Bayesian Estimation of Transition Probabilities for Markovian Jump Systems

Vesselin P. Jilkov, *Member, IEEE,* and X. Rong Li*, *Senior Member, IEEE*

EDICS: 2-ESTM, 2-ADPT

This work was supported in part by NNSF and the National Key Project of China, and by ONR grant N00014-00-1-0677, NSF grant ECS-9734285, and NASA/LEQSF grant (2001-4)-01.
Authors' Address: Department of Electrical Engineering, University of New Orleans, New Orleans, LA 70148, USA (e-mail: vjilkov@uno.edu, xli@uno.edu); * Corresponding Author

Abstract

Markovian jump systems (MJS) evolve in a jump-wise manner by switching among simpler models, according to a finite Markov chain, whose parameters are commonly assumed known. This paper addresses the problem of state estimation of MJS with *unknown* transition probability matrix (TPM) of the embedded Markov chain governing the switching. Under the assumption of constant but random TPM, a recursion for the TPM's posterior probability density function (PDF) within the Bayesian framework is obtained. Based on this recursion four recursive algorithms for minimum mean-square error (MMSE) estimation of the TPM are derived. The first algorithm computes the optimal estimate exactly for the case of a two-state Markov chain. Its computational load is linear in the data length. The other three algorithms, based on PDF moments' truncation, quasi-Bayesian approximation and numerical integration, respectively, compute the optimal estimate approximately and have constant computational cost as data size increases.

The proposed TPM estimation is naturally incorporable into a typical Bayesian estimation scheme for MJS (e.g. Generalized Pseudo-Bayesian (GPB), or Interacting Multiple Model (IMM)). Thus adaptive versions of MJS state estimators with unknown TPM are provided. Simulation results of TMP-adaptive IMM algorithms for a system with failures and maneuvering target tracking are presented.

Index Terms

Adaptive Estimation, Markovian Jump System, Multiple Model, IMM

I. Introduction

Theoretically in the estimation for MJS (or more generally hybrid system [1] or *multiple model* (MM) estimation [2]) the TPM is almost always assumed known. In practice it is a "design parameter" whose choice is done *a priori*. Techniques for design of TPM in target tracking application have been proposed and studied [3], [4], [5], [6], [7], [8]. They are a priori in nature, tailored to the specific application of interest, and do not rely on the on-line received data to *estimate* the TPM. For many applications, however, a priori information about the TPM may be inadequate or even lacking. Using an inaccurate apriori TPM may lead to performance degradation of the state estimation, due to the sensitivity of the MM state estimators to the TPM used. The uncertainty regarding the TPM is a major issues in the application of MM to real-life problems, e.g. maneuvering target tracking, systems' fault detection and isolation, and many others. The "unreasonable" need to provide TPM a priori even in the case of insufficient information has been cited by some as one of the major reservations of using Markov-chain based MM estimation algorithms. Thus, it is highly desirable to have algorithms which can identify (estimate) the TPM recursively during the course of processing measurement data so as to allow adaptation of the MM state estimation. There exist only a few publications in the engineering literature devoted to solving this important and difficult problem.

In [9] the problem was considered for the simplest case of a binary Markov chain modelling measurement failures (interrupted observations). Based on the Bayesian approach, proposed therein is an approximate TPM estimator which requires numerical integration to compute the estimate. It is a particular case of one of the estimators (viz. Algorithm 4) proposed in this paper. For the same restricted case, [10] and [11] developed and analyzed the convergence of a maximum likelihood (ML) estimator of TPM, parameterized over a finite set. Since the

ML algorithm requires exponentially increasing memory and computation, an approximate scheme, referred to as truncated ML estimation, was further proposed in [12] for the general case. It in essence employs the finite Gaussian sum approximation technique to truncate the ever branching likelihood. The unknown TPM is assumed to belong to a compact (practically finite) set of candidate TPMs. Its estimate is obtained at the cost of running a bank of Kalman filters for each candidate TPM to compute the respective truncated likelihood functional. The method was shown theoretically to be convergent, but it is computationally quite involved — with complexity $O(m^{s})$, where m is the number of chain states and s is the truncated data length. The choice of candidate TMPs in some practical applications may also be problematic. Parameter estimation of the Markov chain was also considered in [13] and [14].

Our work is conceptually based on the assumption that the unknown TPM is a *random constant* matrix with a given prior distribution defined (truncated) over the *continuous* set (simplex) of valid TPMs. Within this Bayesian framework we obtain a recursion for the TPM's posterior PDFs in terms of the MM estimator's model probabilities and likelihoods and seek for computationally feasible recursive algorithms to compute the TPM's MMSE-estimate, based on this recursion. As a result we propose four alternative solutions. For the case of a two-state Markov chain we find an algorithm for computing the optimal estimate. It has an increasing computational complexity $O(k)$, where k is the time index (data length). That is why for the general case of a Markov chain with two or more states, three approximate algorithms are further developed by means of alternative approximation techniques — truncation of the higher order moments of the PDF, quasi-Bayesian approximation for estimation of finite mixtures, and numerical integration, respectively. All TPM estimators use the information contained in the posterior model probabilities and likelihoods, supplied by an MM state estimator and are easily coupled with such an estimator for a joint TPM and state estimation. Thus a *generic* scheme for Bayesian TPM-adaptive MM state estimation is also outlined.

The remaining part of the paper is organized as follows. Section 2 describes the problem formulation and the general adaptation scheme. In Section 3 a basic recursion for the posterior PDF of TPM is developed. Subsection 4.A gives the derivation of an exact (MMSE-optimal) TPM estimator for the case of a two-state Markov chain and addresses some implementation issues. In Subsection 4.B a second-order approximated TPM estimator for the general case is obtained. Subsections 4.C and 4.D present respectively a quasi-Bayesian TPM estimator and a TPM estimator algorithm using numerical updating/integration of PDFs. In Section 5, simulations conducted to evaluate and compare the performances of non-adaptive and alternative TPM-adaptive IMM state estimators are presented. Two typical application examples are considered — estimation for systems subject to failures and maneuvering target tracking. Summary and conclusions are given in Section 6.

II. Problem Formulation

Consider the following model of Markovian jump system in discrete time $k = 1, 2, \ldots$

$$x(k) = f[k, m(k), x(k-1)] + g[k, m(k), w[k, m(k)]] \quad (1)$$

$$z(k) = h[k, m(k), x(k)] + v[k, m(k)] \quad (2)$$

where x is the *base (continuous) state* vector with transitional function f, w is the random input vector, z is the measurement vector with measurement function $h,$ and v is the random measurement error vector. All vectors are assumed of appropriate dimensions. The *modal (discrete) state* $m(k) \in \mathbb{M} \triangleq \{1,\ 2,\ \ldots,\ m\}$ is a Markov chain with initial and transition probabilities respectively

$$P\{m_j(0)\} = \mu_j(0) \quad (3)$$

$$P\{m_j(k)|m_i(k-1)\} = P\{m_j(k)|m_i(k-1), z^k\} = \pi_{ij},\ i, j = 1, \ldots m \quad (4)$$

where $m_i(k)$ stands for the event $m(k) = i$ and $z^k = \{z(1),\ \ldots,\ z(k)\}$ is the cumulative measurement set.

Provided all parameters of the model (1)–(4) are *known* the MMSE-optimal estimate $\hat{x}(k) = E\left[x(k)|z^k\right]$ of the base state x can be obtained by the Bayesian *full-hypothesis-tree* (FHT) estimator [15]. The FHT is however infeasible because of its exponentially growing computation and memory, and thus suboptimal approximations with limited complexity are of interest in practice. There exist a number of such approximations, commonly referred to as *multiple model* (MM) algorithms, which can be broadly grouped as *soft, hard*, or *random decision-based*, regarding the approach used to approximate the optimal FHT estimator. (See [2] and the long list of references therein.) Recent progress in MM estimation includes [16], [17], [18], [19], [20], [21], [22], [23], [24]. For the class of Bayesian MM estimation considered here a good trade-off between performance and computational cost is provided by the popular *Interacting MM* (IMM) estimator [25], [26]. To be more specific we will apply our further development to the case of IMM algorithm as an example.

Let us consider now the state estimation problem for the above hybrid system model (1)–(4) without the presumed knowledge of the transition probability matrix $\Pi = [\boldsymbol{\pi}_1, \boldsymbol{\pi}_2, \ldots, \boldsymbol{\pi}_m]'$ with $\boldsymbol{\pi}_i = [\pi_{i1}, \pi_{i2}, \ldots, \pi_{im}]'$, $i = 1, \ldots,\ m$, defined by (4). We assume that Π is an *unknown* but *time-invariant* matrix with some given prior distribution[1] defined over the simplex of valid TPMs. In such a Bayesian framework we consider the problem of finding recursively the posterior MMSE estimate $\overline{\Pi}(k) = E\left[\Pi|z^k\right]$ of Π for $k = 1, 2, \ldots,$ and formulate the following.

Bayesian TPM-Adaptive MM State Estimation. *On receipt of the new measurement at time k:*

- *Run the MM algorithm with the previous state/covariance estimates, model probabilities $\boldsymbol{\mu}(k-1)$ and $\overline{\Pi}(k-1)$ to update the current state/ covariance estimates, model probabilities*

$$\boldsymbol{\mu}(k) = [\mu_1(k), \ldots, \mu_m(k)]' \text{ with } \mu_i(k) \triangleq P\left\{ m_i(k) \,|\, \Pi, z^{k-1}\right\} \quad (5)$$

[1] If no prior information is available, uniform (non-informative) prior distribution may be used.

and model likelihoods

$$\mathbf{\Lambda}(k) = [\Lambda_1(k), \ldots, \Lambda_m(k)]' \text{ with } \Lambda_j(k) \triangleq p\left[z(k) \mid m_j(k), \Pi, z^{k-1}\right] \quad (6)$$

- *Update the TPM estimate* $\overline{\Pi}(k)$ *based on* $\overline{\Pi}(k-1)$ *and the new measurement information contained in the TPM likelihood* $p\left[z(k)|\Pi, z^{k-1}\right]$.

Note that this scheme is not restricted to the IMM algorithm only. Provided a feasible recursive TPM estimator, as specified above, is available the adaptation scheme is applicable to various Bayesian MM state estimation algorithms in a straightforward manner.

III. Posterior PDF of TPM

In this section we obtain recursive relationships for updating the TPM's posterior PDF in terms of the model probabilities and likelihoods, as defined by (5)–(6).

The TPM's likelihood is

$$\begin{aligned} p\left[z(k)|\Pi, z^{k-1}\right] &= \sum_{j=1}^{m} p\left[z(k)|m_j(k), \Pi, z^{k-1}\right] P\left\{m_j(k)|\Pi, z^{k-1}\right\} \\ &= \sum_{j=1}^{m} p\left[z(k)|m_j(k), \Pi, z^{k-1}\right] \sum_{i=1}^{m} P\{m_j(k)|m_i(k-1), \Pi, z^{k-1}\} P\left\{m_i(k-1)|\Pi, z^{k-1}\right\} \\ &= \sum_{j=1}^{m} \Lambda_j(k) \sum_{i=1}^{m} \pi_{ij}\, \mu_i(k-1) = \mathbf{\Lambda}'(k)\, \Pi' \boldsymbol{\mu}(k-1) = \boldsymbol{\mu}'(k-1)\, \Pi \mathbf{\Lambda}(k) \end{aligned} \quad (7)$$

Then $p\left[z(k)|z^{k-1}\right]$ becomes

$$\begin{aligned} p\left[z(k)|z^{k-1}\right] &= \int p\left[z(k)|\Pi, z^{k-1}\right] p\left[\Pi|z^{k-1}\right] d\Pi \\ &= \int \boldsymbol{\mu}'(k-1)\, \Pi \mathbf{\Lambda} p\left[\Pi|z^{k-1}\right] d\Pi = \boldsymbol{\mu}'(k-1)\, \overline{\Pi}(k-1)\, \mathbf{\Lambda}(k) \end{aligned} \quad (8)$$

In the above, $\boldsymbol{\mu}(k-1)$ and $\mathbf{\Lambda}(k)$ are computed by replacing the unknown Π with its best estimate $\overline{\Pi}(k-1)$ (conditioned on z^{k-1}) available at the time. Thus, by the Bayes rule

$$p\left(\Pi|z^k\right) = \frac{\boldsymbol{\mu}'(k-1)\, \Pi \mathbf{\Lambda}(k)}{\boldsymbol{\mu}'(k-1)\, \overline{\Pi}(k-1)\, \mathbf{\Lambda}(k)} p\left(\Pi|z^{k-1}\right) \quad (9)$$

An analogous recursion for the marginal PDF $p\left(\boldsymbol{\pi}_i|z^k\right)$ of each row $\boldsymbol{\pi}_i'$, $i = 1, \ldots, m$ of Π is given by the following.

Theorem 1: *If*

$$\int \boldsymbol{\pi}_l p\left(\boldsymbol{\pi}_1, \ldots, \boldsymbol{\pi}_m|z^{k-1}\right) d\boldsymbol{\pi}_1 \ldots d\boldsymbol{\pi}_{i-1} d\boldsymbol{\pi}_{i+1} \ldots d\boldsymbol{\pi}_m = \overline{\boldsymbol{\pi}}_l(k-1)\, p\left(\boldsymbol{\pi}_i|z^{k-1}\right) \quad (10)$$

for $l, i = 1, 2, \ldots m;\ l \neq i$ *then*

$$p\left(\boldsymbol{\pi}_i|z^k\right) = \{1 + \eta_i(k)\, [\boldsymbol{\pi}_i - \overline{\boldsymbol{\pi}}_i(k-1)]' \mathbf{\Lambda}(k)\}\, p\left(\boldsymbol{\pi}_i|z^{k-1}\right) \quad (11)$$

where

$$\eta_i(k) = \frac{\mu_i(k-1)}{\boldsymbol{\mu}'(k-1)\,\overline{\Pi}(k-1)\,\boldsymbol{\Lambda}(k)} \tag{12}$$

□

Proof: See Appendix.

Remark. It can be easily verified that the condition (10) is satisfied if $\boldsymbol{\pi}_i,\ i = 1, \ldots, m$ are assumed independent. However, it is in general much looser than the independence assumption.

Eq. (11) is the basis for various TPM estimators developed in this paper.

IV. MMSE Estimation of TPM

In view of (11), the estimation objective is to obtain a feasible algorithm for computing the conditional means $\overline{\boldsymbol{\pi}}_i(k) = E\left[\boldsymbol{\pi}_i|z^k\right],\ i = 1, \ldots m$, which constitute the optimal estimate of Π in the MMSE sense.

A. Optimal Algorithm

We consider first the case of a two-state ($m = 2$) Markov chain. Our algorithm is a corollary of the following general result.

Proposition 1: *Let θ be a random variable with linear conditional likelihood function, i.e. given a data sequence* $z(1), z(2), \ldots$

$$p\left[z(k)|\theta,\, z^{k-1}\right] \propto \beta(k) + \gamma(k)\theta, \quad k = 1,\, 2,\, \ldots \tag{13}$$

where $\beta(k) = \beta\left[z(k)\right]$ and $\gamma(k) = \gamma\left[z(k)\right]$ are non-random functions of the observations $z(k)$. Then the optimal MMSE estimator $\overline{\theta}(k) = E\left[\theta|z^k\right]$ can be computed via the following recursive algorithm (with $\overline{\theta}^{(n)}(l) \triangleq E[\theta^n|z^l]$):

- *Initialization:* $\overline{\theta}^{(k)}(0) = E[\theta^k],\ k = 1, 2, \ldots$
- *For* $k = 1,\, 2,\, \ldots$

  For $l = 1,\, 2,\, \ldots,\, k$

$$\overline{\theta}^{(k-l+1)}(l) = \beta(l)\overline{\theta}^{(k-l+1)}(l-1) + \gamma(l)\overline{\theta}^{(k-l+2)}(l-1) \tag{14}$$

□

Proof: By the Bayes rule and (13)

$$p\left[\theta|z^l\right] = \left[\beta(l) + \gamma(l)\theta\right] p\left[\theta|z^{l-1}\right],\ l = 1,\, 2,\, \ldots$$

Then for each $n = 1, 2, \ldots$

$$\begin{aligned}\overline{\theta}^{(n)}(l) &= E\left[\theta^n|z^l\right] = \int \theta^n p\left[\theta|z^l\right] d\theta \\ &= \beta(l)\int \theta^n p\left[\theta|z^{l-1}\right] d\theta + \gamma(l)\int \theta^{n+1} p\left[\theta|z^{l-1}\right] d\theta \\ &= \beta(l)\overline{\theta}^{(n)}(l-1) + \gamma(l)\overline{\theta}^{(n+1)}(l-1)\end{aligned}$$

The algorithm's recursion (14) thus follows. ■

1) Optimal TPM Estimator: Now let $\Pi = \begin{bmatrix} \theta_1 & 1-\theta_1 \\ 1-\theta_2 & \theta_2 \end{bmatrix}$ for $\theta_i \in [0,1]\,,\; i = 1,\, 2$. Then (11) reads in terms of θ_i as

$$p\left[\theta_i | z^k\right] = \left\{1 + \gamma_i\left(k\right)\left[\theta_i - \overline{\theta}_i\left(k-1\right)\right]\right\} p\left(\theta_i | z^{k-1}\right) = \{\underbrace{\left[1 - \gamma_i\left(k\right)\overline{\theta}_i\left(k-1\right)\right]}_{\beta_i(k)} + \gamma_i\left(k\right)\theta_i\} p\left(\theta_i | z^{k-1}\right) \quad (16)$$

with

$$\gamma_i\left(k\right) = \eta_i(k)\left[\Lambda_i\left(k\right) - \Lambda_{\overline{i}}\left(k\right)\right] \quad (17)$$

where $\overline{i} = 2$ if $i = 1$ and $\overline{i} = 1$ if $i = 2$. It is seen from (16) that the likelihoods $p\left[z(k)|\theta_i,\, z^{k-1}\right]$ are linear in θ_i (i.e. in the form of (13)). Thus by applying Proposition 1 one has the following.

Algorithm 1 *(Optimal TPM Estimator)*

- *Initialization:* $\overline{\theta}_i^{(k)}(0) = E[\theta_i^k]$, $k = 1, 2, \ldots$
- *For* $k = 1,\, 2,\, \ldots$

  For $i = 1, 2$ *(possibly in parallel)*

  For $l = 1,\, 2,\, \ldots,\, k$

$$\overline{\theta}_i^{(k-l+1)}\left(l\right) = \left[1 - \gamma_i\left(l\right)\overline{\theta}_i\left(l-1\right)\right]\overline{\theta}_i^{(k-l+1)}\left(l-1\right) + \gamma_i\left(l\right)\overline{\theta}_i^{(k-l+2)}\left(l-1\right) \quad (18)$$

where the measurement functions $\gamma_i\left(k\right)$, $i = 1, 2$ are computed through (17).

2) Implementation Issues: This algorithm requires a knowledge of all moments $\overline{\theta}_i^{(2)}\left(0\right)$, $\overline{\theta}_i^{(3)}\left(0\right)$, ..., $\overline{\theta}_i^{(k)}\left(0\right)$, ... of the TPM's prior PDF, which are in general nonzero. In such cases the algorithm may become infeasible for large k since the recursion (18) apparently requires computation and memory which increase linearly with k (Figure 1). The news is, however, not too bad. First the algorithm is finite for priors with a finite number of nonzero moments, and second, the exact formulation would still allow various approximations to be done in order to provide tractable and efficient implementations for the other cases. Another doubtless use of the optimal algorithm is for *off-line* identification of TMPs in the MM algorithms' design.

Some techniques to limit the computation/memory of the algorithm are:

- *"Sliding window" (SW)*: Run the recursion (18) for $l = k-s+1,\, k-s+2,\, \ldots,\, k$, where the window length s is fixed. This is a common technique in adaptive estimation. In our algorithm it is virtually equivalent to ignoring the moments of order higher than s.
- *"Freezing" the steady estimate*: Run the optimal TPM estimator until steady state of the estimate is reached (depending on the allowable computational resources) and then stop adaptation. This variant relies heavily on the assumption of a constant TPM and the algorithm's convergence rate.
- *Combined (Exact/SW) implementation*: More sophisticated and more accurate is to run the optimal estimator until reaching some allowable level of computation/memory and continue afterwards with sliding window mode. Running the optimal estimator initially is intended to speed-up the convergence of the subsequent SW.

$$
\begin{array}{lllllll}
\overline{\theta}^{(1)}(0) & \overline{\theta}^{(2)}(0) & \overline{\theta}^{(3)}(0) & \ldots & \overline{\theta}^{(k)}(0) & \overline{\theta}^{(k+1)}(0) & \ldots \\
\overline{\theta}^{(1)}(1) & \overline{\theta}^{(2)}(1) & \ldots & \overline{\theta}^{(k-1)}(1) & \overline{\theta}^{(k)}(1) & \ldots & \\
\overline{\theta}^{(2)}(2) & \ldots & \overline{\theta}^{(k-1)}(2) & \overline{\theta}^{(k)}(2) & \ldots & & \\
\ldots & \ldots & \ldots & \ldots & & & \\
\overline{\theta}^{(1)}(k-1) & \overline{\theta}^{(2)}(k-1) & \ldots & & & & \\
\overline{\theta}^{(1)}(k) & \ldots & & & & & \\
\ldots & & & & & &
\end{array}
$$

Fig. 1: Optimal Algorithm

figure

The optimal algorithm derived in this section not only provides a solution for the 2-mode case, but also reveals some inherent computational difficulties of the optimal MMSE estimator, which stem from the dependence of the estimator gain (variance) at time k on all moments of the prior distribution of order up to k. Thus, for the general case of a Markov chain with $m \geq 2$, addressed later, we indispensably focus on various approximation techniques to obtain feasible TPM estimators. Of particular interest for its simplicity is the second-order approximation of the optimal Algorithm 1, which we derive next.

B. *Second-Order Approximate Algorithm*

In the sequel we consider the general case with $m \geq 2$.

First, it follows from (11) that

$$
\begin{aligned}
\overline{\boldsymbol{\pi}}_i(k) &= \int \boldsymbol{\pi}_i p\left[\boldsymbol{\pi}_i | z^k\right] d\boldsymbol{\pi}_i \\
&= \int \boldsymbol{\pi}_i p\left[\boldsymbol{\pi}_i | z^{k-1}\right] d\boldsymbol{\pi}_i + \eta_i(k) \int \boldsymbol{\pi}_i \left[\boldsymbol{\pi}_i - \overline{\boldsymbol{\pi}}_i(k-1)\right]' \boldsymbol{\Lambda}(k)\, p\left[\boldsymbol{\pi}_i | z^{k-1}\right] d\boldsymbol{\pi}_i \\
&= \overline{\boldsymbol{\pi}}_i(k-1) + \eta_i(k) \int \left[\boldsymbol{\pi}_i - \overline{\boldsymbol{\pi}}_i(k-1)\right]\left[\boldsymbol{\pi}_i - \overline{\boldsymbol{\pi}}_i(k-1)\right]' p\left[\boldsymbol{\pi}_i | z^{k-1}\right] \boldsymbol{\Lambda}(k)\, d\boldsymbol{\pi}_i \\
&= \overline{\boldsymbol{\pi}}_i(k-1) + \eta_i(k) \operatorname{cov}\left(\boldsymbol{\pi}_i | k-1\right) \boldsymbol{\Lambda}(k)
\end{aligned}
\tag{19}
$$

Similarly, after straightforward manipulations, we have for the covariance (see Appendix):

$$
\begin{aligned}
\operatorname{cov}\left(\boldsymbol{\pi}_i | k\right) &= \operatorname{cov}\left(\boldsymbol{\pi}_i | k-1\right) - \eta_i \operatorname{cov}\left(\boldsymbol{\pi}_i | k-1\right) \boldsymbol{\Lambda}(k) \boldsymbol{\Lambda}'(k) \operatorname{cov}\left(\boldsymbol{\pi}_i | k-1\right) \eta_i \\
&\quad + \eta_i \int \left[\boldsymbol{\pi}_i - \overline{\boldsymbol{\pi}}_i(k-1)\right]\left[\boldsymbol{\pi}_i - \overline{\boldsymbol{\pi}}_i(k-1)\right]'\left[\boldsymbol{\pi}_i - \overline{\boldsymbol{\pi}}_i(k-1)\right]' \boldsymbol{\Lambda}(k)\, p\left(\boldsymbol{\pi}_i | z^{k-1}\right) d\boldsymbol{\pi}_i
\end{aligned}
\tag{20}
$$

Under the approximating assumption that the third central moments of $p\left(\boldsymbol{\pi}_i | z^{k-1}\right)$ in (20) are zero we obtain the following second-order approximate MMSE estimator for the TPM.

Algorithm 2 *(Second-Order TPME)*

- *Initialization:* $\overline{\boldsymbol{\pi}}_i\left(0\right)= E\left[\boldsymbol{\pi}_i\right], \, \Sigma_i\left(0\right) =cov(\boldsymbol{\pi}_i)$
- *For* $k = 1, \, 2, \, \ldots$

$$\widetilde{\boldsymbol{\Lambda}}\left(k\right) = \frac{\boldsymbol{\Lambda}\left(k\right)}{\boldsymbol{\mu}'\left(k-1\right)\overline{\overline{\Pi}}\left(k-1\right)\boldsymbol{\Lambda}\left(k\right)} \tag{21}$$

For $i = 1, 2, \, \ldots m$, *(possibly in parallel)*

$$\Gamma_i\left(k\right) = \mu_i\left(k-1\right)\Sigma_i\left(k-1\right) \tag{22}$$

$$\overline{\boldsymbol{\pi}}_i\left(k\right) = \overline{\boldsymbol{\pi}}_i\left(k-1\right) + \Gamma_i\left(k\right)\widetilde{\boldsymbol{\Lambda}}\left(k\right) \tag{23}$$

$$\Sigma_i\left(k\right) = \Sigma_i\left(k-1\right) - \Gamma_i\left(k\right)\widetilde{\boldsymbol{\Lambda}}\left(k\right)\widetilde{\boldsymbol{\Lambda}}'\left(k\right)\Gamma_i'\left(k\right) \tag{24}$$

This estimator has a simple interpretation: at time k the old estimate $\left\{\overline{\boldsymbol{\pi}}_i\left(k-1\right), \, \Sigma_i\left(k-1\right)\right\}$ summarizes the previous information; $\widetilde{\boldsymbol{\Lambda}}\left(k\right)$ provides the new data information; $\Gamma_i\left(k\right)$ plays the role of the estimator gain in (23); and the term $\Gamma_i\left(k\right)\widetilde{\boldsymbol{\Lambda}}\left(k\right)\widetilde{\boldsymbol{\Lambda}}'\left(k\right)\Gamma_i'\left(k\right)$ in (24) expresses the estimate's *covariance reduction* resulting from the new information.

Albeit intuitively appealing and practicable, this algorithm may unfortunately have numerical problems in keeping the estimated TPM with rows summing-up to one and the non-negative definiteness of the covariances $\Sigma_i\left(k\right)$. Precautions must be taken to prevent a possible divergence. This will be addressed in the simulation section.

C. Quasi-Bayesian Algorithm

To overcome the nontrivial numerical difficulties of the above moment-based approximation, another approach to limit the computation/memory increase of the exact Bayesian estimator is introduced next. It is based on the so-called Quasi-Bayesian approximation of [27] for estimation of finite mixtures.

Starting from (11)the likelihood of each $\boldsymbol{\pi}_i, \, i = 1, \, \ldots, \, m$ can be represented (see Appendix) in the following *mixture* form

$$f\left[z\left(k\right)|\boldsymbol{\pi}_i\right] = p\left[z\left(k\right)|\boldsymbol{\pi}_i, \, z^{k-1}\right] = \sum_{j=1}^{m}\pi_{ij}g_{ij}\left(k\right) \tag{25}$$

where

$$g_{ij}\left(k\right) = g_{ij}\left[z\left(k\right)\right] = 1 + \eta_i\left(k\right)\left[\Lambda_j\left(k\right) - \overline{\boldsymbol{\pi}}_i'\left(k-1\right)\boldsymbol{\Lambda}\left(k\right)\right] \tag{26}$$

Thus for each $\boldsymbol{\pi}_i, \, i = 1, \, \ldots, \, m$ our TMP-estimation problem can be boiled down to a well known classification problem referred to as *prior probability estimation (PPE)* of finite mixtures [28], [29], or to as *unsupervised learning* [27], in a more general statistical setting.

1) QB Prior Probability Estimation: The PPE problem can be stated as follows: *For the finite mixture model*

$$f\left[z(k)|\boldsymbol{\pi}\right] = \sum_{j=1}^{m}\pi_j g_j\left[z(k)\right], \quad \sum_{j=1}^{m}\pi_j = 1, \, \pi_j \in [0,1] \tag{27}$$

with known mixture component PDFs $g_j\left[z(k)\right]$, *estimate the unknown probabilistic weights* $\pi_j, \, j = 1, \, \ldots, \, m$ *given a sequence of independent observations* $z(1), \, z(2), \, \ldots, \, z(k), \, \ldots$ *with PDF* $f\left[z(k)|\boldsymbol{\pi}\right]$.

There exist a number of publications and results in solving this problem, see e.g. [28], [29], [30], [27]. For the Bayesian framework of the TPM estimation, adopted in this work, the approach of *Quasi-Bayesian* (QB) PPE estimation [30] appears the fittest, since it provides a quasi-posterior *mean* of the PPE, and thus meets our MMSE TPM estimation objective exactly.

The QB approximation relies on the assumption of the *Dirichlet distribution* (DD) [31] for the mixture's prior probabilities $\boldsymbol{\pi} = (\pi_1, \ldots, \pi_m)$. The DD is defined by[2]

$$\mathcal{D}(\boldsymbol{\pi}; \alpha_1, \ldots, \alpha_m) = \frac{\Gamma(\alpha_1 + \alpha_2 + \ldots + \alpha_m)}{\Gamma(\alpha_1)\Gamma(\alpha_2)\ldots\Gamma(\alpha_m)} \prod_{j=1}^{m} \pi_j^{\alpha_j - 1}$$

with parameters $\alpha_j \geq 0$, $j = 1, \ldots, m$. It can be viewed as a multivariate Beta distribution, which is natural for $\boldsymbol{\pi}$ due to the unit simplex requirement $\sum_{j=1}^{m} \pi_j = 1,\ \pi_j \in [0,1]$.

On receipt of a measurement, e.g. $z(1)$, from the mixture (27) the Dirichlet prior $p(\pi)$ yields a posterior $p(\pi|z_1)$ which is a weighted sum of m DDs. Such a splitting leads in time to a weighted sum with an ever increasing number of DD components [27]. The essence of the QB approach is to approximate at each time step the posterior weighted sum of m DDs by a *single* DD. Thus the subsequent updating of the posterior takes place within the family of DDs, and calculating sequentially the quasi-posterior means $\overline{\pi}_j(k)$, $j = 1, 2, \ldots m$ becomes straightforward. This scheme is elaborated in [30] and [27]. The resulting QB procedure for the PPE problem (27) computes $\overline{\pi}_j(k)$ recursively via

$$\overline{\pi}_j(k) = \frac{1}{k + \alpha(0)} \alpha_j(k) \tag{28}$$

$$\alpha_j(k) = \alpha_j(k-1) + \frac{\alpha_j(k-1)\, g_j(k)}{\sum_{l=1}^{m} \alpha_l(k-1)\, g_l(k)} \tag{29}$$

where $\alpha_j(0) > 0$ and $\alpha(0) = \sum_{j=1}^{m} \alpha_j(0)$.

In can be noted that the underlying idea of the QB approximation is similar to that of the GPB1 approach for MM state estimation [15] where at each time-step the branching weighted sum of Gaussians is approximated by a single Gaussian, and thus the state estimate update is limited to the class of single Gaussians.

2) QB TPM Estimation: By applying the QB algorithm of [30] given by (28)–(29) to the TPM estimation problem, formulated via (25)–(26) as mixture estimation for each $\boldsymbol{\pi}_i$, $i = 1, \ldots, m$ we obtain the following.

Algorithm 3 *(Quasi-Bayesian TPME)*

- *A priori parameters:*

$$\boldsymbol{\alpha}_i(0) = [\alpha_{i1}(0), \alpha_{i2}(0), \ldots, \alpha_{im}(0)]' \tag{30}$$

$$\alpha_i(0) = \sum_{j=1}^{m} \alpha_{ij}(0),\ i = 1, 2, \ldots, m; \quad \alpha_{ij}(0) \geq 0 \tag{31}$$

[2]Note that the DD is $m-1$ dimensional (defined over the $(m-1)$-dimensional unit simplex) since π_j are linearly dependent, e.g. $\pi_m = 1 - (\pi_1 + \ldots + \pi_{m-1})$.

- *Initialization:*

$$\overline{\boldsymbol{\pi}}_i\left(0\right)=\frac{1}{\alpha_i\left(0\right)}\boldsymbol{\alpha}_i\left(0\right),\ i=1,2,\ \ldots,\ m \tag{32}$$

- *For* $k=1,\ 2,\ ,\ldots$

  For $i=1,2,\ \ldots m$ *(possibly in parallel)*

  For $j=1,2,\ \ldots m$

$$g_{ij}\left(k\right)=1+\eta_i\left(k\right)\left[\Lambda_j\left(k\right)-\overline{\boldsymbol{\pi}}_i'\left(k-1\right)\boldsymbol{\Lambda}\left(k\right)\right] \tag{33}$$

$$\alpha_{ij}\left(k\right)=\alpha_{ij}\left(k-1\right)+\frac{\alpha_{ij}\left(k-1\right)g_{ij}\left(k\right)}{\sum_{j=1}^{m}\alpha_{ij}\left(k-1\right)g_{ij}\left(k\right)} \tag{34}$$

$$\overline{\pi}_{ij}\left(k\right)=\frac{1}{k+\alpha_i\left(0\right)}\alpha_{ij}\left(k\right) \tag{35}$$

where $\eta_i\left(k\right)$ is given by (12).

Note that the necessary parameter vectors $\boldsymbol{\alpha}_i\left(0\right)$ represent the un-normalized a priori TPM vectors $\overline{\boldsymbol{\pi}}_i\left(0\right)$ and are normalized through (32) so that the initial TMP estimate belongs to the unit simplex of valid stochastic matrices. Regarding the initialization of the algorithm the following property of the DD is very useful in practical application: if its parameters are chosen as $\alpha_1=\alpha_2=\ldots=\alpha_m=1$ it coincides with the uniform distribution over the unit $(m-1)$-dimensional simplex. Thus, if no a priori information about the TPM is available, the QB-TPM estimator is naturally initialized with the non-informative (uniform) prior, which in view of (32) gives $\overline{\pi}_{ij}\left(0\right)=1/m$, $i,j=1,\ldots,m$. The QB algorithm is extremely simple for implementation and requires moderate amount of computation. The convergence properties of the QB approach are discussed in [27] and the relevant references therein.

D. *Numerical-Integration Algorithm*

Finally, we consider another, probably the most straightforward approach to TPM estimation — implement the PDF update Eq. (9) numerically over a finite grid of valid transition probability matrices and then calculate the posterior mean's estimate by numerical integration using the posterior PDF over the grid set. The resulting algorithm for N grid TPMs $\Pi^{(s)},\ s=1,2,\ \ldots\ N$ is:

Algorithm 4 *(Numerical-Integration TPME)*

- *Initialization:*

$$\overline{\Pi}\left(0\right)=\frac{1}{N}\sum_{s=1}^{N}\Pi^{(s)}p^{(s)}\left(0\right)\ \text{ where } p^{(s)}\left(0\right)\triangleq P\left\{\Pi^{(s)}\right\}$$

- *For* $k=1,\ 2,\ ,\ldots$

$$p^{(s)}\left(k\right)=\frac{\boldsymbol{\mu}'\left(k-1\right)\Pi^{(s)}\boldsymbol{\Lambda}\left(k\right)}{\boldsymbol{\mu}'\left(k-1\right)\overline{\Pi}\left(k-1\right)\boldsymbol{\Lambda}\left(k\right)}p^{(s)}\left(k-1\right),\quad s=1,\ 2,\ \ldots,\ N$$

$$\overline{\Pi}\left(k\right)=\frac{1}{N}\sum_{s=1}^{N}\Pi^{(s)}p^{(s)}\left(k\right)$$

Note that this algorithm implements directly the general TPM's recursion (9) rather than its decoupled version (11) valid under the additional assumption (10) in Theorem 1. Apparently the decoupled version of the algorithm (i.e. that based on (11)) can be implemented in completely the same manner. It has a substantially reduced computational load in comparison with the fully coupled Algorithm 4.

V. SIMULATION RESULTS

In this section we evaluate the performances of TPM-adaptive IMM algorithms, based on different TPM-estimators. Two typical applications of Markov switching models are simulated $-$ *systems subject to measurement failures* and *maneuvering target tracking*.

A. Example 1: System with Failures

Under consideration is a scalar dynamic system described by

$$x(k+1) = x(k) + w(k)$$
$$z(k) = m(k)x(k) + [100 - 90m(k)]\, v(k)$$

with $x(0) \sim \mathcal{N}(0, 20^2),\ w(k) \sim \mathcal{N}(0, 2^2),\ v(k) \sim \mathcal{N}(0, 1).$ The Markov model of switching is:

$$m(k) \in \left\{m^{(1)},\, m^{(2)}\right\} = \{0,\ 1\}, \quad \mu_1(0) = \mu_2(0) = 0.5$$
$$\pi_{ij} = P\left\{m(k+1) = m^{(j)} | m(k) = m^{(i)}\right\}, \quad i, j = 1, 2$$

Note that $m(k) = 0$ corresponds to a measurement failure at k.

The following five IMM algorithms were implemented:

- *Exact TPM*: two-model IMM with the true TPM;
- *Non-Adaptive TPM*: an IMM with a typical design of TPM with diagonally-dominant elements: $\pi_{11} = \pi_{22} = 0.9$;
- *Optimal TPM-Adaptive:* the IMM using the optimal TPM estimator (Algorithm 1);
- *Second-Order TPM-Adaptive:* the IMM using Algorithm 2. To avoid divergence the estimate is truncated within $[\epsilon, 1-\epsilon]$, $\epsilon = 0.001$ if necessary[3], and at the instance of truncation the variance $\sigma^2(k)$ is reset to $\sigma^2(0)$.
- *Numerical TPM-Adaptive:* the IMM using the decoupled version of Algorithm 4 with 50 grid vectors for each row.

Two scenarios were simulated. In the first (*fixed-TPM*) scenario the true TPM was chosen with $\pi_{11} = 0.6,\ \pi_{22} = 0.15$. It describes situations with frequent measurement failures. In each run a true mode switching sequence was generated according to the true TPM. A sample sequence generated by this transition model is

010001101010000100001010100000010100101000100000001010001000000010001000100100000101

This can be interpreted as receiving measurements occasionally with unknown (random) sampling rate.

[3] Practically this occurred rarely (about once per 100 runs with 60 scans each).

Results from 1000 Monte Carlo runs of this scenario are shown in Figures 2 and 3. The estimation accuracy of TPM and state estimation is evaluated in terms of *mean absolute error* (MAE) – the absolute value of the actual error, averaged over runs. Figure 2 illustrates the convergence of the TPM estimators. Figure 3 shows that in case of a mismatch regarding the TPM the adaptive algorithms perform better than the typical IMM design (with π_{22} fairly largely mismatched in this scenario).

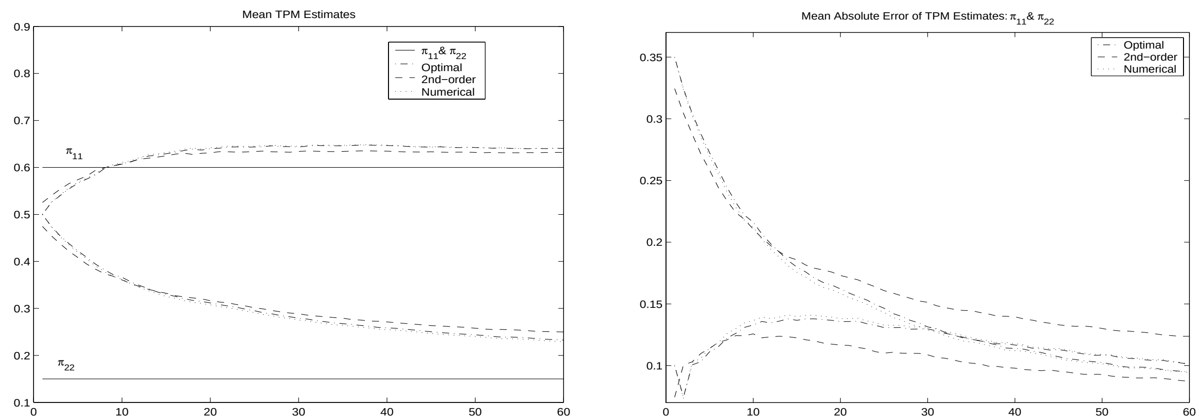

Fig. 2: Convergence of TPM Estimators

figure

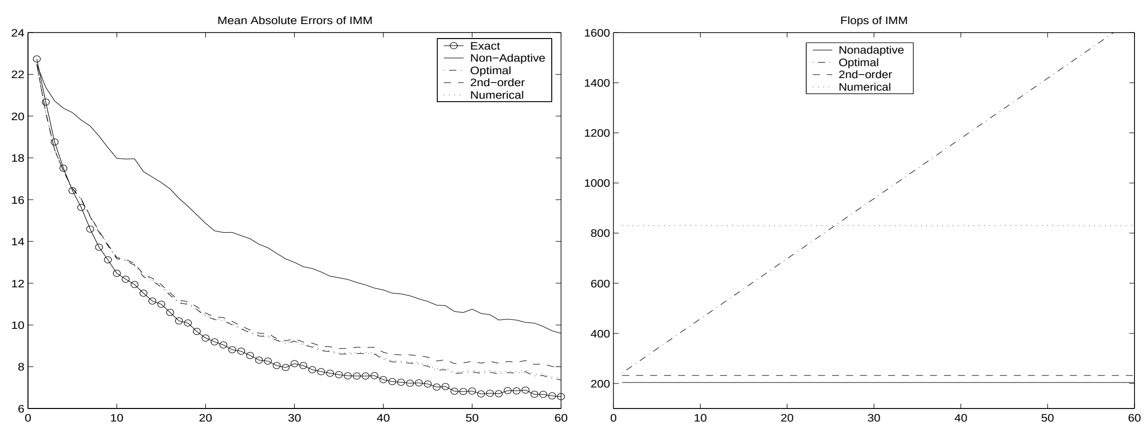

Fig. 3: IMM Algorithms' Errors & Computation

figure

The second scenario is random (*random-TPM*). In each run the true TPM was randomly chosen from the set of valid TPMs. Its sampling was done assuming uniform distribution of the TPMs. The purpose of this scenario is to provide an as fair as possible comparison of the algorithms over an ensemble of different system failure scenarios.

Results from 2000 MC runs are shown in Figure 4[4]. The performances of the adaptive algorithms are better

[4]The errors of the numerical and optimal algorithms are indistinguishable in this plot resolution.

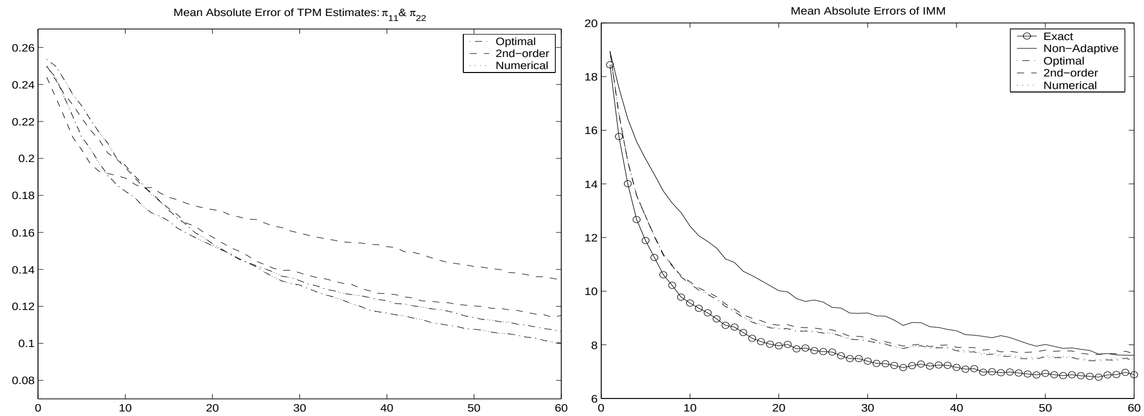

Fig. 4: Random Scenario: TPM & IMM Errors

figure

again.

B. Example 2: Maneuvering Target Tracking

We considered a very simplified illustrative example of moving target with Markov switching acceleration. The one-dimensional target dynamics is given by

$$\begin{bmatrix} p \\ v \end{bmatrix}(k+1) = \begin{bmatrix} 1 & T \\ 0 & 1 \end{bmatrix} \begin{bmatrix} p \\ v \end{bmatrix}(k) + \begin{bmatrix} T^2/2 \\ T \end{bmatrix} [a\,(k+1) + w(k+1)]$$

where p, v and a stand for the target position, velocity and acceleration respectively and w models "small" perturbations in the acceleration. $p\,(0) \sim \mathcal{N}\left[80000, 100^2\right]$, $v(0) \sim \mathcal{N}\left[400, 100^2\right]$ and $w\,(k) \sim \mathcal{N}\left[0, 2^2\right]$ is white noise sequence independent of $p\,(0)$ and $v\,(0)$. The acceleration process $a\,(k)$ is a Markov chain with three states $a_1 = 0$, $a_2 = 20$, $a_2 = -20$ and initial probabilities $\mu_1\,(0) = 0.8$, $\mu_2\,(0) = \mu_3\,(0) = 0.1$. The sampling period is $T = 10$.

The measurement equation is

$$z\,(k) = x\,(k) + \nu(k)$$

where the measurement error $\nu(k) \sim \mathcal{N}\left[0, 100^2\right]$ is white noise.

Four IMM algorithms were implemented:

- *Exact-TPM*: IMM knowing the true TPM;
- *Non-Adaptive*: IMM with a typical design of *fixed* TPM with diagonally-dominant elements: $\pi_{ii} = 0.9, \pi_{ij} = 0.05$ for $j \neq i,\ i, j = 1,\ 2,\ 3;$
- *QB-Adaptive:* IMM using the Quasi-Bayesian TPM estimator (Algorithm 3). The initialization was done with $\alpha_{ij}\,(0) = 0.01,\ i, j = 1, 2, 3.$

- *NI-Adaptive:* IMM using a decoupled version of the Numerical Integration TPM estimator (Algorithm 4) with a number of grid vectors for each row $N = 231$.

We ran first the four algorithms on various scenarios with *fixed* true TPM. The results are scenario-dependent. A specific problem arises sometimes from the small sample size (60 scans in each run) — some of the random switching histories may not contain statistically sufficient information to well estimate *all* elements of the TPM. It was also noted that for some parameters of the dynamic system (e.g. very small jumps, very large measurement noise, or very high sampling rate) the TPM may become not identifiable. See [27] for a discussion on the finite mixture identifiability.

For the above chosen set of model parameters the algorithms converged successfully, although the convergence of the TPM appeared sometimes rather slower than desirable. In all cases the convergence and accuracy of the numerical integration algorithm were more satisfactory. To obtain an overall evaluation we ran the algorithms over an ensemble of TPMs, randomly (uniformly) chosen from the set of *all* valid TPMs.

Results from 1000 MC runs are presented in Figures 5 through 7. The adaptive IMMs show better overall performance than the non-adaptive one. Clearly, in quite a lot of the runs the fixed TPM of the latter is mismatched with the true one, while the adaptive algorithms provide better estimated TPM. This indicates that for situations where the target behavior can not be described by one TPM it is preferable to use an adaptive TPM.

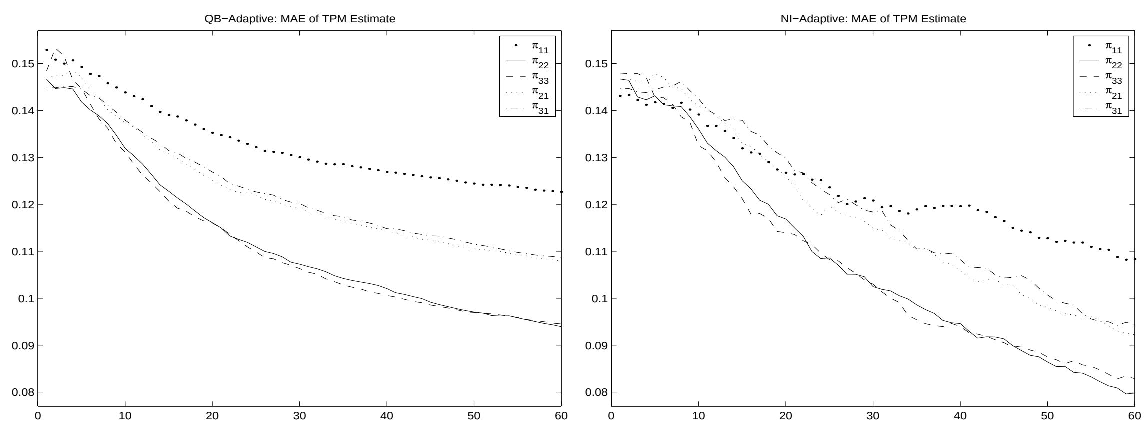

Fig. 5: Random Scenario: Convergence of TPM Estimators

figure

The relative computational load of the simulated IMMs[5] is given in Table I.

IMM	QB	2nd-order	NI
1	1.135	1.137	18.281

Table I: Computational Load

table

[5]This result for the 2-nd order algorithm is obtained, based on extrapolating the data from Example 1.

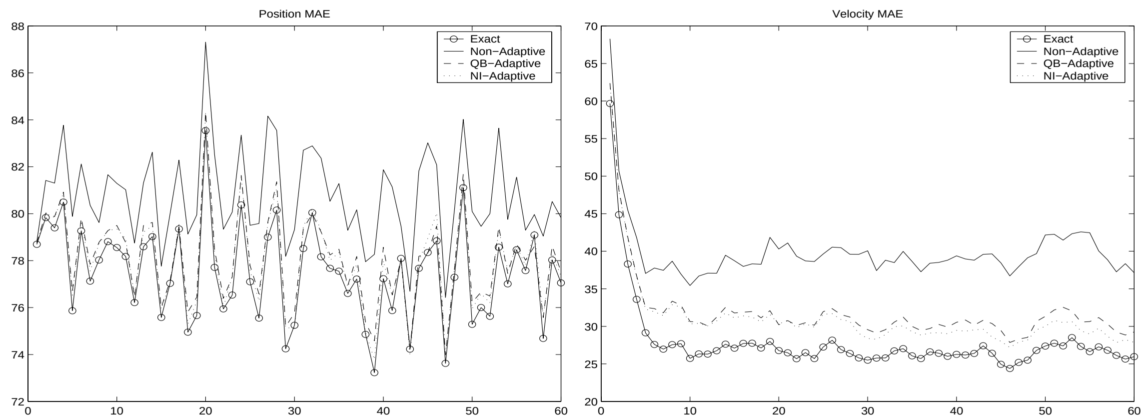

Fig. 6: IMM: Algorithms' Errors

figure

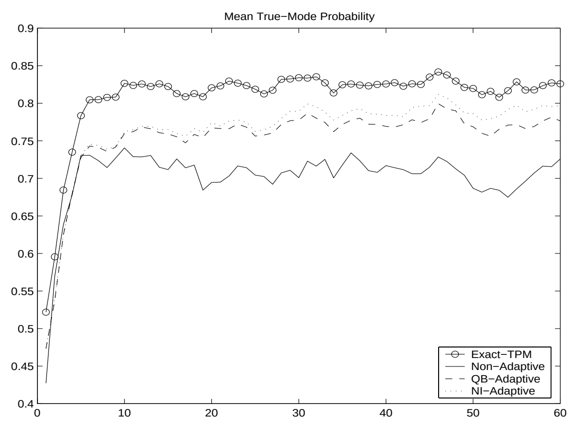

Fig. 7: IMM Probability of the True Model

figure

VI. Summary and Conclusions

In this work we have obtained a feasible recursion for the posterior probability density function, and proposed four different recursive algorithms for estimation/adaptation of the transition probability matrix in multiple model estimation for Markov jump systems. They all utilize the approximation of the TPM likelihood, available from most MM state estimators and differ in the way they compute the MMSE estimate (posterior *mean*) of the TPM.

- The optimal TPM estimator has exhibited the best accuracy in simulation. However its computational complexity increases linearly with time. Techniques for limiting its computation/memory have been discussed.
- The second-order approximate TPM estimator truncates higher order moments of the TPM posterior PDF. The estimator is computationally attractive, but divergence could sometimes occur, because the *computed* covariance may loose non-negative definiteness or the estimate may leave the TPM validity set. Precautions

must be taken in application.

- The Quasi-Bayesian TPM estimator updates the estimates within the class of Dirichlet distributions. It is simple for implementation and has almost negligible computational load. As compared with the numerical integration TPM estimator it is less accurate and its convergence is slower. The differences, however, are not substantial.
- The numerical integration TPM estimator computes the estimates via numerical updating of the posterior TPM PDFs. While showing best accuracy and convergence among the approximate algorithms, even with a very coarse grid, its computational burden may become prohibitive for high dimensional problems.

In general, as illustrated by simulations, the proposed TPM-adaptive MM estimation can in principle be efficient when the TPM of the switching Markov chain is unknown, but *identifiable.* The identifiability of the TPM of MJS is a serious and important issue, which requires a further study.

APPENDIX

Proof of Theorem 1.

Denoting $d\Pi/d\boldsymbol{\pi}_i = d\boldsymbol{\pi}_1 \ldots d\boldsymbol{\pi}_{i-1} d\boldsymbol{\pi}_{i+1} \ldots d\boldsymbol{\pi}_m$ we have, in view of (9), that

$$p\left(\boldsymbol{\pi}_i|z^k\right) = \int p\left(\Pi|z^k\right) d\Pi/d\boldsymbol{\pi}_i = \frac{A_i}{\boldsymbol{\mu}'\left(k-1\right)\overline{\overline{\Pi}}\left(k-1\right)\boldsymbol{\Lambda}\left(k\right)} \tag{40}$$

where

$$A_i = \int \boldsymbol{\mu}'\left(k-1\right)\Pi\boldsymbol{\Lambda}\left(k\right)p\left(\Pi|z^{k-1}\right) d\Pi/d\boldsymbol{\pi}_i$$

Then

$$A_i \;=\; \int \sum_{l=1}^{m} \mu_l\left(k-1\right)\boldsymbol{\pi}_l'\boldsymbol{\Lambda}\left(k\right)p\left(\Pi|z^{k-1}\right) d\Pi/d\boldsymbol{\pi}_i \;=\; \sum_{l=1}^{m} \mu_l\left(k-1\right)\int \boldsymbol{\pi}_l' p\left(\Pi|z^{k-1}\right) d\Pi/d\boldsymbol{\pi}_i\boldsymbol{\Lambda}\left(k\right) \tag{42}$$

Further, accounting for (10)

$$\begin{aligned} A_i &= \left[\sum_{l\neq i} \mu_l\left(k-1\right)\overline{\boldsymbol{\pi}}_l'\left(k-1\right) + \mu_i\left(k-1\right)\boldsymbol{\pi}_i'\right]\boldsymbol{\Lambda}\left(k\right)p\left(\boldsymbol{\pi}_i|z^{k-1}\right) \\ &= \left[\sum_{l=1}^{m} \mu_l\left(k-1\right)\overline{\boldsymbol{\pi}}_l'\left(k-1\right) + \mu_i\left(k-1\right)\boldsymbol{\pi}_i' - \mu_i\left(k-1\right)\overline{\boldsymbol{\pi}}_i'\left(k-1\right)\right]\boldsymbol{\Lambda}\left(k\right)p\left(\boldsymbol{\pi}_i|z^{k-1}\right) \\ &= \left\{\boldsymbol{\mu}'\left(k-1\right)\overline{\overline{\Pi}}\left(k-1\right)\boldsymbol{\Lambda}\left(k\right) + \mu_i\left(k-1\right)\left[\boldsymbol{\pi}_i' - \overline{\boldsymbol{\pi}}_i'\left(k-1\right)\right]\boldsymbol{\Lambda}\left(k\right)\right\}p\left(\boldsymbol{\pi}_i|z^{k-1}\right) \end{aligned} \tag{43}$$

Thus (11) follows from (40) and (43)

■

Proof of (20):

$$\begin{aligned}
\operatorname{cov}(\boldsymbol{\pi}_i|k) &= \int \left[\boldsymbol{\pi}_i - \overline{\boldsymbol{\pi}}_i(k)\right]\left[\boldsymbol{\pi}_i - \overline{\boldsymbol{\pi}}_i(k)\right]' p\left[\boldsymbol{\pi}_i|z^k\right] d\boldsymbol{\pi}_i \\
&= \int \left\{\left[\boldsymbol{\pi}_i - \overline{\boldsymbol{\pi}}_i(k-1)\right] - \eta_i \operatorname{cov}(\boldsymbol{\pi}_i|k-1)\,\boldsymbol{\Lambda}(k)\right\}\left\{\left[\boldsymbol{\pi}_i - \overline{\boldsymbol{\pi}}_i(k-1)\right] - \eta_i \operatorname{cov}(\boldsymbol{\pi}_i|k-1)\,\boldsymbol{\Lambda}(k)\right\}' \\
&\qquad \left\{p\left[\boldsymbol{\pi}_i|z^{k-1}\right] + \eta_i\left[\boldsymbol{\pi}_i - \overline{\boldsymbol{\pi}}_i(k-1)\right]'\boldsymbol{\Lambda}(k)\,p\left(\boldsymbol{\pi}_i|z^{k-1}\right)\right\} d\boldsymbol{\pi}_i \\
&= \operatorname{cov}(\boldsymbol{\pi}_i|k-1) - \eta_i \operatorname{cov}(\boldsymbol{\pi}_i|k-1)\,\boldsymbol{\Lambda}(k)\,\boldsymbol{\Lambda}'(k)\operatorname{cov}(\boldsymbol{\pi}_i|k-1)\,\eta_i \\
&\quad + \eta_i \int \left[\boldsymbol{\pi}_i - \overline{\boldsymbol{\pi}}_i(k-1)\right]\left[\boldsymbol{\pi}_i - \overline{\boldsymbol{\pi}}_i(k-1)\right]'\left[\boldsymbol{\pi}_i - \overline{\boldsymbol{\pi}}_i(k-1)\right]'\boldsymbol{\Lambda}(k)\,p\left(\boldsymbol{\pi}_i|z^{k-1}\right) d\boldsymbol{\pi}_i
\end{aligned}$$

■

Proof of (25) with (26):

From (9) and (12)

$$\begin{aligned}
p\left[z(k)\,|\boldsymbol{\pi}_i,\, z^{k-1}\right] &= 1 + \eta_i(k)\left[\boldsymbol{\pi}_i - \overline{\boldsymbol{\pi}}_i(k-1)\right]'\boldsymbol{\Lambda}(k) = \eta_i(k)\boldsymbol{\pi}_i'\boldsymbol{\Lambda}(k) + \left[1 - \eta_i(k)\overline{\boldsymbol{\pi}}_i(k-1)'\boldsymbol{\Lambda}(k)\right] \\
&= \sum_{j=1}^{m} \eta_i(k)\pi_{ij}\Lambda_j + \left[1 - \eta_i(k)\overline{\boldsymbol{\pi}}_i(k-1)'\boldsymbol{\Lambda}(k)\right] = \sum_{j=1}^{m}\left\{\eta_i(k)\Lambda_j + \left[1 - \eta_i(k)\overline{\boldsymbol{\pi}}_i(k-1)'\boldsymbol{\Lambda}(k)\right]\right\}\pi_{ij} \\
&= \sum_{j=1}^{m}\left\{1 + \eta_i(k)\left[\Lambda_j - \overline{\boldsymbol{\pi}}_i(k-1)'\boldsymbol{\Lambda}(k)\right]\right\}\pi_{ij}
\end{aligned}$$

■

REFERENCES

[1] D. D. Sworder and J. Boyd, *Estimation Problems in Hybrid Systems*, Cambridge University Press, 1999.

[2] X. R. Li, "Hybrid Estimation Techniques," in *Control and Dynamic Systems: Advances in Theory and Applications*, C. T. Leondes, Ed., vol. 76, pp. 213–287. Academic Press, New York, 1996.

[3] W. D. Blair and G. A. Watson, "Interacting Multiple Model Algorithm with Aperiodic Data," in *Proc. SPIE Symp. on Acqisition, Tracking and Pointing*, Orlando, FL, Apr. 1992.

[4] Y. Bar-Shalom and X. R. Li, *Estimation and Tracking: Principles, Techniques, and Software*, Artech House, Boston, MA, 1993, (Reprinted by YBS Publishing, 1998).

[5] M. Busch and S. Blackman, "Evaliation of IMM Filtering for an Air Defence System Application," in *Signal and Data Processing of Small Targets 1995*, 1995, vol. SPIE 2561, pp. 435–447.

[6] Y. Bar-Shalom and X. R. Li, *Multitarget-Multisensor Tracking: Principles and Techniques*, YBS Publishing, Storrs, CT, 1995.

[7] S. S. Blackman and R. F. Popoli, *Design and Analysis of Modern Tracking Systems*, Artech House, Norwood, MA, 1999.

[8] L. Bloomer and J. E. Gray, "Are More Models Better?: The Effect of the Model Transition Matrix on the IMM Filter," in *The 34th Southeastern Simposium on System Theory (SSST)*, Huntsville, AL, March 2002.

[9] Y. Sawaragi, T. Katayama, and S. Fujichige, "Adaptive Estimation for a Linear System with Interrupted Observations," *IEEE Trans. Automatic Control*, pp. 152–154, April 1973.

[10] J. K. Tugnait and A. H. Haddad, "State Estimation under Uncertain Observations with Unknown Statistics," *IEEE Trans. Automatic Control*, vol. 24, pp. 201–210, April 1979.

[11] J. K. Tugnait and A. H. Haddad, "Adaptive Estimation in Linear Systems with Unknown Markovian Noise Statistics," *IEEE. Trans. Information Theory*, vol. 26, pp. 66–78, Jan. 1980.

[12] J. K. Tugnait, "Adaptive Estimation and Identification for Discrete Systems with Markov Jump Parameters," *IEEE Trans. Automatic Control*, vol. AC-27, no. 5, pp. 1054–1065, October 1982.

[13] J. Goutsias and J.M. Mendel, "Optimal Simultaneous Detection and Estimation of Filtered Discrete semi-Markov Chains," *IEEE. Trans. Information Theory*, vol. IT-34, no. 3, pp. 551–568, May 1988.

[14] M. West and J.F. Harrison, *Bayesian Forecasting and Dynamic Models*, Springer Series in Statistics. Springer-Verlag, 2 edition, 1997.

[15] Y. Bar-Shalom, X. R. Li, and T. Kirubarajan, *Estimation with Applications to Tracking and Navigation: Theory, Algorithms, and Software*, Wiley, New York, 2001.

[16] X. R. Li and Y. Bar-Shalom, "Multiple-Model Estimation with Variable Structure," *IEEE Trans. Automatic Control*, vol. AC-41, no. 4, pp. 478–493, Apr. 1996.

[17] X. R. Li, "Multiple-Model Estimation with Variable Structure—Part II: Model-Set Adaptation," *IEEE Trans. Automatic Control*, vol. AC-45, no. 11, pp. 2047–2060, Nov. 2000.

[18] X. R. Li, X. R. Zhi, and Y. M. Zhang, "Multiple-Model Estimation with Variable Structure—Part III: Model-Group Switching Algorithm," *IEEE Trans. Aerospace and Electronic Systems*, vol. AES-35, no. 1, pp. 225–241, Jan. 1999.

[19] X. R. Li, Y. M. Zhang, and X. R. Zhi, "Multiple-Model Estimation with Variable Structure—Part IV: Design and Evaluation of Model-Group Switching Algorithm," *IEEE Trans. Aerospace and Electronic Systems*, vol. AES-35, no. 1, pp. 242–254, Jan. 1999.

[20] X. R. Li and Y. M. Zhang, "Multiple-Model Estimation with Variable Structure—Part V: Likely-Model Set Algorithm," *IEEE Trans. Aerospace and Electronic Systems*, vol. AES-36, no. 2, pp. 448–466, Apr. 2000.

[21] X. R. Li, "Engineer's Guide to Variable-Structure Multiple-Model Estimation for Tracking," in *Multitarget-Multisensor Tracking: Applications and Advances*, Y. Bar-Shalom and D. W. Blair, Eds., vol. III, chapter 10, pp. 499–567. Artech House, Boston, MA, 2000.

[22] A. Logothetis and V. Krishnamurthy, "Expectation Maximization Algorithms for MAP Estimation of Jump Markov Linear Systems," *IEEE Trans. Signal Processing*, vol. 47, no. 8, pp. 2139–2156, August 1999.

[23] A. Doucet, N.J. Gordon, and V. Krishnamurthy, "Particle Filters for State Estimation of Jump Markov Linear Systems," *IEEE Trans. on Signal Processing*, 2000.

[24] L. A. Johnston and V. Krishnamurthy, "An Improvement to the Interacting Multiple Model (IMM) Algorithm," *IEEE Trans. Signal Processing*, vol. 49, no. 12, pp. 2893–2908, 2001.

[25] H. A. P. Blom and Y. Bar-Shalom, "The Interacting Multiple Model Algorithm for Systems with Markovian Switching Coefficients," *IEEE Trans. Automatic Control*, vol. AC-33, no. 8, pp. 780–783, Aug. 1988.

[26] E. Mazor, A. Averbuch, Y. Bar-Shalom, and J. Dayan, "Interacting Multiple Model Methods in Target Tracking: A Survey," *IEEE Trans. Aerospace and Electronic Systems*, vol. AES-34, no. 1, pp. 103–123, 1996.

[27] D.M. Titterington, A.F.M. Smith, and U.E. Makov, *Statistical Analysis of Finite Mixture Distributions*, John Wiley & Sons, 1985.

[28] D. Kazakos, "Recursive Estimation of Prior Probabilities Using a Mixture," *IEEE. Trans. Information Theory*, vol. IT-23, pp. 203–211, March 1977.

[29] G. R. Dattatreya and L. N. Kanal, "Estimation of Mixing Probabilities in Multiclass Finite Mixtures," *IEEE Trans. Systems, Man, and Cybernetics*, vol. SMC-20, no. 1, pp. 149–158, Jan./Feb. 1990.

[30] A. F. M. Smith and U. E. Makov, "A Quasi-Bayesian Sequential Procedures for Mixtures," *J. R. Statist. Soc.*, vol. 40, no. 1, pp. 106–112, 1978.

[31] S. C. Port, *Theoretical Probability for Applications*, John Wiley & Sons, 1994.

Vesselin P. Jilkov (M'01) received the M. S. degree in mathematics (specialization "Operations Research") from the University of Sofia, Sofia, Bulgaria in 1982, the Ph. D. degree in electrical engineering ("Radar Target Tracking") in 1988, and the academic rank "Senior Research Fellow" (Associate Professor) of the Bulgarian Academy of Sciences, Sofia in 1997.

During 1982-1988 he was a researcher at the R&D Institute of Special Electronics, Sofia, where he did research and development of radar target tracking algorithms and software, implemented in radar surveillance systems. He joined the Central Laboratory for Parallel Processing (CLPP) – Bulgarian Academy of Sciences, Sofia in 1989. There he worked as a key researcher in numerous academic and industry projects in the areas of Kalman filtering, hybrid system estimation, target tracking, multi-sensor data fusion, and parallel processing. During 1992-1994 he participated in joint R&D Projects with ELTA Electronics Ltd., Israel. Since June 1999 Dr. Jilkov has been with the Department of Electrical Engineering, University of New Orleans, New Orleans, LA where he is engaged in teaching and conducting research in the areas of hybrid systems estimation and target tracking. His main research interests include stochastic systems, applied estimation and filtering, target tracking, multi-sensor data fusion, parallel processing.

Dr. Jilkov is author/coauthor of over 30 journal and conference papers and one book (In Bulgarian) on parallel Kalman filtering. He is a member of IEEE and ISIF. He also serves as a reviewer of several refereed international journals.

X. Rong Li (S'90-M'92-SM'95) received a B.S. degree and an M.S. degree from Zhejiang University, Hangzhou, Zhejiang, PRC, in 1982 and 1984, respectively, and an M.S. degree and a Ph.D. degree from University of Connecticut, USA, in 1990 and 1992, respectively. He joined the University of New Orleans in 1994. During 1986-1987 he did research on electric power at University of Calgary, Canada. He was an Assistant Professor at University of Hartford, from 1992 to 1994. Professor Li has published four books: "Estimation and Tracking" (Artech House, 1993), "Multitarget-Multisensor Tracking" (YBS Publishing, 1995), "Probability, Random Signals, and Statistics" (CRC Press, 1999), and "Estimation with Applications to Tracking and Navigation" (Wiley, 2001); four book chapters; more than thirty articles; more 100 conference papers. Dr. Li has served as an Editor for IEEE Transactions on Aerospace and Electronic Systems since 1996, an Associate Editor, 1995-1996; received a Career award and an RIA award from the U.S. National Science Foundation; served as Vice President for Technical Activities and member of Board of Directors, International Society of Information Fusion since 1998; served as Steering Chair, General Vice-Chair, and Steering Chair for 1998, 1999, and 2000 International Conference on Information Fusion held in Las Vegas, USA, Sunnyvale, USA, and Paris, France, respectively; received 1995 Early Career Award for excellence in Research from University of New Orleans; given numerous seminars and short courses in U.S., Europe and Asia; won several outstanding paper awards; included in Marquis' Who's Who in America and Who's Who in Science and Engineering; and consulted for several companies. Dr. Li's current research interests include signal and data processing, target tracking and information fusion, stochastic systems, statistical inference, and electric power.