A Paradigm for Class Identification Problems

Sanjeev R. Kulkarni, *Member, IEEE*, and David N. C. Tse, *Student Member, IEEE*

***Abstract*—The following problem arises in many applications involving classification, identification, and inference. There is a set of objects X, and a particular $x \in X$ is chosen (unknown to us). Based on information obtained about x in a sequential manner, we wish to decide whether x belongs to one class of objects A_0 or a different class of objects A_1. We study a general paradigm applicable to a broad range of problems of this type, which we refer to as problems of class identification or discernibility. We consider various types of information sequences, and various success criteria including discernibility in the limit, discernibility with a stopping criterion, uniform discernibility, and discernibility in the Cesaro sense. We consider decision rules both with and without memory. Necessary and sufficient conditions for discernibility are provided for each case in terms of separability conditions on the sets A_0 and A_1. We then show that for any sets A_0 and A_1, various types of separability can be achieved by allowing failure on appropriate sets of small measure. Applications to problems in language identification, system identification, and discrete geometry are discussed.**

***Index Terms*—Classification, discernibility, identification, learning, Cesaro sense, in the limit, stopping rule, uniform, memory, memoryless.**

I. Introduction

The following problem arises in many applications involving classification, identification, and inference. There is a set of objects X, and a particular $x \in X$ is chosen (unknown to us). Based on information obtained about x in a sequential manner, we wish to decide whether x belongs to one class of objects A_0 or a different class of objects A_1. For example, consider the following problems.

1) Consider the set of discrete-time, linear, time-invariant systems where our observation at each time consists of the output of the system due to some known input. Suppose our goal is to decide whether or not the system is stable.

2) Consider the set of languages over some finite alphabet {0, 1}. Suppose our observation at each time consists of whether or not the unknown language contains a particular word. The goal in this case might be to decide whether or not the unknown language is regular.

3) Consider the set of shapes (compact subsets) in the unit square. Suppose our observation at time n consists of the digitization of the unknown set on a lattice with spacing $1/n$, and our goal is to decide whether or not the unknown set is convex (or whether or not the set is connected). Alternatively, our observations might consist of a sequence of points in the unit square labeled as to whether or not they are contained in the unknown set, and our goal is again to decide whether or not the set is convex (or whether or not it is connected).

Although the particular applications are quite different, these problems are all fundamentally very similar. In this paper, we study a general paradigm which can be applied to a broad range of identification problems such as those mentioned above. We refer to problems of this type as class identification problems or discernibility problems (the term discernibility is coined in [5]). The results obtained provide necessary and sufficient conditions for the existence of decision procedures for such problems.

To address the question of whether a successful decision procedure exists for these types of problems, we first need a precise notion of "success." We consider four distinct success criteria. The first is that we simply require "discernibility in the limit"—that is, we require that, from some point, on the correct classification will always be made. This is a common criterion used in various applications. (For example, in the language identification literature, such a criterion is referred to as "identification in the limit." The difference here is that we do not require identifying the unknown exactly, but only identifying the correct class to which the unknown belongs.) With this criterion, we require only that we eventually stop making mistakes, but place no prior bound on when-convergence to the correct decision is made. Furthermore, it is not required that we *know* when convergence has been established.

One problem with not knowing when convergence has been established is that, although we eventually will make the correct classification, we will never be certain that we are correct. In many applications, one may not be satisfied with simply converging, but one would eventually like a guarantee that a correct decision has been made. Hence, the second criterion we consider is obtained by allowing discernibility in the limit, but requiring a "stopping rule." That is, no prior bound is placed on when we stop making mistakes, but we require that we know when we have locked onto the correct classification.

Manuscript received November 3, 1992; revised August 10, 1993. This work was supported in part by the Army Research Office under Grants DAAL03-86-K-0171 (Center for Intelligent Control Systems) and DAAL03-92-G-0320, by NSF Grants ECS-8552419 and IRI-9209577, and by an NSERC 1967 Fellowship from the Canadian Government. This paper was presented in part at the Conference on Information Sciences and Systems, Princeton University, Princeton, NJ, March 1992.

S. R. Kulkarni was with the Laboratory for Information and Decision Systems and the Center for Intelligent Control Systems, Massachusetts Institute of Technology Cambridge, MA 02139. He is now with the Department of Electrical Engineering, Princeton University, Princeton, NJ 08544.

D. N. C. Tse is with the Laboratory for Information and Decision Systems and the Center for Intelligent Control Systems, Massachusetts Institute of Technology, Cambridge, MA 02139.

IEEE Log Number 9401702.

0018-9448/94$04.00

Although with a stopping rule we will eventually know when we are making a correct decision, there is no prior bound on how long we must wait. The third and strongest criterion we consider is that of "uniform discernibility" obtained by requiring convergence to the correct decision after some fixed finite time.

Finally, we consider a success criterion even weaker than that of discernibility in the limit. This fourth and weakest criterion is "discernibility in the Cesaro sense" in which we require only that the asymptotic proportion of correct decisions converge to one. Hence, this criterion allows mistakes to be made infinitely often, but with density zero.

Given a success criterion, the question of whether there exists a decision procedure still depends on the nature of the information sequences as well as the classes A_0 and A_1. Clearly, success is more difficult if we require that classification be accomplished for many different information sequences. Also, classification will be more difficult if members of A_0 are in some sense very intermixed with members of A_1. Intuitively, successful classification will be possible only if the sets A_0 and A_1 are appropriately "separated" with respect to the information sequences. The main results make this idea precise by providing necessary and sufficient separability conditions on A_0 and A_1 for various information sequences.

Our motivation for this work comes from several directions. In [4], Cover considered the problem of deciding whether the mean of an unknown random variable belongs to a countable set A or its complement given a sequence of i.i.d. observations. He showed that, as long as the unknown mean does not belong to a subset of A^c of measure zero, a decision procedure can be constructed which almost surely makes only a finite number of mistakes. Some refinements and extensions of Cover's results were considered by Koplowitz [8] and Kulkarni and Zeitouni [11]. A much more general statistical decision problem was considered by Zeitouni and Kulkarni in [18]. Namely, they considered the problem of deciding whether the distribution itself (rather than just the mean) belongs to a certain class of distributions or its complement. A similar success criterion was used, and sufficient conditions which allow a decision procedure were provided. Dembo and Peres [5] provided necessary and sufficient conditions for the general decision problem considered in [18]. As a tool for the statistical problem, Dembo and Peres considered a completely deterministic formulation in a general metric space setting. Their deterministic formulation and corresponding results serve as the basis for the present work. Our results provide extensions and applications of the basic paradigm formulated in [5], and most of our results without memory were presented in [10]. Ben-David [2] has also considered a paradigm very similar to [5] and the paradigm considered here.

The present work is also motivated by the recent work of Tse *et al.* [16] in system identification, as well as by the large amount of work in language identification (e.g., see [3], [6], [12]). The criteria considered here are very similar to those used in the work mentioned above. One of our goals is to provide a general paradigm which encompasses results of similar flavor in these and other areas.

In the next section, we give precise descriptions of the paradigm, some classes of information sequences, and success criteria. Necessary and sufficient conditions for discernibility are provided in Sections III and IV for various information sequences and for the four success criteria—discernibility in the limit, discernibility with a stopping criterion, uniform discernibility, and discernibility in the Cesaro sense. In Section V, we show that for any (Borel measurable) sets A_0 and A_1, the discernibility conditions can be satisfied if we allow failure on appropriate sets with small measure. In Section VI, we describe a formulation which we show to be a special case of the general formulation, but which displays the broad applicability of the paradigm. Several applications are discussed in Section VII. Namely, we consider examples from system identification, language identification, and geometry. Finally, in Section VIII, we mention some directions for further work.

II. Discernibility Paradigm

A. Problem Setup

We model the set of objects as a metric space X with metric ρ. We have two subsets A_0 and A_1. A particular object $x \in A_0 \cup A_1$ is chosen unknown to us. At each time $i = 1, 2, \cdots$, we observe an element $x_i \in X$ providing information about the unknown x. Such a sequence of observations $x_1 x_2, \cdots$ is called an *information sequence* for x. The choice of which sequences constitute information sequences for x depends on the particular application. For example, in some problems, we may obtain approximations to the unknown x, but may not have bounds on how good our approximation is at any particular time. In this case, we could take the set of valid information sequences to be the set of all sequences converging to x. If bounds on the approximation are available, then the set of valid information sequences might consist of all sequences converging to x at some known rate. We discuss these and other information sequences in more detail below, but we now turn to formalizing the decision rule.

At time n, the observation consists of an element $x_n \in X$. Our decision rule at time n will be described by a function $h_n: X \to \{0, 1\}$. We decide $x \in A_0$ if $h_n(x_n) = 0$ and $x \in A_1$ otherwise. (For convenience, we will refer to a decision rule as h_n with the understanding that h_n represents a sequence of decision rules for $n = 1, 2, \cdots$.) Note that the decision at time n is a function of only x_n, and does not depend on the previous observations $x_1, x_2, \cdots, x_{n-1}$. Thus, in this sense, the decision rule is required to be *memoryless*. Results on memoryless decision rules are provided in section III.

We also consider the use of decision rules *with memory*. In this case, the decision rule at time n will be described by a function $h_n: X^n \to \{0, 1\}$. We decide $x \in A_0$ if

$h_n(x_1, \cdots, x_n) = 0$ and $x \in A_1$ otherwise. Thus, for decision rules with memory, the decision at time n is allowed to depend on all the observations up to time n—namely $x_1, \cdots, x_n$. Results on decision rules with memory are provided in Section IV.

B. Discernibility Criteria

We consider four distinct classification criteria. The first three criteria are successively stronger, while the fourth is the weakest criterion. The first criterion we consider is "classification in the limit"—that is, we require that, from some point on, the correct classification will always be made. Formally, we say that A_0 and A_1 are *discernible in the limit* with respect to a class of information sequences if there is a decision rule h_n such that, for every $x \in A_j$, $j \in \{0, 1\}$, and every information sequence $x_1, x_2, \cdots$ for x, we have $\lim_{n \to \infty} h_n(x_n) = j$. Note that, since $j \in \{0, 1\}$, the limit condition is equivalent to requiring that there exist some N such that $h_n(x_n) = j$ for all $n \geq N$, i.e., that we make the correct classification from some point on. Note that, as mentioned in the Introduction, with this criterion, we require only that we eventually stop making mistakes, but place no prior bound on when convergence to the correct decision is made. Furthermore, it is not required that we *know* when convergence has been established.

This criterion has been used in various applications. This is a deterministic version of the criterion used by Cover [4] and others building on his work [5], [8], [11], [18]. In fact, as mentioned in the Introduction, the present formulation with this criterion is exactly the formulation of Dembo and Peres [5]. A similar criterion has also been used for some time in the language identification literature [3], [6], [12], and more recently in system identification [16].

The second criterion we consider is obtained by allowing discernibility in the limit, but requiring a "stopping rule." That is, no prior bound is placed on when we stop making mistakes, but we require that we know when we have locked onto the correct classification. Formally, a stopping rule will be a set of functions $\tau_n: X \to \{0, 1\}$ such that, for every information sequence $x_1, x_2, \cdots, \tau_n(x_n)$ is nondecreasing and $\lim_{n \to \infty} \tau_n(x_n) = 1$. These conditions require that eventually $\tau_n(x_n)$ becomes 1. We say that A_0 and A_1 are *discernible with a stopping rule* if there is a decision rule h_n and a stopping rule τ_n satisfying the above criteria such that $h_n(x_n)$ makes the correct classification whenever $\tau_n(x_n) = 1$. Thus, $\tau_n(x_n) = 1$ signifies that the learner knows that the correct classification has been achieved. In the case of decision rules with memory, the stopping rule (as with the decision rule h_n) is allowed to depend on all the past observations $x_1, x_2, \cdots, x_n$, so that $\tau_n: X^n \to \{0, 1\}$.

The third and strongest criterion we consider is that of "uniform discernibility" obtained by requiring convergence to the correct decision by some fixed finite time which holds for all unknown elements in the two sets. Namely, we say that A_0 and A_1 are *uniformly discernible* if there is a decision rule h_n and some $N < \infty$ such that, for every $x \in A_j$, $j \in \{0, 1\}$, and every information sequence for x, we have $h_n(x_n) = j$ for all $n \geq N$.

Finally, the fourth and weakest criterion we consider is "discernibility in the Cesaro sense." In this case, we require only that the proportion of correct decisions converges to 1. In mathematical terms, we say that A_0 and A_1 are *discernible in the Cesaro sense* if there is a decision rule h_n such that, for every $x \in A_j$, $j \in \{0, 1\}$, and every information sequence for x, we have $\lim_{n \to \infty} (1/n) \sum_{k=1}^{n} h_k(x_k) = j$. Thus, infinitely many mistakes are allowed, but the percentage of times that an error is made must be zero. One could also consider an even weaker criterion in which the Cesaro sum need not converge to 0 or 1, but rather is allowed to converge to any two different numbers. Our proofs for discernibility in the Cesaro sense are valid for this weaker criterion as well.

C. Information Sequences

For each object x let $I(x)$ be the set of the possible information sequences that can be observed for x. Discernibility is required for all sequences in $I(x)$. There are several natural classes of information sequences that we consider.

1) Convergent sequences with known rate: here, the sequences in $I(x)$ converge to x with known rate $\{\epsilon_n\}(\epsilon_n \to 0)$, i.e., if $(x_1, x_2, \cdots) \in I(x)$, then $\rho(x_n, x) \leq \epsilon_n$ for all n.

2) Convergent sequences with *eventually* known rate: here, every sequence will eventually converge to x with known rate $\{\epsilon_n\}$ (where $\epsilon_n \to 0$), i.e., for every sequence $(x_1, x_2, \cdots) \in I(x)$, there is a time N such that, for all $n > N$, $\rho(x_n, x) \leq \epsilon_{n-N}$. Hence, all sequences will start converging to x with known error rates after an unknown but finite time, which can be treated as transients.

3) All convergent sequences: here, the information sequences are any sequence which converge to x. Since there are no known error bounds, the rate of convergence can be arbitrarily slow.

4) Convergent to within ϵ: here, there is some known ϵ such that, for all sequences $(x_1, x_2, \cdots) \in I(x)$, $\limsup_{n \to \infty} \rho(x_n, x) < \epsilon$. This may serve as a model for noisy observations, where the noise is completely arbitrary but bounded, so that even when there are more and more observations, there is some intrinsic inaccuracy that cannot be removed.

Classes 1) and 3) were considered in [5]. Clearly, these classes of information sequences are progressively more general, and so the difficulty of discernibility increases correspondingly.

III. Results on Discernibility Without Memory

In this section, we give necessary and sufficient conditions for discernibility under the four criteria and for the four possible classes of information sequences when the decision rule is memoryless. The results of this section are summarized in Table I.

TABLE I
SUMMARY OF DISCERNIBILITY RESULTS. A "?" INDICATES THE CONDITIONS ARE UNKNOWN. IN EACH CASE, UNLESS OTHERWISE SPECIFIED, THE CONDITIONS FOR DISCERNIBILITY WITH AND WITHOUT MEMORY ARE THE SAME.

	Known Rate	Eventually Known Rate	Convergent	Convergent to within ϵ
Cesaro Sense	Memoryless: ? Memory: ?	Memoryless: ? Memory: ?	Open Sep.	Nec.: $\rho(A_0, A_1) > \epsilon$ Suff.: $\rho(A_0, A_1) > 2\epsilon$
In the Limit	F_σ Sep. Memory: ?	F_σ Sep. Memory: ?	Open Sep.	Nec.: $\rho(A_0, A_1) > \epsilon$ Suff.: $\rho(A_0, A_1) > 2\epsilon$
Stopping Rule	Open Sep.	Never	Never	Never
Uniform	Positive Sep.	Never	Never	Never

A. Discernibility in the Limit

Proposition 1 [5]: A_0 and A_1 are discernible in the limit for all convergent sequences with known rate if and only if they can be covered by disjoint F_σ sets. (F_σ sets are those which are countable union of closed sets.)

Since the covers are F_σ, they can be decomposed into countable unions of increasing closed sets. These decompositions can be viewed as complexity hierarchies for the sets: objects in the higher numbered subsets of the decomposition are regarded as more "complex." The discerning rule guarantees convergence to the correct hypothesis by always picking the set containing the simplest element consistent with the observed data and the error bound. Conversely, a given discerning rule partitions the two sets into complexity hierarchies, with the simpler elements being those that can be discerned in a shorter time by the given rule.

Although discerning for all convergent sequences with eventually known rate is, in general, more difficult than with exact known rates, the result below says that it is not so if the requirement is identification in the limit.

Proposition 2: A_0 and A_1 are discernible in the limit for all convergent sequences with eventually known rate if and only if they can be covered by disjoint F_σ sets.

Proof: Discernibility for all convergent sequences with eventually known rate implies discernibility for all convergent sequences with known rate, which in turn implies F_σ separability.

For the other direction, suppose $A_0 \subset \bigcup_{i=1}^{\infty} B_i$ and $A_1 \subset \bigcup_{i=1}^{\infty} C_i$ where the B_i and C_i are closed and their unions are disjoint. Suppose $x_1, x_2, \cdots$ converges to x with eventually known rate ϵ_n. Then there exists N such that $\rho(x_n, x) < \epsilon_{n-N}$ for all $n > N$. Define

$$K_0(x_n, n, k) = \inf\{i \geq 1 \mid \rho(x_n, B_i) < \epsilon_{n-k}\}$$

and

$$K_0(x_n, n) = \min_{k \leq n}\{K_0(x_n, n, k) + k\}.$$

We can define $K_1(x_n, n, k)$ and $K_1(x_n, n)$ similarly, with C_i replacing B_i. Our decision rule is then to decide A_0 if $K_0(x_n, n) < K_1(x_n, n)$ and to decide A_1 otherwise. To show that this decision rule is correct in the limit, note that if $x \in A_0$, then $x \in B_m$ for some m. Then $K_0(x_n, n) \leq m + N < \infty$ for all n. On the other hand, for each i, x is a positive distance from C_i so that, eventually, $K_1(x_n, n) > i$ for all sufficiently large n. Hence, for all sufficiently large n, we will make the correct decision that $x \in A_0$. A similar argument holds if $x \in A_1$. □

If the convergent rates are not known at all, then we need stronger separation between the two sets.

Proposition 3 [5]: A_0 and A_1 are discernible in the limit for all convergent sequences if and only if they can be covered by disjoint open sets.

Proposition 4: A distance between A_0 and A_1 of at least 2ϵ is sufficient, and a distance of at least ϵ is necessary for A_0 and A_1 to be discernible in the limit for all sequences which converge to within ϵ. (Distance between two sets is the infimum of the distances between points in the two sets.)

Proof: If $\rho(A_0, A_1) > 2\epsilon$, then consider the decision rule which decides A_0 if $\rho(x_n, A_0) < \epsilon$ and decides A_1 otherwise. This decision rule clearly discerns A_0 and A_1 since every information sequence converges to within ϵ of the true x, and hence will eventually be within ϵ of the correct A_j and will be further than ϵ from the incorrect A_j.

For the other direction, if $\rho(A_0, A_1) < \epsilon$, then there exist $x_0 \in A_0$ and $x_1 \in A_1$ with $\rho(x_0, x_1) < \epsilon$. Thus, the sequence $x_0, x_0, x_0, \cdots$ is a valid information sequence for both x_0 and x_1. For this information sequence, we cannot discern between A_0 and A_1. □

If X is a convex subset of a linear space, then a distance of 2ϵ between A_0 and A_1 is both necessary and sufficient for discernibility in the limit for all sequences converging to within ϵ. However, in the general case, a separation of 2ϵ is not necessary.

Note that, as expected, the separability condition gets progressively stronger as discernibility becomes more difficult.

B. Discernibility with a Stopping Rule

We first prove a lemma which will be useful in several of our later proofs.

Lemma 1: Two sets A_0 and A_1 can be covered by disjoint open sets if and only if there is an open set containing A_0 that is disjoint from A_1 and there is an open set containing A_1 that is disjoint from A_0.

Proof: The "only if" direction is obvious. To prove the other direction, let C_0 be an open set containing A_0 and disjoint from A_1, and let C_1 be an open set containing A_1 and disjoint from A_0. For each $x \in A_i$, $i = 0, 1$, let $B(x, \epsilon_x)$ be an open ball centered at x and radius ϵ_x such that $B(x, \epsilon_x)$ is contained in C_i. Now, consider the sets $D_i \equiv \bigcup_{x \in A_i} B(x, \epsilon_x/2)$, $i = 0, 1$. Clearly, these two sets are open and they cover A_0 and A_1. It can also be seen that they are disjoint, for if they intersect, this implies that there are points $u \in A_0$ and $v \in A_1$ such that $B(u, \epsilon_u/2) \cap B(v, \epsilon_v/2) \neq \emptyset$. But this implies that $v \in B(u, \epsilon_u)$, a contradiction since $B(u, \epsilon_u)$ is contained in C_0 and C_0 is disjoint from A_1. Hence, D_0, D_1 provide the desired disjoint open covers. □

Proposition 5: A_0 and A_1 are discernible with a stopping rule for all convergent sequences with known rate if and only if they can be covered by disjoint open sets.

Proof: Suppose they are covered by disjoint open sets B_0 and B_1, respectively. Then stop and decide A_i when the ϵ_n-ball around x_n lies entirely inside the set B_i. This will eventually happen since the set B_i is open and the true element x lies inside it.

Now, consider the converse. Suppose there is a stopping rule for discerning between A_0 and A_1. Take any $x \in A_0$. Let $\{x_n\}$ be a convergent sequence to x with known rate $\{\epsilon_n\}$, and suppose the stopping rule stops and decides A_0 at time K. Let B_x be the ball of radius ϵ_K around x_K. By definition, this ball contains x. We claim that $B_x \cap A_1 = \emptyset$. If not, let $y \in B_x \cap A_1$. Then $x_1, \cdots, x_K, y, y, \cdots$ is an information sequence converging to y with known rate $\{\epsilon_n\}$ and such that $y_K = x_K$. Then the stopping rule would stop at y_K and decide A_0, which is incorrect. Hence, $B_x \cap A_1 = \emptyset$. Now, $\bigcup_{x \in A_0} B_x$ is an open cover of A_0, and it is disjoint from A_1. Similarly, we can construct an open cover of A_1 that is disjoint from A_0. By Lemma 1, A_0 and A_1 can be covered by disjoint open covers. □

The separability condition weakens if we require the existence of a stopping rule only for elements in one of the two sets.

Proposition 6: A_0 and A_1 are discernible in the limit, and with a stopping rule if the true object is in A_0, for all convergent sequences with known rate if and only if A_0 can be covered by an open set disjoint from A_1.

Proof: This follows directly from the proof of Proposition 5. □

Proposition 7: For the information sequence classes of convergent sequences with eventually known rate, all convergent sequences, and all convergent sequences to within ϵ, discernibility with a stopping criterion between any two distinct nonempty sets is impossible.

Proof: The main idea is that, for these sequences, convergence can start after an arbitrarily long time. Hence, after the stopping rule stops and decides on a particular class, an adversary can have the sequence starting to converge to an element in the other class. □

C. Uniform Discernibility

As expected, even stronger conditions are required for uniform discernibility.

Proposition 8: A_0 and A_1 are discernible uniformly for all convergent sequences with known rate if and only if the distance between them is positive.

Proof: Suppose $\rho(A_0, A_1) = \alpha > 0$. Choose N such that $\epsilon_N < \alpha/2$. Then we can uniformly discern A_0 and A_1 by time N by simply deciding the class which is closest to the observation.

For the other direction, suppose we have a decision rule which uniformly discerns A_0 and A_1 by time N for all sequences converging with known rate ϵ_n. Then an ϵ_N ball around any element of A_0 cannot contain an element of A_1 since, otherwise, we would not be able to discern between these two elements in A_0 and A_1, respectively. Hence, $\rho(A_0, A_1) \geq \epsilon_N > 0$. □

Proposition 9: For the information sequence classes of convergent sequences with eventually known rate, all convergent sequences, and all convergent sequences to within ϵ, uniform discernibility between any two distinct nonempty sets is impossible.

Proof: This follows from Proposition 7 since requiring uniform discernibility is even stronger than having a stopping rule. □

D. Discernibility in Cesaro Sense

Although discerning in the Cesaro sense is, in general, a weaker notion than discerning in the limit, it turns out that they are equivalent for certain types of information sequences.

Proposition 10: A_0 and A_1 are discernible in the Cesaro sense for all convergent sequences if and only if they can be covered by disjoint open sets.

Proof: Since separability by open sets is sufficient for discerning in the limit, it is also sufficient for discerning in the Cesaro sense. So we need only prove that it is also a necessary condition.

Suppose the two sets are not separable by open sets, but there is a discerning rule h_n for all convergent sequences. Then, by Lemma 1, we can assume, without loss of generality, that there is no open set containing A_0 and disjoint from A_1. Hence, the closure of A_1 intersects A_0 (otherwise, the complement of the closure of A_1 would be such an open set). Let x_0 be a point in this intersection, and let $\{y_k\}$ be a sequence of points in A_1 converging to x_0. Given this sequence, construct an information sequence $\{x_n\}$ as follows. Start with $x_1 = y_1$, and continue presenting y_1 until the decision rule h_n says that the unknown point is in A_1 enough times that the Cesaro sense of the number of times the decision is A_1 is larger than $1/2$. (This will eventually happen because y_1 is in A_1.) Now, let the next observation be y_2, and again repeat until the decision rule says that the unknown point is in A_1 enough times that the Cesaro sense is again greater than $1/2$. Continue this procedure. Notice that the resulting information sequence converges to $x_0 \in A_0$, but the discerning rule decides A_1 with Cesaro sense greater than $1/2$ infinitely many times, thus contradicting the discernibility of A_0 and A_1 in the Cesaro sense. □

Proposition 11: A distance between A_0 and A_1 of at least 2ϵ is sufficient and a distance of at least ϵ is necessary for A_0 and A_1 to be discernible with memory in the Cesaro sense for all sequences which converge to within ϵ.

Proof: The proof is identical to that of Proposition 4. □

The problem of finding necessary and sufficient conditions for discernibility in the Cesaro sense for the other two types of information sequences is still open. We have the following general result which may be useful in finding necessary conditions for discernibility in the Cesaro sense. However, as mentioned in the next section, the corresponding problems for discernibility in the limit with memory are also open.

Proposition 12: For any type of information sequence, discernibility in the Cesaro sense without memory implies discernibility in the limit with memory.

Proof: Suppose there is a memoryless rule h_n which discerns in the Cesaro sense. Then one can construct a rule $\tilde{h}_n$ with memory as follows: at time n, set

$$\tilde{h}_n(x_1,\cdots,x_n) = 0 \qquad \text{if } \frac{1}{n}\sum_{k=1}^{n} h_k(x_k) < \frac{1}{2} \quad \text{and}$$

$$\tilde{h}_n(x_1,\cdots,x_n) = 1 \qquad \text{otherwise.}$$

Since the decision of h_n converges in the Cesaro sense, the corresponding decision of $\tilde{h}_n$ converges in the limit. □

IV. Results on Discernibility with Memory

In this section, we give results on discernibility using decision rules with memory. The results of this section (as with those in Section III) are summarized in Table I.

A. Discernibility in the Limit

Proposition 13: A_0 and A_1 are discernible with memory in the limit for all convergent sequences if and only if they can be covered by disjoint open sets.

Proof: Since separability by open sets is sufficient for discerning without memory, it is also sufficient for discerning with memory. So we need only to prove that it is also a necessary condition.

Suppose the two sets are not separable by open sets, but there is a discerning rule h_n with memory for all convergent sequences. By Lemma 1, we can assume, without loss of generality, that there is no open set containing A_0 and disjoint from A_1. Hence, the closure of A_1 intersects A_0 (otherwise, the complement of the closure of A_1 would be such an open set). Let x_0 be a point in this intersection, and let $\{y_k\}$ be a sequence of points in A_1 converging to x_0. Given this sequence, construct an information sequence $\{x_n\}$ as follows. Start with $x_1 = y_1$, and continue presenting y_1 until the decision rule h_n says that the unknown point is in A_1. (This will eventually happen because y_1 is in A_1.) Now, let the next observation be y_2, and again repeat until the decision rule says the unknown point is in A_1. Continue the procedure. Notice that the resulting information sequence converges to $x_0 \in A_0$, but the discerning rule decides A_1 infinitely many times, thus contradicting the discernibility of A_0 and A_1. □

Proposition 14: A distance between A_0 and A_1 of at least 2ϵ is sufficient and a distance of at least ϵ is necessary for A_0 and A_1 to be discernible with memory in the limit for all sequences which converge to within ϵ.

Proof: The proof of the necessary part of Proposition 4, in fact, holds regardless of whether or not the decision rule has memory. Hence, the conditions are the same. □

The problems of discernibility in the limit with memory for sequences with known rate and with eventually known rate are open. However, we conjecture that, even in these cases, the conditions for discernibility are the same as in the memoryless case.

B. Discernibility with a Stopping Rule

Here, we consider the case of discernibility with a stopping rule, when both the discerning rule and the stopping rule have memory and can decide based on all the past observations.

Proposition 15: A_0 and A_1 are discernible with memory with a stopping rule for all convergent sequences with known rate if and only if they can be covered by disjoint open sets.

Proof: If A_0 and A_1 are separable by disjoint open sets, then they are discernible without memory, and hence also with memory.

For the other direction, suppose we have a discerning rule with a stopping criterion. By Lemma 1, we need only show that there is an open set containing A_0 that is disjoint from A_1, and similarly an open set containing A_1 that is disjoint from A_0. Take any $x \in A_0$. Let $\{x_n\}$ be a convergent sequence to x with known rate $\{\epsilon_n\}$, and suppose the decision rule stops and decides A_0 at time K. Let

$$B_x = \bigcap_{i=1}^{K} B(x_i, \epsilon_i)$$

where $B(x_i, \epsilon_i)$ is the ball of radius ϵ_i centered at x_i. B_x is an open set, and by definition of the sequence x_n, B_x contains x. Then, as in the proof of Proposition 5, we claim that $B_x \cap A_1 = \varnothing$. If not, then as before for $y \in B_x \cap A_1$, the sequence $x_1,\cdots,x_K,y,y,\cdots$ is an information sequence for y converging with rate $\{\epsilon_n\}$. But then the decision rule would stop at y_K and decide A_0, which is incorrect. Therefore, $\bigcup_{x\in A_0} B_x$ is an open cover of A_0 that is disjoint from A_1. Similarly, we can construct an open cover of A_1 that is disjoint from A_0. □

Proposition 16: For the information sequence classes of convergent sequences with eventually known rate, all convergent sequences, and all convergent sequences to within ϵ, discernibility with a stopping criterion with memory between any two distinct nonempty sets is impossible.

Proof: The previous adversary arguments hold even for decision rules with memory. □

C. Uniform Discernibility

Proposition 17: A_0 and A_1 are uniformly discernible with memory for all convergent sequences with known rate if and only if the distance between them is positive.

Proof: If A_0 and A_1 are positively separated, the result follows from the corresponding result for the memoryless case (Proposition 8).

For the other direction, we will show that if A_0 and A_1 are not positively separated, then for any finite N, we cannot uniformly discern with memory by time N. Let $\epsilon = \min_{1<i<N} \epsilon_i$. If A_0 and A_1 are not positively separated, then there exist $x_0 \in A_0$ and $x_1 \in A_1$ such that $\rho(x_0, x_1) < \epsilon$. But then an information sequence which consists of $x_0, \cdots, x_0$ up to time N is a valid information sequence with rate ϵ_n (up to time N) for both x_0 and x_1. Hence, no decision rule can discern between A_0 and A_1 (i.e., x_0 and x_1) by time N on this information sequence. □

Proposition 18: For the information sequence classes of convergent sequences with eventually known rate, all convergent sequences, and all convergent sequences to within ϵ, uniform discernibility between any two distinct nonempty sets is impossible.

Proof: Again, the usual adversary arguments hold even with memory. □

D. Discernibility in the Cesaro Sense

Proposition 19: A_0 and A_1 are discernible with memory in the Cesaro sense for all convergent sequences if and only if they can be covered by disjoint open sets.

Proof: The proof of Proposition 10 holds even if the decision rule has memory. □

Proposition 20: A distance between A_0 and A_1 of at least 2ϵ is sufficient and a distance of at least ϵ is necessary for A_0 and A_1 to be discernible with memory in the Cesaro sense for all sequences which converge to within ϵ.

Proof: The proof of Proposition 4 is, in fact, valid whether or not the decision rule has memory and for a Cesaro sense criterion. □

The problems of discernibility in the Cesaro sense for sequences with known rate and with eventually known rate are open.

V. Approximate Discernibility

In this section, we will consider weaker notions of discernibility where we only require identification to succeed for "most" but not all elements of the two sets. First, we state the following result [5] which shows that there are sets which cannot be completely discerned, even when given information sequences with known rates of convergence. This result is based on the Baire Category Theorem (see, for example, [13]).

Proposition 21: Let X be a complete metric space. If a set and its complement are both dense in X, then they cannot both be F_σ.

For example, suppose $X = [0, 1]$. The rationals and the irrationals are both dense in X, and therefore they cannot be both F_σ. (In fact, the irrationals are not F_σ.) Hence, one cannot discern between the rationals and irrationals, even when given information sequences with known rates.

However, suppose now that we have a Borel probability measure P on X which models the likelihood of seeing an element in X. Then we can "almost" separate any two sets by disjoint F_σ sets. Hence, we can discern in the limit any two sets except for a set of measure zero for all sequences with known rate (or even eventually known rates). (There are similar results in [5].)

Proposition 22: Let A_0 and A_1 be any two disjoint Borel-measurable sets in X. Then there is a set C of P-measure 0 such that $A_0 - C$ and $A_1 - C$ are separated by disjoint F_σ sets. Furthermore, if A_0 is F_σ, then we can take C to be a subset of A_1.

Proof: For any finite Borel measure, any Borel-measurable set can be approximated, arbitrarily closely in measure, by a closed set from within [13]. Hence, there exist F_σ sets A'_0 and A'_1, contained in A_0 and A_1, respectively, such that $P(A_0/A'_0) = 0$ and $P(A_1/A'_1) = 0$. Take $C = (A_0/A'_0) \cup (A_1/A'_1)$ to be the required set. In particular, if A_0 is F_σ, then one can take $A'_0 = A_0$, and C is a subset of A_1. □

For example, it follows that, although we cannot discern between the rationals and the irrationals in [0, 1], there is a discerning rule which fails on only a set of Lebesgue measure 0. Furthermore, since the rationals are F_σ, we can choose the failure set to contain only irrationals.

If we want to discern in the limit any two sets using convergent sequences without known rates, then we need to tolerate failure on slightly bigger sets, and similarly if we want to discern with a stopping rule and/or uniformly. The following results show that, if we allow failure on a set of positive measure, then we can get stronger separability on the remaining portions of A_0 and A_1.

Proposition 23: Let A_0 and A_1 be any two disjoint Borel-measurable sets in X. Then for every $\epsilon > 0$, there is a set A_ϵ of P-measure less then ϵ such that $A_0 - A_\epsilon$ and $A_1 - A_\epsilon$ are separable by disjoint open sets.

Proof: As before, we appeal to Lemma 1. Simply pick closed sets A'_0 and A'_1 in A_0 and A_1, respectively, such that $P(A_i/A'_i) < \epsilon/2$, $i = 1, 2$. Clearly by Lemma 1, A'_0 and A'_1 can be covered by disjoint open sets. □

If the space X satisfies stronger conditions, then we can even get uniform discernibility by removing small sets.

Proposition 24: Suppose X is complete and separable. Let A_0 and A_1 be any two disjoint Borel-measurable sets in X. Then for every $\epsilon > 0$, there is a set A_ϵ of P-measure less then ϵ such that $A_0 - A_\epsilon$ and $A_1 - A_\epsilon$ are at a positive distance from each other.

Proof: By Ulam's Theorem, any finite Borel measure is inner regular, so we can approximate A_0 and A_1 by compact sets from the inside, which are at a positive distance from each other. □

VI. Functional Formulation

In information-based complexity theory [15], one is interested in estimating some attributes of an unknown object x from measurements which give partial information about x. The paradigm introduced can also encompass a similar formulation.

For $i = 1, 2, \cdots$, let $f_i: X \to Y$ be a given sequence of functions on X. The f_i can be viewed as *measurements* that provide partial information about x, and Y is the space of possible observed values of the measurement.

Given a measurement $y_i = f_i(x)$ on x, the information obtained is that the unknown x lies in the set $f_i^{-1}(y_i) \equiv \{z \in X:\ f_i(z) = y_i\}$. For the classification problem, we assume that, at time n, all the observations $y_1, y_2, \cdots, y_n$ up to time n are available. As before, the problem of interest is to decide whether the unknown x belongs to $A_0 \subset X$ or $A_1 \subset X$.

For example, x might be a function, while $f_n(x)$ is the value of the function x evaluated at a certain point. Given successively more function values of x, we may be interested in determining a certain attribute of the function x.

This formulation can be cast in the metric space setting which we have been discussing as follows. Consider the space $Y^\infty = Y \times Y \times \cdots$ consisting of all infinite sequences of elements from Y. Each element of X gives rise to a sequence in Y^∞ through the measurement operators—i.e., $x \in X$ gives rise to $(f_1(x), f_2(x), \cdots) \in Y^\infty$. Let $\mathscr{A}_0 \subset Y^\infty$ be the set of all sequences obtained from elements of A_0, so that

$$\mathscr{A}_0 = \{(f_1(x), f_2(x), \cdots) | x \in A_0\}.$$

Similarly, let $\mathscr{A}_1$ be defined by

$$\mathscr{A}_1 = \{(f_1(x), f_2(x), \cdots) | x \in A_1\}.$$

It can be seen that the classification problem in X using the measurement operators is equivalent to discerning between $\mathscr{A}_0$ and $\mathscr{A}_1$ in Y^∞ where our observation at time n consists of the partial sequence $(y_1, y_2, \cdots, y_n)$ of the unknown sequence $(y_1, y_2, \cdots)$.

To view this in the metric space setting, we define a natural metric on Y^∞ as follows. For two sequences $\bar{y} = (y_1, y_2, \cdots)$ and $\bar{z} = (z_1, z_2, \cdots)$ in Y^∞, let $I(\bar{y}, \bar{z}) = \min\{i | y_i \neq z_i\}$. Then define $\rho_{Y^\infty}(\bar{y}, \bar{z}) = 1/I(\bar{y}, \bar{z})$. It is easy to check that this is, in fact, a metric. With this metric, one can see that the observations we are making about the unknown sequence $\bar{y}$ are converging to $\bar{y}$ with a known rate of $1/n$.

VII. Examples

In this section, we will discuss applications of the above paradigm to some examples from discrete geometry, system identification, and language identification.

A. Geometric Problems

There are a number of problems in discrete geometry and image processing in which one is interested in studying discrete approximations of an underlying continuous object and/or extracting properties of an object from its discrete approximations (e.g., see [7], [9], [14], [17]). In the case where some estimate is made based on a digitized version of an object, a natural condition to impose is that the estimates converge as the digitization level gets finer and finer [9]. The classification paradigm can be used to address the existence of decision rules with various properties for many problems of this type.

For a specific example in discrete geometry, suppose we are given a binary image of some object in which a pixel in the digitization has value 1 if the object intersects the pixel and zero otherwise. If we are interested in deciding whether the underlying object is convex or not, a natural question is whether a decision rule can be found which will guarantee a correct decision for sufficiently fine digitizations. To apply the classification paradigm to this problem, we could take X to be the space of all compact subsets of the unit square. It can be easily shown that the digitized observations form a sequence converging with a known rate in Hausdorff metric to the underlying object (assuming we know the digitization level). Furthermore, it is well known that the set of convex sets is closed (and the set of nonconvex sets is open) with respect to the topology induced by the Hausdorff metric. Since, in a metric space, open sets are F_σ, Proposition 1 implies that this problem is discernible in the limit. Furthermore, it is discernible with a stopping criterion if the object is nonconvex, but no stopping criterion exists if the object is convex. The natural decision rule for this problem is to decide the object is convex if there exists some convex set that gives rise to the observed digitization, and to decide the object is nonconvex otherwise. We see that we can stop if we ever decide that the object is nonconvex.

B. System Identification

Let X be the space of all causal, discrete-time, linear, time-invariant systems. Suppose we observe the impulse response of an unknown system over time, and want to decide if the system belongs to a set A_0 or a set A_1, both subsets of X. This falls under the functional formulation discussed earlier since the value of the impulse response at time n can be regarded as a measurement on the unknown system. We can now consider the space $\boldsymbol{R}^\infty$ (all one-sided infinite sequences corresponding to the impulse responses) and define a metric on $\boldsymbol{R}^\infty$: the distance between two sequences is the reciprocal of the index of the first component on which they disagree. This metric is the one induced by the measurement operators, as discussed earlier. At each time n, the observation is a *vector* containing the first n components of the impulse response of the unknown system. Discerning between the two sets A_0 and A_1 in X is now equivalent to discerning between the two sets of corresponding impulse responses in $\boldsymbol{R}^\infty$ using observations converging at a rate of $1/n$. Note that, although at each time n the discerning rule has access to the first n components of the unknown impulse response, the discerning rule is still considered as memoryless in the functional formulation.

Suppose now A_0 is the set of all BIBO stable systems (i.e., those with absolutely summable impulse responses) and A_1 is the set of all unstable systems. Is there a rule which can discern between A_0 and A_1 in the limit by observing the impulse response of the system over time? The answer is no. Since $\boldsymbol{R}^\infty$ is a complete metric space under the above-defined metric and both A_0 and A_1 are dense in this metric space, it follows from Proposition 21 that A_0 and A_1 are not discernible in the limit.

However, consider now that we restrict ourselves to finite-dimensional plants (rational transfer functions), so that A_0 is now the set of all stable finite-dimensional plants and A_1 is the unstable finite-dimensional ones. It can now be seen that A_0 and A_1 *are* discernible in the limit. One needs only to see that the set of all stable (resp., unstable) systems of order $\leq k$ is closed for each k, so that the set A_0 (resp., A_1) is, in fact, F_σ. At each time n, the discerning rule computes the lowest ordered plant consistent with the observed data, and decides A_0 if that plant is stable and A_1 if that plant is unstable. This procedure is guaranteed to converge since, if the true plant is, say, stable, then the order of the lowest ordered stable plant consistent with the data has to be bounded, while that of the lowest ordered unstable plant consistent with the data will grow unbounded.

C. Language Identification

An area which has received much attention in the computer science community is language identification, and it is generally formalized as follows (e.g., see [3], [6], [12]). An unknown language is selected from a countable collection of languages $\mathscr{L} = L_1, L_2, \cdots$ over a finite alphabet, say $\{0, 1\}$. A sequence of strings are presented with labels as to whether or not they belong to the unknown language. The goal is to exactly identify which language from the class $\mathscr{L}$ was selected. The usual criterion is that of "identification in the limit," in which it is only required that we eventually stop making mistakes in our sequence of decisions. A number of variations arise, depending on the particular manner in which data are received (e.g., positive data only, arbitrary order of string presentation, etc.).

Our present class identification formulation is for deciding membership between two (or any finite number) of classes, and hence, can be used to address problems such as deciding whether or not an unknown language is regular (or context-free, recursively enumerable, etc.). For example, suppose the strings are presented in a fixed ordering. In this case, the problem is a special case of our functional setting in which, at time i, the measurement operator f_i returns a value 1 or 0, depending on whether or not the ith string is in the unknown language. (Recall that, as shown in Section VI, the functional setting can be cast in a metric space setting with information sequences converging at a known rate.) Since the class of regular languages and its complement are both dense, by Proposition 21, we cannot discern between regular and nonregular languages. Similarly, we cannot discern whether or not a language is context-free, or whether or not it is r.e., etc. However, by Proposition 22, if given a probability measure on the set of languages, then we *can* discern between these sets if we are allowed to fail on a set of measure zero. Similarly, if we are allowed to fail on a set of positive measure, then we can discern with a stopping criterion and/or uniformly, depending on the failure set.

Extensions of the class identification paradigm to a countable number of classes $A_0, A_1, \cdots$ may be possible. Such extensions would encompass the usual language identification paradigms by taking $A_i = \{L_i\}$. Furthermore, such a paradigm also can be used to reject many languages which are *not* in the countable collection $\mathscr{L}$. This idea is generally not considered in the language identification literature. Finally, we mention that the problem of language identification with arbitrary word orderings cannot be treated with the results obtained thus far. In this case, discernibility is required for information sequences which are *not* convergent with a known rate, but also do not consist of *all* convergent sequences. It would be interesting to characterize the set of convergent sequences in this case, prove general discernibility results for these sequences, and obtain conditions for language identification. In the countable case with $A_i = \{L_i\}$, these conditions would be equivalent to the condition obtained by Angluin [1]. In the case A_0 and A_1 (both possibly uncountable), we conjecture (which is the natural extension of Angluin's result to this case) that the conditions for discernibility are, for every $L \in A_0$, there exists a finite $D \subseteq L$ such that, if $D \subseteq L' \in A_1$, then $L' \subsetneq L$, and the analogous condition with A_0 and A_1 interchanged.

VIII. Concluding Remarks

Our results provide necessary and sufficient conditions for discernibility with respect to various information sequences and success criteria. The results are in terms of topological separation conditions on the two classes A_0 and A_1. It is interesting to note that, in all cases in which we have results, the conditions for discernibility in the limit without memory are identical to those for discernibility in the Cesaro sense with memory. Thus, allowing memory and allowing a Cesaro sense criterion does not make discernibility any easier. Furthermore, as noted in Section II, one could consider a criterion in which the Cesaro sum need only converge to distinct limits, depending on the class of the unknown object. Our proofs hold for this even weaker criterion as well. Hence, if it is possible to correctly decide an arbitrarily small fraction better than random guessing, then it is possible to correctly decide from some point on.

The examples presented in system identification, language identification, and discrete geometry show the broad applicability of the paradigm. It is our hope that this paradigm will serve as a first step towards unifying results in these and other areas.

There are a number of interesting directions for further work. First, there are several cases in which the problem of obtaining necessary conditions for discernibility are open—specifically, those for discernibility in the Cesaro sense with and without memory and in the limit with memory for sequences with known rate and sequences with eventually known rate. For several of these, we conjecture that the necessary condition is F_σ separability of A_0 and A_1. Second, all of our results are for discerning between two sets A_0 and A_1, although they extend immediately to a finite number of sets as well. A natural extension to consider (as done in [4], [8], [11], [18]) is discernibility between a countable number of sets $A_0, A_1, \cdots$. The work in [4], [5], [8], [11], [18] was for a specific stochastic problem. It would be interesting to consider a general probabilistic formulation or a formulation with "noisy" information sequences. Finally, we feel that it would be worthwhile to study the relationships between the present paradigm and other models in machine learning.

Acknowledgment

We would like to thank T. Cover, J. Tsitsiklis, and O. Zeitouni for helpful discussions, and an anonymous referee for some helpful suggestions.

References

[1] D. Angluin, "Inductive inference of formal languages from positive data," *Inform. Contr.*, vol. 45, pp. 117–135, 1980.

[2] S. Ben-David, "Can finite samples detect singularities of real-valued functions?," in *Proc. 24th ACM Symp. Theory of Computing*, May 1992, pp. 390–399.

[3] J. Case and C. Smith, "Comparison of identification criteria for machine inductive inference," *Theoret. Comput. Sci.*, vol. 25, pp. 193–220, 1983.

[4] T. M. Cover, "On determining the irrationality of the mean of a random variable," *Ann. Statist.*, vol. 1, pp. 862–871, 1973.

[5] A. Dembo, and Y. Peres, "A topological criterion for hypothesis testing," *Ann. Statist.*, vol. 22, pp. 106–117, 1994.

[6] E. M. Gold, "Language identification in the limit," *Inform. Contr.*, vol. 10, pp. 447–474, 1967.

[7] C. E. Kim and A. Rosenfeld, "Digital straight lines and convexity of digital regions," *IEEE Trans. Pattern. Anal. Machine Intell.*, vol. PAMI-4, pp. 149–153, 1982.

[8] J. Koplowitz, in *Abstr. of Papers, Int. Symp. Inform. Theory*, Cornell Univ., Ithaca, NY, Oct., 1977, p. 64.

[9] S. R. Kulkarni, "Problems of computational and information complexity in machine vision and learning," Ph.D. dissertation, Dep. E.E.C.S., M.I.T., 1991.

[10] S. R. Kulkarni and D. N. C. Tse, "A paradigm for class identification problem," in *Proc. 1992 Conf. Inform. Sci. Syst.*, Mar., 1992, pp. 442–447.

[11] S. R. Kulkarni and O. Zeitouni, "Can one decide the type of the mean from the empirical measure?," *Statist. Prob. Lett.*, vol. 12, pp. 323–327, 1991.

[12] D. N. Osherson, M. Stob, and S. Weinstein, *Systems that Learn.* Cambridge, MA: M.I.T. Press, 1986.

[13] H. L. Royden, *Real Analysis.* New York: Macmillan, 1988.

[14] D. Shaked, J. Koplowitz, and A. M. Bruckstein, "Star-shapedness of digitized planar shapes," *Contemp. Math.*, vol. 119, pp. 137–158, 1991.

[15] J. F. Traub, G. W. Wasilkowski, and H. Wozniakowski, *Information-Based Complexity.* New York: Academic, 1988.

[16] D. Tse, M. A. Dahleh, and J. N. Tsitsiklis, "Optimal asymptotic identification under bounded disturbances," *IEEE Trans. Automat. Contr.*, vol. 38, pp. 1176–1190, 1993.

[17] X. Wu and J. Rockne, "On properties of discretized convex curves," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 11, pp. 217–223, 1989.

[18] O. Zeitouni and S. R. Kulkarni, "A general classification rule for probability measures," submitted to *Ann. Statist.*, 1991.