A KERNEL BASED APPROACH TO STRUCTURED NONLINEAR SYSTEM IDENTIFICATION PART I: ALGORITHMS

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: We consider interconnected systems consisting of linear time-invariant systems and static nonlinear maps. Under the assumptions that the linear dynamics are known and the input to the nonlinear maps are measurable, it is shown that the identification problem can be reduced to a least squares problem. An identification algorithm utilizing a kernel-based dispersion function is proposed.

Keywords: nonlinear system identification, structured systems, kernel

1. INTRODUCTION

The identification of nonlinear systems is daunting due to the large class of behaviors that a general nonlinear system can exhibit. In practice, application of *a priori* information can yield a more feasible identification problem. This information may be of many different forms. In this paper, we consider the situation in which the interconnection between the various components of the system is known. We refer to this as a structured nonlinear system.

The inclusion of available *a priori* information is of fundamental importance in system identification. Information regarding the structural interconnections is generally known with a high degree of confidence due to an understanding of the underlying relationships between system components. While black-box methods of identification 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, and in addition 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. In particular, we restrict our analysis to those interconnected systems consisting of linear time-invariant (LTI) systems and static nonlinear maps. Furthermore, we assume that the LTI systems are known. The goal is to identify the nonlinear components of the interconnection.

The frequently studied Wiener and Hammerstein block structures are special cases of our formulation. However, it is important to note that the class of problems we wish to identify may involve *complex* interconnections. In this paper, the interconnected systems we consider may contain any finite number of LTI systems and static nonlinear maps.

One approach to the identification of static nonlinearities is to assume parameterizations such as as radial basis functions, polynomial expansions, or finite Fourier series (Greblicki, 1989; Kryzak,

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

1989). While these parameterizations often yield good approximations, the resulting estimates may be sensitive to the proper choice of basis functions. In addition, this technique of enforcing parameterizations on unknown elements in an interconnected sytem can lead to substantially biased estimates. The algorithm presented in this paper will be nonparametric in that an *a priori* basis expansion of the nonlinearity will not be assumed.

Nonparametric methods already exist to treat the identification problem in extreme generality, for example using splines, kernel smoothing 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 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. In this paper, we propose a kernel based approach and propose an identification algorithm. In a companion paper (Hsu *et al.*, 2005*c*), we investigate conditions under which the optimization problem is well posed. In particular, we formulate conditions on the interconnection (identifiability) and on the input signals (persistence of excitation) such that the nonlinearity estimate converges.

NOTATION

x	infinite vector-valued sequence $(x_k)_{k=1}^{\infty}$
$x^{[L]}$	finite vector-valued sequence $(x_k)_{k=1}^{L-1}$
x_k	kth element of x or $x^{[L]}$
x_k^*	complex-conjugate transpose of x_k
π_L	L-sample truncation operator
$\mathcal{L}$	known LTI system
$\mathcal{N}$	static nonlinear function
$\mathbb{N}$	class of static nonlinear functions with a particular input-output structure
$u_k \in \mathbf{R}^{n_u}$	known input to $\mathcal{L}$
$e_k \in \mathbf{R}^{n_e}$	unknown input to $\mathcal{L}$
$y_k \in \mathbf{R}^{n_y}$	measured output of $\mathcal{L}$
$z_k \in \mathbf{R}^{p}$	input to $\mathcal{N}$
$w_k \in \mathbf{R}^{m}$	output of $\mathcal{N}$
$\mathcal{D}^L$	dispersion function
$\mathcal{I}$	interpolant of $(z^{[L]}, w^{[L]})$
$\mathbb{K}$	set of potential interpolants
$\mu^L(e^{[L]})$	metric for $e^{[L]}$: $\mu^L(e^{[L]}) = \frac{1}{L}\sum_{i=1}^{L} e_i^* e_i$
p	number of inputs to $\mathcal{N}$
$\Omega^{[i]} \subseteq \mathbf{R}^{p_i}$	domain where $\mathcal{N}^{[i]}$ is to be identified
$\Omega \subseteq \mathbf{R}^{p}$	$\Omega^{[1]} \times \Omega^{[2]} \times \cdots \times \Omega^{[m]}$

2. PROBLEM DESCRIPTION

To present all structured systems under a common framework, we gather the known linear blocks and unknown nonlinearities into a Linear Fractional Transformation. As depicted in Figure 1, the

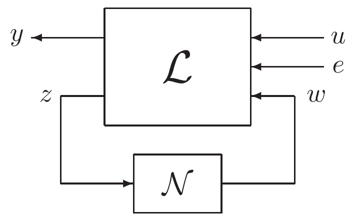


Fig. 1. LFT model structure

linear block $\mathcal{L}$ contains all the LTI systems in the interconnection. The nonlinear block $\mathcal{N}$ consists of the nonparametric static maps. The block $\mathcal{L}$ and the signals u and y are known. The input e is an unknown noise signal. The goal is to identify the components of the static nonlinear map $\mathcal{N}$.

We will refer to the generalized system in Figure 1 as the *LFT model structure*. The LFT model structure also retains structural information by specifying that $\mathcal{N} \in \mathbb{N}$ satisfies a particular known interconnection structure between its inputs and outputs. That is, any given output depends only on a known subset of the vector-valued input z. We will denote the interconnection structure of any $\mathcal{N} \in \mathbb{N}$ as

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

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

In this paper, we will specialize to a class of LFT model structures that satisfy the following property.

Definition 1. z is <u>measurable</u> if there exists an LTI system Ψ_M such that

$$\left[\, \mathcal{L}_{ze} \;\; \mathcal{L}_{zw} \,\right] = \Psi_M \left[\, \mathcal{L}_{ye} \;\; \mathcal{L}_{yw} \,\right].$$

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 u, y, and $\mathcal{L}$ alone.

3. THE IDENTIFICATION ALGORITHM

We submit that when identifying nonparametric static maps, selecting a candidate function is equivalent to selecting a sequence of input-output pairs (z, w) that represent a sampling of that function. However, we desire that these sequences have the following properties.

- Consistency with input-output data.
- Consistency with a priori knowledge on unknown inputs.
- (z, w) must be representative of a static map.

A straightforward method of choosing between candidate (z, w) sequences is to develop a metric for the above properties. However, a metric for staticness perhaps requires some discussion. Let us consider an illustrative example.

Example 1. Consider three systems: a hysteresis, an LTI system, and $\tan^{-1}(z)$. White noise is applied as an input to each system and the output is measured. A scatter plot for each system is shown in Figure 2. It is intuitive that the third plot is most representative of a static function. □

We speculate that the mechanism by which one intuitively evaluates the scatter plot is as follows. One mentally interpolates the points in the scatter plot and judges the resulting interpolant using a metric that is biased towards *simple* functions. Thus, our metric of staticness will measure the complexity of interpolating the points (z, w). The development of this metric will be presented in Section 4.

Let us now introduce the metrics associated with the desired properties listed above.

- (Consistency with input-output data). Let us partition $\mathcal{L}$ conformably as
$$\mathcal{L} \sim \begin{bmatrix} \mathcal{L}_{yu} & \mathcal{L}_{ye} & \mathcal{L}_{yw} \\ \mathcal{L}_{zu} & \mathcal{L}_{ze} & \mathcal{L}_{zw} \end{bmatrix}.$$
We will define
$$\mathbb{C}^L(u, y) = \Big\{ (w^{[L]}, e^{[L]}) : (2) \text{ and } (3) \text{ are satisfied for } (u, y) \Big\} \quad (1)$$
$$\pi_L y = \pi_L(\mathcal{L}_{yu} u + \mathcal{L}_{ye} e^{[L]} + \mathcal{L}_{yw} w^{[L]}) \quad (2)$$
$$\pi_L z = \pi_L(\mathcal{L}_{zu} u + \mathcal{L}_{ze} e^{[L]} + \mathcal{L}_{zw} w^{[L]}). \quad (3)$$
Here, $\mathbb{C}^L(u, y)$ is the set of signals $(w^{[L]}, e^{[L]})$ that are consistent with the measured data (u, y). $\mathbb{C}^L(u, y)$ is assumed to be nonempty.
- (Consistency with a priori information on e.) For the identification problem, we will consider signals $e \in \mathbb{E} = \{e : \|e\|_\infty < \infty\}$ and define the metric
$$\mu^L(e) = \frac{1}{L} \sum_{i=1}^{L} e_i^* e_i.$$
- (Staticness) Let $\mathcal{D}^L(z^{[L]}, w^{[L]})$ be a function that measures the complexity of interpolating the points $(z^{[L]}, w^{[L]})$. Due to its importance, we give it a name that captures its purpose: *the dispersion function.*

In order to solve the structured nonlinear system identification problem, we propose to rate potential solutions $w^{[L]}$ via minimizing the following cost function:

$$J^L(w^{[L]}) = \min_{e^{[L]}} \mu^L(e^{[L]}) + \beta(L)\mathcal{D}^L(z^{[L]}, w^{[L]}) \quad \text{s.t. } (w^{[L]}, e^{[L]}) \in \mathbb{C}^L(u, y). \quad (4)$$

We will use the convention that if there does not exist $e^{[L]}$ such that $(w^{[L]}, e^{[L]}) \in \mathbb{C}(u, y)$, then $J^L(w^{[L]}) = \infty$. Thus, J^L captures the cost of choosing $w^{[L]}$ as a particular estimate as a weighted sum of the dispersion plus the best case noise magnitude.

The identification algorithm is then solved as follows:

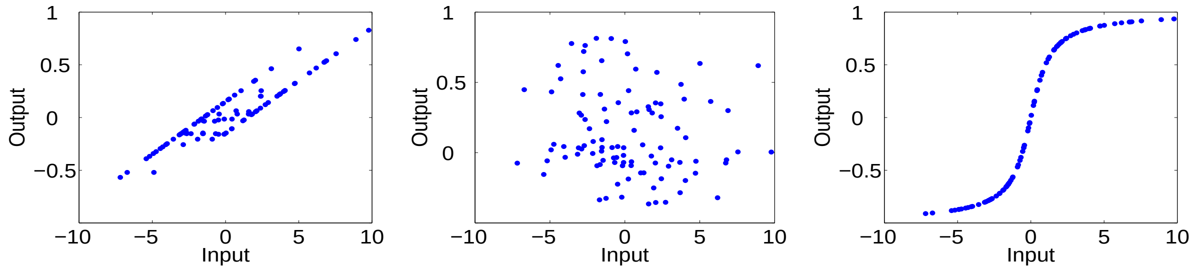
(1) Find $\hat{w}^{[L]} = \arg\min J^L(w^{[L]})$.
(2) Form an estimate of $\mathcal{N}$ via an interpolation of these points, denoted $\hat{\mathcal{I}}^{[L]}$.

Because the linear constraints given by $\mathcal{L}$ are affine in the unknown signals, $\mathbb{C}$ defines a linear variety that can be parameterized as
$$\begin{bmatrix} e \\ w \end{bmatrix} = \begin{bmatrix} e^\circ \\ w^\circ \end{bmatrix} + \begin{bmatrix} \mathcal{B}_e \\ \mathcal{B}_w \end{bmatrix} s$$
where
$$\begin{bmatrix} \mathcal{L}_{ye} & \mathcal{L}_{yw} \end{bmatrix} \begin{bmatrix} \mathcal{B}_e \\ \mathcal{B}_w \end{bmatrix} = 0,$$
(w°, e°) is one element of $\mathbb{C}$, and s is a free signal. Thus, (4) can be formulated as an unconstrained minimization problem over the free signal s.

4. THE $\mathcal{D}_K$ DISPERSION FUNCTION

A class of dispersion functions can be obtained by using concepts from approximation theory (Schölkopf and Smola, 2002). To develop this class of functions, let us first specialize to the single-input single-output case. Suppose that we have the following: a set $\mathbb{K}$ of potential interpolants, an interpolant $\mathcal{I} \in \mathbb{K}$, and a metric $\|\mathcal{I}\|_{\mathbb{K}}$ which measures the complexity of this function. Let us make the following definition.

Definition 2.
$$\mathcal{D}_K^L(z^{[L]}, w^{[L]}) = \min_{\mathcal{I} \in \mathbb{K}} \|\mathcal{I}\|_K^2 \quad \text{s.t. } w_k = \mathcal{I}(z_k).$$



(a) (b) (c)

Fig. 2. Data sampled using white noise input for (a) hysteresis, (b) LTI system, (c) $\tan^{-1}$.

That is, the dispersion of $(z^{[L]}, w^{[L]})$ is the cost of the simplest interpolation of those points by elements from the set $\mathbb{K}$.

One choice for the class of functions $\mathbb{K}$ is defined in the following manner. Let Ω be the interval $[0, 1]$. Let $(\lambda_n)_{n=1}^{\infty}$ be a sequence of nonnegative real numbers such that $\sum_{n=1}^{\infty} \lambda_n < \infty$. Define

$$\mathbb{K} = \left\{ \mathcal{I} : \mathcal{I}(z) = \sum_{n=1}^{\infty} a_n \phi_n(z), \sum_{n=1}^{\infty} \frac{|a_n|^2}{\lambda_n} < \infty \right\},$$

where $\phi_n(z) = \exp(-2\pi jnz)$. That is, the interpolants have a Fourier series expansion with coefficients $\{a_n\}_{n=1}^{\infty}$ that must decay to zero since $\lambda_n \to 0$. This in effect enforces the functions $\mathcal{I} \in \mathbb{K}$ to possess a smoothness property by limiting the influence of the higher frequency components.

For any two functions $\mathcal{I}_1(\cdot) = \sum_{n=1}^{\infty} a_n \phi_n(\cdot) \in \mathbb{K}$ and $\mathcal{I}_2(\cdot) = \sum_{n=1}^{\infty} b_n \phi_n(\cdot) \in \mathbb{K}$, we can define an inner product

$$\langle \mathcal{I}_1, \mathcal{I}_2 \rangle_K = \sum_{n=1}^{\infty} \frac{a_n^* b_m}{\lambda_n}.$$

Using the norm induced by this inner product, we can rewrite $\mathbb{K}$ as

$$\mathbb{K} = \left\{ \mathcal{I} : \mathcal{I}(z) = \sum_{n=1}^{\infty} a_n \phi(z), \|\mathcal{I}\|_K^2 < \infty \right\}. \quad (5)$$

The norm $\|\cdot\|_K$ clearly penalizes variation. If the points $\{z_k, w_k\}_{k=1}^{L}$ induces high frequency oscillations in the interpolant, the "cost" of interpolating is higher.

As defined, it appears that we need to compute the Fourier coefficients of a particular interpolant to determine its cost. However, it turns out that the cost of the minimizing interpolant can be computed in a straightforward manner. The key to computation utilizes a kernel function defined as follows. Let

$$K(z_1, z_2) = \sum_{n=1}^{\infty} \lambda_n \phi_n(z_1)^* \phi_n(z_2). \quad (6)$$

Since $\phi_n(z_2 - z_1) = \phi_n(z_1)^* \phi_n(z_2)$, we have $K(z_1, z_2) = K(z_2 - z_1) = \sum_{n=1}^{\infty} \lambda_n \phi_n(z_2 - z_1)$. Note that the choice of $(\lambda_n)_{n=1}^{\infty}$ is equivalent to a particular choice of a symmetric, real-valued kernel function.

Theorem 1. With $\mathbb{K}$ and K defined as in (5) and (6), $\mathcal{D}_K^L(z^{[L]}, w^{[L]})$ is quadratic in $w^{[L]}$. That is,

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

where

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

We see from above there is an equivalence between choosing $(\lambda_n)_{n=1}^{\infty}$ and $\{\phi_n\}_{n=1}^{\infty}$ and choosing a kernel K. In fact, all calculations can be done using only K. In the following sections, we make the choice of a Gaussian kernel

$$K(z_1, z_2) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(z_1 - z_2)^2}{2\sigma^2} \right),$$

which corresponds to the set $\mathbb{K}$ defined above. By employing a multivariate Gaussian kernel, the multivariate extension is straightforward.

5. COMPUTING THE ESTIMATES

Because of the assumption that z is measurable, the identification algorithm can be reduced to a simple computational problem. In the following sections, we will denote a stabilizable and detectable realization of $\mathcal{L}$ as

$$\mathcal{L} \sim \left[\begin{array}{c|ccc} A & B_u & B_e & B_w \\ \hline C_y & D_{yu} & D_{ye} & D_{yw} \\ C_z & D_{zu} & D_{ze} & D_{zw} \end{array} \right]. \quad (7)$$

For ease of notation, let us define

$$B = \begin{bmatrix} B_e & B_w \end{bmatrix}$$
$$D = \begin{bmatrix} D_{ye} & D_{yw} \end{bmatrix}.$$

5.1 Problem Formulation and Solution

As dicussed earlier, we can parameterize the set of all signals (e, w) consistent with the input-output data (u, y) as

$$\begin{bmatrix} e \\ w \end{bmatrix} = \begin{bmatrix} e^\circ \\ w^\circ \end{bmatrix} + \begin{bmatrix} \mathcal{B}_e \\ \mathcal{B}_w \end{bmatrix} s,$$

1 Solve the Riccati Difference equation for P_t for $t = 0, \cdots, L-1$ with $P_0 = 0$.

$$\begin{aligned} M_t &= AP_tC_y^* + BD^* \\ P_{t+1} &= AP_tA^* + BB^* - M_tQ_tM_t^* \\ Q_t &= (C_yP_tC_y^* + DD^*)^{-1} \end{aligned}$$

2 Compute the matrices

$$\begin{aligned} K_t &= -(AP_tC_y^* + BD^*)Q_t \\ F_t &= A + K_tC_y \\ G_t^u &= B_u + K_tD_{yu} \\ G_t^e &= B_e + K_tD_{ye} \\ G_t^w &= B_w + K_tD_{yw} \end{aligned}$$

3 Compute innovations ν and iterate for $t = 0, \cdots, L-1$ with $m_0 = 0$

$$\begin{bmatrix} m_{t+1} \\ \nu_t \end{bmatrix} = \begin{bmatrix} F_t & G_t^u & -K_t \\ C_y & D_u & -I \end{bmatrix} \begin{bmatrix} m_t \\ u_t^d \\ y_t^d \end{bmatrix}$$

4 Compute adjoint filter (x_o, e) and iterate for $t = L-1, \cdots, 0$ with $\xi_{L-1} = 0$

$$\begin{bmatrix} \xi_{t-1} \\ e_t^\circ \\ w_t^\circ \end{bmatrix} = \begin{bmatrix} F_t^* & C_y^* \\ -(G_t^e)^* & -D_e^* \\ -(G_t^w)^* & -D_w^* \end{bmatrix} \begin{bmatrix} \xi_t \\ Q_t\nu_t \end{bmatrix}$$

Fig. 3. Computation of particular solutions e°, w°.

$$\text{where} \quad \begin{bmatrix} \mathcal{L}_{ye} & \mathcal{L}_{yw} \end{bmatrix} \begin{bmatrix} \mathcal{B}_e \\ \mathcal{B}_w \end{bmatrix} = 0.$$

Here, $[\mathcal{B}_e^* \ \mathcal{B}_w^*]^*$ is a basis for the null space of $[\mathcal{L}_{ye} \ \mathcal{L}_{yw}]$, s is a free signal and (e°, w°) is a particular solution. Since w is an affine function of the decision variable s, the $\mathcal{D}_K$ dispersion function will be a quadratic function of s as well. In this case, we can formulate the optimization problem as

$$\min_s \left\| \begin{bmatrix} \frac{1}{\sqrt{L}} e^\circ \\ \sqrt{\beta(L)} Q^{\frac{1}{2}} w^\circ \end{bmatrix} + \begin{bmatrix} \frac{1}{\sqrt{L}} \mathcal{T}_{\mathcal{B}_e} \\ \sqrt{\beta(L)} Q^{\frac{1}{2}} \mathcal{T}_{\mathcal{B}_w} \end{bmatrix} s \right\|^2,$$

where $\mathcal{T}_{\mathcal{B}_e}$ and $\mathcal{T}_{\mathcal{B}_w}$ are the Toeplitz matrices of the systems $\mathcal{B}_e$ and $\mathcal{B}_w$, respectively, and $Q^{\frac{1}{2}}$ is the Cholesky factorization of Q. To solve the above optimization problem, we need to obtain the solutions e°, w° and the systems $\mathcal{B}_e, \mathcal{B}_w$. Calculations of these can be found in the following sections.

5.2 Particular Solutions (e°, w°)

A particular solution (e°, w°) may be obtained via a solving a Kalman smoothing problem. This will yield the minimum norm solution (Kailath *et al.*, 2000). To ease computation of the Kalman Filter, let as assume that DD^* is invertible. The algorithm is shown in Figure 3.

An approximate solution for a time-invariant $\mathcal{L}$ can be obtained by using the steady-state Kalman Filter gains, which works very well in practice.

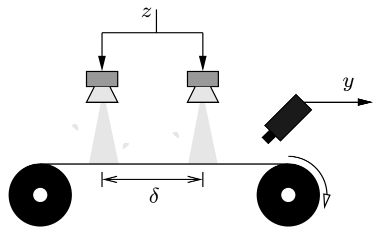


Fig. 5. Spray deposition system.

5.3 Basis Systems $\mathcal{B}_e$ and $\mathcal{B}_w$

To compute the null basis systems $\mathcal{B}_e$ and $\mathcal{B}_w$, let us assume that DD^* is invertible. Define $\Pi = D^*(DD^*)^{-1}$. We can then realize the systems as

$$\begin{bmatrix} \mathcal{B}_e \\ \mathcal{B}_w \end{bmatrix} \sim \left[\begin{array}{c|c} A - B\Pi C_y & B(I - \Pi D) \\ \hline -\Pi C_y & I - \Pi D \end{array} \right].$$

The Toeplitz matrices $\mathcal{T}_{\mathcal{B}_e}$ and $\mathcal{T}_{\mathcal{B}_w}$ can then be formed via this realization.

6. EXAMPLES

6.1 Spray Deposition System

Consider the spray deposition system in Figure 5. Here, there are two spray sources depositing material onto a moving substrate. The sources are separated by a known distance δ, and the total deposited material y is measured. Modeled as a Hammerstein system, the LFT model structure for this system contains the blocks

$$\mathcal{L} = \begin{bmatrix} 0 & 1 & [1 \ z^{-2}] \\ 1 & 0 & 0 \end{bmatrix} \quad \text{and} \quad \mathcal{N} = \begin{bmatrix} f \\ g \end{bmatrix}.$$

The true actuator nonlinearities used were $f = -|z|$ and $g = 0.5(\sin((3z)^2) + (2z)^2 - 2)$. The input signal was a sinusoid of frequency 0.1 radians per second sampled at 1 Hz and e was an i.i.d. Gaussian process with mean 0 and variance 1. The results are shown in Figure 4 for various choices of $\beta(L)$. In practice, one would select $\beta(L)$ by means of a cross-validation procedure (Ljung, 1999).

Note that even though the input signal had very low complexity, we are still able to recover the nonlinearities. The conditions for persistently exciting signals will be discussed in a companion paper (Hsu *et al.*, 2005*c*).

6.2 Multi-Input Single-Output Nonlinearity

We now consider an interconnection with a 2-input 1-output nonlinearity. The linear block $\mathcal{L}$ of the LFT was chosen to be

$$\begin{bmatrix} \mathcal{L}_{yu} & I & \mathcal{L}_{yw} \\ \mathcal{L}_{zu} & 0 & \mathcal{L}_{zw} \end{bmatrix},$$

where the systems $\mathcal{L}_{yu}, \mathcal{L}_{yw}, \mathcal{L}_{zu}$, and $\mathcal{L}_{yw}$ are random stable systems and $\mathcal{L}_{yu}$ is a three state,

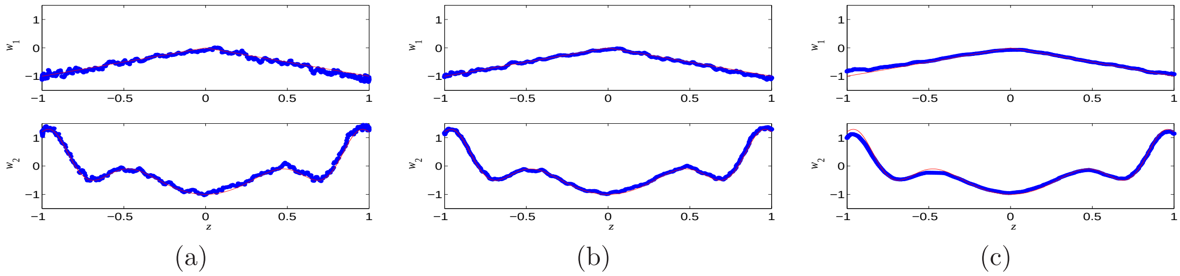


Fig. 4. True (line) and estimated nonlinearities (dots) using (a) $\beta = 0.001$, (b) $\beta = 0.01$, (c) $\beta = 0.1$.

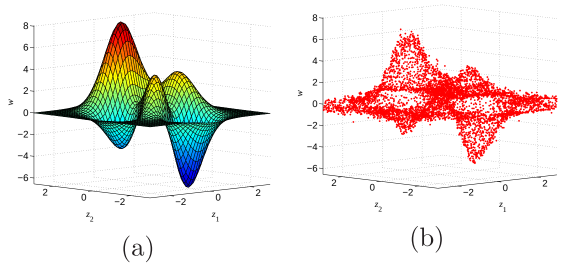


Fig. 6. (a) True and (b) estimated nonlinearity.

two-input system. The two inputs were chosen to be i.i.d uniform processes with mean 0 and variance 1.5. The noise signal is an i.i.d. Gaussian process with mean 0 and variance 1. The resulting estimate is shown in Figure 6.

7. CONCLUSION

We have examined the structured nonlinear system identification problem utilizing Linear Fractional Transformations to systematically incorporate *a priori* structrual information. Although the general structured nonlinear system identification problem is very complex, we have shown that when the linear dynamics are known and the inputs to the nonlinearities are measurable, the system identification algorithms developed here can be formulated as a least squares problem.

REFERENCES

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. (1989). Nonparametric orthogonal series identification of hammerstein systems. *International Journal of Systems Science* **20**, 2355–2367.

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 II: Convergence and consistency. *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.

Kailath, T., A. Sayed and B. Hassibi (2000). *Linear Estimation.* Prentice Hall. Upper Saddle River, New Jersey.

Kryzak, A. (1989). Identification of discrete hammerstein systems by the fourier series regression estimate. *International Journal of Systems Science* **20**, 1729–1744.

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.

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.*