A KERNEL BASED APPROACH TO STRUCTURED NONLINEAR SYSTEM IDENTIFICATION PART II: CONVERGENCE AND CONSISTENCY

Kenneth Hsu [*] **Tyrone Vincent** [**]
Kameshwar Poolla [*]

[*] *Dept. of Mech. Eng., UC Berkeley, CA 94720 USA*
[**] *Div. of Eng., Colorado School of Mines, CO 80401 USA*

Abstract: In (Hsu *et al.*, 2005*c*), an algorithm for the identification of structured nonlinear systems was proposed and its computational properties were explored. In this paper, we continue the investigation and formalize notions of identifiability and persistence of excitation. Conditions under which the estimated nonlinearity converges uniformly to the true nonlinearity are developed for a class of kernel based dispersion functions.

Keywords: nonlinear system identification, kernel, identifiability, persistence of excitation, convergence

1. INTRODUCTION

The inclusion of available *a priori* information is of fundamental importance in system identification. One example of *a priori* knowledge that has not yet been fully explored is that of the signal interconnections between the various components in a system. We refer to the signal interconnections as the system *structure*, and the resulting identification problem is called a structured nonlinear system identification problem. The class of systems that we consider in this paper are those consisting of linear time-invariant (LTI) systems and static nonlinear maps.

Information regarding the structural interconnection is generally known with a high degree of confidence due to an understanding of the underlying relationships between the system components. While black-box identification methods are by nature unbiased and offer a very general approach to system identification problems (Sjöberg *et al.*, 1995; Juditsky *et al.*, 1995), these methods can require large amounts of data and computational resources. In addition, black-box methods generally do not offer the opportunity to obtain physical insight to the system of interest. By including *a priori* structural information, we can restrict our attention to more specific classes of nonlinear system identification problems and offer systematic algorithms and convergence results.

Nonparametric methods already exist to treat the identification problem in extreme generality, for example using splines, kernel smoothing methods, or neural networks (Greblicki, 1997; Sjöberg *et al.*, 1995; Barron, 1993). Asymptotic analysis and local convergence results are available. However, the generality of these methods limit their ability to systematically incorporate *a priori* information and usually result in large computational problems.

We utilize the identification framework developed in (Claassen, 2001; Wemhoff, 2003; Wemhoff *et*

[1] This work is supported by NSF Grants ECS 03-02554 and ECS 01-34132.

al., 1999) but introduce an alternate class of dispersion functions. In (Hsu *et al.*, 2005*b*), a parametric approach to the structured identification problem was considered. In (Hsu *et al.*, 2005*a*; Hsu *et al.*, 2005*d*), a dispersion function motivated by total variation was examined. We now consider the identification problem using a kernel based approach. In a companion paper (Hsu *et al.*, 2005*c*), an identification algorithm is proposed and its computational properties are explored. In this paper, we formalize notions of identifiability and persistence of excitation, and develop conditions under which the estimated nonlinearity converges uniformly to the true nonlinearity.

NOTATION

x	infinite vector-valued sequence $(x_k)_{k=1}^{\infty}$
$x^{[L]}$	finite vector-valued sequence $(x_k)_{k=1}^{L}$
x_k	k^{th} element of x
x_k^*	complex-conjugate transpose of x_k
π_L	L-sample truncation operator
$\mathcal{L}$	known LTI system
$\mathcal{N}$	static nonlinear function
$\mathcal{M}(\mathcal{L},\mathcal{N})$	LFT model structure defined by $\mathcal{L}$,$\mathcal{N}$
$\Omega^{[i]} \subseteq \mathbf{R}^{p_i}$	domain where $\mathcal{N}^{[i]}$ is to be identified
$\Omega \subseteq \mathbf{R}^{p}$	$\Omega^{[1]} \times \Omega^{[2]} \times \cdots \times \Omega^{[m]}$
$\mathcal{C}$, $\mathcal{C}^1$	set of continuous, continuously differentiable functions on Ω
$\mathbb{L}_\gamma$	class of Lipschitz continuous functions with Lipschitz constant γ.
$\mathbb{N}$	class of static nonlinear functions with a particular input-output structure
$\mathbb{K}$	Reproducing Kernel Hilbert Space
K	reproducing kernel for $\mathbb{K}$
$\hat{w}^{[L]}$	estimate of w
$\mathcal{I}^{[L]}$	interpolant of $(z^{[L]}, w^{[L]})$
$\hat{\mathcal{I}}^{[L]}$	interpolant of $(z^{[L]}, \hat{w}^{[L]})$
$\mathcal{D}^L$	dispersion function
$\mu^L(e^{[L]})$	metric for $e^{[L]}$: $\mu^L(e^{[L]}) = \frac{1}{L}\sum_{i=1}^{L} e_i^* e_i$
p_i	number of inputs to $\mathcal{N}^{[i]}$
p	$p = \sum_{i=1}^{m} p_i$
$\mathbb{E}$	class of unmeasured inputs to $\mathcal{L}$ $\mathbb{E} = \{e : \mu(e) < \infty\}$

2. PROBLEM FORMULATION

To present all structured systems under a common framework, we gather the known linear blocks and unknown nonlinearities into a Linear Fractional Transformation (LFT) as shown in Figure 1. The LFT paradigm is a powerful data structure that can represent all finite interconnected systems in one common setting. We will refer to the generalized system in Figure 1 as the *LFT model structure*. In the LFT model structure, the LTI system $\mathcal{L}$ and the signals u and y are known. $\mathcal{N}$ is a static nonlinearity that is unknown and to be identified. The signal e is an unknown signal. In addition, it is known that $\mathcal{N}$ satisfies a particular input-output interconnection structure, denoted

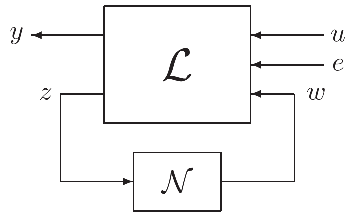

Fig. 1. LFT model structure

$$\mathcal{N} = \begin{bmatrix} \mathcal{N}^{[1]} & & \\ & \ddots & \\ & & \mathcal{N}^{[m]} \end{bmatrix}.$$

Without loss of generality, we assume that each $\mathcal{N}^{[i]}$ is single-output. This can always be realized by introducing redudant copies of the respective inputs.

We will also require that the signal z be *measurable*. This can be realized by requiring the following property of $\mathcal{L}$. Let us partition $\mathcal{L}$ as

$$\mathcal{L} = \begin{bmatrix} \mathcal{L}_{yu} & \mathcal{L}_{ye} & \mathcal{L}_{yw} \\ \mathcal{L}_{zu} & \mathcal{L}_{ze} & \mathcal{L}_{zw} \end{bmatrix}.$$

Definition 1. z is measurable if there exists an LTI system Ψ_M such that

$$\begin{bmatrix} \mathcal{L}_{ze} & \mathcal{L}_{zw} \end{bmatrix} = \Psi_M \begin{bmatrix} \mathcal{L}_{ye} & \mathcal{L}_{yw} \end{bmatrix}.$$

Measurability of z implies that

$$\begin{aligned} z &= \mathcal{L}_{zu}u + \mathcal{L}_{ze}e + \mathcal{L}_{zw}w \\ &= \mathcal{L}_{zu}u + \Psi_M(\mathcal{L}_{ye}e + \mathcal{L}_{yw}w) \\ &= \mathcal{L}_{zu}u + \Psi_M(y - \mathcal{L}_{yu}u), \end{aligned}$$

so that z can be inferred from knowledge of u, y, and $\mathcal{L}$.

In (Hsu *et al.*, 2005*c*), the following identification algorithm was proposed.

(1) Find $\hat{w}^{[L]} = \arg\min_{w^{[L]}} J^L(w^{[L]})$, where

$$J^L(w^{[L]}) = \min_{e^{[L]}} \mu^L(e^{[L]}) + \beta\mathcal{D}^L(z^{[L]}, w^{[L]})$$
$$\text{s.t. } (w^{[L]}, e^{[L]}) \in \mathbb{C}^L(u, y)$$

and

$$\mathbb{C}^L(u,y) = \Big\{(w^{[L]}, e^{[L]}) : (w^{[L]}, e^{[L]}) \text{ satisfies}$$
$$\pi_L y = \pi_L(\mathcal{L}_{yu}u + \mathcal{L}_{ye}e^{[L]} + \mathcal{L}_{yw}w^{[L]})$$
$$\pi_L z = \pi_L(\mathcal{L}_{zu}u + \mathcal{L}_{ze}e^{[L]} + \mathcal{L}_{zw}w^{[L]})\Big\}.$$

(2) Form an estimate of $\mathcal{N}$ via an interpolation of the points $(z^{[L]}, \hat{w}^{[L]})$, denoted $\hat{\mathcal{I}}^{[L]}$.

Here, $\mathcal{D}^L$ is a function that measures the complexity of interpolating the points given by $(z^{[L]}, w^{[L]})$. For ease of analysis, we will focus on a related problem. First, we make the additional assumption that $\mathcal{L}_{ye}^{-1}$ exists so that given $w = \mathcal{N}(z)$, we can determine

$$e = \mathcal{L}_{ye}^{-1}(y - \mathcal{L}_{yu}u - \mathcal{L}_{yw}w).$$

With an appropriate choice of $\lambda > 0$, the following optimization problem is equivalent:

$$\begin{aligned} \hat{w}^{[L]} = \arg\min_{w^{[L]}} \; & \mu^L(e^{[L]}) \\ \text{s.t. } & (w^{[L]}, e^{[L]}) \in \mathbb{C}^L(u, y) \\ & \lambda \mathcal{D}^L(z^{[L]}, w^{[L]}) \leqslant 1. \end{aligned} \quad (1)$$

In what follows, we will establish the convergence properties of the estimated interpolant $\hat{\mathcal{I}}^{[L]}$. Due to space limitations, proofs are removed or only sketched. Some of the ideas from parametric identification, namely identifiability and persistence of excitation, have direct analogs. Formal definitions will come later, but rough definitions are as follows:

(1) The LFT model structure is identifiable if it allows for nonlinear functions to be distinguishable through input-output experiments.
(2) The input is persistently exciting if the response of the model to this input with a non-zero nonlinear function cannot be zero.

Clearly, if either of these conditions are not satisfied, we have no hope of claiming in general that the estimated interpolant converges to the true nonlinearity. However, there is a subtle issue that arises: what is the set of nonlinear functions to which these definitions apply? If we allow for completely arbitrary functions, there is no hope that these conditions can every be satisfied. We are obligated to include an assumption about the class of functions that we will be considering, and we must ensure that the estimated interpolant will fall within this class. In other words, the optimization problem (1) must ensure that the interpolants lie in a set of functions on which the properties of identifiability and persistence of excitation can be successfully evaluated. We have the following definition.

Definition 2. The dispersion function is coercive if for every L, $\mathcal{D}^L(z^{[L]}, \mathcal{N}^{true}(z^{[L]})) < \infty$ and

$$\mathcal{D}^L(z^{[L]}, w^{[L]}) \leq \mathcal{D}^L(z^{[L]}, \mathcal{N}^{true}(z^{[L]}))$$

implies that $\mathcal{I}^{[L]} \in \mathbb{S}$.

Note that this definition depends on a choice for both the dispersion function $\mathcal{D}^L$ and the set $\mathbb{S}$. Appropriate choices for both will be the subject of Section 4.

3. ASSUMPTIONS

The principle assumptions for our analysis are listed below.

A.1 Ω is a bounded, connected, open subset of $\mathbf{R}^p$.
A.2 The input-output data $(u_k, y_k)_{k=1}^L$ is known, z is measurable, and the signal $y - \mathcal{L}_{yu}u$ is bounded.
A.3 $\mathcal{L}$ is known, $\mathcal{L}_{ye}$ is invertible with a stable inverse, and $\mathcal{L}_{yw}$ is stable.
A.4 There is no undermodelling. That is, there exists a true nonlinearity $\mathcal{N}^{true} \in \mathbb{S}$ with $\lambda\mathcal{D}(z^{true}, N^{true}(z^{true})) \leq 1$ and true signals e^{true}, w^{true} such that $w^{true} = \mathcal{N}^{true}(z)$ and $(\pi_L w^{true}, \pi_L e^{true}) \in \mathbb{C}^L(u, y)$.
A.5 $e^{true} \in \mathbb{E} = \{e : \mu(e) < \infty\}$.
A.6 e^{true} is the output of a stable linear system driven by zero mean white noise with bounded fourth moments and independent of z^{true}

For our analysis of identifiability, we will also make use of the following property.

Definition 3. z is co-measurable if there exists an LTI system Ψ_C such that

$$\begin{bmatrix} \mathcal{L}_{ze} & \mathcal{L}_{zw} \end{bmatrix} = \mathcal{L}_{zu}\Psi_C.$$

Co-measurability of z implies that

$$z = \mathcal{L}_{zu} \begin{bmatrix} I & \Psi_C \end{bmatrix} \begin{bmatrix} u \\ e \\ w \end{bmatrix},$$

so that all possible signals z lie in $Range(\mathcal{L}_{zu})$.

4. A CLASS OF DISPERSION FUNCTIONS

A large class of dispersion functions can be developed by utilizing concepts from approximation theory, specifically, Reproducing Kernel Hilbert Spaces (Aronszajn, 1950; Schölkopf and Smola, 2002; Suykens *et al.*, 2002). Let K be a symmetric, bounded, real-valued function of two (possibly vector-valued) variables.

Definition 4. K is positive definite if for all positive integers m and all $z_1, \cdots, z_m \in \Omega$,

$$\sum_{i,j} c_i \bar{c}_j K(z_i, z_j) \geq 0$$

for all $c_i \in \mathbf{C}$. If the inequality is strict, then K is strictly positive definite.

Definition 5. Let $\mathbb{K}$ be a class of functions defined on Ω, forming a Hilbert space with inner product $\langle \cdot, \cdot \rangle_K$. A symmetric, bounded, real-valued function K is a reproducing kernel of $\mathbb{K}$ if

(1) For every $x \in \Omega$, $K(x, \cdot)$ as a function of x belongs to $\mathbb{K}$.

(2) For every $x \in \Omega$ and every $\mathcal{N} \in \mathbb{K}$

$$\mathcal{N}(x) = \langle \mathcal{N}, K(\cdot, x) \rangle_K .$$

Theorem 1. (See (Aronszajn, 1950)). To every positive definite K there corresponds one and only one class of functions with a uniquely determined inner product on it, forming a Hilbert space and admitting K as a reproducing kernel.

Thus, choosing a positive definite kernel K is equivalent to choosing a Hilbert space of functions $\mathbb{K}$. We will define $\mathcal{D}_K^L(z^{[L]}, w^{[L]})$ to be the smallest norm of a function in $\mathbb{K}$ that interpolates the points $(z^{[L]}, w^{[L]})$.

$$\mathcal{D}_K^L(z^{[L]}, w^{[L]}) = \begin{array}{c} \min\limits_{\mathcal{I} \in \mathbb{K}} \quad \|\mathcal{I}\|_K^2 \\ \text{subject to } w_k = \mathcal{I}(z_k). \end{array} \quad (2)$$

It turns out that this optimization problem has a simple solution and thus, the dispersion function can be easily calculated.

Lemma 1. Let K be a strictly positive definite reproducing kernel of $\mathbb{K}$. Then, $\mathcal{D}_K^L(z,^{[L]}, w^{[L]})$ is quadratic in $w^{[L]}$. That is,

$$\mathcal{D}_K^L(z,^{[L]}, w^{[L]}) = (w^{[L]})^* Q w^{[L]},$$

where

$$Q = \begin{bmatrix} K(z_1, z_1) & \cdots & K(z_L, z_1) \\ \vdots & \ddots & \vdots \\ K(z_1, z_L) & \cdots & K(z_L, z_L) \end{bmatrix}^{-1} .$$

The choice of K also imparts some assumed smoothness properties for the functions in $\mathbb{K}$. This behavior is consistent with that desired of a dispersion function.

We can now use these results to demonstrate that the cost function with $\mathcal{D}_K^L$ is coercive.

Theorem 2. Let z be measurable and let K be a positive definite reproducing kernel of $\mathbb{K}$. Let $\mathbb{S} = \{\mathcal{N} : \|\mathcal{N}\|_K < B\}$ for some $B > \|\mathcal{N}^{true}\|_K$ and let $\mathcal{D}^L = \mathcal{D}_K^L$. Then, $\mathcal{D}^L$ is coercive and

(1) If K is uniformly continuous, then $\mathbb{S} \subset \mathcal{C}$.
(2) If $\left.\frac{\partial K(x_1,x_2)}{\partial x_1 \partial x_2}\right|_{x_1=x_2=t}$ is uniformly bounded, then $\mathbb{S} \subset \mathcal{C}^1 \cap \mathbb{L}_\gamma$.

5. IDENTIFIABILITY

We now address the issue of identifiability. Loosely speaking, the LFT model structure in Figure 1 is identifiable if it is possible to determine the static nonlinear block $\mathcal{N}$ uniquely on the basis of noise-free input-output experiments. As is well known, identifiability concepts are of fundamental importance in system identification (Ljung, 1999). Let us begin with the following definition.

1 Parameterize $\mathbb{X}$ as $X = \sum_i^q \theta_i K_i$, where $K_i \in \mathbf{R}^{m \times p}$ and q is the dimension of $\mathbb{X}$.
2 Perform coprime factorization of systems $\mathcal{L}_{yw} = D_{yw}^{-1} N_{yw}$ and $\mathcal{L}_{zu} = N_{zu} D_{nu}^{-1}$ with poles of N_{yw} and N_{zu} at zero.
3 Form Toeplitz matrix $\mathcal{T}$ of (finite) impulse response of $\sum_i^q \theta_i N_{yw} K_i N_{zu}$
4 Check null space of $\mathcal{T}$. If non-trivial, then the LFT model structure is not identifiable.

Fig. 2. Computatable identifiability test.

Definition 6. The LFT model structure in Figure 1 is <u>identifiable</u> if for every $\mathcal{N}_1 \in \mathbb{S}$, there does not exist $\mathcal{N}_1 \neq \mathcal{N}_2 \in \mathbb{S}$ such that $\mathcal{M}(\mathcal{L}, \mathcal{N}_1) = \mathcal{M}(\mathcal{L}, \mathcal{N}_2)$.

In the remainder of this section, we will develop a test for identifiability. Let the following assumptions hold for this section.

I.1 Every $\mathcal{N} \in \mathbb{S}$ is differentiable on Ω.
I.2 z is measurable and co-measurable.
I.3 For every $\mathcal{N} \in \mathbb{S}$, there exists $\omega \in \Omega$ such that $\mathcal{N}(\omega) = 0$.
I.4 The LFT model structure $\mathcal{M}(\mathcal{L}, \mathcal{N})$ is well-posed.

The following theorem demonstrates that under these conditions, the identifiability can be evaluated by considering only the "forward" path from u to y. This allows us to disregard the effect of the unmeasured signal e and the feedback interconnection between z and w. For this, let us define

$$\mathbb{X} = \begin{bmatrix} X^{[1]} & & \\ & \ddots & \\ & & X^{[m]} \end{bmatrix},$$

where $X^{[i]}$ has the same input-output dimensions as $\mathcal{N}^{[i]}$. We have the following result.

Theorem 3. Let Assumptions I.1-I.4 hold. The LFT model structure is identifiable if and only if there does not exist $0 \neq X \in \mathbb{X}$ such that $\mathcal{L}_{yw} X \mathcal{L}_{zu} = 0$.

This result allows us to formulate a computationally simple test for identifiability. The goal is to compute the existence of the required matrix X. Since we know that the mapping from X to $\mathcal{L}_{yw} X \mathcal{L}_{zu}$ is linear, we can determine if this mapping has a nontrivial nullspace on $\mathbb{X}$. A computable test for identifiablity is given in Figure 2.

6. PERSISTENCE OF EXCITATION

Given an identifiable LFT model structure, an input signal is persistently exciting if two different nonlinear functions always produce different output sequences, thus enabling an identification algorithm to distinguish between them. General conditions for persistence of excitation appear to be very complex. However, considerable insight can be obtained in the case where the signal z is measurable.

For the remainder of this section, we will assume that the LFT model structure is identifiable. Our condition for persistence of excitation will be stated in terms of the signal z.

Definition 7. The signal z is persistently exciting if for every $\mathcal{N} \in \mathbb{S}$,

$$\lim_{L\to\infty} \mu^L(\mathcal{L}_{ye}^{-1}\mathcal{L}_{yw}\mathcal{N}(z^{[L]})) = 0 \implies \|\mathcal{N}\|_\infty = 0.$$

Note that the persistence of excitation condition is dependent on the linear block of the LFT model structure. More specifically, a signal z that is persistently exciting for one LFT model structure may be not persistently exciting for another. While persistence of excitation conditions are commonly independent of the model, a notion of persistence of excitation that considers the model structure may be more appropriate in the case of general interconnected systems. One should also note that if the LFT model structure is not identifiable, no persistently exciting signal exists. This explains the necessity of the identifiability assumption throughout this section.

6.1 *Examples of Persistently Exciting Signals*

We now pose the question, "What types of signals are persistently exciting?" In order to present an illustrative example, let the additional assumptions hold.

PP.1 Let $H = \mathcal{L}_{ye}^{-1}\mathcal{L}_{yw}$ be an FIR filter with $t+1$ taps.

PP.2 For every $\mathcal{N} \in \mathbb{S}$, there exists $\omega \in \Omega$ such that $\mathcal{N}(\omega) = 0$.

We begin with the following definition.

Definition 8. $(z_k)_{k=1}^{\infty}$ is dense on Ω if for every $\omega \in \Omega$ and every $\epsilon > 0$, there exists an index j such that

$$\|z_j - \omega\| < \epsilon.$$

In the following example, we demonstrate how to construct a persistently exciting signal.

Example 1. Let $\mathcal{M}(\mathcal{L},\mathcal{N})$ be identifiable and $\mathbb{S} \subset \mathbb{N} \cap \mathcal{C}$. We can write the signal $\tilde{e} = H\mathcal{N}(z)$ as

$$\tilde{e}_k = h_0^*\mathcal{N}(z_k) + h_1^*\mathcal{N}(z_{k-1}) + \cdots + h_t^*\mathcal{N}(z_{k-t}).$$

Define 0_t to be a sequence of t zeros. We will construct the signal z as follows. Let

$$z = (0_t, a_1, 0_t, a_2, 0_t, a_3, 0_t, \dots),$$

where $(a_k)_{k=1}^{\infty}$ is dense on Ω. Then,

$$\tilde{e} = (\dots, h_0^*\mathcal{N}(a_1), h_1^*\mathcal{N}(a_1), \dots, h_t^*\mathcal{N}(a_1), h_0^*\mathcal{N}(a_2), h_1^*\mathcal{N}(a_2), \dots).$$

Suppose that $\lim_{L\to\infty}\mu(\tilde{e}) = 0$. Then, for every $\epsilon > 0$ there exists I such that for almost every $i > I$, $\|h_j^*\mathcal{N}(a_i)\| < \epsilon$ for $j = 0, \dots, t$. Since $\mathcal{N}$ is continuous and $(a_i)_{i=1}^{\infty}$ is dense on Ω, this implies that $h_j\mathcal{N} = 0$ on Ω for $j = 0, \dots, t$. We can then conclude that $H\mathcal{N} = 0$ on Ω. The identifiability of the LFT model structure results in $\mathcal{N} = 0$. As a result, the signal z is persistently exciting.

Sufficient conditions for the persistence of excitation condition will be formalized in the following theorem. In essence, we need an input for which we can search through to find sequences of length $t+1$ whose linear combinations resemble impulses that explore Ω.

Theorem 4. Let Assumptions PP.1 and PP.2 hold. Let the LFT model structure be identifiable. Let $\mathbb{S} \subset \mathbb{N} \cap \mathcal{C} \cap \mathbb{L}_\gamma$. Define $Z_m = [z_m^* \ \cdots \ z_{m+t}^*]^* \in \mathbf{R}^{p(t+1)}$.

Suppose that for any $\delta > 0$, there exists closed sets $S_i \subseteq \Omega$, points $\xi_i \in S_i$, and

(1) For every $s \in S_i$, $M > 0$, and integer $0 \leqslant q \leqslant t$, there exists indices $m > M$ and $n > M$ such that $\|s - z_{n+q}\| < \delta$, $\|\xi_i - z_{m+q}\| < \delta$, and $\|Z_m - Z_n - [0_{pq} \ z_{m+q}^* - z_{n+q}^* \ 0_{p(t-q)}]^*\| < \delta$.
(2) For any two sets S_i and S_j, there exists a sequence of sets (S_{n_k}) such that $S_{n_k} \cap S_{n_{k+1}} \neq \{\phi\}$ and $S_i, S_j \in \{S_{n_k}\}$.
(3) $\overline{\bigcup S_i} = \Omega$.

Then, z is persistently exciting.

Figure 3 illustrates the conditions in the previous theorem. This is a plot of the space of sequences (z_{m-1}, z_m, z_{m+1}). If the input is dense in the regions shown, then the input satisfies condition 1 for $\Omega = [0,1]$, $q = 1$, and $t = 2$.

Unfortunately, the conditions of the previous theorem are somewhat convoluted. However, if we restrict ourselves to considering z such that $[z_{m-t}^* \ \cdots \ z_m^* \ \cdots \ z_{m+t}^*]^*$ are sets in $\Omega \times \Omega \times \cdots \times \Omega$ with non-empty interior, a simpler characterization is possible, albeit with stronger conditions.

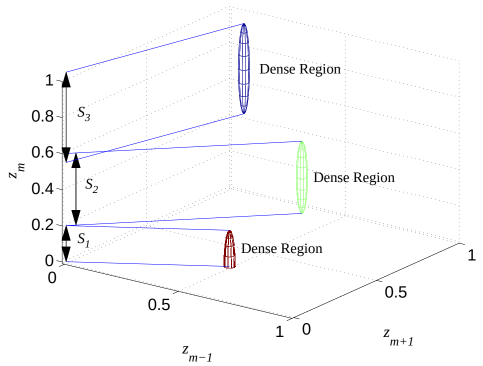

Fig. 3. Illustration of the persistence of excitation condition.

Theorem 5. Let $\Omega \subseteq \mathbf{R}^p$. Define $\Pi : \mathbf{R}^{p(2t+1)} \rightarrow \mathbf{R}^p$ to be the projection operator such that $\Pi[z^*_{m-t} \quad \cdots \quad z^*_m \quad \cdots \quad z^*_{m+t}]^* = z_m$. Let $\overline{\Omega}$ be the closure of Ω. If (Z_m) is dense on a subset $S \in \mathbf{R}^{p(2t+1)}$ with non-empty interior such that $\Pi S \supset \overline{\Omega}$, then z is persistently exciting.

The key property of a persistently exciting signal is that it satisfies a density in both space and time.

7. CONVERGENCE PROPERTIES

Utilizing the preceding results, the following convergence theorem can be stated.

Theorem 6. Let Assumptions A.1 - A.6 hold. Let $\mathcal{D}^L = \mathcal{D}^L_K$ with $\frac{\partial K(t,t)}{\partial x_1 \partial x_2}$ uniformly bounded. If the LFT model structure is identifiable and z is persistently exciting, then

$$\|\hat{\mathcal{I}}^{[L]} - \mathcal{N}^{true}\|_\infty \longrightarrow 0 \quad \text{as } L \longrightarrow \infty.$$

with probability 1.

Since the dispersion function is coercive, we are ensured that $\hat{\mathcal{I}}^{[L]} \in \mathbb{S}$. A persistently exciting signal z then generates an input-output data set for which a unique estimate is obtained as $L \rightarrow \infty$. Assumption A.6 along with the identifiability of the LFT model structure then ensure that in the limit, $\mathcal{N}^{true}$ is almost surely the unique minimizer.

8. CONCLUSION

In (Hsu *et al.*, 2005*c*), an algorithm was developed for the structured nonlinear system identification problem. In this paper, a class of dispersion functions that retains the least squares nature of the algorithm is described. We also explored notions of identifiability and persistence of excitation to provide conditions under which the estimated nonlinearity converges uniformly to the true nonlinearity.

REFERENCES

Aronszajn, N. (1950). Theory of reproducing kernels. *Trans. Amer. Math. Soc.* **68**, 337–404.

Barron, A.R. (1993). Uniform approximation bounds for superpositions of a sigmoidal function. *IEEE Transactions on Information Theory* **39**, 930–945.

Claassen, M. (2001). System identification for structured nonlinear systems. *Ph.D Dissertation, University of California at Berkeley.*

Greblicki, W. (1997). Nonparametric approach to wiener system identification. *IEEE Transactions on Circuits and Systems* **44**, 538–545.

Hsu, K., C. Novara, M. Milanese and K. Poolla (2005*a*). Parametric identification of static nonlinearities in a general interconnected system. *Proceedings of the 16th IFAC World Congress.*

Hsu, K., M. Claassen, C. Novara, P. Khargonekar, M. Milanese and K. Poolla (2005*b*). Nonparametric identification of static nonlinearities in a general interconnected system. *Proceedings of the 16th IFAC World Congress.*

Hsu, K., T. Vincent and K. Poolla (2005*c*). A kernel based approach to structured nonlinear system identification part I: Algorithms. *Proceedings of the 2006 IFAC Symposium on System ID.*

Hsu, K., T. Vincent, C. Novara, M. Milanese and K. Poolla (2005*d*). Identification of nonlinear maps in interconnected systems. *Proceedings of the 2005 Conference on Decision and Control.*

Juditsky, A., H. Hjalmarsson, A. Benveniste, B. Delyon, L. Ljung, J. Sjöberg and Q. Zhang (1995). Nonlinear black-box models in system identification: Mathematical foundations. *Automatica* **31**, 1725–1750.

Ljung, L. (1999). *System Identification Theory for the User, 2nd Edition.* Prentice Hall. Upper Saddle River, N.J.

Schölkopf, B. and A.J. Smola (2002). *Learning with Kernels.* MIT Press. Cambridge.

Sjöberg, J., Q. Zhang, L. Ljung, A. Benveniste, B. Delyon, P.Y. Glorennec, H. Hjalmarsson and A. Juditsdy (1995). Nonlinear black-box modeling in system identification: A unified overview. *Automatica* **31**, 1691–1724.

Suykens, J.A.K., T.V. Gestel, J.D. Brbanter, B. De Moor and J. Vanderwalle (2002). Least squares support vector machines.

Wemhoff, E. (2003). Signal estimation in structured nonlinear systems with unknown functions. *Ph.D Dissertation, University of California at Berkeley.*

Wemhoff, E., A. Packard and K. Poolla (1999). On the identification of nonlinear maps in a general interconnected system. *Proceedings of the 1999 American Controls Conference.*