Nonlinear dynamics identified by multi-index models

David Lindgren, Lennart Ljung

Division of Automatic Control

E-mail: lindgren@isy.liu.se, ljung@isy.liu.se

14th June 2007

Report no.: [LiTH-ISY-R-2791](#)

Accepted for publication in Proc. 16th IFAC World Congress, Prague

Address:

Department of Electrical Engineering

Linköpings universitet

SE-581 83 Linköping, Sweden

WWW: <http://www.control.isy.liu.se>

AUTOMATIC CONTROL
REGLERTEKNIK
LINKÖPINGS UNIVERSITET

Abstract

For a class of nonlinear systems, the residual of a well fitted model has low intrinsic dimensionality. For these systems, a particular low-dimensional linear projection of the regressor will facilitate both visualization of the nonlinearities and subsequent nonlinear modeling. The least squares fit of polynomial and piecewise affine functions are used as criterion by which numerical programs search for the linear projection that gives the best low-dimensional description of the residual. For a simulated water tank and for real life data sampled from an electronic circuit, the regressor can be projected down to 2 dimensions and still yield a model simulation fit by about 99%. The electronic circuit data can be described by a model structure with far less parameters than conventional, nonlinear black-box models.

Keywords: identification

NONLINEAR DYNAMICS IDENTIFIED BY MULTI-INDEX MODELS¹

David Lindgren and Lennart Ljung

*Division of Automatic Control
Linköping University, Sweden*

Abstract: For a class of nonlinear systems, the residual of a well fitted linear model has low *intrinsic* dimensionality. For these systems, a particular low-dimensional linear projection of the regressor will facilitate both visualization of the nonlinearities and subsequent nonlinear modeling. The least squares fit of *polynomials* and *piecewise affine functions* are used as criterion by which numerical programs search for the linear projection that gives the best low-dimensional description of the residual. For a simulated water tank and for real life data sampled from an electronic circuit, the regressor can be projected down to 2 dimensions, and still yield a model simulation fit of about 99%. The electronic circuit data can be described by a model structure with far less parameters than conventional, nonlinear black-box models.

Keywords: System Identification, Multi-Index Models, Projection Pursuit

1. INTRODUCTION

The time-discrete SISO model

$$y_t = f(\varphi_t) + v_t, \quad (1)$$

with the special *multi-index* structure

$$f(\varphi_t) = \mathbf{b}^T \varphi_t + g(\mathbf{S}^T \varphi_t), \quad (2)$$

appears to have very interesting applicability in nonlinear system identification. Here, y_t denotes the system output, φ_t an n -dimensional regression vector (defined in (5)) and v_t accounts for uncertainties and discrepancies. g is thought of as the residual of the linear part $\mathbf{b}$, a *nonlinear* residual that is confined to the column space of the matrix $\mathbf{S}$ in $\mathbb{R}^{n \times k}$. Particularly, when the column space of $\mathbf{S}$ has much lower dimensionality than the regression vector, this implies the complexity of the parameterization and estimation of g will

be greatly reduced. Say φ_t exceeds 10 dimensions, which is certainly not extreme. Not only will a nonlinear structure in 10 dimensions have excessively many parameters, but also as many as 10^6 observations (sample points) will be *very sparse* in this space. The multi-index structure allows the nonlinearity to be modeled in low dimensions, where the point density is larger and where functions may be defined by few parameters.

Not all systems benefit from a low-dimensional regressor projection, however. It is observed that when the nonlinearity enters the differential equation additively, the multi-index structure is well suited. An example of such a system is the drained water tank, to which will be returned below.

Interest attaches to the estimation of $\mathbf{S}$. From one viewpoint, finding $\mathbf{S}$ would be the first step estimating f , estimating g the second. However, explicit parameterizations of g are used here, and programs that estimate $\mathbf{S}$ and g simultaneously by the least squares criterion will be proposed.

¹ This work was supported by the *Swedish Research Council*. It is also part of the VISIMOD project supported by the *Swedish Foundation for Strategic Research*.

Apart from efficient parameterization, knowing $\mathbf{S}$ may also allow visualization of the nonlinear behavior of a data set. If $k = 1$, the curve g can be viewed by plotting the residual points

$$y_t - \mathbf{b}^T \boldsymbol{\varphi}_t \quad (3)$$

against the regressor projection $\mathbf{S}^T \boldsymbol{\varphi}_t$. Special visualization techniques used in this respect are described in (Johansson *et al.*, 2004).

1.1 Example: Drained Water Tank

Theoretically, a drained water tank obeys the nonlinear differential equation

$$\frac{d}{dt}y(t) = -\sqrt{y(t)} + u(t). \quad (4)$$

Sampling the tank gives a data set $\{y_t, u_t\}_1^N$, the same data is studied in (Lindgren and Ljung, 2004). The regression vector takes the general form

$$\boldsymbol{\varphi}_t = \begin{bmatrix} y_{t-1} \\ \vdots \\ y_{t-n_a} \\ u_{t-d} \\ \vdots \\ u_{t-n_b-d+1} \end{bmatrix} \in \mathbb{R}^n. \quad (5)$$

Select the model orders $n_a = 3$, $n_b = 3$, $d = 1$, $k = 1$ yielding regression vector dimensionality

$$n = n_a + n_b = 6,$$

and consider it the user's first attempt. This means that $\mathbf{S}$ is an n -dimensional vector. Study two different choices of regressor projections in Fig. 1. In the left plot, $\mathbf{S}_1$ and $\mathbf{b}$ are from a multi-index model of the tank. The projection $\mathbf{S}_1^T \boldsymbol{\varphi}_t$ is plotted versus the linear model residual $y_t - \mathbf{b}^T \boldsymbol{\varphi}_t$. In the right plot, the first latent variable $\mathbf{S}_2$ of *partial least squares* is plotted versus y_t . Partial least squares finds a projection that maximizes the sample covariance with y_t and is a well-established linear technique to calculate regressor projections. The maximized covariance means that the area of the covariance ellipsoid in $[y_t \ \mathbf{S}_2^T \boldsymbol{\varphi}_t]$ is minimal. The projections in Fig. 1 evaluates as

$$\begin{aligned} \mathbf{S}_1 &= [0.98 \ -0.2 \ 0.025 \ 0.052 \ -0.013 \ 0.001]^T, \\ \mathbf{S}_2 &= [0.22 \ 0.22 \ 0.21 \ 0.53 \ 0.54 \ 0.55]^T. \end{aligned} \quad (6)$$

It is interesting to observe that it actually exists a 1-dimensional linear projection that totally isolates the nonlinearity. Evidently, it is not trivial to find it, however. The benefit of knowing $\mathbf{S}$ in the subsequent modeling is clear. Not only can g be parameterized in low dimensions. Visualization like the plot in Fig. 1 (left) may give the user a clue which nonlinear function family to be used.

Fig. 1. Regressor projections of tank data. To the left $\mathbf{S}_1$ from a multi-index model, to the right $\mathbf{S}_2$ from partial least squares (PLS), see (6)

1.2 Model Structure, Motivation

The model is divided into the linear part $\mathbf{b}^T \boldsymbol{\varphi}_t$ and the nonlinear part $g(\mathbf{S}^T \boldsymbol{\varphi}_t)$. We have concluded that if the linear part is not separated in this way, the nonlinear dependence between the regressors are more or less obscured by strong (linear) correlation between consecutive samples – for instance due to high sampling rate. It is then almost impossible to *see* any nonlinearity in the regressor projection. The same phenomenon is seen fitting a linear model to data sampled from a nonlinear system. If the sampling rate is high, the one step ahead prediction error may appear low, while a subsequent simulation immediately reveals the inappropriateness of the model.

This explains why the multi-index model (2) is so interesting: it refines the nonlinear dependence by filtering out the linear correlation, and then again refines the nonlinear dependence by concentrating it to a linear subspace.

1.3 Related Work

Finding regressor projections for system identification by using a *nonparametric* measure of nonlinear dependence founded on Delaunay triangulation was proposed in (Lindgren and Ljung, 2004). In (Carrol *et al.*, 1997) the estimation of (2) with the constraint $\mathbf{S}^T \mathbf{S} = \mathbf{I}$ is discussed in the framework of generalized linear models with quasi-likelihood criteria. A comprehensive text on nonlinear modeling is (Hastie *et al.*, 2001).

2. PARAMETERIZATION

There are two issues involved in parameterizing the nonlinear part $g(\mathbf{S}^T \boldsymbol{\varphi}_t)$: the linear projection $\mathbf{S}$ from $\mathbb{R}^n$ to $\mathbb{R}^k$ and the nonlinear function g from $\mathbb{R}^k$ to $\mathbb{R}$. In this paper we only consider the special case $k = 1$. Future work will generalize the methods to higher dimensions. Below are described the parameterizations mainly considered to this date: linear projection parameterized by rotations, nonlinearity by polynomial or piecewise affine function.

2.1 Linear Projection $\mathbf{S}$

For identifiability, $\mathbf{S}$ is required to be orthonormal, $\mathbf{S}^T \mathbf{S} = \mathbf{I}$. For the case $k = 1$, this means $\mathbf{S}$ is vector on the unit sphere. This vector is parameterized by the angles $\mathbf{p} = [p_1 \ p_2 \ \dots \ p_{n-1}]$,

$$\mathbf{S}(\mathbf{p}) = \mathbf{e}_1^T \prod_{i=1}^{n-1} \begin{bmatrix} I_{i-1} & 0 & 0 \\ 0 & \cos p_i & \sin p_i & 0 \\ 0 & -\sin p_i & \cos p_i & 0 \\ 0 & 0 & 0 & I_{n-i-1} \end{bmatrix}, \quad (7)$$

where $\mathbf{e}_1^T = [1 \ 0 \ \dots \ 0]$. Thus, $\mathbf{S}$ has $n - 1$ parameters. Generalizations to $k > 1$ are well-known in numerical linear algebra (QR factorization by Givens rotations).

2.2 Polynomial g

Here, $g : \mathbb{R} \rightarrow \mathbb{R}$ is a polynomial,

$$g^{(p)}(x; \mathbf{c}) = [1 \ x^2 \ x^3 \ \dots \ x^l] \mathbf{c}. \quad (8)$$

The figure l denotes the polynomial order and is also used below to denote the complexity of the nonlinear function. Note that (8) is an ordinary polynomial except the linear term x . The well-known drawback modeling dynamics by high-order polynomials is they usually behave uncontrolled outside the support of the estimation data. However, they are very easy both to fit to data, and to differentiate. It is also pointed out that they are here primarily a tool used to estimate the projection $\mathbf{S}$. Once $\mathbf{S}$ is at hand, a more robust function family can of course be used.

With the polynomial, the total number of model parameters is

$$\dim \mathbf{b} + \dim \mathbf{p} + \dim \mathbf{c} = 2n + l - 1. \quad (9)$$

2.3 Piecewise Affine g

Here, $g : \mathbb{R} \rightarrow \mathbb{R}$ is a piecewise affine function,

$$g^{(a)}(x; \mathbf{c}) = \sum_{i=1}^l (\alpha_i + \beta_i x) w_i(x), \quad (10)$$

where

$$\mathbf{c} = [\alpha_1 \ \alpha_2 \ \dots \ \alpha_l \ \beta_1 \ \beta_2 \ \dots \ \beta_l \ z_1 \ z_2 \ \dots \ z_l]^T$$

represents the parameter vector. The weight functions $w_i(\cdot)$ allocate every x in $\mathbb{R}$ to a local sum of affine functions. The affine functions are localized by knots (center points) z_i , one for each affine function. The weight functions are, given z_i , induced by an interpolator $\mathcal{P}$:

$$w_i(x) = \begin{cases} \mathcal{P} \left(\frac{x - z_i}{z_{i+1} - z_i} \right) & \text{if } x > z_i, \\ \mathcal{P} \left(\frac{z_i - x}{z_i - z_{i-1}} \right) & \text{else.} \end{cases} \quad (11)$$

Here, $z_0 = -\infty$ and $z_{l+1} = \infty$. The interpolator is trigonometric, and gives a smooth transition from one affine function to the next,

$$\mathcal{P} = \begin{cases} \cos^2 \frac{\pi}{2} t & \text{if } t < 1, \\ 0 & \text{else.} \end{cases} \quad (12)$$

The knots are chosen with respect to the distribution of x , more on this in Section 4.2.4

Compared to the polynomial, the piecewise affine function is much more complicated to estimate and handle. However, it gives a good balance between adaptivity and stability/variance near the boundaries of the estimation set support.

The total number of affine functions l determines the adaptivity of $g^{(a)}$ and also the model order. The total number of model parameters is

$$\dim \mathbf{b} + \dim \mathbf{p} + \dim \mathbf{c} = 2n + 3l - 1. \quad (13)$$

3. VALIDATION

The one step ahead predictor for a given model $\{\mathbf{S}, \mathbf{b}, g\}$ is

$$\hat{y}_t = \mathbf{b}^T \boldsymbol{\varphi}_t + g(\mathbf{S}^T \boldsymbol{\varphi}_t) \quad (14)$$

for $\boldsymbol{\varphi}_t$ defined in (5). Comparing this model output to the sampled output defines the model one step ahead prediction errors

$$e_t = y_t - \hat{y}_t. \quad (15)$$

The root mean square (RMS) error for a data set $\{y_t, u_t\}_t^N$ is

$$E = \sqrt{\frac{1}{N} \sum_{t=1}^N e_t^2}. \quad (16)$$

The RMS prediction error E is one way to validate a model – small error means, in a sense, that the model fits data well. Usually the dataset used for model validation is independent to that used for parameter estimation (next section).

The one step ahead predictor uses system output samples up to time $t - 1$ comprised in the regression vector $\boldsymbol{\varphi}_t$. The model may be validated in *another* sense by withholding this information, and only supply the predictor with the initial system state at time $t = 0$. The so-obtained prediction, denoted $\hat{y}_{t|0}$, corresponds to the output sequence of a *simulation*, feeding the model with the sampled input data u_t (only). $\hat{y}_{t|0}$ is defined like (14), but with y_{t-r} in (5) replaced by $y_{t-r|0}$. $e_{t|0}$ and E_0 are defined analogously to e_t and E .

The prediction error E and the simulation error E_0 indeed validate a model in different senses. Validation by E_0 is generally validation by higher standards, since the model dynamics are fully tested here. The computation of E_0 is however rather complex, since it founded on a model simulation. In contrast, E can often be calculated mainly by linear matrix operations.

4. ESTIMATION

Given data $\{y_t, u_t\}_1^N$ and a model structure $\{n_a, n_b, d, g(\cdot; \cdot)\}$, the least squares errors defined above are indeed functions of the model parameters, $E = E(\mathbf{p}, \mathbf{b}, \mathbf{c})$ and $E_0 = E_0(\mathbf{p}, \mathbf{b}, \mathbf{c})$. The least squares estimates of the model parameters are

$$\{\hat{\mathbf{p}}, \hat{\mathbf{b}}, \hat{\mathbf{c}}\} = \arg \min_{\mathbf{p}, \mathbf{b}, \mathbf{c}} E(\mathbf{p}, \mathbf{b}, \mathbf{c}), \quad (17)$$

and analogously for E_0 . A model that minimizes E is often termed *prediction error model* and one that minimizes E_0 an *output error model*.

In this section is proposed how to solve (17) for the suggested functions $g^{(p)}$ and $g^{(a)}$. The techniques are based on numerical optimization programs. It should be said immediately that these programs are not guaranteed to converge to the global minimum of E and E_0 , since no results on convexity are yet available.

In the development below, some numerical aspects like data standardization and rank deficiency are not covered.

4.1 Prediction Error Model, Polynomial (8)

First note that minimizing $E(\mathbf{p}, \mathbf{b}, \mathbf{c})$ is equivalent to minimizing

$$V(\mathbf{S}, \mathbf{b}, \mathbf{c}) = \sum_{t=1}^N [y_t - \mathbf{b}^T \boldsymbol{\varphi}_t - g(\mathbf{S}^T \boldsymbol{\varphi}_t; \mathbf{c})]^2 \quad (18)$$

with the constraint $\mathbf{S}^T \mathbf{S} = \mathbf{I}$. In this section, $g(\cdot)$ is assumed differentiable, for instance the the polynomial $g^{(p)}$ in (8).

Given some initial value $\mathbf{S}^{(0)}$, V is minimized locally by repeating until convergence ($j = 1, 2, \dots$):

- (1) Let $\mathbf{b}^{(j)}$ and $\mathbf{c}^{(j)}$ solve $\min_{\mathbf{b}, \mathbf{c}} V(\mathbf{S}^{(j-1)}, \mathbf{b}, \mathbf{c})$
- (2) Let $\tilde{\mathbf{S}}^{(j)}$ solve $\min_{\mathbf{S}} V(\mathbf{S}, \mathbf{b}^{(j)}, \mathbf{c}^{(j)})$.
- (3) Let $\mathbf{S}^{(j)} = \tilde{\mathbf{S}}^{(j)} / \|\tilde{\mathbf{S}}^{(j)}\|_2$.

4.1.1. Initial Value $\mathbf{b}^{(0)}$ is calculated as the solution to the linear least squares problem $\min_{\mathbf{b}} V(0, \mathbf{b}, 0)$, corresponding to a linear ARX model. $\mathbf{S}^{(0)}$ is calculated as the first singular vector of the residual of this ARX model.

4.1.2. Calculate $\mathbf{b}$ and $\mathbf{c}$ given $\mathbf{S}$ For notational purposes, the iteration superscripts are here dropped. Let

$$\boldsymbol{\phi} = \begin{bmatrix} \boldsymbol{\varphi}_1^T \\ \boldsymbol{\varphi}_2^T \\ \vdots \\ \boldsymbol{\varphi}_N^T \end{bmatrix}, \quad \mathbf{Y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_N \end{bmatrix}. \quad (19)$$

Then $\mathbf{b}$ and $\mathbf{c}$ are given as the solution to the linear least squares problem

$$[\boldsymbol{\phi} \ \mathbf{1} \ (\boldsymbol{\phi} \mathbf{S})^2 \ (\boldsymbol{\phi} \mathbf{S})^3 \ \dots \ (\boldsymbol{\phi} \mathbf{S})^l] \begin{bmatrix} \mathbf{b} \\ \mathbf{c} \end{bmatrix} \stackrel{\text{LS}}{=} \mathbf{Y}. \quad (20)$$

By $(\boldsymbol{\phi} \mathbf{S})^r$ is denoted the r th component-wise power.

4.1.3. Calculate $\mathbf{S}$ given $\mathbf{b}$ and $\mathbf{c}$ Consider the first two terms in the Taylor expansion at some location in the regressor space $\mathbf{S} = \mathbf{S}_i$:

$$\begin{aligned} g(\mathbf{S}^T \boldsymbol{\varphi}_t; \mathbf{c}) &\approx g(\mathbf{S}_i^T \boldsymbol{\varphi}_t; \mathbf{c}) + g'(\mathbf{S}_i^T \boldsymbol{\varphi}_t; \mathbf{c})(\mathbf{S} - \mathbf{S}_i)^T \boldsymbol{\varphi}_t \\ &= u(\mathbf{S}_i, \mathbf{c}, \boldsymbol{\varphi}_t) + \mathbf{S}^T \mathbf{v}(\mathbf{S}_i, \mathbf{c}, \boldsymbol{\varphi}_t), \end{aligned} \quad (21)$$

where the u and $\mathbf{v}$ have obvious definitions. Let

$$\mathbf{U}_i = \begin{bmatrix} u(\mathbf{S}_i, \mathbf{c}, \boldsymbol{\varphi}_1) \\ u(\mathbf{S}_i, \mathbf{c}, \boldsymbol{\varphi}_2) \\ \vdots \\ u(\mathbf{S}_i, \mathbf{c}, \boldsymbol{\varphi}_N) \end{bmatrix}, \quad \mathbf{V}_i = \begin{bmatrix} \mathbf{v}(\mathbf{S}_i, \mathbf{c}, \boldsymbol{\varphi}_1)^T \\ \mathbf{v}(\mathbf{S}_i, \mathbf{c}, \boldsymbol{\varphi}_2)^T \\ \vdots \\ \mathbf{v}(\mathbf{S}_i, \mathbf{c}, \boldsymbol{\varphi}_N)^T \end{bmatrix}.$$

Then the *quasi-Newton* update of $\mathbf{S}_i$ is given as the linear least squares solution to

$$\mathbf{V}_i \mathbf{S} \stackrel{\text{LS}}{=} \mathbf{Y} - \boldsymbol{\phi} \mathbf{b} - \mathbf{U}_i, \quad (22)$$

which thus defines $\mathbf{S}_{i+1}$. Starting at $\mathbf{S}_0 = \mathbf{S}^{(j)}$, this inner loop repeated until convergence, ($i = 0, 1, \dots$).

4.2 Prediction Error Model, Piecewise Affine (10)

Now the objective function is expressed

$$\begin{aligned} V(\mathbf{S}, \mathbf{b}, \boldsymbol{\alpha}, \boldsymbol{\beta}, \mathbf{W}) &= \sum_{t=1}^N [y_t - \mathbf{b}^T \boldsymbol{\varphi}_t - g(\mathbf{S}^T \boldsymbol{\varphi}_t; \boldsymbol{\alpha}, \boldsymbol{\beta}, \mathbf{W})]^2. \end{aligned} \quad (23)$$

Here, the parameter vector $\mathbf{c}$ is divided into

$$\boldsymbol{\alpha} = [\alpha_1 \ \alpha_2 \ \dots \ \alpha_l]^T, \quad \boldsymbol{\beta} = [\beta_1 \ \beta_2 \ \dots \ \beta_l]^T$$

and $\mathbf{W}$ in $\mathbb{R}^{N \times l}$ is a weight matrix in which the entry on row t and column i (w_{ti}) is the weight observation t has for the affine function i ,

$$g(\mathbf{S}^T \boldsymbol{\varphi}_t; \boldsymbol{\alpha}, \boldsymbol{\beta}, \mathbf{W}) = \sum_{i=1}^l (\alpha_i + \beta_i \mathbf{S}^T \boldsymbol{\varphi}_t) w_{ti}. \quad (24)$$

A complication here is that any reasonable policy for choosing the knots z_i , that in turn induce the weight functions, must take the distribution of the projected points $\boldsymbol{\phi} \mathbf{S}$ into account, that is, $\mathbf{W} = \mathbf{W}(\boldsymbol{\phi} \mathbf{S})$.

Given some initial values $\mathbf{S}^{(0)}$, $\mathbf{b}^{(0)}$, and some initial weight matrix $\mathbf{W}^{(0)}$, the minimization program repeats until convergence ($j = 1, 2, \dots$):

- (1) Let $\boldsymbol{\alpha}^{(j)}$ and $\boldsymbol{\beta}^{(j)}$ solve $\min_{\boldsymbol{\alpha}, \boldsymbol{\beta}} V(\mathbf{S}^{(j-1)}, \mathbf{b}^{(j-1)}, \boldsymbol{\alpha}, \boldsymbol{\beta}, \mathbf{W}^{(j-1)})$

(2) Let $\tilde{\mathbf{S}}^{(j)}$ and $\tilde{\mathbf{b}}^{(j)}$ solve

$$\min_{\mathbf{S}, \mathbf{b}} V(\mathbf{S}, \mathbf{b}, \boldsymbol{\alpha}^{(j)}, \boldsymbol{\beta}^{(j)}, \mathbf{W}^{(j)})$$

and $\mathbf{S}^{(j)}, \mathbf{b}^{(j)}$ be the normalized solution.

(3) Calculate $\mathbf{W}^{(j)} = \mathbf{W}(\boldsymbol{\phi} \mathbf{S}^{(j)})$.

4.2.1. Initial Values $\mathbf{b}^{(0)}$ is calculated as the solution to the linear least squares problem $\min_{\mathbf{b}} V(0, \mathbf{b}, 0, 0, 0)$, corresponding to a linear ARX model. $\mathbf{S}^{(0)}$ is calculated as the first singular vector of the residual of this ARX model. $\mathbf{W}^{(0)} = \mathbf{W}(\boldsymbol{\phi} \mathbf{b})$ (described below).

4.2.2. Calculate $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ Assume $\mathbf{S}, \mathbf{W}, \mathbf{b}$ are given. Let

$$\mathbf{U} = \begin{bmatrix} \mathbf{W}_1^T \mathbf{S}^T \boldsymbol{\varphi}_1 \\ \mathbf{W}_2^T \mathbf{S}^T \boldsymbol{\varphi}_2 \\ \vdots \\ \mathbf{W}_N^T \mathbf{S}^T \boldsymbol{\varphi}_N \end{bmatrix}, \quad (25)$$

where $\mathbf{W}_t$ is the t th row of $\mathbf{W}$. Then $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ are given as the solution to the linear least squares problem

$$[\mathbf{W} \ \mathbf{U}] \begin{bmatrix} \boldsymbol{\alpha} \\ \boldsymbol{\beta} \end{bmatrix} \stackrel{\text{LS}}{=} \mathbf{Y} - \boldsymbol{\phi} \mathbf{b}. \quad (26)$$

4.2.3. Calculate $\mathbf{S}$ and $\mathbf{b}$ $\mathbf{S}$ and $\mathbf{b}$ are the solution to the linear least squares problem

$$[\boldsymbol{\beta}^T \mathbf{W}^T \boldsymbol{\phi} \ \boldsymbol{\phi}] \begin{bmatrix} \mathbf{S} \\ \mathbf{b} \end{bmatrix} \stackrel{\text{LS}}{=} \mathbf{Y} - \mathbf{W} \boldsymbol{\alpha}. \quad (27)$$

If necessary, $\mathbf{S}$ is normalized to unit length. Then $\mathbf{b}$ is recalculated as the solution to

$$\boldsymbol{\phi} \mathbf{b} \stackrel{\text{LS}}{=} \mathbf{Y} - \mathbf{W} \boldsymbol{\alpha} - \mathbf{S}^T \boldsymbol{\phi}^T \mathbf{W} \boldsymbol{\beta}. \quad (28)$$

4.2.4. The matrix $\mathbf{W}(\mathbf{x})$ Here

$$\mathbf{x} = [x_1 \ x_2 \ \dots \ x_N]^T$$

is a set of points distributed over $\mathbb{R}$. As described earlier, the weight functions are induced by a set of knots z_i . A simple way to locate these knots is to place them on a uniform grid over the support of $\mathbf{x}$. The model parameters are somewhat better utilized, however, if relatively more knots are allocated to dense regions of $\mathbf{x}$ (regions where many x_t are located).

It is known from splines theory that it is difficult to locate knots optimally, see (de Boor, 1978). A fast *ad hoc* procedure has been used, recursively dividing dense regions into subregions in a way that avoids too skew distributions within the subregions. Then a knot is allocated to each region.

How the knots relate to the weight functions has been described earlier. Entry (t, i) in $\mathbf{W}(\mathbf{x})$ corresponds to the weight x_t has for the affine function i .

4.3 Output Error Model

The aim is to find model parameters $\mathbf{p}, \mathbf{b}, \mathbf{c}$ that minimize the simulation error,

$$\{\hat{\mathbf{p}}, \hat{\mathbf{b}}, \hat{\mathbf{c}}\} = \min_{\mathbf{p}, \mathbf{b}, \mathbf{c}} E_0(\mathbf{p}, \mathbf{b}, \mathbf{c}). \quad (29)$$

Here, the regressor projection is given by the parameterization described earlier, $\mathbf{S} = \mathbf{S}(\mathbf{p})$. The above is an unconstrained optimization problem which can be locally solved more or less efficiently by any standard numerical program for unconstrained minimization.

Since the objective function E_0 relies on a model simulation, it takes relatively long time to compute. This approach therefore needs very good start values for the model parameters, for instance given by a minimizer of the prediction error E , which is much faster to compute.

The nonlinear function is $g^{(p)}$ in (8) or $g^{(a)}$ in (10).

5. NUMERICAL EXPERIMENTS

5.1 Silver Box Data

This data set has been studied earlier in (NOLCOS, 2004). It is sampled from an electronic circuit, and in theory, it obeys the nonlinear differential equation

$$m \frac{d^2 y(t)}{dt^2} + d \frac{dy(t)}{dt} + ay(t) + by(t)^3 = u(t). \quad (30)$$

The sampling interval is 0.0016384s. $N = 86916$ samples are available for estimation and 40000 for model validation. The estimation and validation data have different distributions.

The results fitting different models to the Silver data is given in Table 1. Fig. 2 depicts a regressor projection obtained from a multi-index model fitted with polynomial of degree $l = 3$. Here, $n_a = n_b = 5$, so the regressor space is 10-dimensional, cf. (NOLCOS, 2004).

5.2 Simulated Water Tank

The drained water tank mentioned in the introductory example is simulated in SIMULINK. 10000 samples are produced for estimation and 10000 for validation.

A polynomial multi-index model does not fit well, see the regressor projection in Fig. 3(a) with the corresponding polynomial of degree 5. A piecewise affine function with $l = 18$ fits better and behaves a lot better outside the data support, Fig. 3(b). (In the figures, a random subset of 1000 points are used.)

Table 1. Silver box data. Validation of Linear models (Lin), multi-index models (M-I) with polynomial (8) and non-linear ARX models with artificial neural network (ANN) with 1 hidden layer and l nodes. E_0 is the simulation error for the prediction error model, and E_0 , OE for the output error model. The model orders are given as well as the total number of model parameters (Prms).

Model	n_a	n_b	d	l	Prms	$10^3 E_0$	$10^3 E_0, OE$
Lin	2	3	0	-	5	14	15
M-I	2	3	0	3	12	5.0	0.44
M-I	5	5	0	3	22	2.5	0.44
ANN	2	3	0	10	82	4.2	2.6
ANN	2	3	0	75	617	0.52	0.46

Fig. 2. Regressor projection versus linear model residual, silver box data (subset with $N = 2000$). The model polynomial of degree 3 is also drawn.

In Table 2, some different model structures are compared, including a linear model (Lin). E is the *estimation error*.

5.3 Results

For the silver box data the polynomial fits very well. Using the prediction error criterion E , the multi-index model with 12 parameters has comparable error with an ANN model with 82 parameters but ten times the error of an ANN model with 617 parameters. Using E_0 as criterion, there were no significant difference in error, even when comparing with the 617 parameter ANN model.

For the simulated water tank, the polynomial is not appropriate. This motivates the more complex piecewise affine function, which adapts and behaves well.

6. CONCLUSIONS

It is observed that for some systems the nonlinear dynamics can be well modeled in a low-dimensional linear projection of the regressor. These systems are with advantage modeled with a *multi-index structure* with two terms: one linear and one nonlinear.

Table 2. Simulated Water Tank. Comparison between polynomial and piecewise affine multi-index models. Both the estimation error (E , Est) and validation (simulation) error E_0 are given.

Model	n_a	n_b	d	l	Prms	$10^3 E$, Est	$10^3 E_0$
Lin	1	1	1	-	2	7.2	39
M-I, $g^{(p)}$	1	1	1	5	8	0.6	2.0
M-I, $g^{(a)}$	1	1	1	11	36	0.15	0.25
M-I, $g^{(a)}$	5	5	1	18	73	0.063	0.13

Fig. 3. Regressor projection for (a) polynomial and (b) piecewise affine multi-index models, tank data.

For real life data sampled from an electronic circuit (the silver box data), it was seen that the nonlinearity could be well modeled in 1 dimensions with a model total of 12 parameters. In a validation by simulation, this (output error) model performed equally well as an artificial neural network with one hidden layer and 75 nodes (617 parameters).

REFERENCES

- Carrol, R. J., J. Fan, I. Gijbels and M. P. Wand (1997). Generalized partially linear single-index models. *J Am Stat Assoc* **92**, 477–489.
- de Boor, Carl (1978). *A Practical Guide to Splines*. Springer-Verlag. New York.
- Hastie, T., Robert Tibshirani and Jerome Friedman (2001). *The Elements of Statistical Learning*. Springer.
- Johansson, Jimmy, David Lindgren, Matthew Cooper and Lennart Ljung (2004). Interactive visualization as a tool for analysing time-varying and non-linear dynamics. Submitted to the IFAC World Congress 2005.
- Lindgren, D. and L. Ljung (2004). Nonlinear dynamics isolated by Delaunay triangulation criteria. In: *Proceedings of the 43rd IEEE Conference on Decision and Control*, to appear.
- NOLCOS (2004). Special session on identification of nonlinear systems: The silver box study. In: *Proc. NOLCOS 2004 - IFAC Symposium on Nonlinear Control System*.

Avdelning, Institution Division, Department Division of Automatic Control Department of Electrical Engineering		Datum Date 2007-06-14
Språk Language ☐ Svenska/Swedish ☒ Engelska/English ☐ _____	Rapporttyp Report category ☐ Licentiatavhandling ☐ Examensarbete ☐ C-uppsats ☐ D-uppsats ☒ Övrig rapport ☐ _____	ISBN _____ ISRN _____ Serietitel och serienummer ISSN Title of series, numbering 1400-3902
URL för elektronisk version http://www.control.isy.liu.se		LiTH-ISY-R-2791
Titel Nonlinear dynamics identified by multi-index models Title		
Författare David Lindgren, Lennart Ljung Author		
Sammanfattning Abstract <p>For a class of nonlinear systems, the residual of a well fitted model has low intrinsic dimensionality. For these systems, a particular low-dimensional linear projection of the regressor will facilitate both visualization of the nonlinearities and subsequent nonlinear modeling. The least squares fit of polynomial and piecewise affine functions are used as criterion by which numerical programs search for the linear projection that gives the best low-dimensional description of the residual. For a simulated water tank and for real life data sampled from an electronic circuit, the regressor can be projected down to 2 dimensions and still yield a model simulation fit by about 99%. The electronic circuit data can be described by a model structure with far less parameters than conventional, nonlinear black-box models.</p>		
Nyckelord Keywords identification		