A direct data-driven method for frequency response estimation using finite data

Ivan Markovsky [a] and Hamid Ossareh [b]

[a] I. Markovsky is with the Catalan Institution for Research and Advanced Studies (ICREA), Pg. Lluis Companys 23, Barcelona, Spain and the International Centre for Numerical Methods in Engineering (CIMNE), Gran Capitàn, Barcelona, Spain.

[b] H. Ossareh is with Department of Electrical and Biomedical Engineering University of Vermont, 33 Colchester Ave, Burlington, USA.

Abstract

The existing nonparametric frequency response estimation methods suffer from leakage and have limited frequency resolution. Due to the leakage and interpolation errors these methods do not yield the correct result in case of exact data of a linear time-invariant system. Our main contribution is a nonparameteric direct data-driven frequency response estimation method that in case of exact data satisfying standard persistency of excitation condition eliminates the leakage and has infinite frequency resolution. The method is derived in the behavioral setting. It requires solving a system of linear equations and has no hyper-parameters. In case of noisy data, a modification of the method with low-rank approximation result in an effective frequency response estimator.

Key words: direct data-driven methods, frequency response estimation, behavioral approach, leakage elimination.

1 Introduction

Nonparametric frequency response estimation is a classical system identification problem, see [3, Chapter 6]. The basic solution, refered to as the *empirical transfer-function estimate* is to compute the discrete Fourier transforms of the given input/output data sequences and divide per frequency the Fourier transform of the output by the Fourier transform of the input. The resulting method is direct data-driven. It is conceptually simple, computationally efficient, thanks to the fast Fourier transform, and easy to use. Even with exact data, however, due to *leakage errors*, the empirical transfer-function estimate is not guaranteed to deliver the exact frequency response. Also, since the frequencies are evaluated on the discrete Fourier transform's grid the method has limited frequency resolution.

From a system theoretic perspective, the leakage error is the effect of the ignored initial conditions and the resulting transient response [11, Section 6.3.2]. Numerous modifications of the basic empirical transfer-function method aim to reduce the errors due to the leakage. These modifications are based on pre-processing of the data by filtering. They make extra assumptions (apart from the linear time-invariant dynamics) and involve user defined hyper-parameters. The limited frequency resolution can be overcome by interpolation. The interpolation methods being used, however, do not take into account the linear time-invariant dynamics, involve additional hyper-parameters, and introduce additional approximation error [1]. Thus, although nonparametric frequency response estimation methods have been developed for many years, the leakage and limited frequency resolution problems have not been resolved.

First, we address the leakage and limited frequency resolution problems in case of exact (noise free) data. The basic problem, refered to as *frequency response evaluation*, is: Given an exact finite input/output trajectory of a finite order deterministic linear time-invariant system and a set of frequencies, find the frequency response of the system at the given frequencies. Then, we show how the solution can be modified to yield a frequency response estimator in case of inexact (noisy) data. Our main contribution is a nonparameteric direct data-driven method that solves the frequency response evaluation problem. The method has no hyper-parameters and is provably correct under a standard persistency of excitation condition on the data. In case of noisy data in the errors-in-variables setting, we propose a modification of the method based on low-rank approximation, which has as a hyper parameter the order of the system.

The derivation of the method is done in the behavioral set-

⋆ Corresponding author I. Markovsky.

Email addresses: imarkovsky@cimne.upc.edu (Ivan Markovsky), Hamid.Ossareh@uvm.edu (Hamid Ossareh).

ting, which views a dynamical system as a set of trajectories. The key tool at the heart of the data-driven method proposed in the paper is a nonparametric representation of the finite horizon behavior of the system. The result that gives conditions for validity of the data-driven representation became known as the fundamental lemma [17]. The data-driven representation was effectively used in [10] for solving simulation and tracking control problems. The methods were originally developed for exact data but were subsequently generalized to noisy data and some classes of nonlinear time-varying systems, see the overviews [8,9]. The necessary background is given in Section 2. Section 3.1 presents to solution in case of exact data and Section 3.2 presents the modification of the method for the case of noisy data. The method is illustrates and validates in Section 4.

2 Preliminaries and problem statement

Let $(\mathbb{C}^q)^{\mathbb{N}}$ be the set of q-variate complex-valued signals with time-axis the natural numbers $\mathbb{N}$ (*i.e.*, vector sequences). In the *behavioral approach* to systems theory [15,12,16], a dynamical system is defined as a set of trajectories $\mathscr{B} \subset (\mathbb{C}^q)^{\mathbb{N}}$. The important difference from the classical approach is the distinction of the system (set of trajectories) from its representations (algebraic/difference/differential equations).

Note 1 (Complex-valued signals and systems) *In the system identification and data-driven signal processing and control literature using the behavioral approach, trajectories and behaviors have traditionally been real-valued. In this paper, we extend the methods to complex-valued signals w and correspondingly define the system $\mathscr{B}$ as a subset of $(\mathbb{C}^q)^{\mathbb{N}}$. In practice, however, the response to a real-valued input and real-valued initial condition is a real-valued signal. With some abuse of notation, we call systems with such property real-valued. The response $y = y_{real} + \mathbf{i}y_{imag}$ of a real-valued system $\mathscr{B}$ to a complex-valued input $u = u_{real} + \mathbf{i}u_{imag}$ and/or complex-valued initial condition is complex-valued, however, $w_{real} = \left[\begin{smallmatrix} u_{real} \\ y_{real} \end{smallmatrix}\right]$ and $w_{imag} = \left[\begin{smallmatrix} u_{imag} \\ y_{imag} \end{smallmatrix}\right]$ are decoupled,* i.e., *two independent real-valued trajectories are formally put together in one complex-valued trajectory. This formalism is used in Section 3, where we consider complex exponentials instead of sine and cosine trajectories.* □

We use the behavioral approach because of its relevance to the data-driven methods in signal processing and control. For example, it allows us to use the short-hand notation $w \in \mathscr{B}$ for "the signal w is a trajectory of the system $\mathscr{B}$". Since we consider finite signals we use the following notation for restricting the time axis to a finite interval: $w|_T := \big(w(1), \dots, w(T)\big)$ is the restriction of w to the interval $[1,T]$ and $\mathscr{B}|_T \subset (\mathbb{R}^q)^T$ is the restriction of $\mathscr{B}$ to the interval $[1,T]$.

In this paper, we consider linear time-invariant systems, *i.e.*, shift-invariant subspaces. The number of inputs m, lag ℓ, and order n of a linear time-invariant system $\mathscr{B}$ are invariant of the representation of the system and are therefore properties of the system $\mathscr{B}$ [15]. The restricted behavior $\mathscr{B}|_T$ for $T \geq \ell$ admits a representation

$$\mathscr{B}|_T = \text{image} \underbrace{\begin{bmatrix} w_d(1) & w_d(2) & \cdots & w_d(T_d - T + 1) \\ w_d(2) & w_d(3) & \cdots & w_d(T_d - T + 2) \\ \vdots & \vdots & & \vdots \\ w_d(T) & w_d(T+1) & \cdots & w_d(T_d) \end{bmatrix}}_{\mathscr{H}_T(w_d)}, \tag{1}$$

by a trajectory $w_d \in \mathscr{B}|_{T_d}$ that satisfies the *generalized persistency of excitation* condition [7]

$$\text{rank}\ \mathscr{H}_T(w_d) = mT + n. \tag{2}$$

The matrix $\mathscr{H}_T(w_d) \in \mathbb{R}^{qT \times (T_d - T + 1)}$ the Hankel matrix with depth T constructed from the data w_d. Note that (2) is verifiable from the data w_d and the prior knowledge of the system's number of inputs and order.

Note 2 *The data-driven representation (1) and the generalized persistency of excitation condition (2) are derived for real-valued systems and signals. The derivation, however, is also valid in the complex-valued case. If $\mathscr{B}$ is real-valued, it is sufficient to use a real-valued trajectory w_d that satisfies (2). Complex-valued trajectories are represented in (1) by complex g vector.* □

Consider a linear time-invariant system $\mathscr{B}$ with an input/output partitioning of the variables $w = \left[\begin{smallmatrix} u \\ y \end{smallmatrix}\right]$. Let $\mathscr{B}(H)$ be the transfer function representation of the controllable part of $\mathscr{B}$, corresponding to the input/output partitioning $w = \left[\begin{smallmatrix} u \\ y \end{smallmatrix}\right]$. The data-driven frequency response evaluation problem considered is defined as follows.

Problem 1 (Data-driven frequency response evaluation) *Given a finite input/output trajectory (u_d, y_d) of a linear time-invariant system $\mathscr{B}$ and a frequency $\omega \in [0, \pi)$, find the frequency response $H(e^{\mathbf{i}\omega})$ of $\mathscr{B}$ at the frequency ω.*

Trivial generalizations of the problem are to have as data multiple trajectories $\{ w_d^1, \dots, w_d^N \}$, where $w_d^i \in (\mathbb{R}^q)^{T_i}$ (this is achieved by using a mosaic Hankel matrix [5]) and to aim at evaluation of the frequency response at multiple frequencies $\Omega := \{ \omega_1, \dots, \omega_K \}$. Nontrivial generalizations are to consider noisy data in the errors-in-variables and output error setups as well as nonlinear systems. We will address this extension in Section 3.2.

3 The proposed method

Section 3.1 presents a solution of the data-driven frequency response evaluation problem (*i.e.*, assuming exact data w_d) that uses the data-driven representation (1) as the main tool.

Section 3.2 presents a modification of the solution for inexact case, *i.e.*, noisy data and/or data from a nonlinear system, based on preprocessing of the data matrix $\mathscr{H}_L(w_\mathrm{d})$ with low-rank approximation [6].

3.1 Solution of Problem 1

We consider general multivariable linear time-invariant systems. Using the data-driven representation, first, we describe the m complex exponential responses of the system to complex exponential inputs applied separately on the m input channels. Then, we show that the m vector-valued trajectories are compactly described as a matrix valued trajectory.

Let $e_i \in \mathbb{R}^m$ be the i-th unit vector (i-th column of the $m \times m$ identity matrix I_m) and $\exp_z(t) := z^t$ be the complex exponential function with $z \in \mathbb{C}$. From the behavioral point of view, the frequency response $H(e^{\mathbf{i}\omega})$ of a system $\mathscr{B}$ describes the quasi steady-state subbehavior of $\mathscr{B}$, *i.e.*, the set of the periodic trajectories of $\mathscr{B}$. It turns out that for linear time-invariant systems a basis for the quasi steady-state subbehavior is given by trajectories of the form

$$w^i = \big(e_i \exp_z,\ h_{z,i} \exp_z\big), \quad \text{for } i = 1,\dots,m, \quad \text{and} \quad z = e^{\mathbf{i}\omega},\ \omega \in [0,\pi). \tag{3}$$

The input of w_i is the complex exponential $e_i \exp_z$ and the output is a scaled version of the input $h_{z,i} \exp_z$, where the scaling factor $h_{z,i}$ is the i-th column of $H(z)$. The set of trajectories (3) can be written as the matrix-valued trajectory

$$W = \big(I_m \exp_z,\ H_z \exp_z\big), \quad \text{where } H_z = H(z) \quad \text{and} \quad z = e^{\mathbf{i}\omega},\ \omega \in [0,\pi). \tag{4}$$

Using the data-driven representation (1) for the trajectory (4) with length $T \geq \ell$ and denoting with $\mathbf{z} := [z^1\ \dots\ z^T]^\top$ the T-samples long complex exponential signal $\exp_z$, we obtain the following system of equations

$$\begin{bmatrix} \mathscr{H}_T(u_\mathrm{d}) \\ \mathscr{H}_T(y_\mathrm{d}) \end{bmatrix} G = \begin{bmatrix} \mathbf{z} \otimes I_m \\ \mathbf{z} \otimes H_z \end{bmatrix},$$

where $\mathscr{H}_T(u_\mathrm{d})$ is the Hankel matrix defined in (1), $G \in \mathbb{R}^{(T_\mathrm{d}-T+1)\times m}$, and $\otimes$ is the Kroneker product. The system can be rewritten in the following standard form

$$\begin{bmatrix} 0_{mT\times p} & \mathscr{H}_T(u_\mathrm{d}) \\ -\mathbf{z} \otimes I_p & \mathscr{H}_T(y_\mathrm{d}) \end{bmatrix} \begin{bmatrix} H_z \\ G \end{bmatrix} = \begin{bmatrix} \mathbf{z} \otimes I_m \\ 0_{pT\times m} \end{bmatrix}, \tag{5}$$

from which the parameter of interest H_z can be computed by solving for the unknowns H_z and G. The solution based on (5) allows us to evaluate the transfer function $H(z)$ at any complex number z, not only at z on the unit circle $e^{\mathbf{i}\omega}$.

The correctness of the method follows from its derivation and the correctness of the data-driven representation (1) under the assumptions of exact data w_d and (2). We state the result in the following theorem.

Theorem 2 *For exact data $w_\mathrm{d} = (u_\mathrm{d}, y_\mathrm{d}) \in \mathscr{B}|_{T_\mathrm{d}}$ satisfying (2) and for $T \geq \ell + 1$, (5) has a unique solution for H_z, such that $H_z = H(z)$, where H is the transfer function of $\mathscr{B}$ with the input/output partitioning $w = (u, y)$.*

Note 3 (No hyper-parameters) *Although for the verification of the conditions of Theorem 2 the lag ℓ and the order n of the system are needed, the method itself does not use them. Without prior knowledge of ℓ, the parameter T should be chosen as the maximum value for which $\mathscr{H}_T(w_\mathrm{d})$ has at least as many columns as rows,* i.e.,

$$T = T_{\max} = \lfloor (T_\mathrm{d}+1)/(q+1) \rfloor.$$

3.2 Modification for inexact data

One way of modifying the method presented in Section 3.1 for the case of noisy data is to first preprocess the data, aiming to approximate the exact noise-free data, and then apply the method on the preprocessed data. A popular preprocessing heuristic is unstructured low-rank approximation of the data matrix $\mathscr{H}_T(w_\mathrm{d})$ enforcing the prior knowledge that $\operatorname{rank} \mathscr{H}_T(w_\mathrm{d}) = mT + n$. If the model order is known, the rank $mT + n$ approximation can be obtained by truncation of the singular value decomposition of $\mathscr{H}_T(w_\mathrm{d})$. If the model order is not known, it can be estimated from the decay of the singular values, by visual inspection or by a range of rank estimation heuristics.

The solution method using preprocessing by low-rank approximation is summarized in Algorithm 1. An implementation in Matlab, available from

`https://imarkovs.github.io/frest`

is essentially five lines of code. Moreover, it applies to general multivariable systems and can use data from multiple trajectories $\{w_\mathrm{d}^1, \dots, w_\mathrm{d}^N\}$ as well as estimate the transfer function at multiple points in the complex plane. In the next section, the implementation `dd_frest` of Algorithm 1 is tested on simulated data and is compared with alternative direct and indirect frequency response estimation methods.

4 Numerical examples

This section illustrates and empirically validates the proposed method—Algorithm 1, implemented in the function `dd_frest` in case of noisy data (errors-in-variables setup). The benchmark reference for the evaluation is the maximum-likelihood estimator, implemented by an indirect method—the identification method of [4], implemented in the function `ident`. As an alternative direct data-driven method we

Algorithm 1 Data-driven frequency response estimation.

Input: Trajectory $(u_\mathrm{d}, y_\mathrm{d})$, complex number $z \in \mathbb{C}$, and order n.

1: Compute the singular value decomposition

$$\begin{bmatrix} \mathscr{H}_{n+1}(u_\mathrm{d}) \\ \mathscr{H}_{n+1}(y_\mathrm{d}) \end{bmatrix} = U \Sigma V^\top \quad \text{and let} \quad P := U(:, 1:mT+n).$$

2: Solve the system $\left[\begin{bmatrix} 0_{mT\times p} \\ -\mathbf{z} \otimes I_p \end{bmatrix} \; P \right] \begin{bmatrix} H_z \\ G \end{bmatrix} = \begin{bmatrix} \mathbf{z} \otimes I_m \\ 0_{pT\times m} \end{bmatrix}.$

Output: $H_z = H(z)$

use a spectral analysis nonparameteric estimator `spa` with the Welch method to calculate spectral densities [14].

The comparison of the methods is done on the benchmark of [2], which is a 4th order single-input single-output system $\mathscr{B}$ defined by the transfer function

$$H(z) = \frac{0.2826z + 0.5067z^2}{1 - 1.4183z + 1.5894z^2 - 1.3161z^3 + 0.8864z^4}.$$

The data is obtained in the errors-in-variables setting [13], *i.e.*, $w_\mathrm{d} = \overline{w}_\mathrm{d} + \widetilde{w}_\mathrm{d}$, where $\overline{w}_\mathrm{d} \in \mathscr{B}|_{T_\mathrm{d}}$ is the true value and $\widetilde{w}_\mathrm{d}$ is a zero mean white Gaussian noise with variance s^2.

Fig. 1 shows the frequency response estimates by the three methods—`dd_frest`, `ident`, and `spa`—in an experiment with $T_\mathrm{d} = 1000$ samples and noise level 10% (*i.e.*, signal-to-noise ratio 0.1). On the resolution of the figure, the amplitude estimates of `dd_frest` and `ident` are indistinguishable and are close to the true value. The `spa` method is less accurate.

Next, we show the relative percentage estimation errors

$$e_a := 100\% \frac{||\overline{H}_z| - |\widehat{H}_z||}{|\overline{H}_z|} \quad \text{and} \quad e_p := 100\% \frac{|\angle \overline{H}_z - \angle \widehat{H}_z|}{\angle \overline{H}_z},$$

where $\widehat{H}_z$ is the estimated frequency response and $\overline{H}_z$ the true frequency response, averaged over 100 Monte-Carlo repetitions of the simulation experiment. The frequency response is estimated at $\omega = \pi/4$. Fig. 2 shows the relative averaged estimation errors e_a and e_p as a function of the noise level in an experiment with $T_\mathrm{d} = 500$ data samples and Fig. 3 shows the relative averaged estimation errors e_a and e_p as a function of the number of samples T_d for noise level $s = 5\%$.

For exact data (zero noise level) the proposed method and the maximum-likelihood method give exact result (zero errors) and for increasing noise levels the increase of the errors. The fact that the proposed method given the correct result for exact data is an empirical confirmation of Theorem 2. The gap between the error of the proposed method and the errors of the maximum-likelihood method quantifies the sub-optimality of the proposed method due to the the low-rank approximation heuristic. The errors of the classical nonparameteric method are much higher.

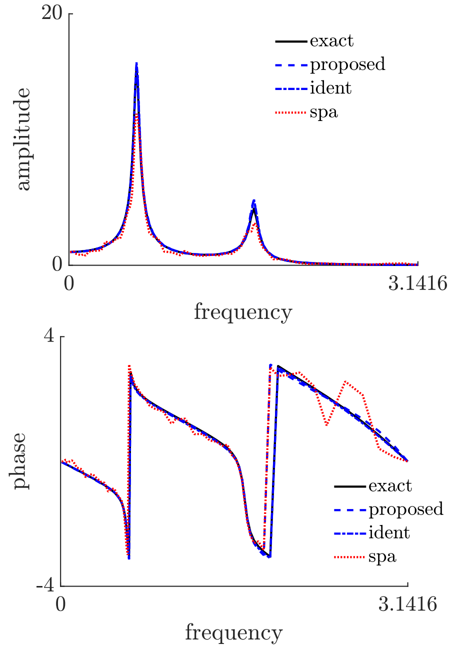

Fig. 1. Frequency response estimates in an experiment with 1000 samples and 10% noise level. The amplitude estimates of the proposed method is nearly identical to the one of the maximum-likelihood method. The classical nonparametric method is less accurate.

5 Conclusions and perspectives

We proposed a direct data-driven method for frequency response evaluation and estimation that does not suffer from leakage errors and has unlimited frequency resolution. The assumption for exact evaluation in case of noise free data is standard persistency of excitation condition on the data. The resulting algorithm has no hyper-parameters and requires solution of a system of linear equations only. Future work will focus on statistical analysis and modification of the method for the case of noise data under different noise assumptions.

6 Acknowledgements

The authors wish to thank Florian Dörfler, Petre Stoica, Torsten Söderström, Alessandro Chiuso, Maarten Schoukens, Keith Godfrey, and Koen Tiels for feedback and discussions on the results in the paper. The research leading to these results has received funding from the Catalan Institution for Research and Advanced Studies (ICREA), the Fond for Scientific Research Vlaanderen (FWO) projects G090117N and G033822N, the Fonds de la Recherche Scientifique FNRS–FWO EOS Project 30468160, and NASA under cooperative agreement VT-80NSSC20M0213.

References

[1] M. Gevers, R. Pintelon, and J. Schoukens. The local polynomial method for nonparametric system identification: Improvements and experimentation. In *IEEE Conf. on Decision and Control*, pages 4302–4307, 2011.

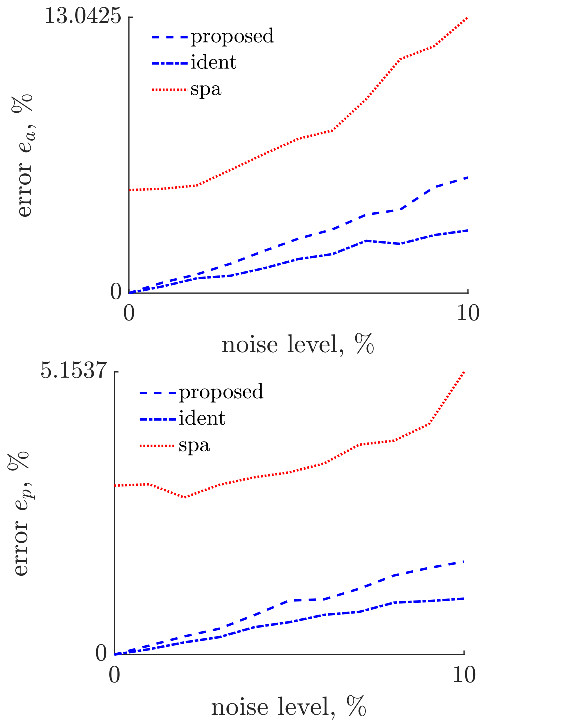

Fig. 2. The relative averaged estimation errors e_a and e_p as a function of the noise level show that for exact data (zero noise level) the proposed method and the maximum-likelihood method give exact result (zero errors) and for increasing noise levels the increase of the errors. The gap between the error of the proposed method and the errors of the maximum-likelihood method quantifies the lack of efficiency of the proposed method due to the the low-rank approximation heuristic. The errors of the classical nonparameteric method are much higher.

[2] I. Landau, D. Rey, A. Karimi, A. Voda, and A. Franco. A flexible transmission system as a benchmark for robust digital control. *European Journal of Control*, 1(2):77–96, 1995.

[3] L. Ljung. *System Identification: Theory for the User*. Prentice-Hall, 1999.

[4] I. Markovsky. A software package for system identification in the behavioral setting. *Control Eng. Practice*, 21:1422–1436, 2013.

[5] I. Markovsky. Recent progress on variable projection methods for structured low-rank approximation. *Signal Processing*, 96PB:406–419, 2014.

[6] I. Markovsky. *Low-Rank Approximation: Algorithms, Implementation, Applications*. Springer, 2nd edition edition, 2019.

[7] I. Markovsky and F. Dörfler. Identifiability in the behavioral setting. Technical report, 2019.

[8] I. Markovsky and F. Dörfler. Behavioral systems theory in data-driven analysis, signal processing, and control. *Annual Reviews in Control*, 52:42–64, 2021.

[9] I. Markovsky, L. Huang, and F. Dörfler. Data-driven control based on behavioral approach: From theory to applications in power systems. Technical report, 2022.

[10] I. Markovsky and P. Rapisarda. Data-driven simulation and control. *Int. J. Control*, 81(12):1946–1959, 2008.

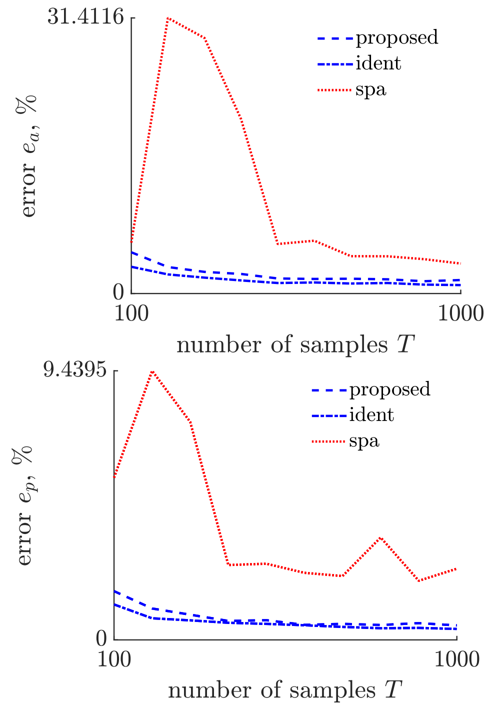

Fig. 3. Relative averaged estimation errors e_a and e_p as a function of the number of samples T_d. The errors of both the proposed method and the maximum-likelihood method show the typical $1/\sqrt{T_\text{d}}$ convergence rate. Again the gap between the error of the proposed method and the errors of the maximum-likelihood method quantifies the lack of efficiency of the proposed method due to the the low-rank approximation heuristic. The errors of the classical nonparameteric method are much higher.

[11] R. Pintelon and J. Schoukens. *System Identification: A Frequency Domain Approach*. IEEE Press, Piscataway, NJ, second edition, 2012.

[12] J. Polderman and J. C. Willems. *Introduction to Mathematical Systems Theory*. Springer-Verlag, 1998.

[13] T. Söderström. Errors-in-variables methods in system identification. *Automatica*, 43:939–958, 2007.

[14] P. Stoica and R. Moses. *Spectral Analysis of Signals*. Prentice Hall, 2005.

[15] J. C. Willems. From time series to linear system—Part I. Finite dimensional linear time invariant systems, Part II. Exact modelling, Part III. Approximate modelling. *Automatica*, 22, 23:561–580, 675–694, 87–115, 1986, 1987.

[16] J. C. Willems. On interconnections, control, and feedback. *IEEE Trans. Automat. Contr.*, 42:326–339, 1997.

[17] J. C. Willems, P. Rapisarda, I. Markovsky, and B. De Moor. A note on persistency of excitation. *Control Lett.*, 54(4):325–329, 2005.