WORST-CASE IDENTIFICATION OF LINEAR SYSTEMS: EXISTENCE AND COMPLEXITY

D.W. HADWIN*, K.J. HARRISON†, AND J.A. WARD‡

Abstract. In the identification of linear systems the aim is to estimate the impulse response to within a given tolerance based on a finite number of noisy observations of the output. Whether this is possible depends upon the model set, that is, the set of impulse responses to which that of the system is assumed to belong. We give conditions on the model set which ensure that such identification is possible and also briefly review recent results concerning the complexity of identification, that is, the minimum number of required output samples.

AMS subject classifications. 62G05,62G20.

1. Introduction. The objective of time-domain system identification is to estimate the unknown impulse response h of the system by measuring its response to a carefully chosen input signal u. The output signal y is the sum of the system response to u and a noise term η, that is,

$$y = u * h + \eta, \tag{1.1}$$

where $*$ denotes a convolution. It is assumed that h belongs to some prescribed model set $\mathcal{M}$, and an estimate h^* is made on the basis of a finite number of output values y and the known input signal u. Worst-case identification requires a prescribed upper bound on the approximation error, that is, the estimate h^* must differ from h in norm by no more than some specified amount. It also differs from stochastic identification [8] in that it presupposes no statistical properties of the noise η other than a uniform bound.

Such identification is possible only for certain types of model sets. In these cases the minimum number of sample output values required for identification has practical significance. This is the *complexity* of identification.

Worst-case identification has been studied in a variety of settings, involving discrete and continuous model sets, and various time domain and frequency domain based norms. We shall give conditions that guarantee that worst-case identification is possible within a given model set, and briefly describe known results concerning the *complexity* of identification in some commonly used model sets. These include the discrete finite impulse response (FIR) systems and discrete exogenous autoregressive (ARX) model sets, and their continuous counterparts.

The reader is also encouraged to read the review articles [10] and [15]. The first of these focuses on convergence issues and on the computation of optimal or near-optimal algorithms, and the second also addresses model validation and control relevancy issues.

2. Identification via functionals. In this section we give the framework for worst-case identification within an arbitrary model set. In the discrete case (1.1)

* Department of Mathematics, University of New Hampshire, USA. email:don@christa

† School of Mathematical and Physical Sciences, Division of Science, Murdoch University, Murdoch, WA 6150, Australia. email:harrison@prodigal.murdoch.edu.au

‡ School of Mathematical and Physical Sciences, Division of Science, Murdoch University, Murdoch, WA 6150, Australia. email:harrison@prodigal.murdoch.edu.au

becomes

$$y(k) = \sum_{j=0}^{k} u(k-j)h(j) + \eta(k) \text{ for } k \geq 0,$$

and for continuous systems it has the form

$$y(t) = \int_0^t u(t-\zeta)h(\zeta)\, d\zeta + \eta(t) \text{ for } t > 0.$$

In each case the estimate h^* is based upon the input u and a finite number of values of the output y. For discrete systems these are usually the first N values, and for continuous systems the sample points are typically successive integer multiples of a fixed number. The maps $h \to (u * h)(k)$ and $h \to (u * h)(t_k)$ are linear functionals, and so worst-case identification in both the discrete and continuous cases fits within a more general framework, which we now describe.

Suppose that $\mathcal{X}$ is a normed linear space and that $\varphi_1, \varphi_2, \ldots, \varphi_N$ are continuous linear functionals on $\mathcal{X}$. Suppose also that

$$y_k = \varphi_k(h) + \eta_k \text{ and } |\eta_k| \leq \delta \text{ for } 1 \leq k \leq N, \tag{2.1}$$

where h is an unknown element of a given subset $\mathcal{M}$ of $\mathcal{X}$. The aim is to estimate h on the basis of the noisy observations y_k of the functional values $\varphi_k(h)$. Problems of this type have been studied in [11].

The *feasibility set* $S(\varphi, y, \delta)$ contains all elements of $\mathcal{M}$ that satisfy (2.1) for some noise term η, and hence are consistent with the data $y = (y_1, y_2, \ldots, y_N)$, the finite set $\varphi = (\varphi_1, \varphi_2, \ldots, \varphi_N)$ of test functionals and the noise bound δ. Thus

$$S(\varphi, y, \delta) = \{h' \in \mathcal{M} : |y_k - \varphi_k(h')| \leq \delta \text{ for } 1 \leq k \leq N\}.$$

The elements of $S(\varphi, y, \delta)$ are the possible candidates for the true system response, and hence the size of this set determines the bounds for the worst-case identification error. Recall that the diameter and radius of a subset $\mathcal{K}$ of the normed space $\mathcal{X}$ are defined by

$$\text{diam}\mathcal{K} = \sup_{x_1, x_2 \in \mathcal{K}} ||x_1 - x_2|| \text{ and } \text{rad}\mathcal{K} = \inf_{c \in \mathcal{X}} \sup_{x \in \mathcal{K}} ||x - c|| .$$

A point $c \in \mathcal{X}$ for which $\sup_{x \in \mathcal{K}} ||x - c|| = \text{rad}\mathcal{K}$ is called a *centre* of $\mathcal{K}$. The diameter and radius are related by the inequalities

$$\frac{1}{2}\text{diam}\mathcal{K} \leq \text{rad}\mathcal{K} \leq \text{diam}\mathcal{K}.$$

Since each element of $S(\varphi, y, \delta)$ is a candidate for h, any bound for the set of possible identification errors $||h - h^*||$ must be at least as large as $\text{rad}S(\varphi, y, \delta)$.. This lower limit is achieved if h^* is a centre of $S(\varphi, y, \delta)$, and algorithms that produce such estimates are called *central.* However central algorithms are often hard to construct, and so we shall assume merely that h^* is given by an *interpolatory* algorithm, that is, $h^* \in S(\varphi, y, \delta)$. In this case $\text{diam}S(\varphi, y, \delta)$ is a bound for the set of possible identification errors. Furthermore, $\text{diam}S(\varphi, y, \delta)$ is the smallest bound that works for all interpolatory algorithms. For this reason we regard $\text{diam}S(\varphi, y, \delta)$ as a measure of the local worst-case identification error. It is local in the sense that it is a bound for the identification error that applies for a given set of output data y.

The (global) *worst-case identification error* $E(\varphi, \delta)$, for a given set of test functionals φ and noise bound δ, is the supremum of the local worst-case errors, taken over all possible output data y. Thus

$$E(\varphi, \delta) = \sup_{y \in \mathbf{C}^N} \operatorname{diam} S(\varphi, y, \delta). \tag{2.2}$$

We say that φ is a (δ, τ)*-identifying set* for $\mathcal{M}$ if $E(\varphi, \delta) \leq \tau$, and that $\mathcal{M}$ is *identifiable* if for each $\tau > 0$ there is a finite (δ, τ)*-identifying set* for $\mathcal{M}$ for some $\delta > 0$. In other words $\mathcal{M}$ is identifiable if it is possible to estimate elements of $\mathcal{M}$ to any specified accuracy using any interpolatory algorithm and noisy outputs of a finite set of test functionals, provided only that the noise is sufficiently small.

We say that a model set $\mathcal{M}$ is *absolutely convex* if $\lambda_1 f_1 + \lambda_2 f_2 \in \mathcal{M}$ whenever f_1 and $f_2 \in \mathcal{M}$ and $|\lambda_1| + |\lambda_2| \leq 1$. For absolutely convex model sets the feasibility set $S(\varphi, 0, \delta)$, which consists of all elements of $\mathcal{M}$ consistent with the zero noisy response to the functionals φ_k, is greatest in diameter.

Proposition 2.1. *If $\mathcal{M}$ is absolutely convex then*

$$\operatorname{diam} S(\varphi, y, \delta) \leq \operatorname{diam} S(\varphi, 0, \delta) \textit{ for any } y \in \mathbf{C}^N.$$

Proof. If $h_1, h_2 \in S(\varphi, y, \delta)$ then $(h_1 - h_2)/2 \in \mathcal{M}$ by absolute convexity. Furthermore,

$$|\varphi_k((h_1 - h_2)/2)| \leq |(\varphi_k(h_1) - y_k)/2)| + |(\varphi_k(h_2) - y_k)/2)| \leq \delta/2 + \delta/2 = \delta$$

for $1 \leq k \leq N$, and so $(h_1 - h_2)/2 \in S(\varphi, 0, \delta)$. A similar argument shows that $(h_2 - h_1)/2 \in S(\varphi, 0, \delta)$. So

$$||h_1 - h_2|| = ||(h_1 - h_2)/2 - (h_2 - h_1)/2|| \leq \operatorname{diam} S(\varphi, 0, \delta),$$

and the result follows. □

It follows from Proposition 2.1 that if $\mathcal{M}$ is absolutely convex then $E(\varphi, \delta) = \operatorname{diam} S(\varphi, 0, \delta)$. Furthermore, in this case $S(\varphi, 0, \delta)$ is also absolutely convex, and so

$$\operatorname{diam} S(\varphi, 0, \delta) = 2\operatorname{rad} S(\varphi, 0, \delta) = \sup\{||h|| : h \in S(\varphi, 0, \delta)\}.$$

This leads to the following useful criterion for identifiability in absolutely convex model sets.

Lemma 2.1. *Linear functionals $\varphi_1, \varphi_2, \varphi_3, \ldots, \varphi_N$ form a (δ, τ)-identifying set for the absolutely convex model set $\mathcal{M}$ if and only if*

$$\sup_{1 \leq k \leq N} |\varphi_k(h)| \geq \delta \textit{ for each } h \in \mathcal{M}_\tau, \tag{2.3}$$

where $\mathcal{M}_\tau = \{h \in \mathcal{M} : \|h\| = \tau\}$.

We say that a finite set $(\varphi_1, \varphi_2, \ldots, \varphi_N)$ satisfying (2.3) is a (δ, τ)-*cover* for $\mathcal{M}$. Clearly each (δ, τ)-identifying set for $\mathcal{M}$ is a (δ, τ)-cover for $\mathcal{M}$, and by Lemma 2.1 the converse is also true if $\mathcal{M}$ is absolutely convex. The following result is a generalization of Lemma 2.1 that applies to model sets that are not necessarily absolutely convex. For any model set $\mathcal{M}$ we write $\mathcal{M}^\#$ for the set

$$\mathcal{M}^\# = \{(x-y)/2 : x, y \in \mathcal{M}\}.$$

Note that $\mathcal{M} = \mathcal{M}^\#$ if $\mathcal{M}$ is absolutely convex.

Lemma 2.2. *Each* (δ, τ)-*cover for* $\mathcal{M}^\#$ *is a* $(\delta, 2\tau)$-*identifying set for* $\mathcal{M}$.

Proof. Suppose that $\varphi = (\varphi_1, \varphi_2, \ldots, \varphi_N)$ is a (δ, τ)-cover for $\mathcal{M}^\#$, $y \in \mathbf{C}^N$ and that $h_1, h_2 \in S(\varphi, y, \delta)$.. Then, as in the proof of Proposition 2.1, $|\varphi_k((h_1 - h_2)/2)| \leq \delta$ for each k. Now $(h_1 - h_2)/2 \in \mathcal{M}^\#$, and since φ is a (δ, τ)-cover for $\mathcal{M}^\#$ it follows that $\|(h_1 - h_2)/2\| \leq \tau$. Therefore $\text{diam} S(\varphi, y, \delta) \leq 2\tau$, and since y is arbitrary it follows that $E(\varphi, \delta) \leq 2\tau$. □

The following theorem describes the relationship between compactness and identifiability in closed, absolutely convex model sets. The proof is given in Section 5.

Theorem 2.1. *An absolutely convex subset* $\mathcal{M}$ *of a normed space is identifiable if and only if each closed and bounded subset of* $\mathcal{M}$ *is compact.*

An easy modification establishes the following criterion for identifiability of model sets which are not absolutely convex.

Theorem 2.2. *A subset* $\mathcal{M}$ *of a normed space is identifiable if each closed and bounded subset of* $\mathcal{M}^\#$ *is compact.*

3. Identification via convolution. We return now to the problem of more practical interest, where the functionals used for identification are of the form $\varphi(h) = (u * h)(t)$, where u is the chosen input signal and t belongs to a specified set of real numbers. We say that $\mathcal{M}$ is $*$-identifiable if $\mathcal{M}$ is identifiable using this restricted set of linear functionals. Clearly any $*$-identifiable model set is identifiable. We give conditions on the underlying normed space which ensure that identifiable model sets are $*$-identifiable. For the sake of simplicity we deal with discrete and continuous model sets separately.

3.1. Discrete systems. Here we assume that the model set $\mathcal{M}$ is a closed subset of a normed sequence space $\mathcal{X} \subseteq \ell^\infty$, where ℓ^∞ is the Banach space of all bounded sequences. We assume also that the norm topology on $\mathcal{X}$ is finer than the topology induced by the ℓ^∞ norm, that is, we assume that there exists $C > 0$ such that

$$\|f\|_\infty \leq C \|f\| \text{ for each } f \in \mathcal{X}. \tag{3.1}$$

Condition (3.1) ensures that the convolution functionals arising from any fixed ℓ^1 sequence are uniformly continuous. To see this suppose that $\psi \in \ell^1$ and that for each $n \geq 0$, Ψ_n is the linear functional defined on $\mathcal{X}$ by

$$\Psi_n(f) = (\psi * f)(n) \text{ for each } f \in \mathcal{X}. \tag{3.2}$$

Then by (3.1),

$$|\Psi_n(f)| = \left| \sum_{m=0}^{n} \psi(m) f(n-m) \right| \leq \|\psi\|_1 \|f\|_\infty \leq C \|\psi\|_1 \|f\|,$$

and so $\|\Psi_n\| \leq C \|\psi\|_1$ for each $n \geq 0$.

The following theorem shows that condition (3.1) on the underlying normed space is enough to ensure that identifiable model sets are $*$-identifiable. The proof is given in Section 5.

Theorem 3.1. *Suppose that $\mathcal{X}$ is a normed subspace of ℓ^∞, and that the norm topology on $\mathcal{X}$ is finer than the topology induced by the ∞-norm. Suppose also that $\mathcal{M}$ is an absolutely convex subset of $\mathcal{X}$. Then $\mathcal{M}$ is $*$-identifiable if and only if each closed and bounded subset of $\mathcal{M}$ is compact.*

Furthermore the proof of Theorem 3.1 can be modified to establish the following result.

Theorem 3.2. *Suppose that $\mathcal{X}$ is a normed subspace of ℓ^∞, and that the norm topology on $\mathcal{X}$ is finer than the topology induced by the ∞-norm. Then a subset $\mathcal{M}$ is of $\mathcal{X}$ is $*$-identifiable if each closed and bounded subset of $\mathcal{M}^\#$ is compact.*

The ℓ^p norm of a sequence f is given by $\|f\|_p = \left(\sum_{n=0}^{\infty} |f(n)|^p\right)^{1/p}$, for $1 \leq p < \infty$, and $\|f\|_\infty = \sup_{n \geq 0} |f(n)|$. Since $\|f\|_\infty \leq \|f\|_p$ for any sequence f and any $1 \leq p \leq \infty$, condition (3.1) is satisfied by each of the ℓ^p norms.

The H^p norm of f is defined in terms of its z-transform $\widehat{f}(z) = \sum_{n=0}^{\infty} f(n) z^n$. If $f \in \ell^\infty$ then $\widehat{f}(z)$ is analytic in the open unit disc $|z| < 1$. The H^p norm of f is defined by

$$\|f\|_{H^p} = \lim_{r \to 1^-} \left(\frac{1}{2\pi} \int_0^{2\pi} \left| \widehat{f}(re^{i\theta}) \right|^p d\theta \right)^{1/p},$$

and the H^∞ norm is defined by $\|f\|_{H^\infty} = \sup_{|z|<1} \left| \widehat{f}(z) \right|$.

To establish (3.1) for the H^p norms we argue as follows. By Cauchy's integral formula we have, for each $n \in \mathbf{Z}^+$ and $0 < r < 1$,

$$f(n) = \frac{1}{2\pi i} \oint_{C_r} z^{-n-1} \widehat{f}(z) \, dz,$$

where C_r is the circle $|z| = r$. Therefore

$$|f(n)| \leq \frac{1}{2\pi r^n} \int_0^{2\pi} \left| \widehat{f}(re^{i\theta}) \right| d\theta \leq r^{-n} \|f\|_{H^1} \leq r^{-n} \|f\|_{H^p}.$$

It follows that $||f||_\infty \leq ||f||_{H^p}$ for any sequence f and any $1 \leq p \leq \infty$. This leads to the following corollary of Theorem 3.2.

Corollary 3.1. *Suppose that $\mathcal{M}$ is a subset of ℓ^p or H^p, where $1 \leq p \leq \infty$. Then $\mathcal{M}$ is $*$-identifiable if and only if each closed and bounded subset of $\mathcal{M}^\#$ is compact.*

If $\mathcal{M}$ is absolutely convex, then $\mathcal{M}^\#$ may be replaced by $\mathcal{M}$ in Corollary 3.1.

3.1.1. Applications. Three important classes of discrete model sets have been studied in the literature, and will be considered in detail below. These are the FIR model sets [1], [4], [7] and [14], the so-called bounded model sets [5], and the $ARX(n, n-1)$ model sets [3].

FIR Models. The appropriate model set for a discrete FIR system is $\mathcal{P}_n$, the set of all 'polynomial' sequences $h = (h(j))_{j=0}^\infty$ for which $h(j) = 0$ for all $j \geq n$, where n is a fixed positive integer. Each $\mathcal{P}_n$ is an n-dimensional linear subspace of ℓ^p and of H^p, for any $1 \leq p \leq \infty$. So $\mathcal{P}_n$ is absolutely convex, and its closed and bounded subsets are compact. It follows from Corollary 3.1 that $\mathcal{P}_n$ is $*$-identifiable in any ℓ^p or H^p norm.

Bounded model sets. The model sets $\mathcal{K}(g)$ are introduced in [13]. For each non-negative sequence g in ℓ^1,

$$\mathcal{K}(g) = \{h = (h(n))_{n=0}^\infty : |h(n)| \leq g(n) \text{ for each } n \geq 0\}.$$

For systems in which the terms of the impulse response sequence are known to decay exponentially a model set $\mathcal{K}(g)$, where $g(n) = \lambda^n$ and $0 < \lambda < 1$, is appropriate. Any model set of the form $\mathcal{K}(g)$ is absolutely convex, and compact in any ℓ^p or H^p norm. So by Corollary 3.1 it is $*$-identifiable in any ℓ^p or H^p norm.

Discrete ARX models. A discrete ARX system is governed by a constant-coefficient difference equation and the z-transform of the impulse response is a rational function. For such systems the sets $\mathcal{V}(n,r)$ form a useful class of model sets. These are defined for $n \geq 1$ and $0 \leq r < 1$ and consist of all $ARX(n, n-1)$ models with poles, if any, lying outside the circle $|z| > r^{-1}$. Thus $\mathcal{V}(n,r)$ is the set of all sequences h, whose transfer functions have the form $\widehat{h}(z) = p(z)/q(z)$, where $p(z)$ is a polynomial of degree less than n, $q(z)$ is a polynomial of degree at most n and the zeros of $q(z)$ lie outside the circle $|z| = r^{-1}$.

The sequences in $\mathcal{V}(n,r)$ are linear combinations of sequences of the form $(k^j \lambda^k)_{k=0}^\infty$, where $|\lambda| \leq r$ and $0 \leq j < n$, and hence their terms decay exponentially. This can be used to show that closed and bounded subsets of $\mathcal{V}(n,r)$ are compact in any ℓ^p or H^p norm.

The model set $\mathcal{V}(n,r)$ is not absolutely convex if $r > 0$, but $\mathcal{V}(n,r)^\# \subset \mathcal{V}(2n,r)$. So by Corollary 3.1 the model sets $\mathcal{V}(n,r)$ are also $*$-identifiable in any ℓ^p or H^p norm.

3.2. Continuous systems. Here we assume that the underlying normed space $\mathcal{X}$ consists of measurable functions defined on $\mathbf{R}^+$, the positive half line. As in the discrete case it will be necessary to impose a growth condition on the functions in $\mathcal{X}$ in order to guarantee the uniform continuity of certain families of convolution functionals.

We define the mixed norm $\|f\|_{(\infty,1)}$ for any measurable function f defined on $\mathbf{R}^+$ by

$$\|f\|_{(\infty,1)} = \sup_{n\geq 0}\left\{\int_n^{n+1} |f(t)|\, dt\right\},$$

and denote by $L^{(\infty,1)}$ the Banach space of all measurable functions with finite $(\infty,1)$ norm.

We assume that $\mathcal{M}$ is a subset of a normed space $\mathcal{X} \subseteq L^{(\infty,1)}$, and that the norm topology on $\mathcal{X}$ is finer than the topology induced by the $(\infty,1)$ norm, that is, we assume that there exists $C > 0$ such that

$$\|f\|_{(\infty,1)} \leq C\,\|f\| \text{ for each } f \in \mathcal{X}. \tag{3.3}$$

Condition (3.3) ensures that convolution functionals arising from an essentially bounded, compactly supported function are uniformly bounded. To see this suppose that ψ is essentially bounded and supported on $[0,T]$, and that for each $t \geq 0$ Ψ_t is the linear functional defined on $\mathcal{X}$ by

$$\Psi_t(f) = (\psi * f)(t) \text{ for each } f \in \mathcal{X}. \tag{3.4}$$

Then

$$|\Psi_t(f)| \leq \int_0^T |\psi(s)f(t-s)|\, ds \leq (T+2)\,\|\psi\|_\infty\,\|f\|_{(\infty,1)},$$

and so by (3.1), $|\Psi_t(f)| \leq (T+2)\,\|\psi\|_\infty\, C\,\|f\|$.

The following result is the continuous analogue of Theorem 3.1.

Theorem 3.3. *Suppose that $\mathcal{X}$ is a normed subspace of $L^{(\infty,1)}$, and that the norm topology on $\mathcal{X}$ is finer than the topology induced by the $(\infty,1)$-norm. Suppose also that $\mathcal{M}$ is an absolutely convex subset of $\mathcal{X}$. Then $\mathcal{M}$ is $*$-identifiable if and only if each closed and bounded subset of $\mathcal{M}$ is compact.*

Condition (3.3) is satisfied by the familiar L^p norms. To see this suppose that $f \in L^p$ for some $1 \leq p < \infty$. Then by Hölder's inequality

$$\int_n^{n+1} |f(t)|\, dt \leq \left(\int_n^{n+1} |f(t)|^p\, dt\right)^{1/p} \leq \left(\int_0^{\infty} |f(t)|^p\, dt\right)^{1/p} = \|f\|_p$$

for each $n \geq 0$, and so $\|f\|_{(\infty,1)} \leq \|f\|_p$. The same inequality applies if $p = \infty$.

The Laplace transform $F(s)$ of a measurable function f defined on $\mathbf{R}^+$ is defined by

$$F(s) = \int_0^\infty f(t)e^{-st}dt. \tag{3.5}$$

If $f \in L^{(\infty,1)}$ then $F(s)$ is analytic in the open right half plane $\mathrm{Re}s > 0$, and the H^p norm of f is defined by

$$\|f\|_{H^p} = \lim_{x\to 0+}\left(\int_{-\infty}^{\infty} |F(x+iy)|^p\, dy\right)^{1/p},$$

for $1 \le p < \infty$, and $\|f\|_{H^\infty} = \sup\{|F(s)| : \text{Re}s > 0\}$.

Consideration of functions of the form $e^{i\alpha t^2}$ where $\alpha > 0$ shows that (3.3) does not hold for every H^p norm. However there are some special cases. Firstly, $\|f\|_2 = \|f\|_{H^2}$ by the Paley-Weiner theorem, and so $\|f\|_{(\infty,1)} \le \|f\|_{H^2}$.

A similar inequality can be established for the H^1 norm. If $f \in H^1$, then $F(s)$ is analytic in $\text{Re}s > 0$, and by the inversion formula for Laplace transforms, for each $t > 0$ and $x > 0$

$$f(t) = \frac{1}{2\pi i} \int_{\Gamma_x} F(s) e^{st}\, ds,$$

where Γ_x is the line $\text{Re}s = x$. Therefore

$$|f(t)| \le \frac{e^{xt}}{2\pi} \int_{-\infty}^{\infty} |F(x+iy)|\, dy \le \frac{e^{xt}}{2\pi} \|f\|_{H^1}. \tag{3.6}$$

Since (3.6) holds for all positive x and t, $\|f\|_{(\infty,1)} \le \|f\|_\infty \le \|f\|_{H^1} / 2\pi$ for each $f \in H^1$. So we have the following corollary of Theorem 3.3.

Corollary 3.2. *Suppose that $\mathcal{M}$ is an absolutely convex subset of L^p for any $1 \le p \le \infty$, or of H^1 or H^2. Then $\mathcal{M}$ is $*$-identifiable if and only if each closed and bounded subset of $\mathcal{M}$ is compact.*

3.2.1. Applications. We apply the above results to three classes of continuous model sets for continuous linear systems.

Bounded model sets. For each decreasing non-negative integrable function g, we define

$$\mathcal{K}(g) = \{f : |f(t)| \le g(t) \text{ for all } t \ge 0\}.$$

These are the natural analogues of their discrete counterparts. Each $\mathcal{K}(g)$ is closed in any L^p or H^p norm, and is absolutely convex. However it is easy to see that $\mathcal{K}(g)$ is not compact in any L^p or H^p norm, unless $g = 0$. So by Theorem 2.1, $\mathcal{K}(g)$ is not identifiable, in any L^p or H^p norm.

Bounded model sets with continuity constraints. Suppose that g is a decreasing, non-negative integrable function defined on $\mathbb{R}^+$ as before, and that γ is an increasing, concave, positive-valued function on $\mathbb{R}^+$ with $\gamma(t) \to 0$ as $t \to 0^+$. Let $\mathcal{K}(g,\gamma)$ consist of all functions $f \in \mathcal{K}(g)$ for which $|f(t) - f(t')| \le \gamma(t - t')$ for all $0 \le t' \le t$. Thus γ is the 'uniform modulus of continuity' of functions in $\mathcal{K}(g,\gamma)$. The set $\mathcal{K}(g,\gamma)$ is absolutely convex and compact in any L^p or H^p norm. So by Corollary 3.2 $\mathcal{K}(g,\gamma)$ is $*$-identifiable in any L^p norm or in H^1 or H^2 norm. It has also been shown that $\mathcal{K}(g,\gamma)$ is $*$-identifiable in the H^∞ norm [6].

Continuous ARX models. A continuous *ARX* system is governed by a constant-coefficient differential equation, and the Laplace transform of the impulse response is a rational function. For such systems the sets $\mathcal{M}(n,r)$ form a useful class of model sets. These are defined for $n \ge 1$ and $0 \le r < 1$, and consist of all functions h whose Laplace transforms have the form $H(s) = p(s)/q(s)$, where $p(s)$ is a polynomial of degree less than n, $q(s)$ is a polynomial of degree n and the zeros of $q(s)$ lie inside the disc $D_r = |(1+s)/(1-s)| \le r$.

The functions in $\mathcal{M}(n,r)$ are linear combinations of functions of the form $(t^j \lambda^t)_{k=0}^{\infty}$, where $\lambda \in D_r$ and $0 \le j < n$. Since D_r is a compact subset of the open left half plane

Re$s < 0$, the functions in $\mathcal{M}(n,r)$ decay exponentially. This can be used to show that closed and bounded subsets of $\mathcal{M}(n,r)$ are compact in any L^p or H^p norm.

The model set $\mathcal{M}(n,r)$ is not absolutely convex if $r > 0$, but $\mathcal{M}(n,r)^{\#} \subset \mathcal{M}(2n,r)$. There is a continuous analogue of Theorem 3.2, and so $\mathcal{M}(n,r)$ is $*$-identifiable in any L^p norm and in the H^1 and H^2 norm. It has also been shown that $\mathcal{M}(n,r)$ is $*$-identifiable in the H^∞ norm [3].

4. Complexity. In the previous section we showed that worst-case identification is possible in various types of model sets and norms using noisy output data. It is also important for practical purposes to obtain estimates of the amount of data required for this type of identification. Problems of this sort have been studied by various authors in the last decade [1] [9] and [16], with most attention directed to identification within $\mathcal{P}_n$ the model set for discrete FIR systems. We summarize here some key results.

For simplicity we shall assume that the output samples are of the form $y_0, \dots, y_{N-1}$ for discrete models, and $y(\Delta), y(2\Delta), y(3\Delta), \dots, y(N\Delta)$, where $\Delta > 0$, for continuous models. We shall also assume that all impulse responses in our model sets are real-valued.

We say that an input sequence (or function in the continuous case) u is a (δ,τ)-*identifying signal of length* N for a model set $\mathcal{M}$, if the functionals $h \to (u*h)(k)$ (or $h \to (u*h)(k\Delta)$) for $0 \le k < N$, form a (δ,τ)-identifying set for $\mathcal{M}$. For given noise and tolerance levels δ and τ, the *complexity of identification* $N(\delta,\tau)$ in a model set $\mathcal{M}$ is the minimum length of (δ,τ)-identifying signals u for which $\|u\|_\infty \le 1$. The numbers $N(\delta,\tau)$ are generally difficult to determine, but useful bounds have been obtained in some cases.

4.1. Discrete model sets. Since $\|f\|_2 = \|f\|_{H^2} \le \|f\|_{H^\infty} \le \|f\|_1$ for any sequence f, the complexity of identification of any model set increases if we change from the ℓ^2 norm to the H^∞ norm, or from the H^∞ to the ℓ^1 norm. We shall see that ℓ^1 identification is typically 'exponential' in complexity, whereas H^∞ and ℓ^2 identification is 'polynomial'.

4.1.1. Identification within $\mathcal{P}_n$.

ℓ^1 norm. It is well known [9] that a sequence that contains as substrings all strings of the form $v_1 v_2 \dots v_n$, where each $v_i \in \{-1,1\}$, is an ℓ^1 (τ,τ)-identifying signal for $\mathcal{P}_n$. The point is that such a sequence u has the property that for each $h \in \mathcal{P}_n$, at least one of the terms of $u*h$ equals $\|h\|_1$. There are 2^n binary strings of length n, and so the length of any such string is at least $2^n + n - 1$. It is a remarkable fact [2] that there are such sequences, called *Galois sequences*, with this minimal length for each positive integer n, and they are easily generated using shift registers. So the complexity of ℓ^1 (δ,τ)-identification in $\mathcal{P}_n$ is no greater than $2^{n/2} + n - 1$ for each $0 \le \delta \le \tau$. In fact it is not difficult to show that the complexity of (τ,τ)-identification is exactly $2^{n/2} + n - 1$. It is shown in [16] that there are shorter identifying signals if $\delta < \tau$, but probabilistic arguments based on the Central Limit Theorem show that in such cases the complexity of ℓ^1 (δ,τ)-identification in $\mathcal{P}_n$ is of order β^n for some $\beta > 1$, that is, there are numbers c_1 and c_2 such that $c_1\beta^n \le N(\delta,\tau) \le c_2\beta^n$ for each n. The numbers β, c_1 and c_2 depend only on the ratio δ/τ.

H^∞ norm. We can obtain H^∞ identifying signals for $\mathcal{P}_n$ by concatenating n-strings of the form

$$(1, \omega^{-1}, \omega^{-2}, \dots, \omega^{-n+1}),$$

where ω is any m^{th} root of unity for some positive integer m. Any sequence u obtained in this way has the property that for each $h \in \mathcal{P}_n$, the terms of $u*h$ include all numbers of the form $\omega^{-n+1}\hat{h}(\omega)$, and so at least one of the terms $|(u*h)(k)|$ is close to $\|h\|_{H^\infty}$ if m is sufficiently large. Since m can be chosen to be of order n [7], we have identifying signals of order n^2. It is shown in [7] that the complexity of H^∞ identification is of order n^2, and so inputs of this type are in a sense best possible.

ℓ^2 norm. It turns out that ℓ^2 identification in $\mathcal{P}_n$ is merely linear in n It is shown in [12] that minimal length ℓ^2 identifying signals for $\mathcal{P}_n$ can be constructed using the coefficients of polynomials which are 'large everywhere on the unit circle'. The interested reader is encouraged to consult [13] for a full description of these signals and some fascinating related problems concerning the behaviour on the unit circle of polynomials whose coefficients are all of modulus 1.

4.1.2. Other discrete systems. The key to identification in $\mathcal{V}(n,r)$ or $\mathcal{K}(g)$ is the fact that there are uniform bounds on the rates of decay of the 'tails' of the elements. In other words, the elements in any such model set $\mathcal{M}$ can be approximated uniformly well by their finite truncations. So $\mathcal{M}$ can be effectively replaced by a subset of a suitable polynomial space $\mathcal{P}_{n'}$, and any (δ,τ')-identifying signal for $\mathcal{P}_{n'}$ for a suitably chosen $\tau' < \tau$ is a (δ,τ)- identifying signal for $\mathcal{M}$.

So to determine bounds for the complexity we need bounds on n' as a function of the parameters n, r and g. These have been obtained in [3] and [5]. For example, it is shown in [3], that for the model sets $\mathcal{V}(n,r)$ the cut-off n' is bounded by a linear function of n, (for fixed r, δ and τ). So it follows that identification in $\mathcal{V}(n,r)$ is no greater than an exponential function of n in the ℓ^1 case, a quadratic function of n in the H^∞ case and a linear function of n in the ℓ^2 case. Similar lower bounds are also obtained in [3].

The complexity of identification in $\mathcal{K}(g)$ is difficult to determine in general. However it is possible to give estimates in some special cases. For example, using the methods presented in [5] we can show that if $g(k) = \lambda^k$ where $0 < \lambda < 1$, then the complexity of identification is essentially exponential in $(1-\lambda)^{-1}$ in the ℓ^1 case, and low order polynomial in $(1-\lambda)^{-1}$ in the H^∞ and ℓ^2 cases.

4.2. Continuous model sets. The complexity of identification in the continuous *ARX* model sets $\mathcal{M}(n,r)$ has been studied in [3], and results similar to those for the discrete analogues $\mathcal{V}(n,r)$ have been obtained. In particular, the exponential and quadratic nature of L^1 and H^∞ identification persists, but the bounds are not as sharp. Interested readers may check [3] for full details. Similar results apply to $\mathcal{K}(g,\gamma)$, but here we merely give the appropriate reference [6].

We conclude with the observation that there is another natural measure of the complexity of identification in continuous model sets, apart from the minimum sample size. This is the *minimum total sampling time.* The two are related if a fixed sampling rate Δ is assumed. What can be said if there is no restriction on the sampling rate? It can be shown that in model sets such as $\mathcal{M}(n,r)$ and $\mathcal{K}(g,\gamma)$, where the rate of change of functions is bounded, there is no point in sampling too fast. In fact in these model sets the sampling time complexity has the same types of bounds for L^1 and H^∞ as the sample size complexity [3] and [6].

5. Proofs.

5.1. Proof of Theorem 2.1. Suppose that $\mathcal{M}$ is absolutely convex and identifiable, and that $\tau > 0$ and $\varepsilon > 0$. Then by Lemma 2.1 there are linear functionals

$\varphi_1, \varphi_2, ..., \varphi_N$ and $\delta > 0$ such that

$$\max_{1 \le j \le N} |\varphi_j(h)| \ge \delta/2 \text{ for every } h \in \mathcal{M}_{\varepsilon/2}. \tag{5.1}$$

Define $\Phi(x) = (\varphi_1(x), ..., \varphi_N(x))$ for each $x \in \mathcal{X}$. Then $\Phi\mathcal{M}_\tau$ is a bounded subset of $\mathbf{C}^N$ (with the supremum norm), and hence its closure is compact. Choose a finite subset G of $\mathcal{M}_\tau$ such that ΦG is a δ-net for $\Phi\mathcal{M}_\tau$. Then for each $h \in \mathcal{M}_\tau$ there exists $g \in G$ such that $||\Phi(g) - \Phi(h)||_\infty < \delta$. Therefore, since Φ is linear, $||\Phi((g-h)/2)||_\infty < \delta/2$. But $(g-h)/2 \in \mathcal{M}$ by the absolute convexity of $\mathcal{M}$, and so $||(g-h)/2)|| < \varepsilon/2$ by (5.1). So G is an ε-net for $\mathcal{M}_\tau$, and it follows that every closed and bounded subset of $\mathcal{M}$ is compact.

Now suppose that $0 < \delta < \tau$, and let $\varepsilon = \tau - \delta$. Since $\mathcal{M}_\tau$ is relatively compact, there is a finite ε-net $G = \{g_1, ..., g_N\}$ in $\mathcal{M}_\tau$. By the Hahn Banach theorem, there are linear functionals $\varphi_1, ..., \varphi_N$ on $\mathcal{X}$, each with norm 1, such that $\varphi_k(g_k) = ||g_k|| = \tau$ for $1 \le k \le N$. Let $h \in \mathcal{M}_\tau$ and choose k so that $||g_k - h|| < \varepsilon$. Then

$$|\varphi_k(h)| \ge |\varphi_k(g_k)| - |\varphi_k(g_k - h)| > \tau - \varepsilon = \delta,$$

and so (2.3) holds.

5.2. Proof of Theorem 3.1. The proof proceeds by an number of preliminary lemmas.

Lemma 5.1. *Suppose that $\psi \in \ell^1$, $\mathcal{K}$ is a compact subset of $\mathcal{X}$, and $\varepsilon > 0$. Then there is an infinite set Σ of positive integers such that*

$$|(\psi * h)(m) - (\psi * h)(m')| < \varepsilon \textit{ for each } h \in \mathcal{K} \textit{ and each } m, m' \in \Sigma.$$

Proof. For each $n \ge 0$, let Ψ_n be given by (3.2), and let

$$\Sigma_n = \{m \in \mathbf{Z}^+ : \sup_{h \in \mathcal{K}} |\Psi_m(h) - \Psi_n(h)| < \varepsilon/2\}.$$

Since $||\Psi_n|| \le C\,||\psi||_1$ for each $n \ge 0$, the linear functionals Ψ_n, for $n \ge 0$, are uniformly bounded, and hence equi-continuous on $\mathcal{X}$. So by the Arzela-Ascoli theorem at least one of the sets Σ_n is infinite, and it is easy to check that any such set has the desired properties. □

The next result shows how a finite number of convolution functionals of the form (3.2) can be approximated by convolution functionals arising from a single ℓ^1 sequence. First we introduce some notation. We denote by S the shift operator defined for any sequence f by

$$(Sf)(0) = 0 \text{ and } (Sf)(n) = f(n-1) \text{ for } n \ge 1.$$

It is easy to verify that $S(\psi * f) = S\psi * f = \psi * Sf$ for any sequences ψ and f.

Lemma 5.2. *Suppose that $\mathcal{K}$ is a compact subset of $\mathcal{X}$, $\varepsilon > 0$, and $\Psi_1, \Psi_2, \ldots \Psi_K$ are continuous linear functionals on $\mathcal{X}$ of the form $\Psi_k(f) = (\psi_k * f)(m_k)$ for each $f \in \mathcal{X}$, where ψ_k is a sequence and $m_k \in \mathbf{Z}^+$ for each $1 \le k \le K$. Then there is a finitely supported sequence u, and $n_1, n_2, \ldots, n_K \in \mathbf{Z}^+$, such that $\|u\|_\infty = \max_{1 \le k \le K} \|\psi_k\|_\infty$, and*

$$|(u * h)(n_k) - \Psi_k(h)| < \varepsilon \textit{ for each } h \in \mathcal{K} \textit{ and } 1 \le k \le K.$$

Proof. We may assume that each ψ_k is finitely supported, in particular that $\psi_k(n) = 0$ if $n \ge m_k$. The proof is by induction on K. If $K = 1$ we choose $n_1 \ge m_1$ and set $u = S^{n_1 - m_2}\psi_1$. Then for each $f \in \mathcal{K}$,

$$(u * h)(n_1) = (S^{n_1 - m_1}\psi_1 * h)(n_1) = (\psi_1 * h)(m_1) = \Psi_1(h).$$

For the inductive part, assume that v is finitely supported, $n_k \ge 1$ for $1 \le k < K$, and that

$$|(v * h)(n_k) - \Psi_k(h)| < \varepsilon/2 \text{ for each } h \in \mathcal{K} \text{ and } 1 \le k < K.$$

Choose M such that $v(n) = 0$ for $n > M$, and such that $M > \max\{n_{K-1}, m_K\}$. By Lemma 5.1 there are positive integers $m > 2M$ and $m' > m + M$ such that

$$|(v * h)(m) - (v * h)(m')| < \varepsilon \text{ for each } h \in \mathcal{K}. \tag{5.2}$$

Let $u = v - S^{m'-m}v + S^{m'-m_K}\psi_K$, and $n_K = m'$. We shall show that u and $n_1, n_2, \ldots, n_{K-1}, n_K$ have the desired properties.

First observe that v, $S^{m'-m}v$, and $S^{m'-m_K}\psi_K$ have disjoint supports, and so

$$\|u_K\|_\infty = \max\{\|v\|_\infty, \|\psi_K\|_\infty\} = \max_{1 \le k \le K} \|\psi_k\| ..$$

Now suppose that $1 \le k < K$ and that $h \in \mathcal{K}$. Since $m' - m > M > n_k$ and $m' - m_K > M > n_k$, $(S^{m'-m}v * h)(n_k) = (S^{m'-m_K}\psi_K * h)(n_k) = 0$, and so

$$|(u * h)(n_k) - \Psi_k(h)| = |(v * h)(n_k) - \Psi_k(h)| < \varepsilon.$$

Also $(S^{m'-m_K}\psi_K * h)(m') = (\psi_K * h)(m_K) = \Psi_K(h)$ and $(S^{m'-m}v * h)(m') = (v * h)(m)$, and so by (5.2),

$$|(u * h)(n_K) - \Psi_K(h)| = |(v * h)(m') - (v * h)(m)| < \varepsilon.$$

□

In order to prove the main theorem it is enough to show that if each $\mathcal{M}_\tau$ is compact then $\mathcal{M}$ is $*$-identifiable. So suppose that $\mathcal{M}_\tau$ is compact. We shall show that there exists $\delta > 0$, $N \ge 0$, and a finitely supported sequence u with $\|u\|_\infty = 1$ such that

$$\max_{1 \le n \le N} |(u * h)(n)| \ge \delta \text{ for each } h \in \mathcal{M}_\tau.$$

Suppose that Ψ_n is the linear functional on $\mathcal{X}$ defined by (3.2), where $n \geq 0$ and where ψ is a finitely supported sequence, with $\|\psi\|_\infty = 1$. Since Ψ_n is continuous, the set $U(\psi, n, \kappa) = \{f \in \mathcal{X} : |\Psi_n(f)| > \kappa\}$ is open in $\mathcal{X}$ for each $\kappa > 0$.. Furthermore, $\mathcal{M}_\tau$ is covered by sets of this form. To see this suppose that $h \in \mathcal{M}_\tau$. Since $h \neq 0$, $|h(n)| \neq 0$ for some $n \geq 0$. Let $\psi = e^{-i\theta}\delta^{(0)}$, where $\theta = \arg h(n)$. Then $(\psi * h)(n) = |h(n)|$, and so $h \in U(\psi, n, \kappa)$ if $0 < \kappa < |h(n)|$.

Since $\mathcal{M}_\tau$ is compact, there is a finite family $\{\Psi_1, \Psi_2, \ldots, \Psi_K\}$ of linear functionals of the form (3.2) and $\kappa > 0$ such that

$$\max_{1 \leq k \leq K} |\Psi_k(h)| > \kappa \text{ for every } h \in \mathcal{M}_\tau.$$

Choose δ such that $0 < \delta < \kappa$, and let $\varepsilon = \kappa - \delta$. Then by Lemma 5.2 there exists a finitely supported sequence u with $\|u\|_\infty = 1$, and $n_1, n_2, \ldots, n_K \in \mathbb{Z}^+$, such that $|(u * h)(n_k) - \Psi_k(h)| < \varepsilon$ for each $h \in \mathcal{M}_\tau$ and each $1 \leq k \leq K$.

Let $N = \max\{n_k : 1 \leq k \leq K\}$, and suppose that $h \in \mathcal{M}_\tau$. Then $|\Psi_k(h)| > \kappa$ for some $1 \leq k \leq K$. So it follows that

$$\max_{1 \leq n \leq N} |(u * h)(n)| \geq |(u * h)(n_k)| \geq |\Psi_k(h)| - |(u * h)(n_k) - \Psi_k(h)| > \kappa - \varepsilon = \delta.$$

REFERENCES

[1] M.A. Dahleh, T. Theodosopoulos and J.N. Tsitsiklis, The sample complexity of worst-case identification of FIR linear systems, *Systems Control Lett.*, **20** (1993), 157-166.

[2] H. Fredricksen and Maiorana, Necklaces of beads in k-colors and k-ary de Bruijn sequences, *Discrete Mathematics*, **23**, (1978), 207-210

[3] K.J. Harrison, J.R. Partington and J.A. Ward, Complexity of identification of linear systems with rational transfer functions, to appear in *Mathematics of Control, Signals and Systems*, (1998).

[4] K.J. Harrison and J.A. Ward, Fractional covers for convolution products, *Result. Math.*, **30** (1996), 67-78.

[5] K.J. Harrison and J.A. Ward, Complexity of identification in compact model sets for discrete linear systems, IASTED International Conference on Applied Modelling and Simulation, Banff Canada, 1997, 6-9.

[6] K.J. Harrison and J.A. Ward, Worst-case identification in compact model sets for continuous linear systems, IASTED International Conference on Applied Modelling and Simulation, Banff Canada, 1997, 10-13.

[7] K.J. Harrison, J.A. Ward and D.K. Gamble, Sample complexity of worst-case H^∞-identification, *Systems Control Lett.*, **27** (1996), 255-260.

[8] L. Ljung, *System identification for the User*, Prentice-Hall, New Jersey, 1987.

[9] P.M. Mäkilä, Robust identification and Galois sequences, *Int. J. Control*, **54** #5 (1991), 1189-1200.

[10] M. Milanese and A. Vicino, Information based complexity and nonparametric worst-case identification, *J. Complexity*, **9** (1993), 427–446.

[11] J.R.Partington, Interpolation in normed spaces from the values of linear functionals, *Bull. London Math. Soc.*, **26** (1994), 165-170.

[12] J.R.Partington, Worst-case identification in ℓ^2 : linear and non-linear algorithms, *Systems Control Lett.* **22** (1994), 93-98.

[13] J.R.Partington, *Interpolation, Identification and Sampling*, Oxford University Press, 1997.

[14] K. Poolla and A. Tikku, On the time-complexity of worst-case system identification, *IEEE Trans. Automat. Control*, **39** (1994), 944-950.

[15] P.M. Mäkilä, J.R. Partington and T.K. Gustafsson (1995), Worst-case Control-relevant identification, *Automatica*, **31**, #12, 1799-1819.

[16] D.N.C. Tse, M.A. Dahleh and J.N. Tsitsiklis, Optimal asymptotics identification under bounded disturbances, *Proc. 30th Conf. on Decision and Control,* England, 1991, 623-628.