Multiple-Model Estimation with Variable Structure Part III: Model-Group Switching Algorithm

X. RONG LI, Senior Member, IEEE
XIAORONG ZHI
YOUMIN ZHANG
University of New Orleans

A general multiple-model (MM) estimator with a variable structure (VSMM), called model-group switching (MGS) algorithm, is presented. It assumes that the total set of models can be covered by a number of model groups, each representing a cluster of closely related system behavior patterns or structures, and a particular group is running at any given time determined by a hard decision. This algorithm is the first VSMM estimator that is generally applicable to a large class of problems with hybrid (continuous and discrete) uncertainties. It is also easily implementable. It is illustrated via a simple fault detection and identification example that the MGS algorithm provides a substantial reduction in computation while having identical performance with the fixed-structure Interacting Multiple-Model (FSIMM) estimator.

Manuscript received June 17, 1997; revised January 22 and August 19, 1998.

IEEE Log No. T-AES/35/1/01503.

This work was supported in part by ONR via Grant N00014-97-1-0570, NSF via Grants ECS-9496319 and ECS-9734285, and LEQSF via Grant (1996-99)-RD-A-32.

Authors' current addresses: X. R. Li, Dept. of Electrical Engineering, University of New Orleans, New Orleans, LA 70148, (E-mail: xli@uno.edu); Youmin Zhang is on leave from the Dept. of Automatic Control, Northwestern Polytechnical University, Xian 710072, P.R. China; his current address is the Dept. of Electrical and Computer Engineering, The University of Western Ontario, London, ON N6A 5B9, Canada, (E-mail: ymzhang@overlord.engga.uwo.ca); X. Zhi, Ultratech Stepper Inc., 3050 Zanker Rd., San Jose, CA 95035.

I. INTRODUCTION

Multiple-model (MM) estimation, as a powerful approach to adaptive estimation, has received a great deal of attention in recent years due to its unique power to handle problems with both structural and parametric uncertainties and/or changes, and to decompose a complex problem into simpler subproblems.

In the MM estimation, a set of models is designed to represent the possible system behavior patterns or structures, called system modes, and the overall estimate is obtained by a certain combination of the estimates from the filters running in parallel based on the individual models that match (or represent) particular system modes. This approach was initiated in [23]. The early work did not consider jumps in system modes and led to the nonswitching MM algorithms. In the more recent and more realistic switching MM estimators, first proposed in [1], the jumping of system modes is modeled by switching among models. Most recently, an exact discrete-time MM estimator was developed [10] with a finite but exponentially increasing dimension.

Most existing MM estimators have a fixed structure (FSMM) in the sense that they use a fixed set of models at all times. They have found great success in solving many state estimation problems compounded with structural or parametric uncertainty in many areas over the past three decades, in particular target tracking (see [4, 5, 14] and the long lists of references therein). Existing MM estimation techniques with a fixed structure have, however, arrived at such a stage that great improvement can no longer be expected within their fixed structure, especially for problems involving many structural modes.

The limitation of MM estimation with a fixed structure has been more or less perceived for some time. Ad hoc remedies were proposed for particular applications but few theoretical attempts were made to break away from the fixed structure. The investigation of the moving-bank MM estimators in [24] was an early meaningful effort to go beyond the fixed structure. A serious attempt was initiated in [17] and continued in [16, 18] to lay down a theoretical foundation for MM estimation without the limitation of the fixed structure. Specifically, MM estimation with a variable structure (VSMM) was proposed in [16–18] to overcome limitations of the FSMM estimators and to increase the cost-effectiveness, in contrast to the existing effort of developing better implementable FSMM estimators. Theoretical results fundamental to the MM estimation with a variable structure were also presented in [16, 18], which include the optimal VSMM estimator, a graph-theoretic formulation of MM estimators, and a criterion for model-set selection.

As manifested by the great impact of the monumental success of the Interacting Multiple-Model (IMM) estimator [6] on the application of MM estimation techniques, it is fair to say that no matter how promising VSMM estimation may appear, its ultimate success relies on the development of good practical VSMM algorithms that can be readily implemented and are general enough to be applicable to a large class of hybrid estimation problems (i.e., those with both continuous and discrete uncertainties) [14]. If succeeded, such development will be a new milestone in the history of MM estimation. A primary value of the theoretical work on the variable structure is to facilitate this development. Development and design of VSMM algorithms have received increasing attention recently due to their great practical significance. Toward this end, three practical VSMM schemes were outlined in Part I of this series [18] that are general enough to be applicable to many problems. Several ad hoc designs for VSMM estimation were reported in [11–13, 22, 24–25] for the particular applications considered.

In Part II of this series [15], fairly satisfactory theoretical results concerning several important problems in VSMM estimation were presented. They include an efficient and near optimal estimation cycle for VSMM estimation, an efficient and optimal fusion formula, and optimal sequential tests for model-set termination. All these results are essential for the development of the new VSMM algorithms presented here and in the subsequent parts.

This work presents a VSMM estimator, called model-group switching (MGS) algorithm, which is the first VSMM algorithm that is generally applicable to many hybrid estimation problems and is easily implementable. It belongs to one of the three VSMM schemes outlined in [16, 18]—the digraph switching scheme. Its development relies heavily on the theoretical results of Part II [5]. In this estimator, a set of model groups is first set up, each representing (or covering) a collection or cluster of closely related system modes (behavior patterns or structures). The set of models is made adaptive by switching among these groups to follow possible jumps (across groups) of the system mode in such a way that balances well between the needs to have the smallest delay in correct switching and to have a minimum false switching rate. Issues associated with such an approach, such as the initialization of the newly activated model group and the termination of a model group, are solved fairly satisfactorily. Simulation results presented in Part IV [21] for target tracking as well as in this paper for fault detection and identification indicate that the MGS algorithm is substantially more cost-effective than the fixed-structure IMM (FSIMM) estimator, especially when the total model-set is large.

The remaining parts of the paper are organized as follows. Section II describes briefly the general problem of VSMM estimation. The MGS estimation algorithm is presented in Section III. Section IV extends the MGS algorithm to overcome some of its potential weaknesses. A simple example of fault detection and identification is simulated in Section V to demonstrate the superiority of the MGS algorithm to the FSIMM algorithm. The last section provides concluding remarks.

II. VARIABLE-STRUCTURE MM ESTIMATION

Consider the following, one of the simplest stochastic hybrid systems, known as the "jump linear systems,"

$$x_{k+1} = F_k(s_{k+1})x_k + G_k(s_{k+1})w_k(s_{k+1}) \tag{1}$$

$$z_k = H_k(s_k)x_k + v_k(s_k) \tag{2}$$

where x is the (conventional) *base state* vector; z is the noisy measurement vector; s_k is the *modal state* or *system mode* (system mode index) at time k, which denotes the mode in effect during the sampling period ending at t_k; w and v are the mode-dependent process and measurement noise sequences, respectively. The system mode sequence $\langle s_k \rangle$ is assumed to be a first-order homogeneous Markov chain with the following transition probabilities

$$P\{m_{k+1}^j \mid m_k^i\} = \pi_{ij} \qquad \forall \quad m_i, m_j \in \mathbf{S} \tag{3}$$

where the event that mode m_j is in effect at time k is denoted as

$$m_k^j \triangleq \{s_k = m_j\}.$$

$\mathbf{S}$ is the set of all possible system modes.

In the MM approach, it is assumed that the nonlinear system (1)–(2) at any given time can be adequately approximated by one or (a combination of) more models in a set M consisting of the following conventional linear models:

$$x_{k+1} = F_k^j x_k + G_k^j w_k^j \qquad \forall \quad m_j \in M \tag{4}$$

$$z_k = H_k^j x_k + v_k^j \qquad \forall \quad m_j \in M \tag{5}$$

and a jump between the system modes can be modeled by switching from one model to another, governed by, say, the Markov law (3). Here superscript j denotes quantities pertaining to model m_j. It is thus clear that the MM approach fits well into problems that can be characterized by structural as well as parametric uncertainties and/or changes.

The state estimate and its associated covariance matrix may be calculated in an MM estimator using the minimum mean square error (MMSE) criterion as

follows:

$$\hat{x}_{k|k} = \sum_j \hat{x}^j_{k|k} P\{H^j_k \mid z^k\} \tag{6}$$

$$P_{k|k} = \sum_j [(\hat{x}_{k|k} - \hat{x}^j_{k|k})(\hat{x}_{k|k} - \hat{x}^j_{k|k})' + P^j_{k|k}] P\{H^j_k \mid z^k\} \tag{7}$$

where $\hat{x}^j_{k|k}$ is the optimal estimate at time k under the hypothesis $H^j_k \triangleq$ {model sequence j through time k matches the sequence of the true system modes through time k}, and $P^j_{k|k}$ is the associated covariance; $z^k = \langle z_\kappa \rangle_{\kappa \le k}$ is the measurement sequence through time k, with z_0 denoting the initial information. The summations in (6)–(7) of a full-hypothesis-tree (FHT) MM estimator are over all possible hypotheses (model sequences). Due to the exponential increase in computation and memory of the FHT estimator, one of the following approximate hypothesis management techniques is usually used in practice to limit the growth in the number of hypotheses: 1) elimination of the hypotheses unlikely to be true, which yields a partial-hypothesis-tree (PHT) algorithm, 2) merging of the hypotheses having a common history, which yields a merged-hypothesis-tree (MHT) algorithm, and 3) a combination of hypothesis elimination and merging, which yields a reduced-hypothesis-tree (RHT) algorithm.

Most existing MM algorithms have a fixed structure (FSMM) in the sense that the model-set M in (4)–(5) is assumed time invariant, even though the models themselves may be time varying or adaptive. A VSMM estimator is one with a time-varying set M of models.

The FSMM approach has certain inherent limitations, which stem from its fundamental assumption that the system mode at any time can be represented (with a sufficient accuracy) by one of a fixed set of models that can be determined before measurements are received in real time, and its inability to incorporate certain types of a priori information.

The FSMM estimators perform reasonably well for problems that can be handled with a small set of models. When they are applied to solve real-world problems, however, it is often the case that use of only a few models is not good enough. The computational complexity increases considerably as the number of models increases. More importantly, as shown in [16, 18], the use of more models does not necessarily improve the performance; in fact, the performance will deteriorate if too many models are used due to the excessive "competition" from the "unnecessary" (excess) models. Thus one may face a dilemma: more models have to be used to improve the accuracy, but the use of too many models will degrade the performance, let alone the increase in computation.

To find a way out of this dilemma, [17, 18] proposed the use of a variable structure in MM estimation. Not only can a VSMM estimator take advantage of the real-time system mode information contained in the measurement sequence, but it also is able to incorporate certain a priori information of the system mode that would be difficult or impossible for an FSMM estimator to include. An example of such a priori information is that the system mode is unlikely (but not impossible) to be in a certain subset of the total mode set (mode space) for the particular application under consideration. Another example is the a priori knowledge of the magnitude of jump in system mode.

As shown in Part I [18], a key difference between the optimal VSMM and FSMM estimators is that the former is a probabilistically weighted sum of all estimators based on *admissible mode-set* sequences,[1] rather than of all estimators based on *possible mode* sequences as in the latter. This optimal VSMM estimator provides a theoretical basis for the adaptation of the model-set. It also suggests the use of variable structures as a direction of improving performance, in contrast to the existing effort of developing better implementable FSMM estimators.

Although the probabilistically weighted sum of all estimators based on the admissible *mode-set* sequences, as required by the optimal VSMM estimator, is computationally infeasible, a practical VSMM estimator may take advantage of its suggested two-level hierarchical structure: multiple *model-set* sequences at the higher level and multiple *model* sequences at the lower level. For many applications, the higher level with multiple model-set sequences should be replaced, due to computational constraints, with a single model-set sequence (most likely along with a proper merging of model-set sequences) obtained by model-set adaptation, which is hopefully one of the best sequences. This is the recursive adaptive model-set (RAMS) approach.

The RAMS approach has two functional components: model-set adaptation and model-sequence conditioned estimation. Part II [15] deals with both components in a general setting. A fairly satisfactory solution to the problem of the model-sequence conditioned estimation was presented there. For the model-set adaptation, however, it seems that no solution may be found that is complete and generally applicable. Nevertheless, the theoretical results presented in [15] are important in that they provide at least general principles and guidelines. In fact, a major component of the algorithm proposed here is an application of some relevant theoretical results given there.

[1] A mode-set sequence is admissible if it is one that may contain the true mode sequence as its element.

Foreseeing the difficulty in obtaining a general solution to model-set adaptation, [18] proposed and outlined the following three *practical* VSMM schemes: active digraph (active model-set), digraph switching (model-set switching), and adaptive grid schemes.

The *active model-set scheme* is quite general and powerful. Its basic idea is to use a subset of the total model-set as the "active" set for any given time. A simple implementation is as follows. All models in the current active set can be classified into three categories: unlikely, significant, and principal. As such, a reasonable set of rules for model-set adaptation is: 1) discard the unlikely ones; 2) keep the significant ones; and 3) activate the models adjacent *from* the principal ones. A model is adjacent from another one if it is allowed to be switched from the latter model. By an appropriate design at any given time at most a couple of models may be principal, but the number of unlikely models could be very large. This indicates that the active model-set is usually much smaller than the total model-set and, therefore, it may lead to a substantial saving in computation and possibly improvement in performance. Part V [20] of this series presents the only algorithm available in this category. It was developed following the ideas given in [17, 18] and utilizing the theoretical results obtained in [15].

In the *adaptive grid scheme*, the space of the parameters that characterize the possible system modes is quantized (i.e., represented by a set of models) unevenly and adaptively. A coarse grid is set up initially and then the grid is adjusted recursively according to an adaptation scheme based possibly on the current estimates, model probabilities, model likelihoods, and/or measurement residuals. This approach is particularly advantageous in cases where the set of possible system modes is large. In this scheme, the total model-set need not (and usually cannot) be specified in advance. Designs of such algorithms were reported in [9, 11–13, 24]. While following the same idea of making grid adaptive, different adaptation rules were proposed in different designs for applications characterized by different parameter regions. While valuable, they have a common weakness—they are more or less *ad hoc* and are not generally applicable.

The simple and efficient *model-set switching scheme* is the topic of this work. The designs presented in [8, 22, 13] also belong to this class. In [22], a so-called selected filter IMM design was proposed. Simulation results showed that the proposed design with a substantial reduction in computation yields almost as good performance as that of the FSIMM algorithm for a maneuvering target tracking problem, similar to the one considered in Part IV [21]. In [13], a digraph-switching (as well as an adaptive-grid) IMM design for a maneuvering target tracking problem is developed based on coordinated turn models. It was shown that great improvement in performance as well as computation can be achieved by the proposed design relative to the FSIMM algorithm. All these designs are, however, *ad hoc* and thus valid only for the problems considered in that no general techniques were proposed for, e.g., when and how a model subset should be switched to another one and how new models should be initialized.

An *ad hoc* layered IMM architecture was proposed in [8] for maneuvering target tracking in the presence of glint noise under a fundamental assumption that model-sets are independent. It is theoretically equivalent to the FSIMM algorithm but may have a reduced computational complexity. This algorithm belongs to the soft switching of model subsets, to be discussed in the next section. A more general and effective layered IMM algorithm was developed in [19], which does not require the assumption that the layers are independent as made in [8] but is not really of a variable structure.

Compared with the above available results (some of which were developed simultaneously as the algorithm proposed here), the superiority and difference of our algorithm can be summarized as follows.

1) It is the first algorithm of variable structure that is generally applicable to a large class of hybrid estimation problems, rather than being ad hoc for special problems considered.

2) The switching between model subsets is done in two stages. A candidate subset deemed very probable is activated without forcing the termination of the one currently in effect. The algorithm then runs their union until sufficient information is gathered to terminate one of them. This two-stage switching provides a good strategy for completing the conflicting tasks of simultaneously minimizing the chance of false switching and the delay in correct switching. The only cost paid for this is an insignificant increase in computation.

3) The initialization of the newly activated model subset is done automatically using the variable structure IMM (VSIMM) recursion of Part II [15]. This recursion assigns the probabilities to the newly activated models and initializes the filters based on these models in a natural, systematic, and near optimal manner. It also eliminates the need to design the transition probabilities for each model-subset.

4) Model-subset termination is done based on a combination of the sequential mode-set probability ratio test and the sequential model-set likelihood ratio test of Part II, which are optimal in the sense of having the quickest decision subject to prespecified lower bounds on type I and type II error probabilities.

5) The incorporation of the optimal fusion rule of Part II makes the proposed algorithm computationally

even more efficient when using the estimates based on the newly activated models.

With these techniques that are generally applicable, what is proposed here is an algorithm, rather than ideas of such an algorithm, as was done in [17, 18], or ad hoc designs valid only for particular problems.

III. MODEL-GROUP SWITCHING MM ESTIMATOR

The model-set used for an MM estimator can be made adaptive by switching among a number of predetermined model groups according to certain rules. This is the basic idea of the model-group switching (MGS) scheme. In this scheme, a predetermined cover or partition of the total model-set is designed first. This cover or partition consists of model groups representing certain collections (clusters) of closely related system modes. Since a model "group" implies that its member models are (closely) connected in some sense, "model-group switching" is a more proper name for this scheme than "model-set (or model-subset) switching."[2] A partition consists of mutually exclusive (disjoint) and collectively exhaustive model groups (i.e., any model in the total set is in one and only one of these groups). A cover has its model groups collectively exhaustive but not necessarily disjoint (some models might belong to more than one group).

Two types of model-group switching are of major interest: *hard switching* and *soft switching*. They both have the following two-level hierarchical structure: a higher level consisting of model groups and a lower level consisting of models within a group.

The soft MGS scheme is based on the assumption that each of the model groups at any time has a certain probability of having a member model matching (closely) the system mode currently in effect and the overall estimate is the probabilistically weighted sum of the MM estimators based on these model groups. Assuming the sequence of the model group in effect is a Markov or semi-Markov process, it is possible to design the transition probability matrix for this process and to apply the FSMM estimator to it. This can be referred to as the *soft switching* of the model group, just like the soft model switching in a decision-free MM estimator [14]. In this scheme, the FSMM approach with soft decision is applied in two levels: model (lower) level and model group (higher) level. Since probabilistic weights for the model groups are used in this scheme, the model groups have to be disjoint and thus a partition (instead of a cover) has to be used. In the strict sense of variable structure, this scheme is not really one with MGS, or not even a VSMM algorithm since it is nothing but a probabilistically weighted sum of a *fixed* number of *fixed-structure* estimators, each based on a model group. In other words, this soft switching algorithm is a degenerated VSMM (or disguised fixed-structure) scheme. A proper name for it seems "*multiple model-group algorithm*" rather than "model-group soft switching algorithm." Note also that in order to use soft switching, all model groups have to be run at all times and thus its computational complexity is comparable with that of the corresponding FSMM estimator based on the total model-set.

[2]An even more appropriate name is "digraph switching," as described in [18], which is unlikely to become popular since it uses graph-theoretic concepts, unfamiliar for practitioners.

A. Hard Switching of Model Groups

In the *hard switching* of model groups, the switching is done according to a set of certain "hard" rules (i.e., by a hard decision). As such, only one model group has to be run at any time and thus it may provide a substantial saving in computation over the FSMM estimator based on the total model-set. The adjacent model groups should have common models since these models are valuable for the design of adaptation logics and the initialization of newly activated filters. It is thus clear that a hard switching MGS algorithm is in general based on a cover rather than a partition of the total model-set.

Following the idea proposed in Part II [15], it is advantageous to decompose model-set adaptation (switching) into two separate tasks: model-set *activation* and *termination*. As such, MGS is completed in two stages: activation and termination. The model group that is deemed likely to be a good candidate at the time is activated first without forcing the termination of the one currently in effect. The algorithm then runs their union until sufficient information is gathered to terminate one of them. The advantage of such a two-stage switching is clear: it can reduce the chance of false switching dramatically at the cost of an insignificant increase in computation, rather than a significant increase in the delay in correct switching, which would be the case for a single-stage switching algorithm. The use of model-group activation and termination, rather than a single-stage switching, is a key idea of the MGS algorithm. It becomes clear that part of the superiority of the proposed MGS algorithm comes from this decomposition.

General speaking, one cycle of an MGS algorithm consists of the following conceptual steps.

S1. *Model-group adaptation.* Decide whether a candidate model group should be activated. If it is activated, check if it should be maintained for the next time cycle. The model group used in the previous cycle should be retained for the current cycle but it may be terminated for the next cycle by a logic.

S2. *Initialization of newly activated models/filters.* If a candidate model group is activated, then assign proper initial probabilities to the newly activated

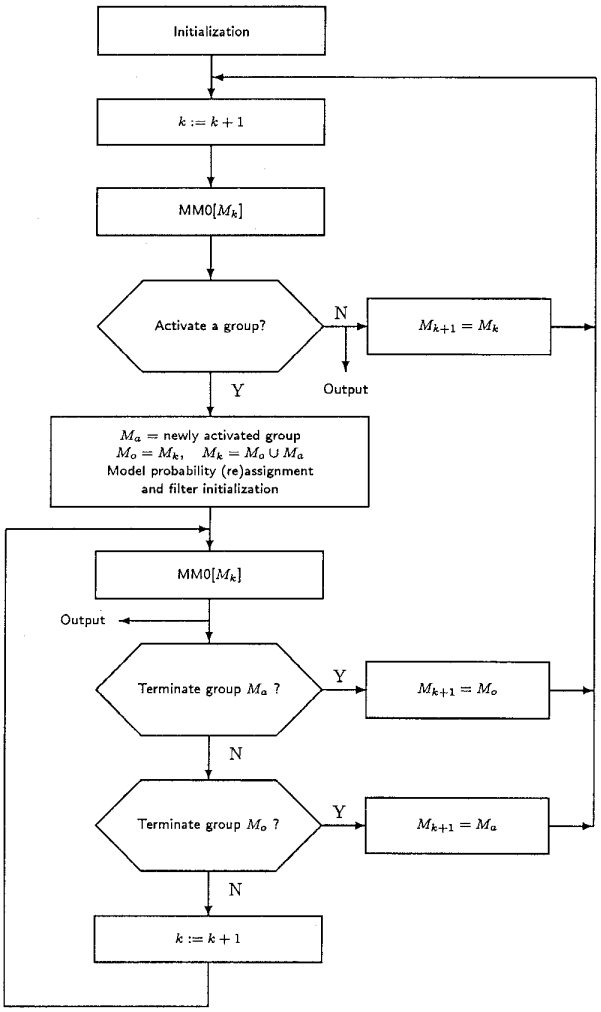


Fig. 1. Introductory flowchart of MGS algorithm.

models and determine the initial estimates and error covariances of the filters based on these models.

S3. *MM estimation.* Make MM estimate for the current cycle based on all the models in the selected group(s).

Step 3 is identical to a cycle of a recursive FSMM estimator. Steps 1 and 2 are unique to the MGS algorithm and thus are investigated in the sequel. Steps 2 and 3 are actually integrated in our MGS algorithm by using the VSIMM recursion of [15].

It seems virtually impossible to really derive rigorously any nonoptimal algorithm with a hard decision, although its properties and performance may be derived rigorously. Fig. 1 gives an introductory flowchart of the MGS algorithm proposed here.

In the flowchart, MM0 [M_k] stands for one cycle of an MM estimator of zero memory depth based on model group M_k in effect at time k. An MM estimator is of a zero memory depth if it calculates all its quantities in the current time cycle using only the quantities obtained in the most recent cycle as well as a priori information. Notable examples of the zero-memory-depth MM estimators include the IMM and GPB1 (the first-order generalized pseudo-Bayesian) algorithms. Even though MM0 is used in the flowchart, the algorithm actually works for other MM estimators as well (with minor changes).

Note that once a model group is activated, it is used starting right at that time, rather than from the next time on (i.e., there is no time delay here). This proves to be very important in reducing the peak estimation error during mode transitions. This is done in such a way that the computation of the old model group (M_k in the figure) is not wasted due to the use of an optimal fusion rule, to be presented in Subsection IIIE.

In this paper, M_k stands for either the model-set in effect at time k or the event $\{s_k \in M\}$ for some set M, which should be clear from the context; and the discrete time is denoted *exclusively* by either k or κ and conversely, k and κ are used *exclusively* for discrete time. For example, M_j stands for the jth model-set, rather than the model-set in effect at time j.

Each of the functional steps in the proposed MGS algorithm is discussed below.

B. Cover of Total Set

In the MGS algorithm, a cover is set up first. This is closely related to the so-called set covering problem, for which a solution can be obtained by solving an integer linear programming problem [2, 7]. The establishment of this cover is best integrated into the model-set design process. For VSMM estimation, however, such a cover can be obtained in most cases from the physical meaning of the system modes without too much difficulty. The resultant family of model groups then act as the range of the time-varying model group of the MGS algorithm. Each group in this family is made up of a number of models in the total set that are "close" to each other in, e.g., an information distance measure [3] (such as Kullback information) or some other measure of similarity as in cluster analysis in statistics. It appears that any effective closeness measure has to capture the essence of the following: two models should be deemed close if the state estimates from the filters based on them are "close" given the system mode in effect and the same measurement sequences (see Part II [15]). For many practical problems, the meaning of closeness is clear, as illustrated in Part IV.

C. Model-Group Adaptation

With the decomposition of model-group switching into activation and termination, the decision for model-group adaptation includes the following.

a) Decision for candidate model-group activation,

b) Decision for the termination of the newly activated candidate model group M_a,

c) Decision for the termination of the model group M_o currently in effect.

1) *Model-Group Activation*: In general, the decision (logic) for model-group activation should consist of a set of rules based on both a priori and a posteriori information about the current system mode in effect. The a priori mode information is incorporated mostly in the topology of the total model-set (more rigorously, digraph) and the corresponding transition probability matrix, even though they are also possible to be made time varying or adaptive. Use of heuristic rules or tricks can also be justified as utilizing the a priori information, which is problem dependent. The a posteriori mode information comes only from the measurement sequence and is most compactly summarized (but not in full) in the sequences of model probabilities and likelihoods (since a mode may actually be different from every model). As such, it can be expected that a more or less general logic should rely primarily on the model probabilities and/or likelihoods, based on the model-set topology and possibly with the transition probabilities taken into account.

Although some useful theoretical results were presented in Part II [15], the actual candidate group activation logic should be problem dependent. How to obtain it is a design task. Its design relies on the topology of the total model-set, the particular cover used, the physical meaning of the models, desired performance versus computational constraints, etc. This design should be integrated with the selection of the thresholds for the model-group termination. A detailed discussion of this logic is given in Part IV.

2) *Model-Group Termination*: Part II [15] proposed the sequential model-set likelihood ratio test and the sequential mode-set probability ratio test for model-subset termination and showed their optimality. Based on these results, the following tests are proposed for the termination of a candidate or a current model group, which are generally applicable, simple, and fairly satisfactory.

Candidate model-group termination: The candidate model group M_a is terminated for the next time $(k+1)$ while competing with the current model group M_0 if and only if *either* of the following two conditions (criteria) is satisfied

$$\text{probability condition:} \quad \frac{\mu_k^{M_a}}{\mu_k^{M_0}} < t_1^{\mu}$$

$$\text{likelihood condition:} \quad \prod_{\kappa=k_0}^{k} \frac{L_\kappa^{M_a}}{L_\kappa^{M_o}} < t_1^{L}$$

where k_0 is the time at which the candidate model group is activated. In the above, the *model-group probability* μ^{M_l} is the sum of all probabilities of the models in group M_l calculated based on $M_a \cup M_o$, given by

$$\mu_k^{M_l} \triangleq P\{M_k^l \mid M_k^{a+o}, z^k\} = \sum_{m_n \in M_l} P\{m_k^n \mid M_k^{a+o}, z^k\},$$

$$l = a \quad \text{or} \quad o$$

where M_k^l (or M_k^{a+o}, respectively) is the event that one of the models in M_l (or $M_a \cup M_o$, respectively) matches the system mode in effect at k:

$$M_k^l = \{s_k \in M_l\}$$
$$M_k^{a+o} = \{s_k \in (M_a \cup M_o)\}.$$

The (marginal) *model-group likelihood* L^{M_l} at time κ is the probabilistically weighted sum of all (marginal) likelihoods at κ of the models in group M_l, given by, for $l = a$ or o,

$$\begin{aligned} L_\kappa^{M_l} &\triangleq p[z_\kappa \mid M_\kappa^l, z^{\kappa-1}] \\ &= \sum_{m_n \in M_l} p[z_\kappa \mid m_\kappa^n, M_\kappa^l, z^{\kappa-1}] P\{m_\kappa^n \mid M_\kappa^l, z^{\kappa-1}\} \\ &= \sum_{m_n \in M_l} p[z_\kappa \mid m_\kappa^n, z^{\kappa-1}] \frac{P\{m_\kappa^n, M_\kappa^l \mid M_\kappa^{a+o}, z^{\kappa-1}\}}{P\{M_\kappa^l \mid M_\kappa^{a+o}, z^{\kappa-1}\}} \\ &= \sum_{m_n \in M_l} L_\kappa^n \hat{\mu}_{\kappa|\kappa-1}^{n|M_l} \end{aligned}$$

where L_κ^n is the likelihood of model m_n at time κ, defined by

$$L_\kappa^n = p[z_\kappa \mid m_\kappa^n, z^{\kappa-1}]$$

$\hat{\mu}_{k|k-1}^{n|M_l}$ is the predicted probability of model m_n at time k based on group M_l:

$$\begin{aligned} \hat{\mu}_{k|k-1}^{n|M_l} &\triangleq P\{m_k^n \mid M_k^l, z^{k-1}\} \\ &= \frac{1}{\hat{\mu}_{k|k-1}^{M_l}} \hat{\mu}_{k|k-1}^n, \qquad \forall \quad m_n \in M_l \end{aligned}$$

$\hat{\mu}_{k|k-1}^n$ is the predicted probability of model m_n at time k based on the total model-set $\mathbf{M}$:

$$\begin{aligned} \hat{\mu}_{k|k-1}^n &\triangleq P\{m_k^n \mid z^{k-1}\} = P\{m_k^n \mid \mathbf{M}_k, z^{k-1}\} \\ &= \sum_{m_m \in M_{k-1}} \pi_{mn} \mu_{k-1}^m \end{aligned}$$

and $\hat{\mu}_{k|k-1}^{M_l}$ is a normalization factor for $\hat{\mu}_{k|k-1}^{n|M_l}$ within M_l, given explicitly by

$$\hat{\mu}_{k|k-1}^{M_l} \triangleq P\{M_k^l \mid \mathbf{M}_k, z^{k-1}\} = \sum_{m_n \in M_l} \hat{\mu}_{k|k-1}^n.$$

Note that if the IMM algorithm is used as the MM0 cycle, then $\hat{\mu}_{k|k-1}^n$ has been obtained in the IMM cycle. See Part II [15] for details.

The thresholds t_1^{μ} and t_1^{L} are design parameters, which in most cases should be chosen (slightly) smaller than unity. They may be chosen to be equal in many cases.

Current model-group termination: The current model-group M_o is terminated for the next time $(k+1)$ while competing with the candidate model-group M_a if and only if *both* of the following two conditions (criteria) are satisfied

$$\text{probability condition:} \quad \frac{\mu_k^{M_a}}{\mu_k^{M_o}} > t_2^{\mu}$$

$$\text{likelihood condition:} \quad \prod_{\kappa=k_0}^{k} \frac{L_\kappa^{M_a}}{L_\kappa^{M_o}} > t_2^{L}.$$

The thresholds t_2^{μ} and t_2^{L} are design parameters, which in most cases should be chosen (slightly) greater than unity. They may be chosen to be equal in many cases.

Based on theoretical results of Part II [15], the two separate tasks of termination of the candidate and current model groups can be combined by using the Sequential Model-Set Likelihood Ratio Test and/or the Sequential Mode-Set Probability Ratio Test, which is optimal in the sense of most efficient use of the information in the observations with guaranteed expected type I and type II error (false alarm and miss) probabilities.

The rationale for the use of "*and*" logic (for the probability and likelihood) in the current group termination and "*or*" logic in the candidate group is because the model-group probabilities and likelihoods are calculated more accurately for the current group than for the candidate group in the MGS algorithm since the candidate group is newly initialized.

In the above tests, the probability ratio $\mu_k^{M_a}/\mu_k^{M_o}$ and the likelihood ratio $\prod_{\kappa=k_0}^{k} L_\kappa^{M_a}/L_\kappa^{M_o}$ may be replaced by the following probability and likelihood ratios of the model groups with the corresponding *core models* of their competing group deleted:

$$\frac{\overline{\mu}_k^{M_a}}{\overline{\mu}_k^{M_o}} = \frac{P\{s_k \in (M_a - C_o) \mid M_k^{a+o}, z^k\}}{P\{s_k \in (M_o - C_a) \mid M_k^{a+o}, z^k\}} \tag{8}$$

$$\prod_{\kappa=k_0}^{k} \frac{\overline{L}_\kappa^{M_a}}{\overline{L}_\kappa^{M_o}} = \prod_{\kappa=k_0}^{k} \frac{p[z_\kappa \mid s_\kappa \in (M_a - C_o), z^{\kappa-1}]}{p[n_k \mid s_\kappa \in (M_o - C_a), z^{\kappa-1}]} \tag{9}$$

where C_o and C_a are the core or kernel model(s) of groups M_o and M_a, respectively. Core models of a group are those that are central and most essential of the group. For example, the core model of the group M_3^1, M_3^2, or M_3^3 of Fig. 3 of Part IV [21] is model m_3.

With such a replacement, the performance and/or computation of the model-group termination should be improved more or less in most cases at the price of extra design effort.

D. Initialization of a Model Group

Initialization of the newly activated filters consists of the following.

Assign the probabilities, at one time cycle before, to the newly activated models,

Determine the state estimates and the associated error covariances for these filters, at one time cycle before.

The concept of state-dependent system mode set, established in Part I [18], is a powerful one, particularly useful for filter initialization. It implies that in plain terms, given the current system mode, the set of possible system modes at the next time is a subset of the total mode-set, which is determined by the Markovian mode transition law (i.e., the graph-theoretic adjacence of the mode sets). As applied to filter initialization, the assignment of the initial probability to a model should account only for the probabilities of those models from which the former model is allowed to be switched; and the initial state estimate of a filter based on a model should be determined only from the estimates (and the probabilities) of those filters based on the models from which the former model is allowed to be switched.

Specifically, if m_n is a *newly* activated model at time k, then its probability of being true (more precisely, matching to the system mode in effect) at $k-1$ is (has to be) zero:

$$\mu_{k-1}^n = P\{m_{k-1}^n \mid M^{k-1}, z^{k-1}\} = 0, \qquad \forall \quad m_n \in M_n$$

where M_n is the set of new and only new models; $M^k = \langle M_\kappa \rangle_{\kappa \le k}$ is the sequence of the model-sets through time k. The expected probability of m_n being true at k without information in the observation at k (i.e., the predicted mode probability) is given by

$$\begin{aligned}\hat{\mu}_{k|k-1}^n &= P\{m_k^n \mid M^{k-1}, z^{k-1}\} \\ &= \sum_{m_j \in E_n} P\{m_k^n \mid m_{k-1}^j\} P\{m_{k-1}^j \mid M^{k-1}, z^{k-1}\} \\ &= \sum_{m_j \in E_n} \pi_{jn} \mu_{k-1}^j\end{aligned}$$

where E_n is the set of models in M_{k-1} that are allowed to switch to m_n:

$$E_n = \{m_l : m_l \in M_{k-1},\ \pi_{ln} \neq 0\}. \tag{10}$$

The initial probabilities of all the models in the current group M_o should be kept unchanged.

The previous estimates and associated covariances of the filters based on models in the current group M_o should be kept unchanged:

$$\hat{x}_{k-1|k-1}^{l,a} = \hat{x}_{k-1|k-1}^{l,b}, \qquad \forall \quad m_l \in M_o$$

TABLE I
VSIMM $[M_k, M_{k-1}]$ Recursion for RAMS Estimation with Zero Memory Depth

1. Model-conditional initialization and reinitialization $[\forall m_j \in M_k]$:

predicted mode probability:

$$\hat{\mu}^j_{k|k-1} \triangleq P\{m^j_k | M_k, M_{k-1}, z^{k-1}\} = \sum_{m_i \in M_{k-1}} \pi_{ij}\mu^i_{k-1}$$

mixing weight: $\mu^{i|j} \triangleq P\{m^i_{k-1} | m^j_k, M_{k-1}, z^{k-1}\} = \pi_{ij}\mu^i_{k-1}/\hat{\mu}^j_{k|k-1}$

mixing estimate: $\hat{x}^{0j} \triangleq E[x_{k-1} | m^j_k, M_{k-1}, z^{k-1}] = \sum_{m_i \in M_{k-1}} \hat{x}^i_{k-1|k-1}\mu^{i|j}$

mixing covariance:

$$P^{0j} \triangleq E\left[(x_k - \hat{x}^{0j})(x_k - \hat{x}^{0j})' | m^j_k, M_{k-1}, z^{k-1}\right] = \sum_{m_i \in M_{k-1}} \left[P^i_{k-1|k-1} + (\hat{x}^{0j} - \hat{x}^i_{k-1|k-1})(\hat{x}^{0j} - \hat{x}^i_{k-1|k-1})'\right]\mu^{i|j}$$

2. Model-conditional filtering $[\forall m_j \in M_k]$:

predicted state:

$$\bar{x}^j \triangleq E[x_k | m^j_k, M_{k-1}, z^{k-1}] = F^j_{k-1}\hat{x}^{0j} + G^j_{k-1}\bar{w}^j_{k-1}$$

predicted covariance:

$$\bar{P}^j \triangleq E\left[(x_k - \bar{x}^j)(x_k - \bar{x}^j)' | m^j_k, M_{k-1}, z^{k-1}\right] = F^j_{k-1}P^{0j}(F^j_{k-1})' + G^j_{k-1}Q^j_{k-1}(G^j_{k-1})'$$

measurement residual: $\tilde{z}^j \triangleq z_k - E[z_k | m^j_k, M_{k-1}, z^{k-1}] = z_k - H^j_k\bar{x}^j - \bar{v}^j_k$

residual covariance: $S^j \triangleq \text{cov}[\tilde{z}^j | m^j_k, M_{k-1}, z^{k-1}] = H^j_k\bar{P}^j(H^j_k)' + R^j_k$

filter gain: $K^j = \bar{P}^j(H^j_k)'(S^j)^{-1}$

updated state: $\hat{x}^j_{k|k} \triangleq E[x_k | m^j_k, M_{k-1}, z^k] = \bar{x}^j + K^j\tilde{z}^j$

updated covariance:

$$P^j_{k|k} \triangleq E\left[(x_k - \hat{x}^j_{k|k})(x_k - \hat{x}^j_{k|k})' | m^j_k, M_{k-1}, z^k\right] = \bar{P}^j - K^jS^j(K^j)'$$

3. Mode probability update $[\forall m_j \in M_k]$:

likelihood function: $L^j \triangleq p[\tilde{z} | m^j_k, M_{k-1}, z^{k-1}] \overset{\text{assume}}{=} \mathcal{N}[\tilde{z}^j; 0, S^j]$

mode probability: $\mu^j_k \triangleq P\{m^j_k | M_k, M_{k-1}, z^k\} = \frac{\hat{\mu}^j_{k|k-1}L^j}{\sum_{m_i \in M_k}\hat{\mu}^i_{k|k-1}L^i}$

4. Combination:

overall estimate: $\hat{x}_{k|k} \triangleq E[x_k | M_k, M_{k-1}, z^k] = \sum_{m_j \in M_k} \hat{x}^j_{k|k}\mu^j_k$

overall covariance:

$$P_{k|k} \triangleq E\left[(x_k - \hat{x}_{k|k})(x_k - \hat{x}_{k|k})' | M_k, M_{k-1}, z^k\right] = \sum_{m_j \in M_k} \left[P^j_{k|k} + (\hat{x}_{k|k} - \hat{x}^j_{k|k})(\hat{x}_{k|k} - \hat{x}^j_{k|k})'\right]\mu^j_k$$

$$P^{l,a}_{k-1|k-1} = P^{l,b}_{k-1|k-1}, \qquad \forall \quad m_l \in M_0$$

where superscripts b and a denote quantities before and after the reinitialization, respectively.

The initialization of every filter based on a newly activated model in the candidate group should be done as follows. In general, when calculating $E[x_k \mid m^n_k]$, only the previous estimates in the set $\{\hat{x}^l_{k-1|k-1}, m_l \in E_n\}$ should be (and need be) used, where E_n is the set of models in M_{k-1} that are allowed to switch to m_n, given by (10). Specifically, the initial estimate for the filter based on model m_n at time k can be obtained by, similarly to the mixing step in the IMM estimator,

$$\begin{aligned}
\hat{x}^{0n}_{k-1|k-1} &= E[x_{k-1} \mid m^n_k, M^{k-1}] \\
&= E[E[x_{k-1} \mid m_{k-1}, m^n_k, M^{k-2}] \mid m^n_k, M^{k-1}] \\
&= \sum_{m_l \in E_n} E[x_{k-1} \mid m^l_{k-1}, M^{k-2}]P\{m^l_{k-1} \mid m^n_k, M^{k-1}\} \\
&= \sum_{m_l \in E_n} \hat{x}^l_{k-1|k-1}\mu^{l|n}_{k-1|k-1}
\end{aligned}$$

where conditioning on z^{k-1} has been dropped for simplicity; the mixing weight is given by

$$\mu^{l|n}_{k-1|k-1} = P\{m^l_{k-1} \mid m^n_k, M^{k-1}\} = \frac{\pi_{ln}\mu^l_{k-1}}{\sum_{m_i \in E_n}\pi_{in}\mu^l_{k-1}}.$$

The associated error covariance can be calculated accordingly. In simpler words, if model m_n can only be switched from model m_l in M_{k-1}, then $\hat{x}^l_{k-1|k-1}$ and $P^l_{k-1|k-1}$ should clearly be used as the initial estimate and covariance for the filter based on m_n at time k. If m_n can be switched from several models in M_{k-1}, then the probabilistically weighted sum of the estimates at $k-1$ from the filters based on these models should be used as the initial estimate for the filter based on m_n at k and the initial covariance can be determined accordingly.

Incorporation of group initialization into estimation cycle: If the VSIMM recursion of [15] is used for the MM0 cycle, then the initialization discussed above can actually be omitted because it is incorporated by the VSIMM recursion automatically. For the completeness of the MGS algorithm, the VSIMM is given in Table I, where $\mathcal{N}(y; \bar{y}, P)$ stands for the

(multivariate) Gaussian (normal) density of y with mean $\bar{y}$ and covariance P. Due to space limitation, the reader is referred to Part II [5] for details.

For some practical problems, to avoid a high false model-group activation rate so as to acheive good accuracy and maintain a low computational burden, it may be inevitable to have a delay in activating the correct model group. This will introduce errors in state estimate (and the associated covariance) and mode identification, for which little can be done other than to design a good activation mechanism. An indirect consequence of this delay is that the initial conditions used for the newly activated models (and possibly some other models) may not be accurate enough, which will introduce additional errors. Three techniques may be used to reduce these latter errors. A generally good and systematic technique is to go back several steps in time to initialize the newly activated models and their filters. Specifically, if a model is activated at k, then n-step back means that it has zero probability at $k-n-1$ but in general non-zero probability at $k-n$. Thus, we may run VSIMM $[M_{k-n}, M_{k-n-1}]$, VSIMM $[M_{k-n+1}, M_{k-n}], \ldots,$ VSIMM$[M_k, M_{k-1}]$ to obtain the current estimates. Note, however, that old overall estimate (and anything else that has already been sent out) may not be overridden unless the problem at hand is smoothing, instead of filtering. Our design used one-step back.

A theoretically more appealing technique is to replace the VSIMM $[M_k, M_{k-1}]$ of zero memory depth with a generic VSMM $[M_k, M_{k-1}, \ldots, M_{k-n-1}]$ cycle of n memory depth for such an initialization. This, however, requires the implementation of such a VSMM cycle.

A less systematic technique is to increase all the (re)initialized error covariances $P^n_{k-1|k-1}$, whenever a model group is activated, to some degree (e.g., by using a fudge or forgetting factor) to account for the extra uncertainties of the estimates of the filters based on the current group caused by possible delay in activating the correct model group. For example, a time-varying forgetting factor λ_k may be used as follows:

$$P^j_{k-1|k-1} := P^j_{k-1|k-1}/\lambda_{k-1}, \qquad m_j \in (M_k \cap \overline{M}_{k-1})$$

where $\overline{M}_{k-1}$ is the complement of M_{k-1} and λ_k may be given by

$$\lambda_k = 1 - \lambda^o + \lambda^o \lambda_{k-1}$$

and $0 < \lambda_0, \lambda^o < 1$ are design parameters that control the initial and variation rate of the forgetting factor. Note that this technique does not alter the estimates directly. However, no such technique was used in our design of the MGS algorithm for the results presented here or in Part IV.

E. Fusion of Estimates Based on Current and Newly Activated Models

In general, the current and newly activated group may have models in common. When this is indeed the case, the MGS algorithm as outlined in Fig. 1 has some waste in computation because the estimates based on these common models as calculated in the second MM0 $[M_k]$ should not differ from those calculated in the first MM0 $[M_k]$. But how to combine the estimates from two model-sets to yield the overall estimate? The following theorem on optimal fusion provides a satisfactory answer.

THEOREM 1. (Optimal Fusion, from Part II [15]) *Given two MM estimators based on two model-sets M_1 and M_2 at time k, respectively:*

$$\{\hat{x}^{i|M_1}_{k|k}, P^{i|M_1}_{k|k}, L^{i|M_1}_k, \hat{\mu}^{i|M_1}_{k|k-1}\}_{m_i \in M_1}$$

$$\{\hat{x}^{j|M_2}_{k|k}, P^{j|M_2}_{k|k}, L^{j|M_2}_k, \hat{\mu}^{j|M_2}_{k|k-1}\}_{m_j \in M_2}$$

where $L^{n|M_i}_k$ and $\hat{\mu}^{n|M_i}_{k|k-1}$ are model likelihood and predicted mode probability, respectively, of model m_n in model-set M_i. The optimal MM estimator based on M^{k-1} and $M_k = M = f[M_1, M_2]$, where f is some set operation (i.e., union, intersection, or difference), is given by

$$\hat{x}^M_{k|k} = \sum_{m_i \in M} \hat{x}^{i|M}_{k|k} \mu^{i|M}_k \tag{11}$$

$$P^M_{k|k} = \sum_{m_i \in M} [P^{i|M}_{k|k} + (\hat{x}^{i|M}_{k|k} - \hat{x}^M_{k|k})(\hat{x}^{i|M}_{k|k} - \hat{x}^M_{k|k})'] \mu^{i|M}_k \tag{12}$$

where

$$\{\hat{x}^{i|M}_{k|k}, P^{i|M}_{k|k}\}_{m_i \in M} = f[\{\hat{x}^{i|M_1}_{k|k}, P^{i|M_1}_{k|k}\}_{m_i \in M_1}, \{\hat{x}^{i|M_2}_{k|k}, P^{i|M_2}_{k|k}\}_{m_i \in M_2}] \tag{13}$$

$$\mu^{i|M}_k \triangleq P\{m^i_k \mid M_k, M^{k-1}, z^k\} = \frac{1}{c} L^i_k \hat{\mu}^i_{k|k-1}, \qquad \forall \quad m_i \in M \tag{14}$$

where c is a normalization factor given by

$$c = p[z_k, M_k \mid M^{k-1}, z^{k-1}] = \sum_{m_i \in M_k} L^i_k \hat{\mu}^i_{k|k-1}.$$

See [15] for details. In our case, set union is of interest, that is, $f[M_1, M_2] = M_1 \cup M_2$. Thus (13) becomes

$$\hat{x}^M_{k|k} = \sum_{m_i \in M} \hat{x}^{i|M}_{k|k} \mu^{i|M}_k. \tag{15}$$

Note that if $m_i \in (M_1 \cap M_2)$, then, under the stated assumptions,

$$\hat{x}_{k|k}^{i|M_1} = \hat{x}_{k|k}^{i|M_2}$$
$$P_{k|k}^{i|M_1} = P_{k|k}^{i|M_2}$$
$$L_k^{i|M_1} = L_k^{i|M_2}$$
$$\hat{\mu}_{k|k-1}^{i|M_1} = \hat{\mu}_{k|k-1}^{i|M_2}.$$

Following this theorem, the above-mentioned computational waste can be eliminated as follows.

1) Replace the second MM0 $[M_k]$ in Fig. 1 by MM0 $[M_n]$, where $M_n = M_a - M_o$ is the set of new and only new models.

2) Use the optimal fusion of Theorem 1 to obtain the overall estimate from $M_a \cup M_o = M_o \cup M_n$.

F. Initialization of MGS Algorithm

How the MGS algorithm should be initialized depends on what a priori information is available concerning the initial system mode. If a priori information indicates that the initial system mode is likely to be in a certain subset of the total mode-set, then the MGS algorithm should start from the corresponding model group. In general, one of the following procedures may be used for initialization if no a priori information for the initial system mode is available.

1) Run an FSMM estimator using all models for a few cycles and then choose the model group which has the highest group probability as the initial one.

2) Run a set of the MGS algorithms using the same total model-set but different initial model groups for a few cycles and then choose the one with the highest group likelihood as the initial model group.

3) Choose as the initial one the "nominal" model group that represents the nominal system mode(s) with a mechanism to switch to other nonnominal groups. A good example of such nominal model groups is M_1, defined by (5) of Part IV [21].

4) Calculate all model likelihoods and choose the set having the largest set likelihood as the initial set. This is equivalent to assuming all models have an equal prior probability, which differs from the assumption that all groups have an equal probability.

G. MGS Algorithm

Summarizing the results of this section and assuming that an IMM cycle is used as the MM0 estimation cycle, the flowchart of the MGS algorithm is given in Fig. 2 and one cycle of the proposed MGS algorithm is listed in Table II, where VSIMM cycle was given in Table I. Note the following.

TABLE II
One Cycle of MGS Algorithm

S1. Increase the time counter k by 1. Run the VSIMM $[M_k, M_{k-1}]$ cycle.

S2. Check if a candidate model-group is activated. If no model group is activated, then output $\hat{x}_{k|k}$, $P_{k|k}$, and $\{\mu_k^i\}_{m_i \in M_k}$ obtained from the VSIMM $[M_k, M_{k-1}]$ cycle. Let $M_{k+1} = M_k$ and go to Step 1.

S3. If a model group M_a is activated, then let $k_0 = k$, $M_o = M_k$ and

- Run the VSIMM $[M_n, M_{k-1}]$ cycle, where $M_n = M_a - M_o$ is the set of new and only new models.
- Let $M_k = M_n \cup M_o = M_a \cup M_o$.
- Fusion: Calculate the estimates, error covariances, and mode probabilities for the union set M_k:

$$\mu_k^i = \frac{L_k^i \hat{\mu}_{k|k-1}^i}{\sum_{m_i \in M_k} L_k^i \hat{\mu}_{k|k-1}^i}, \qquad \forall \ m_i \in M_k \tag{16}$$

$$\hat{x}_{k|k} = \sum_{m_i \in M_k} \hat{x}_{k|k}^i \mu_k^i \tag{17}$$

$$P_{k|k} = \sum_{m_i \in M_k} [(\hat{x}_{k|k}^i - \hat{x}_{k|k})(\hat{x}_{k|k}^i - \hat{x}_{k|k}) + P_{k|k}^i + ']\mu_k^i \tag{18}$$

where the estimates $\{\hat{x}_{k|k}^i\}$, error covariances $\{P_{k|k}^i\}$, likelihoods $\{L_k^i\}$, and predicted probabilities $\{\hat{\mu}_{k|k-1}^i\}$ were obtained in the above VSIMM $[M_k, M_{k-1}]$ and VSIMM $[M_n, M_{k-1}]$ cycles.

S4. Output $\hat{x}_{k|k}$, $P_{k|k}$, and $\{\mu_k^i\}_{m_i \in M_k}$.

S5. For model groups $M_l = M_o$, M_a, compute

$$\mu_k^{M_l} = \sum_{m_i \in M_l} \mu_k^i \tag{19}$$

$$\hat{\mu}_{k|k-1}^{M_l} = \sum_{m_i \in M_l} \hat{\mu}_{k|k-1}^i \tag{20}$$

$$L_k^{M_l} = \frac{1}{\hat{\mu}_{k|k-1}^{M_l}} \sum_{m_i \in M_l} L_k^i \hat{\mu}_{k|k-1}^i \tag{21}$$

If

$$\frac{\mu_k^{M_a}}{\mu_k^{M_o}} < t_1^{\mu} \quad \text{or} \quad \prod_{\kappa=k_0}^{k} \frac{L_\kappa^{M_a}}{L_\kappa^{M_o}} < t_1^L \tag{22}$$

then terminate group M_a: Let $M_{k+1} = M_0$ and go to Step 1.

If

$$\frac{\mu_k^{M_a}}{\mu_k^{M_o}} > t_2^{\mu} \quad \text{and} \quad \prod_{\kappa=k_0}^{k} \frac{L_\kappa^{M_a}}{L_\kappa^{M_o}} > t_2^L \tag{23}$$

then terminate group M_o: Let $M_{k+1} = M_a$ and go to Step 1.

S6. Increase the time counter k by 1 and let $M_{k+1} = M_k$. Run the VSIMM $[M_k, M_{k-1}]$ cycle. Go to Step 4.

1) The use of VSIMM cycle eliminates the need to initialize the newly activated models and filters.

2) Equations (16)–(18) are based on the optimal fusion of Subsection IIIE.

3) Equations (19)–(23) are based on the sequential model-set likelihood ratio test and the sequential

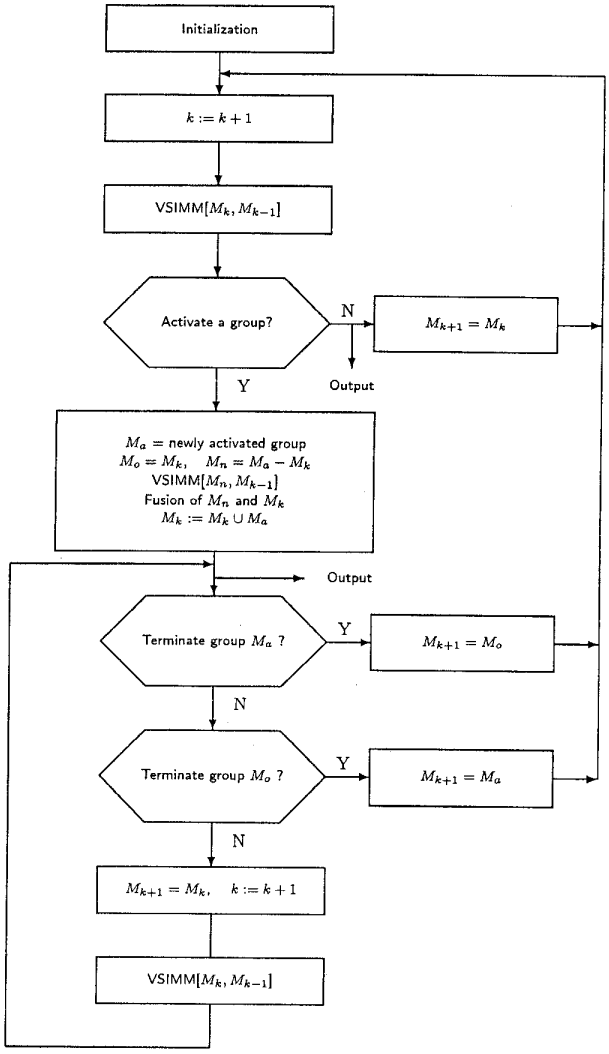


Fig. 2. Flowchart of MGS algorithm.

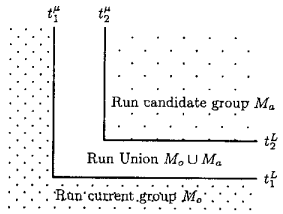


Fig. 3. Graphic illustration of model-group termination logic.

mode-set probability ratio test of Subsection IIIC2. Fig. 3 illustrates the sequential test of (22)–(23).

IV. EXTENDED MODEL-GROUP SWITCHING ALGORITHM

The MGS algorithm of Table II has the following potential weakness. At most, one model group may be activated at any time and no model group may be activated while running the union of two model groups. This may be a problem if one of the following situations occurs.

1) The system may stay in a mode with a very brief sojourn time before it jumps to another mode that is outside the coverage of the above-mentioned union. Note, however, that such a jump usually has to have a large jumping distance in the mode space and should usually be avoided by model-group design.

2) The MGS algorithm may activate a wrong model group. An incorrect model group usually will be terminated right away and then the correct group may be activated. Nevertheless, this would introduce an undesirable delay in correct activation, which is the most serious potential weakness of the MGS algorithm.

3) Sometimes there is a need to activate more than one model group simultaneously. For example, this would be the case if the system mode jumps to a place that is in the intersection of two or more candidate model groups.

To enhance the performance of the MGS algorithm under such situations, an extended version, called extended MGS (EMGS) algorithm, is proposed. It allows the activation of one or more candidate model groups even while running a union of model groups. The flowchart of this extended MGS algorithm is given in Fig. 4.

In Fig. 4, $\mathcal{M}_a$ stands for the set of newly activated one or more model groups; M_n is the set of new and only new models; $\mathcal{M}_k$ is the set of model groups in effect at k; $\mathcal{M}_t$ is the set of newly terminated one or more model groups, which is a subset of $\mathcal{M}_k$.

Note that the ability of the EMGS algorithm to activate more than one model group may reduce the delay in activating the right group, which is the most serious potential weakness of the MGS algorithm. For example, this may be done by using a significantly lower activation threshold that would be inappropriate for the MGS algorithm since it would have increased the risk of activating a wrong group significantly. How and when more than one model group should be activated is in general application dependent.

Termination of a set of model groups in $\mathcal{M}_k$ may be done simply by the sequential ranking tests proposed in Part II [15].

S1. At any time k, rank and order the members of $\mathcal{M}_k$ such that the set of their model-group probabilities $\mu_k^{M_i}$, $\forall M_i \in \mathcal{M}_k$, is in a decreasing order:

$$\mu_k^{M_1} \geq \mu_k^{M_2} \geq \cdots$$

S2. Terminate those model groups whose ratios of probabilities to the highest one are below a certain threshold t^μ; that is, for any $i \neq 1$,

$$\frac{\mu_k^{M_i}}{\mu_k^{M_1}} < t^\mu \Rightarrow \text{terminate } M_i$$

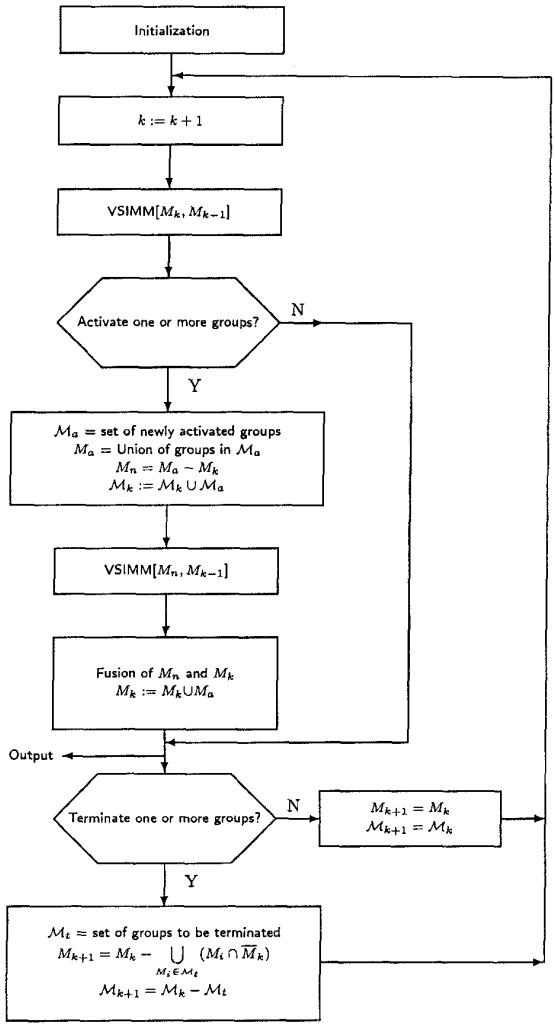


Fig. 4. Flowchart of EMGS algorithm.

Alternatively, the following may be done based on the theoretical results of Part II.

S1. At any time k, rank and order the members of $\mathcal{M}_k$ such that the set of their model-group joint likelihoods $\prod_{\kappa=k_0}^{k} L_\kappa^{M_i}$, $\forall M_i \in \mathcal{M}_k$ is in a decreasing order, where k_0 is the activation time of the most recently activated group.

S2. Terminate those model groups whose ratios of joint likelihoods to the highest one are below a certain threshold t^L; that is, for any $i \neq 1$,

$$\frac{\prod_{\kappa=k_0}^{k} L_\kappa^{M_i}}{\prod_{\kappa=k_0}^{k} L_\kappa^{M_1}} < t^L \Rightarrow \text{terminate } M_i.$$

There are two reasons for the use of k_0 as the starting time for the joint likelihoods: 1) Unlike probabilities, to compare joint likelihoods, they should have the same number of terms in the product (i.e., they should be over the same time period), and 2) an activation is a strong indication that mode may have jumped and thus old (marginal) likelihoods are better ignored. In view of this second reason, it seems that the second procedure may outperform the first one because the model-group probabilities in general carry (no matter how little) information of the system mode in effect before k_0, which could be misleading. However, if the activation has a delay (i.e., if k_0 is not the first time at which a mode change is reflected in the observation), then the use of the mode-set probabilities may be better than the likelihoods.

It is also possible to terminate the groups based on a combination of these two procedures, such as done in Subsection IIIC2.

In the above, t^μ and t^L are design parameters, which control the probabilities of making termination errors (and indirectly, the number of model groups in effect). They should be chosen according to the available computation resource and the amount of overlap among model groups.

Based on the above, one cycle of the extended MGS algorithm is given in Table III.

The advantage of the EMGS algorithm over the basic MGS algorithm is achieved at the cost of a slightly heavier computational burden and complexity of the algorithm.

Note that even when the activation logic activates at most one model group and the termination thresholds are such that at most two model groups are not terminated, the EMGS algorithm is still better than the MGS algorithm because activation is allowed while two model groups are in effect. The EMGS

TABLE III
One Cycle of EMGS

S1.	Increase the time counter k by 1. Run the VSIMM $[M_k, M_{k-1}]$ cycle.
S2.	Check if one or more candidate model groups are activated. If no model group is activated, then go to Step 4.
S3.	If one or more model groups are activated, then let their union be M_a. Let $k_0 = k$, $M_n = M_a - M_k$. • Run the VSIMM $[M_n, M_{k-1}]$ cycle. • Combine estimates from the VSIMM $[M_k, M_{k-1}]$ and VSIMM $[M_n, M_{k-1}]$ cycles using (16)–(18). Let $M_k := M_k \cup M_a$.
S4.	Output $\hat{x}_{k\|k}$, $P_{k\|k}$, and $\{\mu_k^i\}_{m_i \in M_k}$.
S5.	Rank and order the model groups currently in effect in an ordered set such that their probabilities $\mu_k^{M_i}$ are in a decreasing order.
S6.	Rank and order the model groups currently in effect in another ordered set such that their joint likelihoods $\prod_{\kappa=k_0}^{k} L_\kappa^{M_i}$ are in a decreasing order, where k_0 is the activation time of the most recently activated group.
S7.	Terminate those model groups if • their ratios of probabilities to $\mu_k^{M_1}$ are below a threshold t^μ; and/or • their ratios of joint likelihoods to $\prod_{\kappa=k_0}^{k} L_\kappa^{M_1}$ are below a threshold t^L. Go to Step 1.

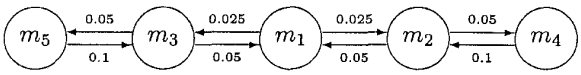


Fig. 5. Digraph associated with total model-set.

algorithm reduces to the MGS algorithm only if in addition to the above conditions, the activation logic is such that it may activate a model group only when no more than one model group is in effect.

V. AN ILLUSTRATIVE EXAMPLE OF FAULT DETECTION AND IDENTIFICATION

A simple example of fault detection and isolation is simulated here. A much more careful and complete design and evaluation of the MGS algorithm for a generic problem of tracking a maneuvering target is given in Part IV [21].

Consider the following simple scalar system with dynamics

$$x_{k+1} = 1.1x_k + w_k \tag{24}$$

and two-dimensional measurements

$$z_k = \begin{bmatrix} 0.5 \\ 1.2 \end{bmatrix} x_k + v_k \tag{25}$$

with $w_k \sim \mathcal{N}(0, 0.1^2)$, $v_k \sim \mathcal{N}(0, I)$, and initial condition $x_0 = 100$. Suppose that a total failure or partial fault may occur to sensor 1; that is, 0.5 may become 0 (total failure) or any number in between 0 and 0.5 (partial fault). Suppose that sensor 2 may also suffer from a total or partial fault.

Assume that an IMM estimator based on the following 5 models is used for the detection and identification of the sensor fault:

$$m_1 : \begin{bmatrix} 0.5 \\ 1.2 \end{bmatrix} \quad m_2 : \begin{bmatrix} 0.25 \\ 1.2 \end{bmatrix} \quad m_3 : \begin{bmatrix} 0.5 \\ 0.6 \end{bmatrix}$$

$$m_4 : \begin{bmatrix} 0 \\ 1.2 \end{bmatrix} \quad m_5 : \begin{bmatrix} 0.5 \\ 0 \end{bmatrix}.$$

The digraph (topology of the model-set with model transition probabilities) assumed in the IMM algorithm is shown in Fig. 5, where the transition probabilities to the models themselves can be obtained from the requirement that all transition probabilities from a model have to sum up to unity. A simple MGS algorithm with the following model groups was designed and evaluated

$$M_1 = \{m_1, m_2, m_3\},$$
$$M_2 = \{m_1, m_2, m_4\},$$
$$M_3 = \{m_1, m_3, m_5\}$$

where the VSIMM cycle is used for each model group. The choice of these model groups is intuitively appealing. The following simple activation logics were used.

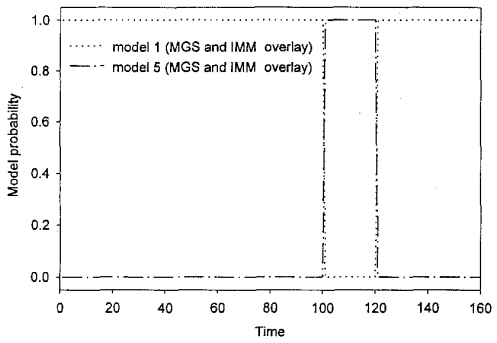


Fig. 6. Model probabilities of MGS and IMM algorithms under deterministic scenario S1.

1) Model group M_2 is activated while M_1 is in effect if $\mu_k^2 > t_1$ and similarly for the activation of M_3 from M_1.

2) Model group M_1 is activated while M_2 or M_3 is in effect if $\mu_k^1 > t_2$.

In our design, $t_1 = t_2 = 0.3$. It was found that the results of this example are virtually unchanged with respect to different activation logics provided that they are not unreasonable.

The rationale for such a design of model groups and activation logic is discussed in Part IV.

Consider the following simple scenarios.

S1. The system has no fault from $k = 0$ to $k = 160$, except that a sensor 2 total failure occurs at $k = 100$ and lasts until $k = 120$.

S2. The system has no fault from $k = 0$ to $k = 160$, except that a sensor 1 or 2 (with equal probability) total failure occurs at a random time $k = \tau_1$ and lasts for a random period of time τ_2, where $\tau_1 \sim \mathcal{N}(100, 5^2)$ and $\tau_2 \sim \mathcal{N}(20, 2^2)$.

Fig. 6 shows the model probabilities of the MGS and the FSIMM (with all 5 models) algorithms under the deterministic scenario S1 over 100 Monte Carlo runs.

Figs. 7 and 8 and Table IV show the rms errors of state estimation of the MGS and the IMM algorithms under the deterministic scenario S1 and the random scenario S2, respectively, over 100 Monte Carlo runs.

Note that the MGS and IMM algorithms have identical performance for this example. However, the computational complexities, in terms of FLOPs, of the MGS algorithm for the two scenarios are only about 54% of that of the IMM algorithm. This reduction in computation is much more significant if more models are used, as demonstrated in Part IV [21]. This is very important because the applications of MM estimators to most practical problems, especially to fault detection and identification, require the use of a large number of models.

GPB2 and the nonswitching MM algorithms were also implemented for this example. GPB2

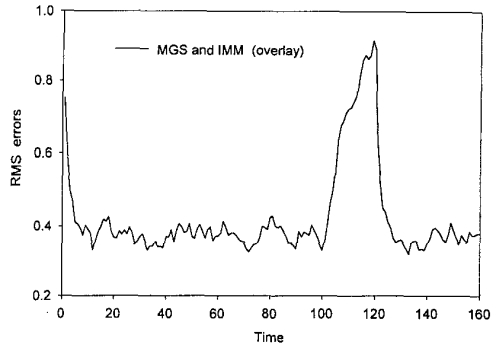


Fig. 7. RMS errors of MGS and IMM algorithms under deterministic scenario S1.

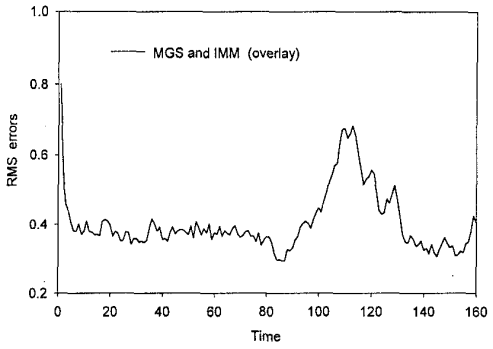


Fig. 8. RMS errors of MGS and IMM algorithms under random scenario S2.

TABLE IV
Comparison Between MGS and IMM Algorithms

	flops	RMS Error	Probability Error
Deterministic scenario S1:			
MGS	0.5375	0.4041	0
IMM	1	0.4041	0
Random scenario S2:			
MGS	0.5375	0.4224	0
IMM	1	0.4223	0

algorithm has the same performance as that of the IMM algorithm, while the nonswitching MM algorithm is substantially worse. Note also that there is no need to use the extended MGS algorithm for this example since there is virtually no delay for the MGS algorithm to switch to the right group.

VI. CONCLUSIONS

An MM estimator, called MGS algorithm, has been presented. It is the first MM estimator with variable structure that is general enough to be applicable to a large class of state estimation problems compounded with structural and/or parametric uncertainties or changes and is easily implementable. The algorithm runs at any given time a time-varying set (group) of closely related models of the total model-set. The determination of which model group to use is done by a hard decision, which consists of two-stage switching logics (hypothesis tests) for activation and termination of model groups. The model-group termination is done by a combination of the sequential mode-set probability ratio test and the sequential model-set likelihood ratio test of Part II [15], which have some desirable optimality properties. The general VSIMM recursion of Part II [15] is adopted in the MGS algorithm, which initializes the newly activated filters and assigns the initial probabilities to the corresponding models naturally, systematically, and near optimally. An extended version, called EMGS algorithm, has also been presented to overcome a potential weakness of the MGS estimator. For the simple simulation example of fault detection and identification presented, the proposed MGS algorithm was shown to be significantly superior to the fixed-structure IMM estimator using all models. The same performance was achieved by the former at a substantially reduced computational complexity. A more careful and complete design and evaluation of the MGS algorithm for a maneuvering target tracking example is presented in Part IV [21], which also supports this conclusion.

REFERENCES

[1] Ackerson, G. A., and Fu, K. S. (1970)
On state estimation in switching environments.
IEEE Transactions on Automatic Control, **AC-15** (Jan. 1970), 10–17.

[2] Balas, E., and Padberg, M. W. (1976)
Set partitioning: A survey.
SIAM Review, **18** (Oct. 1976), 710–761.

[3] Baram, Y., and Sandell, N. R., Jr. (1978)
An information theoretic approach to dynamical systems modeling and identification.
IEEE Transactions on Automatic Control, **AC-23** (Feb. 1978), 61–66.

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

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

[6] Blom, H. A. P., and Bar-Shalom, Y. (1988)
The interacting multiple model algorithm for systems with Markovian switching coefficients.
IEEE Transactions on Automatic Control, **33** (Aug. 1988), 780–783.

[7] Christofides, N. (1975)
Graph Theory: An Algorithmic Approach.
London: Academic Press, 1975.

[8] Daeipour, E., and Bar-Shalom, Y. (1998)
IMM for tracking maneuvering targets in the presence of glint.
IEEE Transactions on Aerospace and Electronic Systems, **34** (July 1998), 996–1003.

[9] Efe, M., and Atherton, D. P. (1996)
Maneuvering target tracking using adaptive turn rate models in the interacting multiple model algorithm.
In *Proceedings of the 35th IEEE Conference on Decision and Control*, Kobe, Japan, Dec. 1996, 3151–3156.

[10] Elliott, R. J., Dufour, F., and Sworder, D. D. (1996)
Exact hybrid filters in discrete time.
IEEE Transactions on Automatic Control, **41** (Dec. 1996), 1807–1810.

[11] Gauvrit, M. (1984)
Bayesian adaptive filter for tracking with measurements of uncertain origin.
Automatica, **20** (Mar. 1984), 217–224.

[12] Jilkov, V. P., and Angelova, D. S. (1996)
Performance evaluation and comparison of variable structure multiple-model algorithms for tracking maneuvering radar targets.
In *Proceedings of the 26th European Microwave Conference*, Prague, Czechaslovakia, Sept. 1996.

[13] Jilkov, V. P., Angelova, D. S., and Semerdjiev, T. A.
Mode-set adaptive IMM for maneuvering target tracking.
IEEE Transactions on Aerospace and Electronic Systems, **35** (Jan. 1999).

[14] Li, X. R. (1996)
Hybrid estimation techniques.
In C. T. Leondes (Ed.), *Control and Dynamic Systems: Advances in Theory and Applications*, Vol. 76.
New York: Academic Press, 1996.

[15] Li, X. R.
Multiple-model estimation with variable structure—Part II: Recursive adaptive model-set approach. Submitted.
An early brief version of its two separate parts has appeared respectively in *Proceedings of the 1998 American Control Conference*, Philadelphia, PA, June 1998, 2439–2443, and in *Proceedings of the 1998 SPIE Conference on Signal and Data Processing of Small Targets*, **3373-58**, Orlando, FL, Apr. 1998, 546–558.

[16] Li, X. R. (1994)
Multiple-model estimation with variable structure: Some theoretical considerations.
In *Proceedings of the 33rd IEEE Conference on Decision and Control*, Orlando, FL, Dec. 1994, 1199–1204.

[17] Li, X. R., and Bar-Shalom, Y. (1992)
Mode-set adaptation in multiple-model estimators for hybrid systems.
In *Proceedings of the 1992 American Control Conference*, Chicago, IL, June 1992, 1794–1799.

[18] Li, X. R., and Bar-Shalom, Y. (1996)
Multiple-model estimation with variable structure.
IEEE Transactions on Automatic Control, **41** (Apr. 1996), 478–493.

[19] Li, X. R., and Dezert, J. (1998)
Layered multiple-model algorithm with application to tracking maneuvering and bending extended target in clutter.
In *Proceedings of the 1st International Conference on Multisource-Multisensor Information Fusion*, Las Vegas, NV, July 1998, 207–214.

[20] Li, X. R., and Zhang, Y. M. (1998)
Multiple-model estimation with variable structure—Part V: Likely-model-set algorithm.
Submitted for publication. An early brief version appeared in *Proceedings of the 1998 SPIE Conference on Signal and Data Processing of Small Targets*, **3373-26**, Orlando, FL, Apr. 1998.

[21] Li, X. R., Zhang, Y. M., and Zhi, X. R. (1999)
Multiple-model estimation with variable structure—Part IV: Design and evaluation of model-group switching algorithm.
IEEE Transactions on Aerospace and Electronic Systems, **35** (Jan. 1999). This issue (242–254).

[22] Lin, H., and Atherton, D. P. (1993)
An investigation of the SFIMM algorithm for tracking maneuvering targets.
In *Proceedings of the 32nd IEEE Conference on Decision and Control*, San Antonio, TX, Dec. 1993, 930–935.

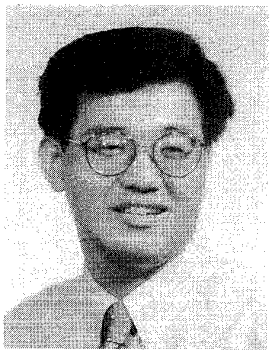
[23] Magill, D. T. (1965)
Optimal adaptive estimation of sampled stochastic processes.
IEEE Transactions on Automatic Control, **AC-10** (1965), 434–439.

[24] Maybeck, P. S., and Hentz, K. P. (1987)
Investigation of moving-bank multiple model adaptive algorithms.
AIAA Journal of Guidance, Control, and Dynamics, **10** (Jan.–Feb. 1987), 90–96.

[25] Munir, A., and Atherton, D. P. (1995)
Adaptive interacting multiple model algorithm for tracking a manoeuvring target.
IEE Proceedings of Radar, Sonar, and Navigation, **142** (Feb. 1995), 11–17.

[26] Li, X. R., and He, C. (1999)
Model-set choice and design for multiple-model estimation.
Submitted.

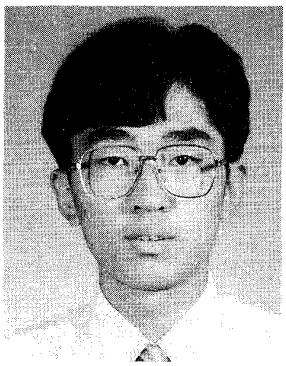
[27] Kirubarajan, T., Bar-Shalom, Y., Pattipati, K. R., and Kadar, I. (1998)
Ground target tracking with topograph-based variable structure IMM estimator.
In *Proceedings of the 1998 SPIE Conference on Signal and Data Processing of Small Targets*, 3373-25, Orlando, FL, Apr. 1998, 222–233.



X. Rong Li (S'90—M'92—SM'95) received the B.S. and M.S. degrees from Zhejiang University, Hangzhou, Zhejiang, PRC, in 1982 and 1984, respectively, and the M.S. and Ph.D. degrees from the University of Connecticut, Storrs, in 1990 and 1992, respectively, all in electrical engineering.

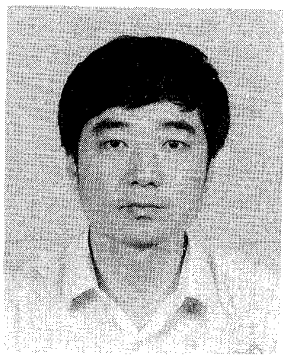
Dr. Li joined the Department of Electrical Engineering, University of New Orleans, LA, in August 1994. During 1986–1987 he did research in electric power engineering at the University of Calgary, Alberta, Canada. He was an Assistant Professor in the Department of Electrical Engineering, University of Hartford, West Hartford, Connecticut, from 1992 to 1994. His current research interests include signal and data processing, target tracking and information fusion, stochastic systems theory and its applications, and electric power.

He has published more than 20 refereed journal articles, more than 50 conference papers, three invited book chapters, coauthored a monograph *Multitarget-Multisensor Tracking* (YBS Publishing, Storrs, CT, 1995) and a graduate text *Estimation and Tracking: Principles, Techniques, and Software* (Artech House, Boston, MA, 1993). He serves as an Editor of the IEEE *Transactions on Aerospace and Electronic Systems*. He received the CAREER award and the RIA award from National Science Foundation, and the 1996 Faculty Early Career Achievement Award for Excellence in Research from University of New Orleans. He is the Steering Committee Chair of the First International Conference on Information Fusion. He has been invited to give more than 30 seminars at research institutions in U.S. and other countries. He has won several Outstanding Paper awards.



Xiaorong Zhi was born in Gansu, China on August 7, 1972. He received the B.S. degree in automatic control from Northwestern Polytechnical University, Xi'an, China, in 1993 and the M.S. degree in electrical engineering from University of New Orleans, New Orleans, in 1997.

He is currently an engineer at Ultratech Stepper Inc., San Jose, CA. His research interests include estimation, target tracking, signal processing, and real time control system.



Youmin Zhang was born in Dali, Shaanxi Province, P.R. China in April 1963. He received the B.S., M.S. and Ph.D. degrees in electrical engineering from Northwestern Polytechnical University, Xian, P. R. China, in 1983, 1986 and 1995, respectively.

Since 1986, he has been worked in the Department of Automatic Control, Northwestern Polytechnical University, Xian, P. R. China, where he has been an Associate Professor since 1992 and an Associate Chairman of the Department since 1994. From Sept. 1995 to Oct. 1998, he held several research positions in the Dept. of Electrical Engineering, University of New Orleans, Dept. of Electrical and Computer Engineering, Louisiana State University, and the Dept. of Electrical Engineering, State University of New York at Binghamton, respectively. He is currently with the Dept. of Electrical and Computer Engineering, The University of Western Ontario as a Postdoctoral Research Fellow. His main research interests include fault detection, diagnosis and fault-tolerant control, filtering, estimation and identification and applications to target tracking, aircraft flight data processing, neural networks and applications to modeling, identification and control, robust filtering, identification and control.

Dr. Youmin Zhang was the recipient of several research achievement awards from the National Education Commission of China, Aeronautics and Astronauts Industry Ministry of China; an outstanding teaching achievement award from Shaanxi Province of China; several outstanding teaching awards from Northwestern Polytechnical University; an award of outstanding young teacher from Northwestern Polytechnical University. He served as the General Chairman of the 10th Youth Conference on Automatic Control and Automation in China in 1994 and has served as the international program committee member of several international conferences. He has published more than 20 journal papers and edited one book. He has also received several Outstanding Paper awards.