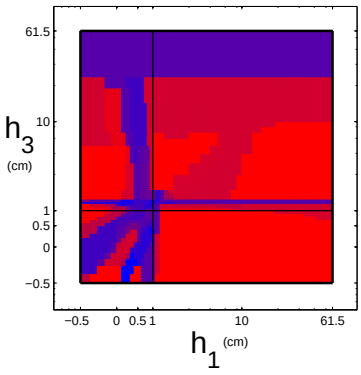

Figure 8: Map of probability of correct mode estimation, with 1s delay, when fault $f_{V_{10}}$ is active. (Red - probability of correct mode estimation 0, Blue - probability of correct mode estimation 1)

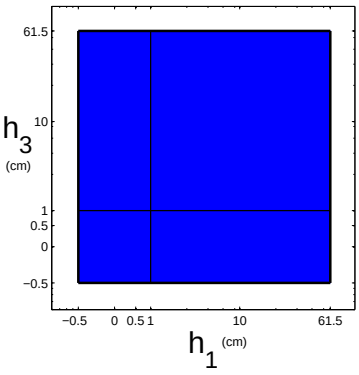

Figure 9: Map of probability of correct mode estimation, with 1s delay, considering that fault $f_{V_{10}}$ has maximum intensity, $w_{V_{10}} = 0.4$. (Red - probability of correct mode estimation 0, Blue - probability of correct mode estimation 1)

4 that the valve $\mathbf{V_{10}}$ is more closed during the time interval $[50\,;100[$s than in $[150\,;200[$s. This fact is reflected in a higher number of incorrect mode sequence estimations in case the valve remains closer to its nominal closed position. The following case will further explore this dependence.

4.1.3 Case 3 - Fault Active with Maximum Intensity

If valve $\mathbf{V_{10}}$ is fully open it becomes much easier to determine its position, thus allowing the IMM algorithm to provide correct estimates for the discrete mode sequence during time intervals $[100\,;150[$s and $[200\,;250[$s. This is quite obvious since the effect on the water level of tank 1 is very intense and can not be mistaken by any other source of uncertainty. This result is depicted in figure 9.

This map of probability of correct mode estimation was computed considering an hypothetical model for the system where valve $\mathbf{V_{10}}$ can only be fully open or fully closed.

Figure 9 shows that when the fault $f_{V_{10}}$ has maximum intensity, $w_{V_{10}} = 0.4$, it is always correctly estimated. However, further results have shown that for very low water levels in tank 1 the difference between a fully open or fully closed valve are reduced, being even undetectable when the tank is empty. This is explained by the fact that the maximum fault intensity allowed by the model, $w_{V_{10}} = 0.4$, can not be achieved in practice when tank 1 is almost empty but rather when it is full.

5 CONCLUSIONS

This paper presented an efficient hybrid estimation algorithm based on an IMM setup composed by a set of least-squares filters. The computational efficiency is obtained by some algorithmic procedures that discard many candidate *dms* before performing heavy computations. These procedures rely on the early determination of good estimates, on the separation of constrained and unconstrained estimates and on some bounding parameters for the squared errors.

The IMM was able to provide accurate online estimates for both continuous states and discrete variables when applied to the hybrid model of the benchmark AMIRA DTS200 three-tanks system experimental setup. The potential of the IMM algorithm was demonstrated when comparing its computational efficiency with the MHE with unknown inputs algorithm for a fault detection problem.

One of the most relevant issues that influence the computational efficiency of hybrid methodologies has to do with the high number of discrete modes that are tipically involved in a medium size hybrid system model. This fact eventually turns most of the problems untractable. For the case of the three-tanks system experimental setup, it was noticed that the consideration of all three tanks in the same hybrid model requires huge computational resources. Thus, authors believe that a multi-agent modeling architecture can significantly simplify the all model complexity while being able to retain its full hybrid dynamical flavour. As the size of the problems to be solved with hybrid systems grows exponentially with the number of discrete modes involved, multi-agent architectures may be the solution to the huge complexity of hybrid

methodologies, thus being a very interesting and possibly fruitful research topic.

REFERENCES

Alessandri, A. and Coletta, P. (2003). Design of observers for switched discrete-time linear systems. In *Proc. American Control Conference*, pages 2785–2790, Denver, Colorado.

Alizadeh, F. and Goldfarb, D. (2001). Second-order cone programming. Technical Report RRR Report number 51-2001, RUTCOR, Rutgers University, Piscataway, New Jersey.

Amira (2002). *DTS200 - Laboratory Setup Three-tank-system*. Amira, Duisburg, Germany.

Antsaklis, P. (2000). A brief introduction to the theory and applications of hybrid systems. *Proc. IEEE, Special Issue on Hybrid Systems: Theory and Applications*, 88(7):879–886.

Athans, M. and Chang, C. (1976). Adaptive estimation and parameter identification using multiple model estimation algorithm. Technical Report 28, M.I.T. - Lincon Laboratory, Lexington, Massachusetts.

Balluchi, A., Benvenuti, L., Benedetto, M. D., and Sangiovanni-Vincentelli, A. (2002). Design of observers for hybrid systems. In *Hybrid Systems: Computation and Control*, volume 2289 of *Lecture Notes in Computer Science*, pages 76–89. Springer Verlag.

Bemporad, A. and Morari, M. (1999). Control of systems integrating logic, dynamics, and constraints. *Automatica*, 35(3):407–427.

Blom, H. A. P. and Bar-Shalom, Y. (1988). The interactive multiple model algorithm for systems with markovian switching coefficients. *IEEE Trans. on Automatic Control*, 33(8):780–783.

Böker, G. and Lunze, J. (2002). Stability and performance of switching Kalman filters. *International Journal of Control*, 75(16/17):1269–1281.

Ferrari-Trecate, G., Mignone, D., and Morari, M. (2002). Moving horizon estimation for hybrid systems. *IEEE Trans. on Automatic Control*, 47(10):1663–1676.

Fletcher, R. (1987). *Practical methods of optimization*. A Wiley Interscience Publication, Chichester, New York, 2^{nd} edition.

Heemels, W., Schutter, B. D., and Bemporad, A. (2001). Equivalence of hybrid dynamical models. *Automatica*, 37(7):1085–1091.

Kamen, E. (1992). Study of linear time-varying discrete-time systems in terms of time-compressed models. In *Proc. 31th IEEE Conf. on Decision and Control*, pages 3070–3075, Tucson, Arizona.

Mazor, E., Averbuch, A., Bar-Shalom, Y., and Dayan, J. (1998). Interacting multiple model methods in target tracking: A survey. *IEEE Trans. on Aerospace and Electronic Systems*, 34(1):103–123.

Pina, L. (2007). *Hybrid state estimation*. PhD thesis, Instituto Superior Técnico, Universidade Técnica de Lisboa, Portugal.

Pina, L. and Botto, M. A. (2006). Simultaneous state and input estimation of hybrid systems with unknown inputs. *Automatica*, 42(5):755–762.

Sontag, E. (1981). Nonlinear regulation: The piecewise linear approach. *IEEE Trans. on Automatic Control*, 26(2):346–358.

Torrisi, F. and Bemporad, A. (2001). Discrete-time hybrid modeling and verification. In *Proc. 40th IEEE Conf. on Decision and Control*, pages 2899–2904, Orlando, Florida.

BRIEF BIOGRAPHY

Miguel Ayala Botto received the master degree in Mechanical Engineering in 1992 and the Ph.D. in Mechanical Engineering in 1996 from Instituto Superior Tcnico, Tecnhical University of Lisbon, Portugal. He spent the year of 1995 at the Control Laboratory, Department of Electrical Engineering, Delft University of Technology, Holland. Further, in the winter semester of the academic year 1999/2000 he held a postdoctoral position at the same laboratory. Since 2001 he is Associate Professor at the Department of Mechanical Engineering, Instituto Superior Tcnico, Portugal. He is currently coordinator of the research group on Systems and Control from the Center of Intelligent Systems of IDMEC - Institute of Mechanical Engineering. Since 2005 he is the head of the Portuguese Association on Automatic Control, the National Member Organization from IFAC. He has published more than 70 journal papers, book chapters, and communications in international conferences. He has been awarded in 1999 with ”The Heaviside Premium”, attributed by the Council IEE - The Institution of Electrical Engineers, UK. Currently he is Associate Editor of the International Journal of Systems Science (Taylor & Francis) and member of the IFAC Technical Committee on Discrete Event and Hybrid Systems. His main research interest is in the field of estimation and control of hybrid dynamical systems.