Proceedings of the 5th International Conf. Information Fusion
Annapolis, MD, USA, July 2002, pp. 26–33

Model-Set Design for Multiple-Model Method—Part I

X. Rong Li*
Department of Electrical Engineering
University of New Orleans
New Orleans, LA 70148, USA
Phone: 504-280-7416, Fax: 504-280-3950, xli@uno.edu

Abstract – *The most important problem in the application of the multiple-model approach to estimation is the design of the model set. This paper deals with this challenging topic in a general setting. Modeling of models as well as true mode as random variables is proposed. Several general methods for design of model sets, along with the initial model probabilities, are presented. They include distribution approximation, minimizing mismatch between mode and models, and moment matching. Examples that demonstrate how the general results presented here can be applied are presented in Part II.*

Keywords: Multiple models, model-set design, variable structure, adaptive estimation, target tracking

1 Introduction

Model-set design is the most important issue in the application of MM estimation. The performance of an MM algorithm for a given problem depends largely on the set of models used and the primary difficulty in the application of the MM method is the design of the model set. Numerous publications have appeared in which ad hoc designs were presented. Unfortunately, very limited theoretical results on this important issue are available. It was shown theoretically in [5] that the use of too many models is as bad as the use of too few models. A circular criterion for model-set choice was presented in [5]. When the mode space is a continuous region, a necessary and sufficient condition was presented in [6] for a convex combination of estimators to be superior to each individual estimators, based on respective model sets. In order to apply the MM method to problems with uncertain *parameters*, two important questions are: (a) which quantity is best selected as the estimatee (i.e., the quantity to be estimated) and (b) how to quantize the parameter space optimally. [2] provides theoretical results on the optimal selection of the estimatee. A procedure to determine the choice of the quantization points was presented in [8] given the number of quantization points. A necessary condition for the effective performance of MM estimation was presented in [1] for a jump linear time-invariant system in terms of its dc gain.

This paper presents theoretical results on the model-set design in a general setting. General results for probability-distribution based, minimum-distance based, and moment-matching design are presented. These results provide concrete designs as well as guidelines and insights that are helpful for model-set design.

2 Probabilistic Modeling of Models and Modes

In this paper, a **mode** refers to the physical behavior pattern or structure of a system/process (or its precise mathematical model), and a **model** refers to the (possibly simplified) mathematical representation or description of the system or process on which an estimator is based (see [3] for a more detailed explanation). Such a distinction is necessary where mismatch between the model and mode is of concern.

Denote by $\mathbf{S}$ the **mode space**, that is, the set of possible modes under consideration. In general, mode space $\mathbf{S}$ may be either a discrete (finite or countable) set or a continuous region. In the latter case it is assumed that a system mode may only *jump* from a point in $\mathbf{S}$ to another one, rather than vary *continuously*.

A contribution of this paper is the recognition of the need for and introduction of probabilistic modeling of models as well as the true mode.

The need to have a proper description of the true mode is evident: Without such a description, model-set design and performance evaluation of MM algorithms are essentially groundless — we can always find a scenario under which any given realizable "optimal" model set is worse

*Research supported in part by ONR grant N00014-00-1-0677, NSF grant ECS-9734285, and NASA/LEQSF grant (2001-4)-01.

than some other model set. Deterministic descriptions of the true mode in the form of "typical" or "representative" scenarios are prevailing in the literature of MM estimation, particularly for performance evaluation. Such deterministic descriptions have certain drawbacks. For example, the choice of particular scenarios is fairly arbitrary, and thus the corresponding performance evaluation results are less objective or convincing since the performance of MM algorithms is highly dependent on test scenarios. (The scenario dependence of the performance of a hybrid estimation algorithm is elaborated in [4].) It is impossible to develop general, systematic methods for model-set design on the basis of such "arbitrary" descriptions of the true mode.

We propose to model the true mode as a random variable $s : \Omega \to \mathbf{S}$, where $\mathbf{S}$ is the mode space and Ω is the sample space. The random variable s may be continuous, discrete, singular, or hybrid. Let $F_s(x)$ and $f_s(x)$ be its cumulative distribution function (cdf) and probability density function (pdf) if exists, respectively. In practice, they can be obtained by past data using statistical techniques or simply from experience. For example, a transposed (i.e., symmetrical) three-phase overhead transmission line in a power system has three simple modes (i.e., normal, single-phase to ground fault, and phase-to-phase fault) and several composite modes (e.g., two-phase to ground fault and three-phase to gound fault). Data of the past operation records (e.g., fault rate and percentage of fault type) provide the required probability distribution of the mode. For a particular application of MM estimation, if $F_s(x)$ is not available at this stage, the benefit of having such a cdf — as presented in this and future papers — suggests that it may be worthwhile to obtain such a cdf. This is a manifestation of guidance of theory to practice. Without such guidance, most practical probabilistic models (e.g., Gaussian models, Poisson models) would not have been developed and probability theory would have very limited practical value.

Similarly, *we also propose that the problem of designing a model set* M *(and the corresponding initial model probabilities) be formulated as that of designing a* ***random model*** m *with range* M; that is, design a random variable $m : \Omega \to M$, where Ω is the sample space. As such, the following needs to be determined: (a) cardinality $|M|$ (i.e., number of models); (b) all elements m_i of $M = \{m_1, m_2, \ldots, m_{|M|}\}$ (i.e., model locations/values); (c) prior (or initial) model probabilities $P\{m = x\}$. Note that cdf $F_m(x)$ of m, or equivalently, probability mass function (pmf) $p_m(x) = P\{m = x\}$ summarizes all information needed. While this concept of *random* model may appear alien to a practitioner, we need only to recall that a random variable is (corresponds to) in fact nothing but a properly defined set of *deterministic* numbers. It is exactly in this way that a set of deterministic models used in the MM method, along with the above constraints (a)–(c), defines a random model.

More generally, the second and third generations of MM algorithms require design of (Markovian) laws governing model transitions based on transitions of the true mode. Even more generally, the true mode is better modeled as a random process $s(t) : (\Omega, \mathcal{F}, P) \times T \to \mathbf{S}$; that is, $s(t)$ is a family of random variables, indexed by $t \in T$ and defined on a common probability space $(\Omega, \mathcal{F}, P)$. Similarly, the problem of model-set design is better formulated as the determination of a random process $m(t) : (\Omega, \mathcal{F}, P) \times T \to \mathbf{M}$, where $\mathbf{M}$ is the total model set. These more general formulations are useful for model-set adaptation and design of model transitions. For offline model-set design (the topic of this paper), however, it usually suffices to consider s and m as random variables, completely described by their cumulative distribution functions.

For simplicity, we assume that the true mode is continuous in this paper. The same approach works for other cases, although modification is sometimes needed. We always assume that the model is discrete (in fact, finite).

For many applications, the true mode s has a real physical meaning directly (see Part II [7] for examples) and the above probabilistic modeling is clearly reasonable. For many other applications, however, s is an index of underlying structures (or behavior patterns) and it is difficult, if not impossible, to define a proper distance metric directly for $\mathbf{S}$ convincingly with a clear interpretation. In such cases, cdf of s may possibly be defined over an abstract space where the elements of s are arranged such that the neighboring elements correspond to the neighboring structures in the physical world. Then a question is how to define the neighbor concept for structures in the physical world? This question can be answered by using, e.g., Kullback-Leibler distance between the distribution or likelihood functions of any two structures s_i and s_j.

3 Formulation of Model-Set Design

Following the previous section, the true mode (at any time) can be reasonably modeled as a continuous random variable in many cases, while it is better modeled as a discrete (or hybrid) random variable in many other cases. In any case, its sample space $\mathbf{S}$ is usually much larger than the model set M affordable in practice.

From the probabilistic modeling of the true mode and models, it is clear that the *model-set design is essentially a problem of finding a discrete random variable* m *to approximate a given random variable* s, which can be continuous, discrete, singular, or hybrid, depending on the application. [1] Unfortunately, to our knowledge, there is no generally acceptable solution to this problem in the literature.

[1] This probabilistic view also makes it quite intuitive the fundamental finding of [5] that the optimal model set M for the MM approach is $M = \mathbf{S}$—the performance of MM estimators deteriorates if either extra models are used ($M \supset \mathbf{S}$) or some models are missing ($M \subset \mathbf{S}$)—and the deterioration worsens as M and $\mathbf{S}$ become more mismatched.

We propose three classes of systematic solutions below.

4 Distribution-Based Design

The first solution is based on the idea of finding the cdf $F_m(x)$ of a discrete random variable (model) m to approximate the cdf $F_s(x)$ of any given random variable (mode) s. We describe this solution in the scalar case (i.e., for scalar s and m) here and extend it to the vector case in Part II.

Assume that the cdf $F_s(x)$ of true mode s is known. Given a tolerance ϵ, we want to construct the cdf $F_m(x)$ of a discrete random variable (i.e., model set) such that $|F_s(x) - F_m(x)| \leq \epsilon$ for all x.

It can be shown that for any given cdf $F_1(x)$ we can find the cdf $F_2(x)$ of some discrete random variable that is arbitrarily close to $F_1(x)$ in terms of the following distance metric

$$d(F_1, F_2) = \max_{x \in R} |F_1(x) - F_2(x)| \qquad (1)$$

where $R = [-\infty, \infty]$. In other words, the problem under consideration always has a solution. What is the minimum number of models needed? The following lemma answers this question.

Lemma 4.1. Given a tolerance ϵ in the above distance metric, the minimum number of models needed is given by

$$|M| = \lceil 1/2\epsilon \rceil = \text{smallest integer not smaller than } 1/2\epsilon$$

A proper tolerance ϵ is not always easy to come by. In some cases, the number of models $|M|$ is predetermined directly from, say, resource for processing or computation.

Theorem 4.1 (Minimum-set design**)**. Given $|M|$, the model set M^*, along with the pmf p^*, (i.e., the random model) that minimizes the distance metric defined by (1), that is,

$$\{M^*, p^*\} = \arg \inf_{\{M,p\} \text{ given } |M|} \sup_{x \in \mathbf{S}} |F_s(x) - F_m(x)|, \ m \in M$$

is given by

$$\begin{aligned} m_i &= \arg \min_{x \in \mathbf{S}} \left[F_s(x) = \frac{i - 1/2}{|M|} \right], \ i = 1, \ldots, |M| \qquad (2) \\ M^* &= \{m_1, m_2, \ldots, m_{|M|}\} \end{aligned}$$

along with the following evenly distributed pmf (i.e., initial model probabilities): for $i = 1, \ldots, |M|$,

$$p_m(x)|_{x=m_i} = P\{m = m_i | m \in M^*\} = \frac{1}{|M|} \qquad (3)$$

This design is depicted in Fig. 1. Note that m_i is chosen to satisfy (2) only from the elements of $\mathbf{S}$, and thus $M \subset \mathbf{S}$.

This approach to model-set design is intuitively appealing. In essence, it partitions the mode space into equally probable regions and places a model at the "center" (in fact, median) of each region. As such, all models are equally loaded in that they are equally likely to take effect and cover an region of equal probability. It uses the minimum number of models. It is also perfectly consistent with the common practice of assigning equal initial probability to every model.

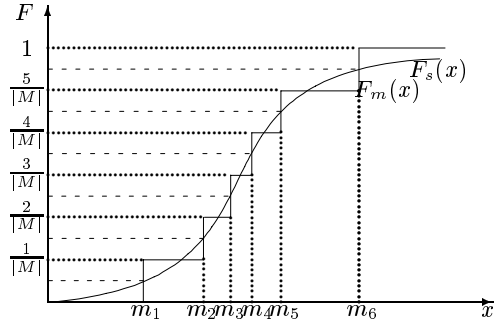


Fig. 1: Approximating a cdf by a stair-case type cdf with given tolerance.

5 Minimum–Distance Design

In the previous section, we design a model m to approximate the true mode s by constructing a cdf $F_m(x)$ that is close to the cdf $F_s(x)$ — the design is actually done in the space of the distribution functions. Alternatively, the design can also be done in the vector space of random variables; that is, find an m that is close to s in their vector space directly. In order to do this, a metric of the closeness between model and mode is needed.

Closeness metric between model and mode. The distance metric in the vector space of random variables is most often defined as the square root of the mean-square value $\bar{d}(s, m) = (E[d(s, m)])^{1/2}$, where $d(s, m) = (s - m)'(s - m)$. Of course, other metrics can also be defined, such as $\bar{d}(s, m) = (E[d(s, m)])^{1/p}$, where $d(s, m) = [(s - m)'(s - m)]^{p/2}$. When s and m are vectors, $d(s, m)$ is actually a scalar metric of the families of vectors, since a random vector actually corresponds to a family of vectors in a linear space. We will consider the more general metric with an arbitrary p but we are more interested in the case $p = 1, 2$.

For $s \in \mathbf{S}$ and $m \in M$, we have

$$\begin{aligned} & E[d(s, m)] = E[E[d(s, m)|s]] \\ = & \int_{\mathbf{S}} \sum_{m_i \in M} d(s, m_i) P\{m = m_i | s\} dF(s) \qquad (4) \end{aligned}$$

It is thus seen that the closeness of m and s depends on $w_i(s) = P\{m = m_i|s\}$, the model probability conditioned on the true mode s. A study of the conditional probability $P\{m = m_i|s\}$ will be reported later.

In this paper, for simplicity, we assume

$$P\{m = m_i|s\} = 1(s; S_i) = \begin{cases} 1 & s \in S_i \\ 0 & s \notin S_i \end{cases} \qquad (5)$$

This is equivalent to assuming that

$$\{\{m = m_1\}, \ldots, \{m = m_N\}\} = \{\{s \in S_1\}, \ldots, \{s \in S_N\}\}$$

is a partition of the mode space $\mathbf{S}$; that is, each model covers a subset (region) of the mode space *exclusively*, which is often perceived in practice. With this assumption, (4) becomes

$$E[d(s,m)] = \sum_i \int_{S_i} d(s,m_i) dF(s) \tag{6}$$

We now present several general results under this assumption.

Theorem 5.1 (Optimality conditions of model set). Assume that $\mathcal{S} = \{S_1, \ldots, S_N\}$ is a partition of the mode space $\mathbf{S}$, where S_i is covered by model m_i *exclusively* in the sense $\{s \in S_i\} = \{m = m_i\}$. Then, the following conditions hold for the optimality in the sense of minimizing distance metric $\bar{d}(s,m)$ defined above.

A. Given any partition $\mathcal{S} = \{S_1, \ldots, S_N\}$ of mode space $\mathbf{S}$, a model set $M = \{m_1, \ldots, m_N\}$ is optimal if each model m_i is a (generalized) centroid of the corresponding partition member S_i:

$$m_i = s_i^* \triangleq \arg\min_m E[d(s,m)|s \in S_i] \tag{7}$$

B. Given any model set M, a partition is optimal if and only if points in any partition member S_i are closer to m_i than to any other $m_j \in M$ (almost surely):

$$S_i = \{s : d(s,m_i) < d(s,m_j), \forall m_j \neq m_i, m_i, m_j \in M\}$$

that is, a point s must be assigned to its nearest neighbor m_i among all $m \in M$; the set of equal-distance points

$$S_{ij} = \{s : d(s,m_i) = d(s,m_j) \leq d(s,m_k), \forall m_k \in M\}$$

may be assigned to either S_i or S_j.

Remarks. (a) This theorem basically states that under the stated assumption, if exists, the optimal model set is within the class in which models are located at the (generalized) centroids of members of a nearest-neighbor partition of the mode space. (b) The generalized centroid reduces to the conditional mean (i.e., the centroid (mean) of S_i) $s_i^* = E[s|s \in S_i]$ if $d(s,m) = (s-m)'(s-m)$ or the conditional median (i.e., the median of S_i) if $d(s,m) = |s-m|$. (c) Both conditions are quite intuitive. (d) This theorem does not address the issue whether an optimal model set that minimizes the above metric is existent or unique, or whether a solution that meets conditions A and B is existent or unique. (e) The optimality conditions of this theorem actually hold for closeness metrics more general than defined above.

Most importantly, this theorem provides a theoretical basis for iteration procedures to find an optimal model set under the stated assumption. For example, we may start with an initial partition of mode space; find a candidate of the model set as the (generalized) centroid of each partition member; use the nearest-neighbor rule to obtain the corresponding (updated) partition; and repeat this process until convergence. Alternatively, we may start with an initial model set; use the nearest-neighbor rule to obtain the corresponding partition; obtain an update of the model set as the (generalized) centroid of each partition member; and repeat this process until convergence.

The above centroid model set has several nice and intuitive properties, as presented in the next theorem.

Theorem 5.2 (Properties of optimal model set). Any model set that covers each S_i by its centroid $m_i = E[s|s \in S_i]$ exclusively (i.e., $\{s \in S_i\} = \{m = m_i\}$) has the following properties:

(a) The (random) model and mode have the same mean: $E[m] = E[s]$.

(b) The modeling error is orthogonal to model: $E[m(s-m)'] = 0$.

(c) $E[ms'] = E[sm'] = E[mm']$ and thus $E[m's] = E[s'm] = E[m'm]$, meaning that cross power of the mode and model is equal to the power of the model.

(d) $E[(s-m)(s-m)'] = E[ss'] - E[mm']$ and thus $E[(s-m)'(s-m)] = E[s's] - E[m'm]$, meaning that minimum MSE is the power of the mode minus the power of the (optimal) model.

(e) $E[s(s-m)'] = E[(s-m)(s-m)']$ and thus $E[s'(s-m)] = E[(s-m)'(s-m)]$.

Remarks. It follows from Theorem 5.1 that given a partition of the mode space, a model set that covers S_i by $m_i = E[s|s \in S_i]$ exclusively is optimal in the sense of minimizing MSE matrix $E[(s-m)(s-m)']$ and thus minimizing MSE scalar $E[(s-m)'(s-m)]$.

6 Moment-Matching Design

In some practical situations, some moments, but not the complete distribution, of the true mode s are known. In some other situations, we do not have a good knowledge of a proper tolerance $|F_s(x) - F_m(x)| \leq \epsilon$, but only want to match the moments of m to the known moments of s.

Given up to the qth moments of s, we want to find a discrete random variable m (i.e., the number and locations of points m_i with the associate probability mass p_i) such that

$$E[m^n] = E[s^n], \quad n = 1, \ldots, q$$

Several questions arise immediately. For example, what is the minimum number of models such that up to the qth

moments of s and m are matched? How to design the corresponding pmf (locations m_i and probability masses p_i) of m? Given the number of models $|M|$, how to design pmf of m that matches as many as possible the lowest moments of s? For simplicity, we will consider only matching mean and covariance in this section since it is the common practice.

Let the pmf of m be

$$p_i = \{m = m_i | m \in M\} > 0, \quad \forall i \in J = \{1, \ldots, |M|\}$$

where $M = \{m_1, \ldots, m_{|M|}\}$. Then, the mean and covariance of m are

$$\bar{m} = \sum_{i \in J} m_i p_i, \quad C_m = \sum_{i \in J} (m_i - \bar{m})(m_i - \bar{m})' p_i$$

6.1 Minimum Model-Set Design

The following theorem answers the first question above for $q = 2$.

Theorem 6.1 (Minimum models). The minimum number of models needed for m to match the *mean $\bar{s}$ and covariance C_s* of the true mode s is rank of C_s plus one:

$$\text{minimum number of models} = \text{rank}(C_s) + 1$$

Now consider the problem of design $\{m_i^*, p_i, i \in J\}$ such that

$$\sum_{i \in J} p_i = 1, \ \sum_{i \in J} m_i^* p_i = \bar{s}, \ \sum_{i \in J} (m_i^* - \bar{s})(m_i^* - \bar{s})' p_i = C_s \tag{8}$$

In fact, we only need to design $\{m_i, p_i, i \in J\}$ such that

$$\sum_{i \in J} p_i = 1, \quad \sum_{i \in J} m_i p_i = \mathbf{0}, \quad \sum_{i \in J} m_i m_i' p_i = I_{n \times n} \tag{9}$$

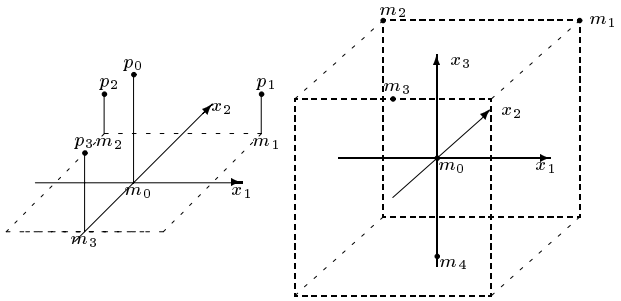
where $n = \text{rank}(C_s)$. All designs presented below are for this standard problem. Given a problem with known mean $\bar{s}$ and covariance C_s, the design $\{m_i, p_i, i \in J\}$ can be converted to design $\{m_i^*, p_i, i \in J\}$ by $m_i^* = A[m_i', \mathbf{0}]' + \bar{s}$, which satisfies (8), where $C_s = A\text{diag}(I_{n \times n}, \mathbf{0})A'$.

Theorem 6.2 (Minimal-set design). The design $\{m_i, p_i\}_{i=0}^{n+1}$ with

$$\begin{aligned}
&0 \le p_0 < 1, \quad p_0^1 = p_0, \quad p_1^1 = p_2^1 = (1 - p_0)/2 \\
&m_0^1 = 0, \ m_1^1 = (1 - p_0)^{-1/2}, \ m_2^1 = -(1 - p_0)^{-1/2} \\
&\quad \vdots \\
&p_0^j = p_0, \quad p_i^j = p_i^{j-1}/2, \quad i = 1, \ldots, j \\
&p_{j+1}^j = (1 - p_0)/2, \quad m_{j+1}^j = \left[\mathbf{0}, -(1 - p_0)^{-1/2}\right]' \\
&m_0^j = \mathbf{0}, \ m_i^j = \left[(m_i^{j-1})', (1 - p_0)^{-1/2}\right]', i = 1, \ldots, j
\end{aligned}$$

satisfies (9), where $m_i = m_i^n, p_i = p_i^n, i = 0, 1, \ldots, n+1$, and the superscript denotes dimension of a vector.

Fig. 2 illustrates this design with a minimal model set for $n = 2, 3$, respectively. For $n = 3$, m_0 is at the center of the cube, while all other models are on the surface of the cube; m_4 is at the center of the bottom square. Note that the coordinates of every model are either 0 or $\pm(1 - p_0)^{-1/2}$. The mean and covariance are matched by the probability mass: $\sum_{i=1}^{j} p_i^j = p_{j+1}^j$.



(a) 2D case (b) 3D case

Fig. 2: Illustration of a minimal-set design.

In this design, $0 \le p_0 < 1$ is a free parameter for us to choose. If we choose $p_0 = 0$ (i.e., delete m_0), we actually have $n + 1$ models, which by Theorem 6.1 is the smallest possible number of models to match mean and covariance.

Remarks. (a) Although the model m_0 is not needed to match mean and covariance, in practice, such a model located at the expected true mode is usually very beneficial for MM estimation. (b) The value of p_0 affects higher order moments — a greater p_0 implies that the distribution of m is more concentrated around the mean. (c) This minimal-set design depends very much on not only the choice of the coordinate system but also the artificial labeling of each coordinate (e.g., the locations and the probability masses of the models would vary if x_1 and x_3 of Fig. 2(b) were interchanged). The latter dependence is entirely artificial and is better eliminated, while the former dependence is inevitable because the coordinate directions (after transformation from m^* to m) are actually eigenvector directions.

Minimal-set designs are not unique. Fig. 3 illustrates another simple minimal-set design in the 3D case. Its extension to a higher dimension is straightforward. In this design, a model with probability mass p is placed on each positive semi-axis of equal distance α from the origin (i.e., $m_i = \alpha e_i, \forall i$, where $e_i = [\mathbf{0}_{1 \times (i-1)}, 1, \mathbf{0}_{1 \times (n-i)}]'$ is the ith coordinate vector); the last model is $m_{n+1} = \beta[-1, -1, \ldots, -1]'$ with probability mass q. It is clear that the mean and covariance are $\bar{m} = 0$ and $C_m = I_{n \times n}$ if $q = p$ and $\alpha = \beta = 1/\sqrt{p}$. As for the design of Fig. 2, if desirable, an additional model may be placed at the origin with probability p_0 without affecting mean and covariance. Then $p = (1 - p_0)/(n + 1)$. In Sec. 6.3, $q \ne p$ and $\alpha \ne \beta$ are chosen to obtain a minimal set with an equal distance between models.

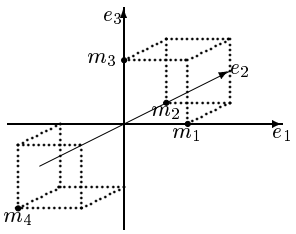


Fig. 3: Illustration of another minimal-set design.

The minimal-set design of Fig. 2 has attractive features that the model locations and probability mass are determined recursively as dimension increases and that all self skewnesses are equal to zero in the design of Fig. 2: $E\{[m(j) - \bar{s}(j)]^3\} = 0, \forall j \leq n$.

6.2 Symmetric Model-Set Design

All the above minimal-set designs clearly have an asymmetrical distribution spatially and possibly probabilistically. For many applications in practice, it is appealing that the models are symmetrically distributed and invariant to the artificial labeling of coordinates. For this reason, we present the following theorem.

Theorem 6.3 (Minimal symmetric-set design). The design $\{m_i, p_i\}_{i=0}^{2n}$ with the following *symmetric* distribution satisfies (9)

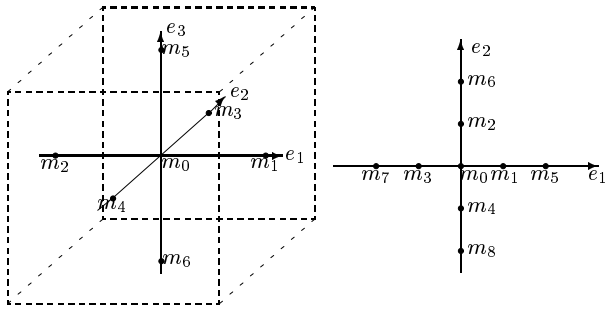
$$0 \leq p_0 < 1, \qquad p_i = (1-p_0)/(2n), \quad i = 1, \ldots, 2n$$
$$m_0 = \mathbf{0}, \qquad m_i = -m_{n+i} = e_i \sqrt{\frac{n}{1-p_0}},\ i = 1, \ldots, n$$

As for the design of Theorem 6.1, $0 \leq p_0 < 1$ is a free parameter for us to choose whose value affects higher-order moments. If we choose $p_0 = 0$ (i.e., delete m_0), we actually have $2n$ models. In practice, however, the use of model m_0 is usually very beneficial for MM estimation.

Fig. 4(a) illustrates this symmetric-set design for $n = 3$, where m_0 is at the center of the cube, while all other models are at the center of a boundary square of the cube. Note that if m_0 is not used, all models are located symmetrically on an axis (representing an eigenvector direction) with an equal distance from the origin; thus, the mean is matched provided an equal probability mass is assigned to all models and the covariance is matched by such a special assignment that all models on each axis have a total contribution of 1 to the covariance.

In this design, there are only two models along each axis direction, excluding m_0. In many applications, more models are needed for an MM estimator to perform well. Therefore, we present the following extension of Theorem 6.3.

Theorem 6.4 (Symmetric-set design). The design



(a) 3D case (b) 2D case

Fig. 4: Illustration of a symmetric-set design.

$\{m_i, p_i\}_{i=0}^{2kn}$ with the following *symmetric* distribution

$$\begin{aligned} 0 \leq p_0 &< 1, m_0 = \mathbf{0} \\ p_{2(j-1)n+i} &= p_{(2j-1)n+i} = \frac{1-p_0}{2\alpha_j \beta_j n} \\ m_{2(j-1)n+i} &= -m_{(2j-1)n+i} = e_i \sqrt{\frac{\alpha_j n}{1-p_0}} \\ & \qquad i = 1, \ldots, n,\ j = 1, \ldots, k \end{aligned}$$

satisfies (9), where $\alpha_k \geq \alpha_{k-1} \geq \cdots \geq \alpha_1 > 0$ and $\beta_j > 0$ satisfy

$$\sum_{j=1}^{k} \frac{1}{\beta_j} = 1, \quad \sum_{j=1}^{k} \frac{1}{\alpha_j \beta_j} = 1$$

A simple and meaningful choice for α_j and β_j is

$$\alpha_j = j\alpha_1, \quad \beta_j = k, \quad j = 1, \ldots, k$$

which yields $\alpha_1 = 3/4$, $\alpha_2 = 3/2$ for $k = 2$, and $\alpha_1 = 11/18$, $\alpha_2 = 22/18$, $\alpha_3 = 33/18$ for $k = 3$. Fig. 4(b) illustrates this symmetric-set design for $n = 2$ and $k = 2$.

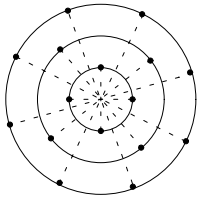


Fig. 5: Illustration of a more evenly distributed model-set design.

A possible drawback of this symmetric design is that the models are distributed highly unevenly in space, albeit symmetrically. We now present a design that is much more evenly distributed. This can be accomplished by rotating models $m_{2(j-1)n+i}$ and $m_{(2j-1)n+i}$ for $j \geq 2$ such that they are more evenly distributed. We only consider $n = 2$ and $k = 4$ with $\alpha_4 = \alpha_3$, as shown in Fig. 5. It can be extended to the general case.

Let

$$e_i^1 = e_i, \quad e_i^2 = (e_i^1 + e_{i+1}^1)/\sqrt{2}, \quad e_i^3 = (e_i^1 + e_i^2)/\sqrt{2},$$
$$e_i^4 = (e_{i+1}^1 + e_i^2)/\sqrt{2}, \quad e_{2+i}^j = -e_i^j$$

for $i = 1, 2$ and $j = 1, 2, 3, 4$. Note first that a key to the design of Theorem 6.4 is

$$\begin{aligned}\text{cov}(m) &= \sum_{i=1}^{n} \text{diag}\left(\mathbf{0}_{(i-1)\times(i-1)}, \sum_{j=1}^{k}\frac{1}{\beta_j}, \mathbf{0}_{(n-i)\times(n-i)}\right) \\ &= I\sum_{j=1}^{k}\frac{1}{\beta_j} = \sum_{j=1}^{k}\frac{1}{\beta_j}[e_1, \ldots, e_n] \\ &= \left[\sum_{j=1}^{k}\frac{1}{\beta_j}e_1, \ldots, \sum_{j=1}^{k}\frac{1}{\beta_j}e_n\right]\end{aligned}$$

Similarly, we may use

$$\text{cov}(m) = \left[\sum_{j=1}^{k}\frac{1}{\beta_j}e_1^j, \ldots, \sum_{j=1}^{k}\frac{1}{\beta_j}e_n^j\right]$$

In our simple case with $n = 2$ and $k = 4$, it becomes

$$\text{cov}(m) = \left[\frac{e_1^1}{\beta_1} + \frac{e_1^2}{\beta_2} + \frac{e_1^3}{\beta_3} + \frac{e_1^4}{\beta_4}, \ \frac{e_2^1}{\beta_1} + \frac{e_2^2}{\beta_2} + \frac{e_2^3}{\beta_3} + \frac{e_2^4}{\beta_4}\right]$$

We may choose

$$\alpha_1 = 13/18, \ \alpha_2 = 2\alpha_1, \ \alpha_4 = \alpha_3 = 3\alpha_1, \ \beta_j = 4$$
$$0 \le p_0 < 1, \ p_{4(j-1)+i} = p_{2(2j-1)+i} = (1 - p_0)/(16\alpha_j)$$
$$i = 1, 2, \quad j = 1, \ldots, 4$$

Note, however, that while this design has zero mean, its covariance is no longer equal to the identity matrix.

6.3 Equal-Distance Model-Set Design

The above symmetric-set designs do not have an even model distribution in space. In practice, it is sometimes desirable to have a set of models that are evenly distributed. For instance, this may be the case when each model is considered to be able to cover a region of the same size.

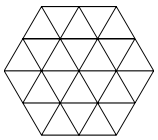


Fig. 6: Illustration of a diamond model-set design.

Diamond set. For this purpose, consider the diamond-set design illustrated in Fig. 6 for the 2D case. Note that the set of models on the whole diamond consists of hexagonal layers of models: 1 at the center (0th layer), 6 on the first layer (i.e., those on the unit circle), 12 on the second layer (6 of them are on the circle of radius 2), 18 on the third layer (6 of them are on the circle of radius 3), and so on. Alternatively, the model set may also be viewed as consisting of even finer (circle) layers of models: Models on each layer have equal distance from the origin (i.e., are on a circle of radius $0, 1, \sqrt{3}, 2, \sqrt{7}, 3, 2\sqrt{3}, \sqrt{13}, 4$, and so on, respectively). In general (the square of) the radii of these circles are given by

$$r_{ij}^2 = \begin{cases} (i\sqrt{3}/2)^2 + ((2j-1)/2)^2, i \text{ odd}, 1 \le j \le \frac{i+1}{2} \\ (i\sqrt{3}/2)^2 + (j-1)^2, i \text{ even}, \ 1 \le j \le i/2 + 1 \end{cases}$$

where the double subscript ij stands for the jth circle that passes through the models on the ith hexagonal layer, for example:

$$\begin{aligned} r_{11}^2 &= (1\sqrt{3}/2)^2 + (1/2)^2 = 1 \\ r_{21}^2 &= (2\sqrt{3}/2)^2 + 0^2 = 3, \ r_{22}^2 = (2\sqrt{3}/2)^2 + 1^2 = 2^2 \\ r_{31}^2 &= (3\sqrt{3}/2)^2 + (1/2)^2 = 7, \\ r_{32}^2 &= (3\sqrt{3}/2)^2 + (3/2)^2 = 3^2 \\ r_{41}^2 &= (4\sqrt{3}/2)^2 + 0^2 = 12, \ r_{42}^2 = (4\sqrt{3}/2)^2 + 1^2 = 13 \\ r_{43}^2 &= (4\sqrt{3}/2)^2 + 2^2 = 4^2 \end{aligned}$$

Clearly, this diamond set is symmetric and has equal distance between any two adjacent models. Furthermore, the following theorem states that this diamond-set design can also be used to match arbitrarily given mean and covariance of the mode by simply assigning each model on the same (hexagonal or circle) layer equal probability.

Theorem 6.5 (Diamond-set design). Consider a diamond-set design as illustrated in Fig. 6. Assign each model on the lth (hexagonal or circle) layer an equal probability p_l such that all probability masses sum up to unity. Let the total contribution to the covariance from the models on the lth layer be C_l. Then this diamond-set design satisfies (9) if $\sum_{l=1}^{k} C_l = I$, where k is the number of layers.

Remark. In particular, p_l and C_l can be chosen so that $C_l = C_r = I/k$ and every model has the same probability or the total probability mass of models on different layers are equal.

The simplest possible diamond-set design (with one at the center and six on the first layer) was implemented in [6] for an example of maneuvering target tracking using MM algorithms.

There are many equal-distance sets. In 3D for example, the well-known regular tetrahedron, cube, regular octahedron, regular dodecahedron, and regular icosahedron each leads to an equal-distance set design by placing a model at every vertex. However, the above diamond-set design is, on top of its regularity, attractive for several other nice properties, such as the ease for design (as stated in Theorem 6.5) and its economy in the sense of using a small number of models to cover a large region.

In reality, each model is effective only over a finite region. Call this region the *effective coverage region* of the model. Two natural questions are: Given the mode space $\mathbf{S}$ and the effective coverage region R_m of each model, what is the minimum number of models needed and where should the models be placed? Clearly, a lower bound on the number of models needed is $|M| \geq V_{\mathbf{S}}/V_m$, where $V_{\mathbf{S}}$ and V_m are the volumes of $\mathbf{S}$ and R_m, respectively. Assume that $\mathbf{S}$ and R_m are (n-dimensional) balls of radii $r_{\mathbf{S}}$ and r_m, respectively. Consider a diamond set in which for every diamond cell, each cell vertex to the cell center is r_m. Then every point in the inscribed ball B of the union of all models' R_m is covered by at least one R_m. It appears that this diamond set covers B using model coverage regions with the smallest number of models in general.

More generally, the diamond set has a small Hausdorff distance to the mode space relative to other (equal-distance) sets of the same number of models (vertices). For two (finite) sets A and B with a distance metric $d(x, y)$, $x \in A$ and $y \in B$, the Hausdorff distance between A and B is defined as $d(A, B) = \max\{\rho(A, B), \rho(B, A)\}$, where $\rho(A, B) = \sup_{x \in A} \inf_{y \in B} d(x, y)$. Note that the use of Hausdorff distance here — which corresponds to the worst case in distance between model and mode — is more reasonable than the more popular distance between two sets: $d(A, B) = \inf_{x \in A, y \in B} d(x, y)$ (which corresponds to the best case and is often zero for model-set design).

Equal-distance minimal-set design. The diamond set has many nice features, but it is not a minimal set. A minimal set with equal distance between models can be obtained by the minimal-set design of Fig. 3 with a special choice of $\{p, q, \alpha, \beta\}$ such that all models are separated by an equal distance. Clearly, $m_1, \ldots, m_n$ have an equal distance of $\sqrt{2}\alpha$. So, we need only to place m_{n+1} in a place such that its distance to every model in $\{m_1, \ldots, m_n\}$ is $\sqrt{2}\alpha$. Specifically, choose the set $\{p, q, \alpha, \beta\}$ of nonnegative numbers to satisfy

$$\begin{aligned} np + q + p_0 &= 1 \quad \text{(unity probability)} \\ \alpha p - \beta q &= 0 \quad \text{(zero mean)} \\ \alpha^2 p + \beta^2 q &= 1 \quad \text{(identity covariance)} \\ (\alpha + \beta)^2 + \beta^2(n-1) &= 2\alpha^2 \quad \text{(equal distance)} \end{aligned}$$

(the last equation above follows from setting $\|m_i - m_{n+1}\|^2 = \|m_i - m_j\|^2, \forall i, j \leq n$), which yields

$$q = \frac{1 - p_0}{\sqrt{n+1}},\ p = \frac{q + p_0 - 1}{n},\ \alpha = \frac{1}{\sqrt{p(1 + p/q)}},\ \beta = \frac{p}{q}\alpha$$

Then, the design of Fig. 3 has a minimal set that satisfies (9) and has equal distance between models. As such, this design places a model at each vertex of a convex $(n+1)$-hedron with equal edge length $\sqrt{2}\alpha$ (e.g., an equilateral triangle in 2D and a regular tetrahedron in 3D). Note, however, that m_{n+1} is closer to the origin than m_i ($i \leq n$) (i.e., the polyhedron is not centered at the origin) because $\|m_{n+1} - 0\|^2 = n\beta^2 < \alpha^2 = \|m_i - 0\|^2$.

7 Conclusions

Model-set design has been considered in a general setting. We have not only argued for the need for and the benefit of probabilistic modeling of the models to be designed as well as the true mode, but also proposed that they be modeled as random variables for offline model-set design. Based on such probabilistic models, we have presented (a) distribution-based, (b) minimum-distance based, and (c) moment-matching general and systematic design methods. Many theoretical results have been presented.

Several examples that demonstrate how these theoretical results can be used as well as their effectiveness are given in Part II [7].

Many of the general results presented in this paper are also useful for performance evaluation of MM algorithms.

References

[1] M. J. Caputi, "A Necessary Condition for Effective Performance of the Multiple Model Adaptive Estimator," *IEEE Trans. Aerospace and Electronic Systems*, AES-31(3):1132–1139, July 1995.

[2] X. R. Li, "Optimal Selection of Estimatee for Multiple-Model Estimation with Uncertain Parameters," *IEEE Trans. Aerospace and Electronic Systems*, AES-34(2):653–657, Apr. 1998.

[3] 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, ch. 10, pp. 499–567, Boston, MA: Artech House, 2000.

[4] X. R. Li and Y. Bar-Shalom, "Performance Prediction of Hybrid Algorithms," in *Control and Dynamic Systems: Advances in Theory and Applications*, vol. 72 (C. T. Leondes, ed.), vol. 72, pp. 99–151, New York: Academic Press, 1995.

[5] X. R. Li and Y. Bar-Shalom, "Multiple-Model Estimation with Variable Structure," *IEEE Trans. Automatic Control*, AC-41(4):478–493, Apr. 1996.

[6] X. R. Li and V. P. Jilkov, "Expected-Mode Augmentation for Multiple-Model Estimation," in *Proc. 2001 International Conf. on Information Fusion*, (Montreal, QC, Canada), pp. WeB1–3–WeB1–10, Aug. 7-10 2001.

[7] X. R. Li, Z.-L. Zhao, P. Zhang, and C. He, "Model-Set Design for Multiple-Model Estimation—Part II: Examples," in *Proc. 2002 International Conf. on Information Fusion* (Paper P1), (Annapolis, MD, USA), July 2002.

[8] S. N. Sheldon and P. S. Maybeck, "An Optimizing Design Strategy for Multiple Model Adaptive Estimation and Control," *IEEE Trans. Automatic Control*, AC-38(4):651–654, Apr. 1993.