2017 25th Mediterranean Conference on Control and Automation (MED)
July 3-6, 2017. Valletta, Malta

Application of low-rank approximation for nonlinear system identification

Ivan Markovsky

Abstract— **The paper considers the class of discrete-time, single-input, single-output, nonlinear dynamical systems described by a polynomial difference equation. This class, call polynomial time-invariant, is a proper generalization of the linear time-invariant model class. The identification data is assumed to be generated in the errors-in-variables setting, where the input and the output noise is zero mean, white, and the noise variances is known up to a scaling factor. The identification problem has two sub-problems**

1) **structure selection: find the monomials appearing in the difference equation representation of the system, and**
2) **parameter estimation: estimate the coefficients of the equation.**

The main result shows that the parameter estimation by minimization of the 2-norm of the equation error leads to unstructured low-rank approximation of an extended data matrix. The resulting method is computationally robust and efficient due to the use of the singular value decomposition. However, it requires knowledge of the model structure and even when the correct model structure is used, it leads to biased results. For the structure selection, the use 1-norm regularization is proposed. For the bias removal an adjustment of the ordinary least squares estimator is proposed. The resulting adjusted low-rank approximation methods defines an unbiased estimator for the model parameters of the polynomial time-invariant model.

I. Introduction

Nonlinear system identification is currently an active research topic in signal processing and control. The aim is to increase the model accuracy for applications where linear models are not adequate. Although the linear time-invariant assumption is dropped, alternative assumptions about the model are still needed. These assumptions are commonly phrased as different types of nonlinear model classes, which strike different points on the accuracy vs complexity trade-off curve. Four main classes of nonlinear models, listed in decreasing level of generality, are:

- Volterra series [1],
- Nonlinear state space [2],
- Nonlinear Auto Regressive Exogenous (NARX) [3],
- Block-oriented [4].

The Volterra series represent the output of the system as a sum of convolutions of what are called the system kernels' with the input. This representation is an universal approximation of a nonlinear system's dynamics as the number of terms in the sum grows [1]. The challenge in system identification with this class of systems is the explosion of the system parameters (the kernel functions). A possible solution is to use regularization and therefore prior knowledge about the kernels.

The NARX model class represents the system through a nonlinear difference equation. The difference equation can be specified by expansion in a basis, neural network, *etc*. Using expansion in a basis and truncating the infinite series to a finite one, turns the model identification problem into a parameter estimation problem—find the expansion coefficients from the data. As with the Volterra series, the problem of estimating the parameters is ill-posed, so that prior knowledge about the model is needed.

The block-oriented models represent the system as a connection of linear time-invariant dynamic and nonlinear static subsystems. Depending on the topology of the connection network and the types of blocks being used, there are different types of block-oriented models (Hammerstein, Wiener, Wiener-Hammerstein, ...). The problem of structure selection for a block-oriented model is to discover from data the topology of the network and the type of models in the nodes. Even with given structure, however, the problem of estimating the model parameters may be challenging.

In this paper we consider the NARX model class with a difference equation defined by a polynomial with constant coefficients. We call this class of models polynomial time-invariant in analogy with the classical linear time-invariant model class. The model structure selection problem is the problem of choosing the monomials that appear in the difference equation representation of the model. We address the structure selection problem by adding a sparsity inducing 1-norm regularizer in the cost function. The 1-norm regularizer imposes the prior knowledge of a small number of monomials, which is often physically meaningful. The number of monomials in the model structure is also a measure for the complexity of the model, so that the regularized cost function reflects the accuracy vs complexity trade-off. Other regularizers (used in a Bayesian setting) impose smoothness on the estimated function of coefficients which is more difficult to quantify and may be harder to justify from a physical point of view.

Once the model structure is selected, the model parameters are re-estimated using the adjusted least-squares method of [5], [6]. This method corrects for the bias of the ordinary least squares estimator and tends to produce more accurate estimates even for small sample sizes. The adjusted least squares method is derived in the errors-in-variables setting, *i.e.*, both the input and the output are observed with additive noise. We assume that the noise is zero mean white and

A full version of this paper is in preparation and will be available from `http://homepages.vub.ac.be/~imarkovs`.

Vrije Universiteit Brussel (VUB), Department ELEC, Pleinlaan 2, 1050 Brussels, Belgium; Email: `imarkovs@vub.ac.be`

978-1-5090-4532-7/17/$31.00 ©2017 IEEE

uncorrelated between the input and the output. Also we assume that the input-output noise variance ratio is know up to a scaling factor. (The scaling factor is estimated by the method.) Under these assumptions, the adjusted least-squares method is consistent.

We show numerical examples that confirm the improved performance of the adjusted least-squares method in comparison with the ordinary least squares method for small sample size estimation problem.

II. The model class: polynomial time-invariant models

Denote by $(\mathbb{R}^q)^{\mathbb{Z}}$ the set of functions (sequences) from the set of integers $\mathbb{Z}$ to the set of q-dimensional real vectors $\mathbb{R}^q$. The behavior (set of trajectories) $\mathscr{B}$ of a discrete-time dynamical system with q external variables is a subset of $(\mathbb{R}^q)^{\mathbb{Z}}$. In practice, we specify $\mathscr{B}$ by an equation, *e.g.*, a higher order difference equation

$$\mathscr{B} := \{ w \mid R\big(w, \sigma w, \ldots, \sigma^{\ell} w\big) = 0 \}, \tag{1}$$

where σ is the backwards shift operator

$$(\sigma w)(t) := w(t+1)$$

and R is a multivariable polynomial. The representation (1) of the system $\mathscr{B}$ is refered to as the *kernel representation*.

In the paper, we study a special case of (1) when $q = 2$ and

$$R(w, \sigma w, \ldots, \sigma^{\ell} w) = f(x) - y$$

with

$$w = \operatorname{vec}(u, y) := \begin{bmatrix} u \\ y \end{bmatrix}$$

and

$$x := \operatorname{vec}(w, \sigma w, \ldots, \sigma^{\ell-1} w, \sigma^{\ell} u).$$

I.e., the model class considered is defined by the nonlinear difference equation

$$\sigma^{\ell} y = f(x). \tag{2}$$

In (2), the variable u can be chosen freely and the variable y is determined by u and the initial conditions

$$w_{\text{ini}} := \big(w(-\ell+1), \ldots, w(0)\big).$$

In this sense, u is an input, y is an output, and (2) is an input/output representation of the model. The vector $x(t) \in \mathbb{R}^{n_x}$, where

$$n_x := 2\ell + 1$$

contains the variables at ℓ past samples (the state of the system) and the input at the current moment of time.

The function f is a n_x variate polynomial

$$\begin{aligned} f(x) &= \theta_1 \underbrace{x_1^{n_{11}} \cdots x_{n_x}^{n_{1 n_x}}}_{\phi_1(x)} + \cdots + \theta_{n_\theta} \underbrace{x_1^{n_{n_\theta 1}} \cdots x_{n_x}^{n_{n_\theta n_x}}}_{\phi_{n_\theta}(x)} \\ &= \begin{bmatrix} \theta_1 & \cdots & \theta_{n_\theta} \end{bmatrix} \begin{bmatrix} \phi_1(x) \\ \vdots \\ \phi_{n_\theta}(x) \end{bmatrix} = \theta^{\top} \phi(x). \end{aligned}$$

The model structure (*i.e.*, the vector of monomials ϕ) is specified by the $n_\theta \times n_x$ matrix $\mathtt{N} = \big[n_{ij}\big]$ of their degrees n_{ij}.

A particular model is specified by the model structured and the model parameter vector θ. With a given model structure, the model

$$\mathscr{B}(\theta) := \{ w = \left[\begin{smallmatrix} u \\ y \end{smallmatrix}\right] \mid \text{(2) holds } \}, \tag{3}$$

depends on the parameter vector θ only. For the input/output model (2), the function R is

$$\begin{aligned} R(w, \sigma w, \ldots, \sigma^{\ell} w) &= \begin{bmatrix} \theta^{\top} & -1 \end{bmatrix} \begin{bmatrix} \phi(x) \\ \sigma^{\ell} y \end{bmatrix} \\ &= \theta_{\text{ext}}^{\top} \phi_{\text{ext}}(x_{\text{ext}}) = R(x_{\text{ext}}) \end{aligned}$$

with

$$x_{\text{ext}} := \begin{bmatrix} x \\ \sigma^{\ell} y \end{bmatrix}$$

and the degrees matrix specifying R, $\mathtt{N}_{\text{ext}} = \left[\begin{smallmatrix} \mathtt{N} & 0 \\ 0 & 1 \end{smallmatrix}\right]$.

The special structure of (2), allows us to compute the response of the model to a given input and initial condition w_{ini} by recursive evaluation of (2) forward in time.

The model class defined by (3) is called polynomial time-invariant and is denoted by $\mathscr{P}$. Informally, we associate the complexity of a model $\mathscr{B}(\theta) \in \mathscr{P}$ as the pair of integers: lag ℓ and number of nonzero coefficients θ_i in the parameter vector θ. Similarly, the complexity of the model class, defined by ϕ, is the pair of integers: lag ℓ and number of monomials, *i.e.*, $n_\phi = \dim(\phi)$.

In what follows we will consider model structures consisting of all monomials with a bound $\mathtt{n}_{\max}$ on the degree. The corresponding model class is denoted by $\mathscr{P}_{\ell, \mathtt{n}_{\max}}$. The number of monomials in a difference equation representation of a model from $\mathscr{P}_{\ell, \mathtt{n}_{\max}}$ is equal to the number of combinations of $n_x - 1$ objects out of $n_x + \mathtt{n}_{\max} - 1$, *i.e.*,

$$n_\phi = \left(\begin{array}{c} n_x + \mathtt{n}_{\max} - 1 \\ n_x - 1 \end{array} \right) = \frac{(n_x + \mathtt{n}_{\max} - 1)!}{(n_x - 1)! \mathtt{n}_{\max}!}.$$

The complexity of the mode class $\mathscr{P}_{\ell, \mathtt{n}_{\max}}$ is n_ϕ.

III. Identification with known structure

The considered identification problem is defined as follows. Given data

$$w = \big(w(1), \ldots, w(T)\big)$$

generated in the errors-in-variables setup

$$w = \bar{w} + \tilde{w}, \qquad \text{where} \quad \bar{w} \in \mathscr{B}(\bar{\theta}) \text{ and } \tilde{w} \sim \mathrm{N}(0, \bar{s}^2 I) \tag{4}$$

find an estimate $\widehat{\theta}$ of the true parameter vector $\bar{\theta}$. Here, $\bar{w}$ is a trajectory (the true data) of $\mathscr{B}(\bar{\theta})$ (the true system) and $\tilde{w}$ is a zero mean white Gaussian measurement noise with variance $\bar{s}^2$. The true noise variance $\bar{s}^2$ is unknown. It is assumed that the true model $\mathscr{B}(\bar{\theta})$ is in the model class of bounded complexity polynomial time-invariant models $\mathscr{P}_{\ell, n_\theta}$.

A. Maximum-likelihood estimator

The maximum-likelihood estimator for the parameter $\bar{\theta}$ in the errors-in-variables model (4) is defined by the optimization problem

$$\begin{aligned} &\text{minimize} \quad \text{over } \theta \text{ and } \widehat{w} \quad \|w-\widehat{w}\| \\ &\text{subject to} \quad \widehat{w} \in \widehat{\mathscr{B}} \in \mathscr{P}_{\ell,n_\theta} \end{aligned} \tag{5}$$

As shown in [7], (5) is equivalent to a polynomially structured low-rank approximation problem

$$\begin{aligned} &\text{minimize} \quad \text{over } \widehat{w} \quad \|w-\widehat{w}\| \\ &\text{subject to} \quad \operatorname{rank}\bigl(\Phi(\widehat{w})\bigr) \leq n_\phi - 1, \end{aligned}$$

where

$$\Phi(\widehat{w}) := \begin{bmatrix} \phi_{\text{ext}}\bigl(\widehat{x}_{\text{ext}}(\ell+1)\bigr) & \cdots & \phi_{\text{ext}}\bigl(\widehat{x}_{\text{ext}}(T)\bigr) \end{bmatrix}.$$

B. Suboptimal method based on low-rank approximation

Since the model (2) is linear in the parameters, the parameter vector θ satisfies the system of linear equations

$$\underbrace{\begin{bmatrix} \theta^\top & -1 \end{bmatrix}}_{\theta_{\text{ext}}} \Phi(w) = 0. \tag{6}$$

In case of exact data $w = \bar{w}$, $\bar{\theta}_{\text{ext}} := \begin{bmatrix} \bar{\theta}^\top & -1 \end{bmatrix}$ is in the left kernel of the extended data matrix $\Phi(w)$. Moreover, provided that the left kernel of $\Phi(w)$ is one dimensional, the true system's parameter vector $\bar{\theta}$ can be computed from a nonzero vector $\widehat{\theta}_{\text{ext}}$ in the left kernel of $\Phi(w)$ by suitable scaling. The condition that $\Phi(w)$ has one dimensional left kernel, *i.e.*,

$$\operatorname{rank}\bigl(\Phi(w)\bigr) = n_x \tag{7}$$

is the nonlinear equivalent of the persistency of excitation assumption in linear system identification [8]. The normalization needed to determine the parameter vector $\bar{\theta}$ from an extended parameter vector $\widehat{\theta}_{\text{ext}}$ in the left kernel is

$$\widehat{\theta} := -\frac{\widehat{\theta}_{\text{ext}}(1:n_x)}{\widehat{\theta}_{\text{ext}}(n_x+1)}. \tag{8}$$

Here we use the Matlab notation $x(1:n)$ to extract the sub-vector of the first n elements of a vector x.

Note 1 (Link to total least squares). The normalization (8) is used in solution of total least squares problem [9].

With noisy data, a heuristic identification method is proposed in this paper by computation of an approximate left kernel, *e.g.*, by minimization of the residual error

$$\begin{aligned} &\text{minimize} \quad \text{over } \theta_{\text{ext}} \quad \|\theta_{\text{ext}}^\top \Phi(w)\| \\ &\text{subject to} \quad \|\theta_{\text{ext}}\| = 1. \end{aligned} \tag{9}$$

Problem (9) is equivalent to minimization of the Rayleigh quotient

$$\text{minimize} \quad \text{over } \theta_{\text{ext}} \quad \frac{\theta_{\text{ext}}^\top \Phi(w)\Phi^\top(w)\theta_{\text{ext}}}{\theta_{\text{ext}}^\top \theta_{\text{ext}}}. \tag{10}$$

It is well known that a global minimum of the Rayleigh quotient is given by the smallest eigenvalue of the matrix $\Phi(w)\Phi^\top(w)$ and a corresponding minimum point $\widehat{\theta}_{\text{ext}}$ is an eigenvector related to the smallest eigenvalue. Equivalently, the solution of (9) can be found by the singular value decomposition.

Lemma 2. *Let*

$$\begin{aligned} &\Phi(w) := U\Sigma V^\top \\ &= \begin{bmatrix} u_1 & \cdots & u_{n_\theta} & u_{n_\theta+1} \end{bmatrix} \operatorname{diag}(\sigma_1, \ldots, \sigma_{n_\theta}, \sigma_{n_\theta+1}) V^\top \end{aligned} \tag{11}$$

be the (reduced) singular value decomposition of the extended data matrix $\Phi(w)$, defined in (6). Then, problem (9) has a unique solution $\widehat{\theta}_{ext} = u_{n_\theta+1}$ if and only if $\sigma_{n_\theta} \neq \sigma_{n_\theta+1}$.

Proof: Substituting (11) in (10) and defining

$$z := U^\top \theta_{\text{ext}},$$

we obtain an equivalent problem

$$\text{minimize} \quad \text{over } z \quad \frac{\sum_{i=1}^{n_\theta+1} \sigma_i^2 z_i^2}{\sum_{i=1}^{n_\theta+1} z_i^2}. \tag{12}$$

A minimum of (12) is achieved at

$$z = \begin{bmatrix} 0 & \cdots & 0 & 1 \end{bmatrix}^\top.$$

It is unique if and only if $\sigma_{n_\theta} \neq \sigma_{n_\theta+1}$. The corresponding minimum point of (9) is

$$\widehat{\theta}_{\text{ext}} = Uz = u_{n_\theta+1}.$$

The method for identification of polynomial time-invariant systems, based on solution of problem (9), is performing unstructured low-rank approximation of $\Phi(w)$, [10].

C. Bias corrected low-rank approximation

The low-rank approximation method yields an inconsistent estimator in the errors-in-variables setup. A bias correction method, called adjusted least squares method, is proposed in [10, Chapter 7].

The estimate $\widehat{\theta}$ obtained by the low-rank approximation method (9) is biased in the errors-invariables setting (4), *i.e.*, $\mathbf{E}(\widehat{\theta}) \neq \overline{\theta}$. We derive a bias correction, which depends on the noise variance s^2. The noise variance, however, can also be estimated from the data. Simulation results show that the resulting bias corrected model $\mathscr{B}(\widehat{\theta}_{\text{c}})$ achieves better fit also for small sample sizes.

Define the matrices

$$\Psi := \Phi(w)\Phi^\top(w) \qquad \text{and} \qquad \overline{\Psi} := \overline{\Phi}(w)\overline{\Phi}^\top(w).$$

The low-rank approximation method computes the extended parameter estimate $\widehat{\theta}_{\text{ext}}$ as an eigenvector related to the smallest eigenvalue of Ψ. We construct an adjusted matrix Ψ_{c}, such that the expected value of Ψ_{c} equals the true value $\overline{\Psi}$

$$\mathbf{E}(\Psi_{\text{c}}) = \overline{\Psi}. \tag{13}$$

This property ensures that the adjusted estimate $\widehat{\theta}_{\text{c}}$, obtained from an eigenvector related to the smallest eigenvalue of Ψ_{c}, is a consistent estimator in the errors-in-variables setting, *i.e.*, the estimator $\widehat{\theta}_{\text{c}}$ converges to the true parameter value $\overline{\theta}$ as the sample size T goes to infinity.

The key tool to achieve bias correction is the sequence of the Hermite polynomials, defined by the recursion

$$h_0(x) = 1, \quad h_1(x) = x, \quad \text{and}$$
$$h_k(x) = xh_{k-1}(x) - (k-2)h_{k-2}(x), \quad \text{for } k = 2,3,\dots \quad (14)$$

The Hermite polynomials have the deconvolution property

$$\mathbf{E}\left(h_k(\overline{x}+\widetilde{x})\right) = \overline{x}^k, \qquad \text{where} \quad \widetilde{x} \sim \mathrm{N}(0,1). \quad (15)$$

Let d_t be the tth column of D_{ext}. We have,

$$\Psi = \sum_{t=1}^{T} \phi(d_t)\phi^\top(d_t) = \sum_{t=1}^{T} \left[\phi_i(d_t)\phi_j(d_t)\right]_{i,j=1}^{n_\theta,n_\theta}.$$

The (i,j)th element of Ψ is

$$\psi_{ij} = \sum_{t=1}^{T} d_{1t}^{\mathtt{d}_{i1}+\mathtt{d}_{j1}} \cdots d_{\mathtt{n}t}^{\mathtt{d}_{i\mathtt{n}}+\mathtt{d}_{j\mathtt{n}}} = \sum_{t=1}^{T}\prod_{k=1}^{\mathtt{n}} (\overline{d}_{kt} + \widetilde{d}_{kt})^{\mathtt{d}_{i\mathtt{n}}+\mathtt{d}_{j\mathtt{n}}}.$$

Assuming that $\widetilde{d}_{kt}$ are independent, zero mean, normally distributed and using the deconvolution property (15) of the Hermite polynomials, we have that

$$\psi_{\mathrm{c},ij} := \sum_{t=1}^{T}\prod_{k=1}^{\mathtt{n}} h_{\mathtt{d}_{ik}+\mathtt{d}_{jk}}(d_{kt})$$

has the unbiasedness property (13), *i.e.*,

$$\mathbf{E}(\psi_{\mathrm{c},ij}) = \sum_{t=1}^{T}\prod_{k=1}^{\mathtt{n}} \overline{d}_{kt}^{\mathtt{d}_{ik}+\mathtt{d}_{jk}} =: \overline{\psi}_{ij}.$$

The elements $\psi_{\mathrm{c},ij}$ of the corrected matrix are even polynomials of s of degree less than or equal to

$$\mathtt{d}_\psi = \left\lceil \frac{\mathtt{n}\mathtt{d}+1}{2} \right\rceil,$$

where $\lceil\cdot\rceil$ denotes rounding to the nearest bigger integer. It is possible to construct the $1 \times (\mathtt{d}_\psi + 1)$ vector of the coefficients of $\psi_{\mathrm{c},ij}$ as a polynomial of s^2. Note that the product of Hermite polynomials is a convolution of their coefficients [10, Chapter 6].

The corrected matrix

$$\Psi_{\mathrm{c}}(s^2) = \overline{\Psi}_{\mathrm{c}} + s^2\Psi_{\mathrm{c},1} + \cdots + s^{2\mathtt{d}_\psi}\Psi_{\mathrm{c},\mathtt{d}_\psi}$$

is then obtained by computing its elements in the lower triangular part.

The rows of the parameter $\widehat{\theta}_{\mathrm{c}}$ form a basis for the p-dimensional (approximate) null space of $\Psi_{\mathrm{c}}(s^2)$

$$\Theta\Psi_{\mathrm{c}}(\sigma^2) = 0.$$

Computing simultaneously s and θ is a *polynomial eigenvalue problem:* the noise variance estimate is the minimum eigenvalue and the parameter estimate is a corresponding eigenvector.

IV. Structure selection

The method for structure selection proposed is based on sparse approximation within a larger model class, *e.g.*, all monomials ϕ with bounded total degree.

We assume that the true model structure $\bar{\phi}$ is included in ϕ, *i.e.*, the monomials in $\bar{\phi}$ are a subset of the monomials in ϕ. The parameter vector of a less complex model, represented within a larger model class is a sparse vector. Only the coefficients corresponding to the monomials that are part of the model's representation are nonzero. In practice, many real-life systems have sparse representations. Therefore, sparsity enforcing prior has a wide range of applications.

Sparsity can be enforced by an ℓ_1-norm regularizer:

$$\text{minimize} \quad \text{over } \theta \quad \underbrace{\left\| \begin{bmatrix}\theta^\top & -1\end{bmatrix} \Phi(w) \right\|_2}_{\text{fitting error}} + \underbrace{\gamma\|\theta\|_1}_{\text{regularizer}}. \quad (16)$$

Problem (16) is LS-SVM [11] problem. It is convex and can be solved globally and efficiently by existing methods. We use the `CVX` package [12].

The regularization parameter γ is selected, so that the computed parameter vector $\widehat{\theta}$ has a specified sparsity, *i.e.*, the number of nonzero elements in $\widehat{\theta}$ is equal to a specified number n_x.

V. Identification experiments

In order to validate the methods presented in the paper, we perform Monte Carlo experiments with data simulated in the errors-in-variables setting (4).

A. Simulation setup

The true model

$$\bar{\mathscr{B}} = \{ w = \left[\begin{smallmatrix} u \\ y \end{smallmatrix}\right] \mid \sigma^2 y + a_1\sigma y + a_0 y = c\alpha y^3 + b_0 u + b_1\sigma u + b_2\sigma^2 \} \quad (17)$$

is polynomial time-invariant with

$$f(x) = \theta_1 u(t) + \theta_2 u(t+1) + \theta_3 u(t+2) + \theta_4 y(t) + \theta_5 y^3(t) + \theta_6 y(t+1). \quad (18)$$

It belongs to the class $\mathscr{P}_{2,3}$, *i.e.*, $\ell = 2$ and $\mathtt{n}_{\max} = 3$.

The degrees matrix `N` is

	$u(t)$	$y(t)$	$u(t+1)$	$y(t+1)$	$u(t+2)$
ϕ_1	1	0	0	0	0
ϕ_2	0	1	0	0	0
ϕ_3	0	0	1	0	0
ϕ_4	0	0	0	1	0
ϕ_5	0	0	0	0	1
ϕ_6	0	3	0	0	0

The true parameter vector is

$$\bar{\theta} = \begin{bmatrix} -0.5 & 0.25 & -1 & -0.25 & 0.3 & 0.1 \end{bmatrix}^\top$$

The true input signal $\bar{u}$ is zero mean white Gaussian.

B. Parameter estimation with simulated data

The relative parameter estimation error

$$e := \frac{\|\bar{\theta} - \widehat{\theta}\|}{\|\bar{\theta}\|},$$

is computed and averaged over $K = 100$ Monte Carlo experiments. It is plotted as a function of the noise level (noise standard deviation s) in Figure 1. The result shows that both the low-rank approximation and bias corrected low-rank approximation methods recover the exact model from noise free data and identify more accurate models than the best linear approximation in case of noisy data, but for noise level above 0.05, the model identified by the bias corrected low-rank approximation method is more accurate than the model identified by the low-rank approximation method. This is an empirical validation of the main result of the paper: the accuracy of the low-rank approximation method is improved by the bias correction procedure.

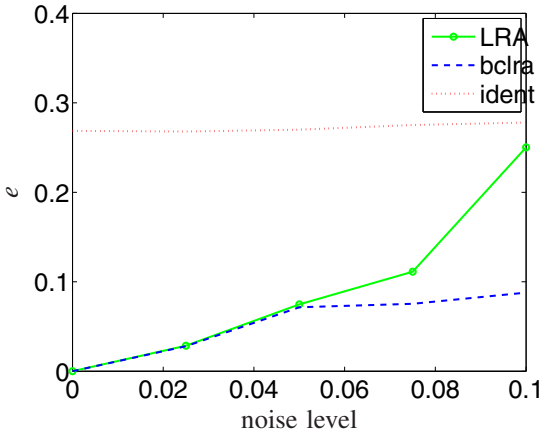

Fig. 1. The plot of the estimation error e as a function of the noise variance shows that both the low-rank approximation and bias corrected low-rank approximation methods recover the exact model from noise free data and identify more accurate models than the best linear approximation in case of noisy data. The main result of the paper is illustrated empirically but the improve accuracy of the low-rank approximation method by the bias correction: indeed for noise level above 0.05, the model identified by the bias corrected low-rank approximation method is more accurate than the model identified by the low-rank approximation method.

VI. Conclusions

The main result of the paper is establishing a link between identification of polynomial time-invariant systems and low-rank approximation. This result makes possible to use estimating methods, developed in the low-rank approximation setting, for nonlinear system identification. More specifically, we address the structure selection problem (determining the monomials that appear in a difference equation representation of the system) by ℓ_1-norm regularization and apply a bias correction procedure for the parameter estimation step. The resulting identification method is computationally cheap due to the use of convex optimization only. Comparison of the new method with existing state-of-the-art methods on nonlinear identification benchmark problems and real-data is a topic of future research.

Acknowledgements

The research leading to these results has received funding from the European Research Council under the European Union's Seventh Framework Programme (FP7/2007-2013) / ERC Grant agreement number 258581 "Structured low-rank approximation: Theory, algorithms, and applications", fund for Scientific Research (FWO-Vlaanderen), FWO projects G028015N "Decoupling multivariate polynomials in nonlinear system identification" and G090117N "Block-oriented nonlinear identification using Volterra series"; and the Belgian Network DYSCO (Dynamical Systems, Control, and Optimization), funded by the Interuniversity Attraction Poles Programme initiated by the Belgian Science Policy Office.

References

[1] S. Boyd, L. Chua, and C. Desoer, "Analytical foundations of volterra series," *IMA Journal of Mathematical Control and Information*, vol. 1, no. 3, pp. 243–282, 1984.

[2] J. Paduart, L. Lauwers, J. Swevers, K. Smolders, J. Schoukens, and R. Pintelon, "Identification of nonlinear systems using polynomial nonlinear state space models," *Automatica*, vol. 46, no. 4, pp. 647–656, 2010.

[3] S. Billings, *Nonlinear system identification: NARMAX methods in the time, frequency, and spatio-temporal domains*. John Wiley & Sons, 2013.

[4] F. Giri and E.-W. Bai, *Block-oriented nonlinear system identification*. Springer, 2010, vol. 1.

[5] A. Kukush, I. Markovsky, and S. Van Huffel, "Consistent estimation in an implicit quadratic measurement error model," *Comput. Statist. Data Anal.*, vol. 47, no. 1, pp. 123–147, 2004.

[6] I. Markovsky, A. Kukush, and S. Van Huffel, "Consistent least squares fitting of ellipsoids," *Numerische Mathematik*, vol. 98, no. 1, pp. 177–194, 2004.

[7] I. Markovsky and K. Usevich, "Nonlinearly structured low-rank approximation," in *Low-Rank and Sparse Modeling for Visual Analysis*, Y. R. Fu, Ed. Springer, 2014, pp. 1–22.

[8] J. C. Willems, P. Rapisarda, I. Markovsky, and B. De Moor, "A note on persistency of excitation," *Systems & Control Lett.*, vol. 54, no. 4, pp. 325–329, 2005.

[9] I. Markovsky and S. Van Huffel, "Overview of total least squares methods," *Signal Proc.*, vol. 87, pp. 2283–2302, 2007.

[10] I. Markovsky, *Low Rank Approximation: Algorithms, Implementation, Applications*. Springer, 2012. [Online]. Available: http://homepages.vub.ac.be/~imarkovs/book.html

[11] J. Suykens and J. Vandewalle, "Least squares support vector machine classifiers," *Neural processing letters*, vol. 9, no. 3, pp. 293–300, 1999.

[12] M. Grant and S. Boyd, "CVX: Matlab software for disciplined convex programming," 2008. [Online]. Available: stanford.edu/~boyd/cvx

[13] Khalil, *Nonlinear Systems*. Prentice Hall, 1996.

[14] M. Vidyasagar, *Nonlinear System Analysis*. Prentice Hall, 1993.

[15] J. Suykens, "Artificial neural networks for modeling and control of nonlinear systems," Ph.D. dissertation, K.U.Leuven, 1995.

[16] G. Vandersteen, "Identification of linear and nonlinear systems in an errors-in-variables least square," Ph.D. dissertation, Vrije Universiteit Brussel, 1997.