Proceedings of the 17th World Congress
The International Federation of Automatic Control
Seoul, Korea, July 6-11, 2008

An Efficient Hybrid Estimator ⋆

Luís Pina * **Miguel Ayala Botto** *

* *Department of Mechanical Engineering, IDMEC, Instituto Superior Técnico, Technical University of Lisbon, Portugal (e-mail: luispina@dem.ist.utl.pt, ayalabotto@ist.utl.pt).*

Abstract: This paper presents an efficient state estimation algorithm for hybrid systems based on a least-squares Interacting Multiple-Model setup. Under some conditions, the proposed algorithm is shown to perform similarly and even better than the Kalman filter. However, due to the possibility of incorrect estimation of the discrete mode, performance deterioration may occur. The computational efficiency of the proposed algorithm is obtained by discarding as many discrete mode sequences as possible while performing the least computations possible. This is done by rapidly computing good estimates, separating the constrained and unconstrained estimates and using some auxiliary coefficients computed off-line. A numerical example shows the characteristics of the proposed algorithm and compares it with the Kalman filter.

Keywords: Hybrid systems, Hybrid Estimation, Interacting multiple-model estimation, Least-squares, Constrained optimal estimation

1. INTRODUCTION

The present work focuses on the state estimation problem for hybrid systems. In recent years the industry and research community have shown an increasing interest on hybrid systems due to their capability of describing the interaction between dynamical and logical components Antsaklis [2000]. This interaction can be found in many real world systems, embedded control systems and in the control of many complex industrial systems via the combination of classical continuous control laws with supervisory switching logic.

The class of hybrid systems considered in this paper is Piecewise Affine (PWA) systems. These are basically composed by a set of affine dynamics and a discrete mode that defines the active dynamics. In Heemels et al. [2001] PWA systems are proven to be equivalent, under some mild assumptions, to many other classes of hybrid systems, and so, the proposed techniques can be interchanged among all the referred classes.

The estimation problem for hybrid systems has already been tackled by several authors. For instance Bemporad et al. [1999] and Pina and Botto [2006] proposed different Moving Horizon Estimation (MHE) schemes that rely on the application of brute force optimization algorithms to estimate both the discrete mode and the continuous state of the system. Some other approaches consider that the discrete mode is known in advance, which greatly simplifies the problem. 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.

Most of the synthesis and analysis problems involving hybrid systems are in general NP-complete, as shown in Torrisi and Bemporad [2001], and the estimation problem is no exception. Researchers have then focused on deriving algorithms that despite having a NP-complete complexity are relatively efficient in practical applications. The present work is an example of such research direction.

The derivation of the truly optimal filter for systems with switching parameters was first presented in Athans and Chang [1976]. However, the requirement for exponentially growing computational resources prohibits its practical application. 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. The proposed algorithm uses least-squares filters since they require very mild assumptions to be applied and can easily coupe with constraints. Moreover, in Sayed and Kailath [1994], where a very detailed description of least-squares filters is presented, their equivalence to truly optimal filters such as the Kalman filter was proven for some special setups.

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. In Section 4 a demonstrative numerical example is presented and 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:

⋆ This work was supported in part by projects POCI/EME/59522/2004 and PTDC/EME-CRO/69117/2006 co-sponsored by FEDER, Programa Operacional Ciência e Inovação 2010, and by the grant SFRH/BD/12208/2003, from FCT, Portugal.

$$x(k+1) = A_{i(k)}x(k)+B_{i(k)}u(k)+f_{i(k)}+L_{i(k)}w(k) \quad (1a)$$
$$y(k) = C_{i(k)}x(k)+D_{i(k)}u(k)+g_{i(k)}+v(k) \quad (1b)$$
$$\text{iff } \begin{bmatrix} x(k) \\ u(k) \\ w(k) \end{bmatrix} \in \Omega_{i(k)} \quad (1c)$$

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, $i(k) \in \mathcal{I} = \{1, \ldots, s\}$ 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:

$$H_{\mathbb{W}_{i(k)}}\, w(k) \leq h_{\mathbb{W}_{i(k)}} \quad , \quad \forall_{k\in\mathbb{N}_0} \quad (2)$$
$$H_{\mathbb{V}_{i(k)}}\, v(k) \leq h_{\mathbb{V}_{i(k)}} \quad , \quad \forall_{k\in\mathbb{N}_0} \quad (3)$$

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:

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

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:

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

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 $\Omega_{\mathbf{i}_T}$:

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

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):

$$\mathbf{H}_{\mathbb{W}_{\mathbf{i}_T}}\, W_T(k) \leq \mathbf{h}_{\mathbb{W}_{\mathbf{i}_T}} \quad (7)$$
$$\mathbf{H}_{\mathbb{V}_{\mathbf{i}_T}}\, V_T(k) \leq \mathbf{h}_{\mathbb{V}_{\mathbf{i}_T}} \quad (8)$$

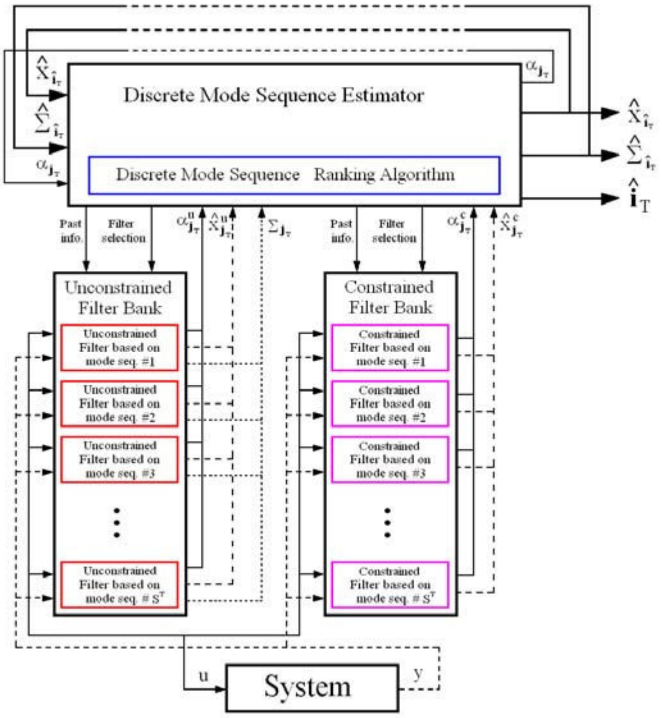


Fig. 1. Interacting Multiple-Model Estimation Algorithm.

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* $\mathbf{j}_T$:

$$\hat{x}^u_{\mathbf{j}_T}(k|t) = \hat{x}_{\mathbf{j}_T}(k|t{-}1)+ \quad (9)$$
$$\mathbf{K}_{\mathbf{j}_T}(k|t{-}1)\big[\,(Y_T(k){-}\mathbf{D}_{\mathbf{j}_T}U_T(k){-}\mathbf{g}_{\mathbf{j}_T}){-}\mathbf{C}_{\mathbf{j}_T}\hat{x}_{\mathbf{j}_T}(k|t{-}1)\big]$$

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:

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

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

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

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

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 $\hat{x}_{\mathbf{j}_T}(k{+}1|t)$:

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

When computing the unconstrained state estimate, no *a priori* information may be available or one may be interested in discarding it, then $\mathbf{\Sigma}^{-1}_{x_{\mathbf{j}_T}}(k|t-1)$ 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:

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

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:

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

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

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:

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

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

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

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

The constrained least-squares filter gain is defined as:

$$\mathbf{K}_{\mathbf{j}_T}(k|t) = \quad (18)$$
$$\left(\begin{bmatrix} \mathbf{\Sigma}_{x_{\mathbf{j}_T}}(k|t-1) & 0 & 0 \\ 0 & \mathbf{\Sigma}_{W_{\mathbf{j}_T}} & 0 \\ 0 & 0 & \mathbf{\Sigma}_{V_{\mathbf{j}_T}} \end{bmatrix}^{-1} + \begin{bmatrix} \mathbf{H_e} \\ \mathbf{H_i} \end{bmatrix}^{\mathrm{T}} \mathbf{Z}_{\mathbf{j}_T}(k|t) \begin{bmatrix} \mathbf{H_e} \\ \mathbf{H_i} \end{bmatrix}\right)^{-1} \begin{bmatrix} \mathbf{H_e} \\ \mathbf{H_i} \end{bmatrix}^{\mathrm{T}} \mathbf{Z}_{\mathbf{j}_T}(k|t)$$

where $\mathbf{\Sigma}_{x_{\mathbf{j}_T}}(k|t-1)$ 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 $\mathbf{\Sigma}^{-1}_{x_{\mathbf{j}_T}}(k|t{-}1)$ 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, $\alpha^c_{\mathbf{j}_T}$:

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

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

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

where:

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

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

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:

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

Equation (20) should be modified to:

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

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 $\mathbf{\Sigma}^{-1}_{x_{\mathbf{j}_T}}(k|t-1)$ 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:

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

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 $\mathbf{j}_T = \big\{j(k+1), \dots, j(t), j(t+1)\big\}$ 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. NUMERICAL EXAMPLE

Consider the following stochastic PWA system:

$$x(k+1)=\begin{cases}\begin{bmatrix}0.6 & 0.9\\ 0 & 0.8\end{bmatrix}x(k)+\begin{bmatrix}1\\1\end{bmatrix}w(k)+\begin{bmatrix}-3\\1\end{bmatrix} \\ \qquad\qquad \text{iff } [\,1\ 0\,]\,x(k)\le 0 \qquad \text{(mode 1)}\\ \begin{bmatrix}0.7 & -0.9\\ 0 & 0.8\end{bmatrix}x(k)+\begin{bmatrix}1\\1\end{bmatrix}w(k)+\begin{bmatrix}-3\\-1\end{bmatrix} \\ \qquad\qquad \text{iff } [\,1\ 0\,]\,x(k)> 0 \qquad \text{(mode 2)}\end{cases}$$

$$y(k) = [\,1\ 0\,]\,x(k) + v(k)$$

$$x(k) \in \mathbb{X} \triangleq \big[-10\ ,\ 10\big] \times \big[-10\ ,\ 10\big]$$

$$w(k) \in \mathbb{W} \triangleq \big[-2\ ,\ 2\big] \quad , \quad \mathrm{E}\{w\} = 0 \quad , \quad \Sigma_w = 0.3$$

$$v(k) \in \mathbb{V} \triangleq \big[-2\ ,\ 2\big] \quad , \quad \mathrm{E}\{\,v\,\} = 0 \quad , \quad \Sigma_v = 0.3$$

The input disturbances and measurement noises are assumed to be Gaussian with expected value 0 and variance 0.3. The disturbance sets are not restrictive since they contain over 99.7% of the possible values.

The developed IMM estimation algorithm will now be applied considering a window of length $T = 2$. The admissible discrete mode sequences are $\big\{[1\ 1], [1\ 2], [2\ 1], [2\ 2]\big\}$.

The $\mathcal{B}$ matrix, whose coefficients were introduced in equation (25), is given by:

$$\mathcal{B} = \begin{bmatrix} 0 & 0 & 0 & 4.395\\ 0 & 0 & 17.578 & 0\\ 0 & 4.395 & 0 & 0\\ 17.578 & 0 & 0 & 0\end{bmatrix} \tag{28}$$

The following examples will show the main characteristics of the proposed IMM estimation algorithm.

4.1 Example 1

This example will show that the proposed algorithm behaves similarly to the Kalman filter. However, it can enforce the constraints of the system, providing consistent estimates at every time instants and possibly improving the estimation. The system will be simulated from the initial position $x(1) = [\,6\,,6\,]^{\mathrm{T}}$ and, the estimators are initialized with initial guess $\bar{x}_0 = [\,4\,,4\,]^{\mathrm{T}}$ and corresponding covariance matrix $\Sigma_{x_0} = I_{n_x}$. The initial time instant ($k = 0$) is only used to represent the initial conditions. Figure 2 shows the evolution of both real and estimated continuous states and *dms*.

As can be seen from figures 2(a) and 2(c), the estimate of the continuous state x_1 from the IMM algorithm is always in the same region as the real one, while the estimate from the Kalman filter is sometimes in the wrong region, despite having *a priori* knowledge of the discrete mode sequence.

The Kalman filter is clearly unable to enforce consistency in the estimates, for instance at time instants $k = 6$ and $k = 7$. Nevertheless, no significant deterioration of the continuous estimates occurs, as shown by the mean squared errors of table 1. The IMM algorithm, on the other hand, is able to enforce the constraints, slightly reducing the estimation error at those time instants.

4.2 Example 2

This last example will provide some insight on the behavior of the IMM estimation algorithm when incorrect *dms* are estimated, and explore the benefits of considering the information of past estimates in the estimation of the *dms*.

The system will be simulated from $x(1) = [-1\,,-4]^{\mathrm{T}}$ and the estimators are initialized with initial guess $\bar{x}_0 = [-2\,,-5]^{\mathrm{T}}$ and corresponding covariance matrix $\Sigma_{x_0} = I_{n_x}$. Figure 3 shows the evolution of real and estimated continuous states and *dms*. Figure 3(c) also shows the *dms* estimates when past information is not discarded, that is, the matrix $\mathbf{\Sigma}^{-1}_{x_{\mathbf{j}_T}}(k|t-1)$ is not set to zero.

As can be seen from figures 3(a) and 3(c), the IMM estimation algorithm can estimate incorrect *dms*, which becomes very likely to occur when the continuous state lies near the boundaries separating the two discrete modes. In this example, the incorrect estimation of last mode of the *dms* enforces the constraints of the incorrect *dms* at that time instant, while incorrect estimation of the first mode of the *dms* induces large errors in the continuous state estimates since these are estimated considering incorrect dynamics. This can be clearly seen by the deviation of the estimate of x_2 at time instants $k = 31$ and $k = 32$, in figure 3(b).

	MSE (example 1)	MSE (example 2)
Kalman filter	0.3806	0.4002
IMM algorithm	0.3778	5.1734

Table 1. Mean squared estimation errors (examples 1 and 2).

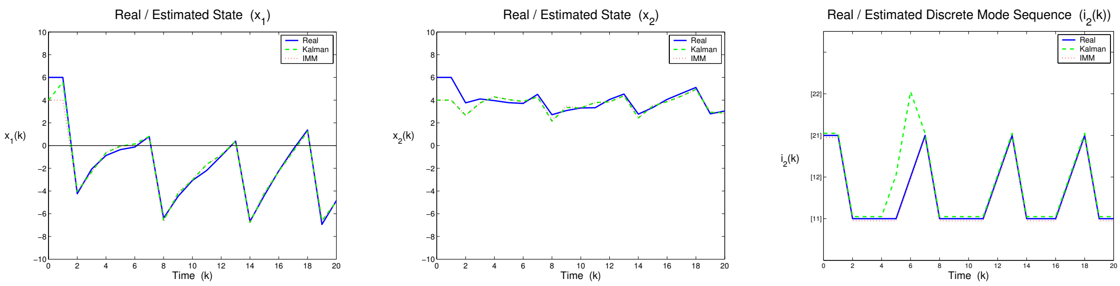
It would be expected that the consideration of past information could lead to significant improvements on the *dms* estimation. This example, however, shows that this may not be the case since the consideration of past information led to an increase of the number of time instants where the discrete mode sequence is incorrectly estimated. This problem has not been thoroughly addressed and the cause of such result has not been investigated. For this specific example, a possible explanation may be related to the behavior of the state x_1, which varies smoothly in the discrete mode 1 and jumps abruptly when the boundary to the discrete mode 2 is crossed.

	(example 1)	(example 2)
Unconstrained	3.12	3.16
Constrained	2.42	2.76

Table 2. Mean number of computed estimates (examples 1 and 2).

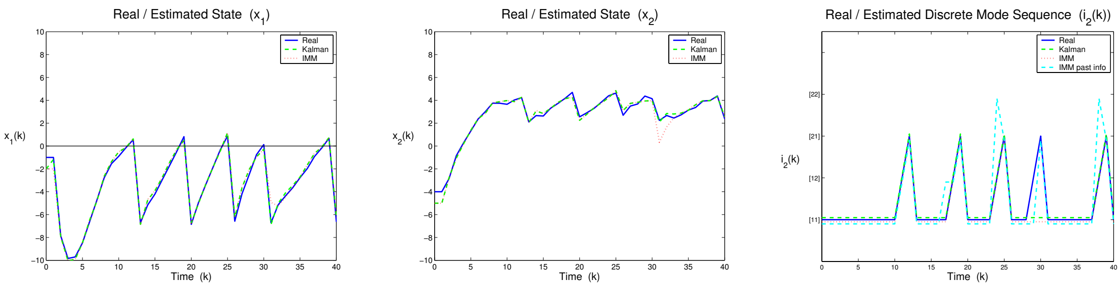
5. CONCLUSIONS AND FUTURE WORK

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.



(a) x_1 (b) x_2 (c) $\hat{\mathbf{i}}_2$

Fig. 2. Hybrid estimation using the Kalman filter and the IMM estimation algorithm (example 1).



(a) x_1 (b) x_2 (c) $\hat{\mathbf{i}}_2$

Fig. 3. Hybrid estimation using the Kalman filter and the IMM estimation algorithm (example 2).

Despite being very simple, the numerical example showed that some heavy computations are avoided. This reduction becomes more evident for higher complexity systems.

The comparison of the proposed algorithm with the Kalman filter showed that when the *dms* is correctly estimated, both performances are very similar. Moreover, the possibility of enforcing constraints not only enhances the performance of the proposed IMM estimation algorithm in the vicinity of constraints but the computed continuous state and *dms* estimates are always coherent.

In hybrid estimation there is always the possibility of computing incorrect *dms* estimates which deteriorates the continuous state estimates. It would be expected that the consideration of past information in the estimation of the *dms* could lead to better results, however, the numerical example showed that this may not always be the case.

Future work should focus on the consideration of past information in the *dms* estimation and, on determining quantitative measures for the associated uncertainty.

REFERENCES

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

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

P.J. Antsaklis. 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, July 2000.

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

A. Bemporad, D. Mignone, and M. Morari. Moving horizon estimation for hybrid systems and fault detection. In *Proc. American Control Conference*, pages 2471–2475, San Diego, California, June 1999.

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

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

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

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

E.W. Kamen. 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, 1992.

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

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

A. Sayed and T. Kailath. A state-space approach to adaptive RLS filtering. *IEEE Signal Processing Magazine*, (July 1994):18–60, July 1994.

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

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