



Non-asymptotic model quality assessment of transfer functions at multiple frequency points[☆]



Sangho Ko^{a,1}, Erik Weyer^b, Marco Claudio Campi^c

^a School of Aerospace and Mechanical Engineering, Korea Aerospace University, Goyang, Gyeonggi-do, 412-791, South Korea

^b Department of Electrical and Electronic Engineering, The University of Melbourne, Parkville, VIC 3010, Australia

^c Department of Information Engineering, University of Brescia, Via Branze 38, 25123 Brescia, Italy

ARTICLE INFO

Article history:

Received 8 June 2014

Received in revised form

28 January 2015

Accepted 17 June 2015

Keywords:

System identification

Frequency response

Confidence regions

Model quality

Finite sample properties

ABSTRACT

In this paper we develop methods for evaluating uncertainties in the frequency response of a dynamical system based on finitely many input–output data points. We extend the “Leave-out Sign-dominant Correlation Regions” (LSCR) algorithm to deliver confidence regions with a guaranteed probability for the frequency response at multiple frequencies, and we introduce a computationally efficient scheme that enables the confidence regions to be constructed frequency by frequency. Simulation examples illustrating the usefulness of the developed algorithm are provided.

© 2015 Elsevier Ltd. All rights reserved.

1. Introduction

In system identification, providing a description of the uncertainties associated with the nominal system model is as important as obtaining the nominal model itself, especially for the synthesis of robust controllers. A popular technique for evaluating the model quality is based on constructing confidence regions using asymptotic system identification theory. This is a mature approach and the confidence regions can be computed relatively easily (see [Ljung, 1999](#)). However, in some cases using asymptotic theory may lead to unreliable results (see [Garatti, Campi, & Bittanti, 2004](#)) when applied to a finite number of data points.

In this paper, we consider a method for constructing confidence regions based on finitely many data points as, e.g., considered in [Bayard \(1993\)](#), [Campi and Weyer \(2005\)](#), [den Dekker, Bombois, and Van den Hof \(2008\)](#), [Goodwin, Gevers, and Ninnes \(1992\)](#) and [Hjalmarsson and Ninness \(2006\)](#). Unlike methods based on asymptotic theory, the developed method generates guaranteed confidence regions for a finite number of data points. The developed approach is based on the LSCR method introduced in [Campi and Weyer \(2005\)](#) (see also [Campi, Ko, & Weyer, 2009](#) and [Campi & Weyer, 2010](#)), and it is extended to produce guaranteed confidence regions for the frequency response of a dynamical system. As a finite number of data points does not provide any information about the tail of the impulse response, prior information, such as exponentially decaying bounds, is introduced and incorporated in the algorithm in order to deal with tail effects. Moreover, an experimental scheme is derived that allows the confidence regions to be constructed separately frequency by frequency. This reduces the computational burden significantly.

In the next subsection we give simple preview examples that illustrate the main ideas of the proposed approach. Then, in Section 2, the procedure used in the preview examples is generalized to construct simultaneous confidence regions when the system is excited by a multi-sine input signal. In Section 3 an experimental scheme and an algorithm that allow the confidence regions to be constructed at low computational costs are introduced. Two simulation examples demonstrating the usefulness of the proposed approach are given in Section 4.

1.1. Preview examples

In this section we first introduce a simple example illustrating the main ideas of LSCR by generating a confidence interval for the

[☆] The material in this paper was partially presented at the 2007 International Conference on Control, Automation and Systems, October 17–20, 2007, Seoul, Korea and at the 17th IFAC World Congress, July 6–11, 2008, Seoul, Korea. This paper was recommended for publication in revised form by Associate Editor Alessandro Chiuso under the direction of Editor Torsten Söderström.

E-mail addresses: sanghoko@kau.ac.kr (S. Ko), ewey@unimelb.edu.au (E. Weyer), marco.campi@ing.unibs.it (M.C. Campi).

¹ Tel.: +82 2 300 0119; fax: +82 2 3158 2191.

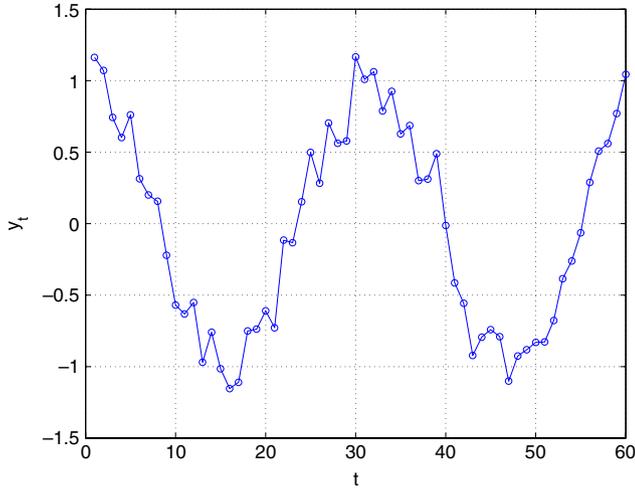


Fig. 1. Observed signal.

amplitude of a sinusoid, before moving on to the construction of a confidence set for the frequency response of a dynamical system at a given frequency. For further descriptions of the main ideas in the LSCR algorithm, the reader is referred to Campi and Weyer (2006) and Section 1.2 of Campi et al. (2009).

1.1.1. Confidence interval for the amplitude of a sinusoid

The signal of interest is a sinusoid observed in noise

$$y_t = A^0 \cos \omega t + n_t.$$

We have observations y_t , $t = 1, \dots, N = 60$. n_t is a sequence of zero mean independent random variables, symmetrically distributed about zero. The frequency $\omega = 0.2$ is known, but the amplitude A^0 is unknown. The observed signal is shown in Fig. 1. We wish to construct a confidence interval for A^0 . Given the signal model

$$\hat{y}_t(A) = A \cos \omega t,$$

we compute the observation error

$$\varepsilon_t(A) = y_t - \hat{y}_t(A) = (A^0 - A) \cos \omega t + n_t,$$

and correlate it with $\cos \omega t$, which gives

$$f_t(A) = \varepsilon_t(A) \cos \omega t = (A^0 - A) \cos^2 \omega t + n_t \cos \omega t.$$

We note that $E\{\sum_{t=1}^N f_t(A)\} = 0$ for $A = A^0$, and is different from zero for $A \neq A^0$. The idea is now to use random subsamples of $f_t(A)$ to form empirical estimates of the correlation between the observation error and $\cos \omega t$. To this end we compute $M = 20$ empirical subsample estimates

$$g_i(A) = \sum_{t=1}^N h_{i,t} f_t(A) = \sum_{t=1}^N h_{i,t} \varepsilon_t(A) \cos \omega t, \quad i = 0, 1, \dots, M - 1,$$

where $h_{i,t}$ are independent and identically distributed (i.i.d.) random variables taking on the values 0 and 1 with probability 1/2 each. The exception is $h_{0,t}$ which is equal to zero for all t , and hence $g_0(A) \equiv 0$. This means that $h_{i,t}$ determines whether sample t is used when $g_i(A)$ is computed.

The $M - 1$ non-zero $g_i(A)$ functions are shown in Fig. 2. Corresponding to the true amplitude A^0 , $g_i(A^0)$ is a sum of zero mean random variables. It is therefore unlikely that nearly all of the $g_i(A)$ functions are positive or negative for $A = A^0$, and hence we exclude those values of A where all the $g_i(A)$ functions take on positive or negative values. Thus, the confidence interval marked with a thick line in Fig. 2 is obtained by keeping those values of

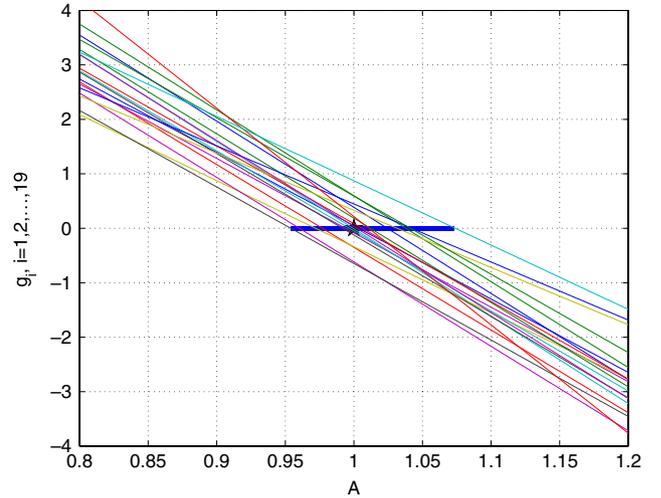


Fig. 2. $g_i(A)$ functions together with the confidence interval (thick line) and the true amplitude (\star).

A for at which at least $q = 1$ of the $g_i(A)$ functions are positive and at least $q = 1$ are negative. It is shown in Theorem 1 that the constructed confidence interval contains the true amplitude ($A^0 = 1$) with probability $1 - 2q/M = 0.9$.

Next we move onto a more realistic situation where also the phase is unknown and transient effects need to be taken into account.

1.1.2. Confidence set for frequency response

Suppose that the true continuous-time system is given by

$$y(t) = \int_0^\infty g^0(\tau) u(t - \tau) d\tau + v(t), \quad (1)$$

where $g^0(\tau)$ is the impulse response function, and $v(t)$ is additive noise. The transfer function $G^0(s)$ of the system (1) is the Laplace transform of $g^0(\tau)$ given by

$$G^0(s) = \int_0^\infty g^0(t) e^{-st} dt \quad (2)$$

and in this example it is given by

$$G^0(s) = \frac{2.5}{s + 2.5}. \quad (3)$$

This information about the true system is given for completeness of description but is unknown to the user.

The input to the system is a sinusoid

$$u(t) = \begin{cases} \cos(t), & t \geq 0 \\ 0, & t < 0. \end{cases} \quad (4)$$

The output is given by

$$\begin{aligned} y(t) &= \int_0^t g^0(\tau) \cos(t - \tau) d\tau + v(t) \\ &= \operatorname{Re} \left\{ \int_0^\infty g^0(\tau) e^{-j\tau} d\tau e^{jt} - \int_t^\infty g^0(\tau) e^{-j\tau} d\tau e^{jt} \right\} + v(t) \\ &= \operatorname{Re} \left\{ G^0(j) \cdot e^{jt} - \int_t^\infty g^0(\tau) e^{-j\tau} d\tau \cdot e^{jt} \right\} + v(t) \\ &= a^0 \cos t - b^0 \sin t + \bar{y}(t) + v(t), \end{aligned}$$

where $a^0 \triangleq \operatorname{Re}\{G^0(j \cdot 1)\}$, $b^0 \triangleq \operatorname{Im}\{G^0(j \cdot 1)\}$ and $\bar{y}(t) \triangleq -\operatorname{Re}\{\int_t^\infty g^0(\tau) e^{-j\tau} d\tau \cdot e^{jt}\}$ represents the transient effects due to initial conditions.

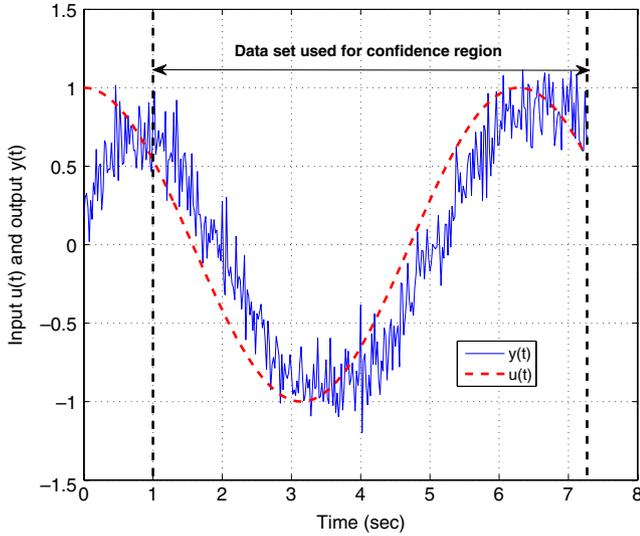


Fig. 3. Input and output data.

Our task is to construct a confidence region for a^0 and b^0 , the frequency response parameters at 1 rad/s based on a finite number of data points obtained by sampling the input and output at time instants $t = kT$ for $k = 1, 2, \dots, 365$ with $T = 0.02$ s. We assume that the sampled noise $v(kT)$, $k = 1, 2, \dots, 365$, is a sequence of independent (but not necessarily identically distributed) random variables with symmetric distributions around zero and all $v(kT)$ admit densities.

One way to estimate the frequency response parameters and to construct a confidence region is to measure the output once the transients have died out. In order to avoid the transient phase of the response we wait $\ell = 50$ samples before starting the measurements of the output $y(kT)$ (see Fig. 3). We take 315 samples of the input and output. This corresponds to approximately one cycle of the input signal.

From finite-length input and output data we cannot obtain *full* information about the frequency response of the system since the data do not carry any information about the tail of the impulse response. The only way we can bound the uncertainty due to the tail is via *a priori* knowledge and assumptions. Here we assume that a bound on the impulse response is available, i.e., parameters M_g and ρ are known such that

$$|g^0(\tau)| \leq M_g e^{-\rho\tau}, \quad \text{for some } 0 < M_g < \infty \text{ and } \rho > 0.$$

For this example, we use the following prior information

$$M_g = 3, \quad \rho = 1.7, \quad (5)$$

which is shown in Fig. 4. Using this information, we can bound the unknown value $\bar{y}(t)$ as follows

$$|\bar{y}(t)| \leq \int_t^\infty |g^0(\tau)| d\tau \leq \int_t^\infty M_g e^{-\rho\tau} d\tau = \frac{M_g e^{-\rho t}}{\rho} \triangleq \gamma(t). \quad (6)$$

We can compute the predictions of the output and the prediction error for $k = 51, \dots, 365$ as follows

$$\begin{aligned} \hat{y}_k(\theta) &= a \cdot \cos(kT) - b \cdot \sin(kT), \\ \varepsilon_k(\theta) &= y(kT) - \hat{y}_k(\theta) \\ &= \tilde{a} \cdot \cos(kT) - \tilde{b} \cdot \sin(kT) + \bar{y}(kT) + v(kT), \end{aligned}$$

where $\tilde{a} \triangleq a^0 - a$, $\tilde{b} \triangleq b^0 - b$ and $\theta = [a, b]^T$ denotes the parameter vector. Using random subsamples of the data set we calculate the following $M = 400$ scaled empirical correlation functions $C_i^a(\theta)$

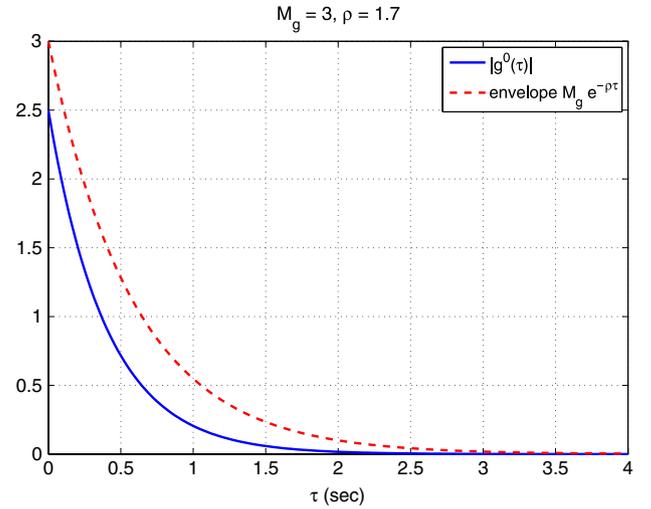


Fig. 4. Impulse response and its envelope.

and $C_i^b(\theta)$ between the prediction error and sines and cosines of the same frequency as the input signal

$$\begin{aligned} C_i^a(\theta) &= \sum_{k=51}^{365} h_{i,k} \varepsilon_k(\theta) \cos(kT) \\ &= \sum_{k=51}^{365} h_{i,k} [\tilde{a} \cdot \cos^2(kT) - \tilde{b} \cdot \sin(kT) \cos(kT) \\ &\quad + \bar{y}(kT) \cos(kT) + v(kT) \cos(kT)], \end{aligned}$$

$$\begin{aligned} C_i^b(\theta) &= \sum_{k=51}^{365} h_{i,k} \varepsilon_k(\theta) \sin(kT) \\ &= \sum_{k=51}^{365} h_{i,k} [\tilde{a} \cdot \cos(kT) \sin(kT) - \tilde{b} \cdot \sin^2(kT) \\ &\quad + \bar{y}(kT) \sin(kT) + v(kT) \sin(kT)], \end{aligned}$$

where $h_{i,k}$ for $i = 1, \dots, 399$ and $k = 51, \dots, 365$ are i.i.d. with distribution

$$h_{i,k} = \begin{cases} 0, & \text{with probability } 0.5 \\ 1, & \text{with probability } 0.5, \end{cases}$$

and are independent of the noise sequence $v(kT)$. The first string is given by $h_{0,k} = 0$ for $k = 51, \dots, 365$. Note that, using (6), we have

$$\begin{aligned} \left| \sum_{k=51}^{365} h_{i,k} \bar{y}(kT) \cos(kT) \right| &\leq \sum_{k=51}^{365} h_{i,k} \gamma(kT) |\cos(kT)| \triangleq \Gamma_i^a, \\ \left| \sum_{k=51}^{365} h_{i,k} \bar{y}(kT) \sin(kT) \right| &\leq \sum_{k=51}^{365} h_{i,k} \gamma(kT) |\sin(kT)| \triangleq \Gamma_i^b, \end{aligned}$$

and, hence, at the true parameter, $\theta = \theta^0$, for all i we obtain

$$C_i^a(\theta^0) - \Gamma_i^a \leq \sum_{k=51}^{365} h_{i,k} v(kT) \cos(kT) \leq C_i^a(\theta^0) + \Gamma_i^a, \quad (7)$$

$$C_i^b(\theta^0) - \Gamma_i^b \leq \sum_{k=51}^{365} h_{i,k} v(kT) \sin(kT) \leq C_i^b(\theta^0) + \Gamma_i^b. \quad (8)$$

Since $v(kT)$ is zero-mean, it is unlikely that nearly all of the sums $\sum_{k=51}^{365} h_{i,k} v(kT) \cos(kT)$, $i = 1, \dots, 399 = M - 1$, take on positive values or that nearly all of them take on negative values, hence, it is unlikely that nearly all $C_i^a(\theta^0) + \Gamma_i^a$ take on negative values or that nearly all $C_i^a(\theta^0) - \Gamma_i^a$ take on positive values, and the same holds

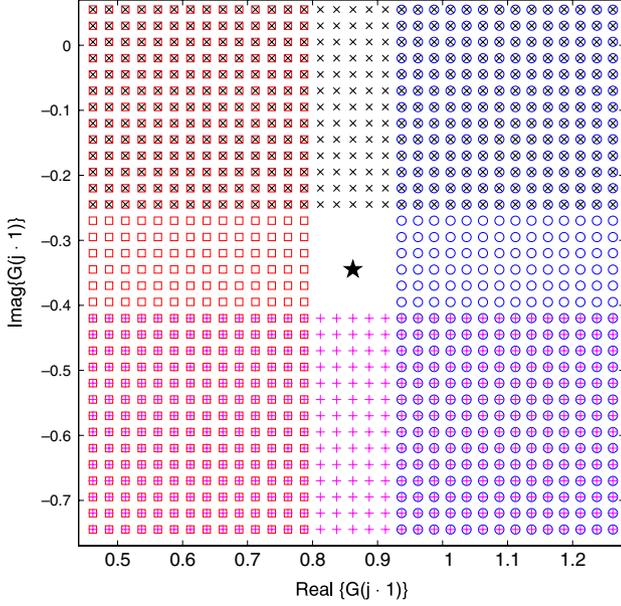


Fig. 5. Confidence region for $G^0(j \cdot 1)$ is the blank area, and \star is the true parameter.

for $C_i^b(\theta^0) + \Gamma_i^b$ and $C_i^b(\theta^0) - \Gamma_i^b$. Based on this observation, in order to construct a confidence region we discard those regions in the parameter space where $C_i^a(\theta) + \Gamma_i^a$ or $C_i^b(\theta) + \Gamma_i^b$ take on negative value too many times and also the regions where $C_i^a(\theta) - \Gamma_i^a$ or $C_i^b(\theta) - \Gamma_i^b$ take on positive value too many times, and hence the name of the algorithm “Leave-out Sign-dominant Correlation Regions”.

Therefore, in order to find a confidence set, we excluded the regions in the parameter space where less than $q = 10$ empirical correlation function satisfies either $C_i^a(\theta) - \Gamma_i^a < 0$, $C_i^b(\theta) - \Gamma_i^b < 0$, $C_i^a(\theta) + \Gamma_i^a > 0$ or $C_i^b(\theta) + \Gamma_i^b > 0$. The obtained confidence set is shown as the blank area in Fig. 5, and according to Theorem 2 it contains the true parameters with probability at least $1 - \frac{2-2q}{M} = 0.9$. In the figure, the region where at most 9 of the $C_i^a(\theta) - \Gamma_i^a$ functions were negative is marked with \square , and the region where at most 9 of the $C_i^a(\theta) + \Gamma_i^a$ were positive is marked with \circ . Likewise, \times and $+$ represents the regions where at most 9 values of $C_i^b(\theta) - \Gamma_i^b$ and $C_i^b(\theta) + \Gamma_i^b$ were negative and positive, respectively. As we can see, each correlation excludes a particular region of the parameter space. Note that in order to construct the confidence region we have not made any assumptions on the noise other than it should be symmetrically distributed around 0. Still the algorithm constructs a confidence region with guaranteed probability with a finite number of data points.

2. Main algorithm

Here we extend the approach in the preview example to a multi sine input signal.

2.1. Problem definition

Data generating system and input–output data:

Consider the following linear continuous-time system with additive noise

$$y(t) = \int_0^\infty g^0(\tau)u(t - \tau)d\tau + v(t), \quad (9)$$

where $g^0(\tau)$ is the impulse response function of the true system.

The following multi-sine input is applied to the system

$$u(t) = \begin{cases} \sum_{m=1}^L A_m \cos \varphi_m(t), & t \geq 0 \\ 0, & t < 0, \end{cases} \quad (10)$$

where

$$\varphi_m(t) \triangleq \Omega_m t + \psi_m. \quad (11)$$

Let the input and output be sampled with sampling period T , and we collect input–output data $\{u(kT), y(kT)\}$ for $k = 1, 2, \dots, N_1$.

We can express the output $y(t)$ in (9) when the input is the multi-sine input in (10) as

$$\begin{aligned} y(t) &= \sum_{m=1}^L A_m \int_0^t g^0(\tau) \cos \varphi_m(t - \tau) d\tau + v(t) \\ &= \sum_{m=1}^L A_m \left[a_m^0 \cos \varphi_m(t) - b_m^0 \sin \varphi_m(t) + \bar{y}_m(t) \right] + v(t), \end{aligned} \quad (12)$$

where

$$\begin{aligned} a_m^0 &\triangleq \operatorname{Re} \{G^0(j\Omega_m)\}, & b_m^0 &\triangleq \operatorname{Im} \{G^0(j\Omega_m)\}, \\ \bar{y}_m(t) &\triangleq -\operatorname{Re} \left\{ \int_t^\infty g^0(\tau) e^{-j\Omega_m \tau} d\tau \cdot e^{j\varphi_m(t)} \right\}. \end{aligned} \quad (13)$$

Here $\bar{y}_m(t)$ are the transients due to that $u(t) = 0$ for $t < 0$. $\bar{y}_m(t)$ is unknown, but can be bounded by

$$|\bar{y}_m(t)| \leq M_g e^{-\rho t} / \rho = \gamma(t) \quad (14)$$

using the finite-length data argument in the preview example and the assumption (A1) below. An iterative method for estimating the bounds is proposed in de Vries and Van den Hof (1995).

Assumptions:

- (A1) $|g^0(\tau)| \leq M_g e^{-\rho \tau}$, for some $0 < M_g < \infty$ and $\rho > 0$, where M_g and ρ are known *a priori*.
- (A2) The sampled noise $v(kT)$ is an independent random variable with symmetric distribution around zero, and all $v(kT)$ admit densities.

To keep the presentation simple we have imposed assumption (A2) on the sampled noise. The assumption that the noise admits densities can be dispensed with (see Campi et al., 2009), but here it has been kept as it simplifies the presentation.

Objective:

The goal is to provide guaranteed confidence regions for $\theta^0 \triangleq [a_1^0, b_1^0, \dots, a_L^0, b_L^0]^T$ using the data set $\{u(kT), y(kT)\}_{k=\ell+1, \dots, N_1}$ consisting of $N = N_1 - \ell$ input–output data measured after waiting $\ell \cdot T$ seconds to reduce the effect of the transients. Confidence regions for the magnitude and phase can subsequently be obtained using (13).

2.2. Construction of confidence regions

This section describes the procedure for constructing confidence regions for the parameter θ^0 .

Procedure for the construction of confidence regions:

- (P1) Compute the prediction and the corresponding prediction error for $k = \ell + 1, \dots, N_1$

$$\hat{y}_k(\theta) = \sum_{m=1}^L A_m [a_m \cos \varphi_m(kT) - b_m \sin \varphi_m(kT)], \quad (15)$$

$$\varepsilon_k(\theta) = y(kT) - \hat{y}_k(\theta), \quad \theta \triangleq [a_1, b_1, \dots, a_L, b_L]^T. \quad (16)$$

- (P2) Compute the correlation functions for $r = 1, \dots, L$ and $k = \ell + 1, \dots, N_1$

$$\begin{aligned} f_{r,k}^a(\boldsymbol{\theta}) &\triangleq \varepsilon_k(\boldsymbol{\theta}) \cos \varphi_r(kT), \\ f_{r,k}^b(\boldsymbol{\theta}) &\triangleq \varepsilon_k(\boldsymbol{\theta}) \sin \varphi_r(kT). \end{aligned} \quad (17)$$

- (P3) Select a positive integer M and construct M binary (0,1) stochastic strings of length $N \triangleq N_1 - \ell$ as follows: Let $h_0 = h_{0,\ell+1}, \dots, h_{0,N_1}$ be the string of all zeros. Every element of the remaining $M - 1$ strings takes the value 0 or 1 with probability 0.5 each, and the elements are independent of each other. However, if a string turns out to be equal to an already constructed string, this string is removed and another string is constructed according to the same rule to be used in its place. Name the constructed non-zero strings $h_{1,\ell+1}, \dots, h_{1,N_1}; h_{2,\ell+1}, \dots, h_{2,N_1}; \dots; h_{M-1,\ell+1}, \dots, h_{M-1,N_1}$. Each of the constructed stochastic strings determines a set of time indices to be used for calculating the empirical correlation functions in Step (P4).

- (P4) Compute the scaled empirical correlation functions for $i = 0, \dots, M - 1$

$$C_{r,i}^a(\boldsymbol{\theta}) \triangleq \sum_{k=\ell+1}^{N_1} h_{i,k} f_{r,k}^a(\boldsymbol{\theta}), \quad C_{r,i}^b(\boldsymbol{\theta}) \triangleq \sum_{k=\ell+1}^{N_1} h_{i,k} f_{r,k}^b(\boldsymbol{\theta}). \quad (18)$$

- (P5) For a fixed $r \in \{1, \dots, L\}$ select an integer q in the interval $[1, (M+1)/2]$ and find the region Θ_r^a (Θ_r^b) such that for all $\boldsymbol{\theta} \in \Theta_r^a$ ($\boldsymbol{\theta} \in \Theta_r^b$) at least q of the empirical correlation functions $C_{r,i}^a(\boldsymbol{\theta})$ ($C_{r,i}^b(\boldsymbol{\theta})$) satisfy $C_{r,i}^a(\boldsymbol{\theta}) - \Gamma_{r,i}^a < 0$ and $C_{r,i}^a(\boldsymbol{\theta}) + \Gamma_{r,i}^a > 0$ ($C_{r,i}^b(\boldsymbol{\theta}) - \Gamma_{r,i}^b < 0$ and $C_{r,i}^b(\boldsymbol{\theta}) + \Gamma_{r,i}^b > 0$) where

$$\Gamma_{r,i}^a \triangleq A \sum_{k=\ell+1}^{N_1} h_{i,k} \gamma(kT) |\cos \varphi_r(kT)|, \quad (19)$$

$$\Gamma_{r,i}^b \triangleq A \sum_{k=\ell+1}^{N_1} h_{i,k} \gamma(kT) |\sin \varphi_r(kT)|,$$

and $A \triangleq \sum_{m=1}^L A_m$. $\gamma(kT)$ is obtained using (14). \square

Note that (18) can be expressed as (20) given in Box 1.

The intuitive idea behind Step (P5) is that, for the true parameter $\boldsymbol{\theta} = \boldsymbol{\theta}^0$, the terms in the parenthesis $\{\cdot\}$ in (20) disappear, and the next term after each parenthesis can be bounded using (14) and (19)

$$\begin{aligned} \left| \sum_{k=\ell+1}^{N_1} h_{i,k} \sum_{m=1}^L A_m \bar{y}_m(kT) \cos \varphi_r(kT) \right| &\leq \Gamma_{r,i}^a \\ \left| \sum_{k=\ell+1}^{N_1} h_{i,k} \sum_{m=1}^L A_m \bar{y}_m(kT) \sin \varphi_r(kT) \right| &\leq \Gamma_{r,i}^b. \end{aligned}$$

Therefore, the empirical correlation functions evaluated at the true parameter satisfy, for $i = 0, \dots, M - 1$, the relations

$$C_{r,i}^a(\boldsymbol{\theta}^0) - \Gamma_{r,i}^a \leq \sum_{k=\ell+1}^{N_1} h_{i,k} v(kT) \cos \varphi_r(kT) \leq C_{r,i}^a(\boldsymbol{\theta}^0) + \Gamma_{r,i}^a, \quad (21)$$

$$C_{r,i}^b(\boldsymbol{\theta}^0) - \Gamma_{r,i}^b \leq \sum_{k=\ell+1}^{N_1} h_{i,k} v(kT) \sin \varphi_r(kT) \leq C_{r,i}^b(\boldsymbol{\theta}^0) + \Gamma_{r,i}^b.$$

Since $v(kT)$ is symmetrically distributed around zero, it is unlikely that nearly all of $C_{r,i}^a(\boldsymbol{\theta}^0) + \Gamma_{r,i}^a$ (or $C_{r,i}^b(\boldsymbol{\theta}^0) + \Gamma_{r,i}^b$) take on negative values or nearly all of $C_{r,i}^a(\boldsymbol{\theta}^0) - \Gamma_{r,i}^a$ (or $C_{r,i}^b(\boldsymbol{\theta}^0) - \Gamma_{r,i}^b$) take on positive values. In Step (P5) we exclude the regions in parameter space where all $C_{r,i}^a(\boldsymbol{\theta}) + \Gamma_{r,i}^a$'s (or $C_{r,i}^b(\boldsymbol{\theta}) + \Gamma_{r,i}^b$'s) are negative or all $C_{r,i}^a(\boldsymbol{\theta}) - \Gamma_{r,i}^a$'s (or $C_{r,i}^b(\boldsymbol{\theta}) - \Gamma_{r,i}^b$'s) are positive except for a small

number q . We therefore expect that $\boldsymbol{\theta}^0 \in \Theta_r^a$ ($\boldsymbol{\theta}^0 \in \Theta_r^b$) with high probability which is indeed the case as shown in the following theorem.

Theorem 1. Under assumptions (A1) and (A2), the sets Θ_r^a and Θ_r^b constructed above have the properties that

$$\Pr\{\boldsymbol{\theta}^0 \in \Theta_r^a\} \geq 1 - \frac{2q}{M}, \quad \Pr\{\boldsymbol{\theta}^0 \in \Theta_r^b\} \geq 1 - \frac{2q}{M}.$$

Proof. See Appendix A. \square

Each one of the sets Θ_r^a and Θ_r^b can be unbounded in some directions of the parameter space, and they are therefore not particularly useful. A useful confidence set can be constructed by intersecting all of the confidence regions

$$\hat{\Theta}_N = \bigcap_{r=1}^L (\Theta_r^a \cap \Theta_r^b).$$

The following theorem is immediate from Theorem 1 using the Bonferroni inequality.

Theorem 2. Under assumptions (A1) and (A2),

$$\Pr\{\boldsymbol{\theta}^0 \in \hat{\Theta}_N\} \geq 1 - 2L \frac{2q}{M}.$$

Theorem 2 shows that the constructed confidence sets contain the true frequency response with a guaranteed user chosen probability for a finite number of data points. Moreover, the confidence set shrinks around the true parameters in the sense that any $\boldsymbol{\theta} \neq \boldsymbol{\theta}^0$ will eventually be excluded from the confidence set, provided that an additional mild assumption on $v(kT)$ is satisfied.

Assumption (A3):

$$\sum_{k=1}^{\infty} \frac{E\{v^2(kT)\}}{k^2} < \infty.$$

Theorem 3 (Convergence). Under assumptions (A1), (A2) and (A3), for every fixed $\boldsymbol{\theta} \neq \boldsymbol{\theta}^0$

$$\Pr\{\exists \bar{N} \mid \boldsymbol{\theta} \notin \hat{\Theta}_N, \forall N > \bar{N}\} = 1.$$

Proof. See Appendix B. \square

Hence, for any fixed $\boldsymbol{\theta} \neq \boldsymbol{\theta}^0$ there exists a realization dependent \bar{N} such that $\boldsymbol{\theta}$ is excluded from the confidence set $\hat{\Theta}_N$ for all $N > \bar{N}$.

Remark 1 (Classical Correlation Method). The proposed method for constructing confidence regions is closely connected to the classical frequency analysis by the correlation method Ljung (1999, p.171), where estimates of the frequency response are obtained by considering the correlations between the output signal and cosines and sines of the same frequency as the input signal. Here, in order to evaluate the uncertainties in the frequency response, we use the correlations between the output prediction error and cosines and sines of the frequencies in the input signal. \square

3. Computational aspects

Using the procedure in the previous section, we can construct non-asymptotic confidence regions for the frequency response at multiple frequencies. However, each of the empirical correlation functions (20) depends on the whole set of parameters

$$\begin{aligned}
C_{r,i}^a(\boldsymbol{\theta}) &= \sum_{m=1}^L A_m \left\{ (a_m^0 - a_m) \left[\sum_{k=\ell+1}^{N_1} h_{i,k} \cos \varphi_m(kT) \cos \varphi_r(kT) \right] - (b_m^0 - b_m) \left[\sum_{k=\ell+1}^{N_1} h_{i,k} \sin \varphi_m(kT) \cos \varphi_r(kT) \right] \right\} \\
&\quad + \sum_{k=\ell+1}^{N_1} h_{i,k} \sum_{m=1}^L A_m \bar{y}_m(kT) \cos \varphi_r(kT) + \sum_{k=\ell+1}^{N_1} h_{i,k} v(kT) \cos \varphi_r(kT), \quad i = 0, \dots, M-1, \\
C_{r,i}^b(\boldsymbol{\theta}) &= \sum_{m=1}^L A_m \left\{ (a_m^0 - a_m) \left[\sum_{k=\ell+1}^{N_1} h_{i,k} \cos \varphi_m(kT) \sin \varphi_r(kT) \right] - (b_m^0 - b_m) \left[\sum_{k=\ell+1}^{N_1} h_{i,k} \sin \varphi_m(kT) \sin \varphi_r(kT) \right] \right\} \\
&\quad + \sum_{k=\ell+1}^{N_1} h_{i,k} \sum_{m=1}^L A_m \bar{y}_m(kT) \sin \varphi_r(kT) + \sum_{k=\ell+1}^{N_1} h_{i,k} v(kT) \sin \varphi_r(kT), \quad i = 0, \dots, M-1.
\end{aligned} \tag{20}$$

Box I.

$a_1, b_1, \dots, a_L, b_L$, and thus the resulting confidence regions Θ_r^a and Θ_r^b are not only dependent on a_r and b_r , but also on all other parameters.

In this section we develop an experiment procedure and a method for the generation of decoupling binary strings such that $C_{r,i}^a(\boldsymbol{\theta}) = C_{r,i}^a(a_r)$ and $C_{r,i}^b(\boldsymbol{\theta}) = C_{r,i}^b(b_r)$, thus we can construct the confidence regions for a_r^0 and b_r^0 at frequency ω_r independent of the other parameters $\{a_m, b_m\}_{m=1, \dots, L, (m \neq r)}$. Initially we assume that

Experimental assumption:

(E1) We can select the sampling period T and the total experiment time T_{exp} .

In Remark 6 we show how the experiment design can be carried out when the experiment time and the sampling period are fixed a priori.

(P0) Experiment design:

(a) Choose the set of frequencies in the multi-sine input (10) as integer multiples of a baseline frequency Ω_0

$$\Omega_m = i_m \cdot \Omega_0 \quad \text{for } i_m \in \mathbb{N}, m = 1, 2, \dots, L. \tag{22}$$

where $i_1 = 1 < i_2 < \dots < i_L = i_{\max}$.

(b) Let S_{\min} be the minimum number of desired samples per period for the highest frequency $\Omega_{i_{\max}} = i_{\max} \cdot \Omega_0$. Let $S = \lfloor S_{\min}/2 \rfloor$ where $\lfloor (\cdot) \rfloor$ indicates that (\cdot) is rounded down to the nearest integer. Choose the sampling period T according to the following procedure which guarantees that there will be between S_{\min} and $2S_{\min}$ samples per period of $\Omega_{i_{\max}}$.

Let $T_0 = 2\pi/\Omega_0$ be the period of the baseline frequency and $P = \lfloor \log_2(2 \cdot i_{\max}) \rfloor + 1$. Let the sampling period be

$$T = \frac{T_0}{S \cdot 2^P}. \tag{23}$$

(c) Choose the total number of samples such that the correlation sums are computed over n periods of the baseline frequency, i.e. let the total number of samples be $N_1 = N + \ell = n \cdot N_0 + \ell$ where $N_0 = S \cdot 2^P$ is the number of samples in one period of the baseline frequency. ℓ is the number of samples to be discarded in order to reduce the effects of the initial conditions. \square

In order to compute the confidence regions for each parameter separately, it can be seen from (20) that we need for $m, r = 1, \dots, L$ with $r \neq m$

$$\sum_{k=\ell+1}^{N_1} h_{i,k} \cos \varphi_m(kT) \cos \varphi_r(kT) = 0,$$

$$\sum_{k=\ell+1}^{N_1} h_{i,k} \sin \varphi_m(kT) \sin \varphi_r(kT) = 0,$$

and for $m, r = 1, \dots, L$

$$\sum_{k=\ell+1}^{N_1} h_{i,k} \sin \varphi_m(kT) \cos \varphi_r(kT) = 0,$$

where $\varphi_m(kT) = i_m \Omega_0 kT + \psi_m$.

Expressing each product of two trigonometric functions as a sum of two trigonometric functions, we find that the highest frequency generated from these products of trigonometric functions is $\Omega_{\max} = 2 \cdot i_{\max} \cdot \Omega_0$. For the decoupling-string generation, it suffices to find a set of time indices $\{k_j\} \subset \{\ell+1, \ell+2, \dots, N_1\}$ such that for all $i_m \in \{1, \dots, 2 \cdot i_{\max}\}$

$$\begin{aligned}
\sum_{\{k_j\}} \sin(i_m \Omega_0 T k_j) &= 0 \quad \text{and} \\
\sum_{\{k_j\}} \cos(i_m \Omega_0 T k_j) &= 0.
\end{aligned} \tag{24}$$

For this to happen, instead of Step (P3) in Section 2.2 we use the new step (P3') below for generating a set of binary strings. The idea behind (P3') is as follows: We divide each period of the baseline frequency into 2^P segments consisting of S time indices each. Since we have n periods of the baseline frequency, we get $n \cdot 2^P$ segments. For the first segment in each period, we randomly select a set of time indices (out of the S time indices), and we denote these sets as $\mathbf{K}_{1,p}$ for $p = 1, \dots, n$. We determine the binary string corresponding to $\mathbf{K}_{1,p}$ and then use this string for all the $2^P - 1$ remaining segments in the p th period. This way we obtain one binary string for the whole sample length. This procedure is repeated $M - 2$ times and a binary string of all zeros is added. The procedure is summarized below.

(P3') Algorithm for decoupling string generation:

(1) Determine n index sets $\mathbf{K}_{1,p}$ for $p = 1, \dots, n$ such that each index set $\mathbf{K}_{1,p}$ consists of the elements from

$$\{(p-1)N_0 + 1, \dots, (p-1)N_0 + S\}$$

by randomly choosing the elements such that for all $k \in \{(p-1)N_0 + 1, \dots, (p-1)N_0 + S\}$,

$$\begin{cases} k \in \mathbf{K}_{1,p}, & \text{with probability } 0.5 \\ k \notin \mathbf{K}_{1,p}, & \text{with probability } 0.5. \end{cases} \tag{25}$$

Let $\mathbf{K}_{1,p} = \{k_{1,p}, \dots, k_{q_p,p}\}$ with $q_p \leq S$ and

$$\begin{aligned}
\mathbf{K}_{j,p} &= \{k_{1,p} + (j-1)S, k_{2,p} + (j-1)S, \dots, k_{q_p,p} \\
&\quad + (j-1)S\}
\end{aligned} \tag{26}$$

for $j = 2, 3, \dots, 2^P$ and $p = 1, \dots, n$. Then, construct

$$\mathbf{J}_p = \{\mathbf{K}_{1,p}, \mathbf{K}_{2,p}, \mathbf{K}_{3,p}, \dots, \mathbf{K}_{2^P,p}\} \tag{27}$$

for $p = 1, \dots, n$. By concatenating the sets \mathbf{J}_p , we generate

$$\mathcal{J}_1 = \{\mathbf{J}_1, \mathbf{J}_2, \dots, \mathbf{J}_n\}. \quad (28)$$

This is a set of time indices which satisfies the decoupling requirement (24) (for the proof see Appendix D.).

(2) By repeating Step (1) $M - 2$ times and adding a null set $\mathcal{J}_0 = \emptyset$, we construct the set

$$\mathcal{J} = \begin{Bmatrix} \mathcal{J}_0 \\ \mathcal{J}_1 \\ \vdots \\ \mathcal{J}_{M-1} \end{Bmatrix}. \quad (29)$$

However, if an index set turns out to be equal to an already constructed set, remove this set and construct another set according to Step (1) to be used in its place. From \mathcal{J} , construct the corresponding binary (0, 1) strings $h_i = h_{i,\ell+1}, h_{i,\ell+2}, \dots, h_{i,N_1}$ of length N such that

$$\begin{cases} h_{i,\ell+k} = 1, & \text{if } k \in \mathcal{J}_i \\ h_{i,\ell+k} = 0, & \text{if } k \notin \mathcal{J}_i \end{cases} \quad (30)$$

holds for $k = 1, \dots, N$ and $i = 0, 1, \dots, M - 1$.

Remark 2 (Theorems 1 and 2). Note that using the procedure (P3') $C_{r,i}^a(\theta)$ can be written as

$$\begin{aligned} C_{r,i}^a(\theta) &= A_r(a_r^0 - a_r) \sum_{k=\ell+1}^{N_1} h_{i,k} \cos^2 \varphi_r(kT) \\ &\quad + \sum_{k=\ell+1}^{N_1} h_{i,k} \sum_{m=1}^L A_m \bar{y}_m(kT) \cos \varphi_r(kT) \\ &\quad + \sum_{k=\ell+1}^{N_1} h_{i,k} v(kT) \cos \varphi_r(kT), \end{aligned}$$

which can be rewritten as

$$\begin{aligned} C_{r,i}^a(\theta) &= A_r(a_r^0 - a_r) \sum_{p=1}^n \sum_{k=\ell+1}^{\ell+S} h_{i,(p-1)2^p+k} \\ &\quad \times \sum_{j=1}^{2^p} \cos^2 \varphi_r((k + (p-1)2^p + (j-1)S)T) \\ &\quad + \sum_{p=1}^n \sum_{k=\ell+1}^{\ell+S} h_{i,(p-1)2^p+k} \\ &\quad \times \sum_{j=1}^{2^p} A_m \bar{y}_m((k + (p-1)2^p + (j-1)S)T) \\ &\quad \times \cos \varphi_r((k + (p-1)2^p + (j-1)S)T) \\ &\quad + \sum_{p=1}^n \sum_{k=\ell+1}^{\ell+S} h_{i,(p-1)2^p+k} \\ &\quad \times \sum_{j=1}^{2^p} v((k + (p-1)2^p + (j-1)S)T) \\ &\quad \times \cos \varphi_r((k + (p-1)2^p + (j-1)S)T) \\ &= A_r(a_r^0 - a_r) \sum_{p=1}^n \sum_{k=\ell+1}^{\ell+S} h_{i,(p-1)2^p+k} \bar{A}_r(p, k) \\ &\quad + \sum_{p=1}^n \sum_{k=\ell+1}^{\ell+S} h_{i,(p-1)2^p+k} (\bar{B}_r(p, k) + \bar{C}_r(p, k)), \end{aligned}$$

where

$$\begin{aligned} \bar{A}_r(p, k) &= \sum_{j=1}^{2^p} \cos^2 \varphi_r((k + (p-1)2^p + (j-1)S)T), \\ \bar{B}_r(p, k) &= \sum_{j=1}^{2^p} A_m \bar{y}_m((k + (p-1)2^p + (j-1)S)T) \\ &\quad \times \cos \varphi_r((k + (p-1)2^p + (j-1)S)T), \\ \bar{C}_r(p, k) &= \sum_{j=1}^{2^p} v((k + (p-1)2^p + (j-1)S)T) \\ &\quad \times \cos \varphi_r((k + (p-1)2^p + (j-1)S)T). \end{aligned}$$

From the construction of the decoupling strings $h_{i,(p-1)2^p+k}$, $p = 1, \dots, n$, $k = \ell + 1, \dots, \ell + S$ are i.i.d. with replacement in case of identical strings and $\bar{C}_r(p, k)$ are independent and symmetrically distributed around zero. A similar expression can be obtained for $C_{r,i}^b(\theta)$. Using the above expression it follows from an inspection of the proofs that Theorems 1 and 2 still hold. \square

Remark 3 (Shape of The Confidence Regions and Computational Load). As shown in the previous remark each correlation function depends only on one parameter, i.e., $C_{r,i}^a(\theta) = C_{r,i}^a(a_r)$, $C_{r,i}^b(\theta) = C_{r,i}^b(b_r)$. This means that each correlation function determines the maximum and minimum values of the corresponding parameter in the confidence regions. Hence, the shape of confidence regions is rectangular, as illustrated in the simulation example in Section 4.1. This fact significantly reduces the computational load. The regions can be determined by evaluating candidate parameter values on a grid, and the number of grid points now increases linearly rather than exponentially in the number of parameters which corresponds to a linear increase in the number of frequencies. Another aspect that comes into play when assessing the computational complexity is the required resolution in the LSCR region evaluations which essentially determines the size of the grid in the θ -space. Hence for a given domain of exploration, the computational load turns out to be proportional to $M \cdot N \cdot (\text{number of parameters in } \theta) \cdot (\text{inverse of grid size})$. \square

Remark 4 (Magnitude and Phase Formulation). The procedures for construction of confidence regions in terms of the real and imaginary parts of the frequency response can be easily modified to produce confidence regions for the magnitude and phase by expressing the predictor in terms of the magnitude and phase instead of (15), as remarked in Ko, Weyer, and Campi (2007) and Ko, Weyer, and Campi (2008). However, the magnitude and phase at each frequency cannot be decoupled as above in the calculations of the empirical correlation functions. Therefore, computationally it is better to construct confidence regions for the magnitude and phase by converting the confidence regions for the real and imaginary parts using (13). \square

Remark 5 (Algorithm Implementation).

- (1) Initial parameter estimation: In Step (P1), one needs candidate values of the parameter θ to compute the prediction error. In some applications, prior knowledge can be used to select a suitable range for θ . In other cases, a parametric or non-parametric identification technique (see e.g. Ljung (1999) and Pintelon and Schoukens (2001)), can be applied to find an initial estimate, and a search for the LSCR region can be conducted by exploring the θ -domain using this estimate as a starting point.
- (2) Binary string generation: As we do not allow for string repetition, the maximum number of strings is 2^N . While this number increases rapidly with N , for small data sets it can pose some practical limits on the number of empirical correlation functions that can be constructed in Step (P3) of the procedure. \square

Remark 6 (Approximate Method for Fixed Sampling Period and Experiment Time). For a fixed *a priori* sampling period T and a total experiment time T_{exp} , we can find a set of approximate frequency points

$$\hat{\Omega} = \{\hat{\Omega}_m = \hat{i}_m \cdot \hat{\Omega}_0 \text{ for } \hat{i}_m \in \mathbb{N}, \\ m = 1, 2, \dots, L \text{ (with } \hat{i}_1 = 1)\}, \quad (31)$$

by minimizing errors between the frequencies in (31) and the desired frequencies. The resulting approximate baseline frequency $\hat{\Omega}_0$ must satisfy the constraint corresponding to (23). That is, we obtain an approximate set of frequencies by solving the following minimization problem:

$$\min_{\mathbf{x}} \sum_{m=1}^L \alpha_m \left[\Omega_m - \hat{\Omega}_m(\mathbf{x}) \right]^2 \\ \text{s.t. } \hat{\Omega}_0 = \frac{2 \cdot \pi}{\hat{S} \cdot 2^{\hat{P}} \cdot T}, \\ \hat{\Omega}_m = \hat{i}_m \hat{\Omega}_0, \quad m = 1, \dots, L, \text{ with } \hat{i}_1 = 1 \\ \hat{P} = \lfloor \log_2(2 \cdot \hat{i}_L) \rfloor + 1, \\ S_\ell \leq \hat{S} \leq S_u, \\ 2\pi / \hat{\Omega}_0 < T_{\text{exp}}, \\ 1 < \hat{i}_2 < \hat{i}_3 < \dots < \hat{i}_L < i_u, \\ \mathbf{x} = [\hat{S}, \hat{i}_2, \hat{i}_3, \dots, \hat{i}_L]^T \in \mathbb{N}^L, \quad (32)$$

where α_m is the weighting factor for the m th frequency, S_ℓ, S_u and i_u are given by the user. Once the approximate frequencies have been obtained, the procedure (P3') can be used to generate decoupling strings. This kind of problem is called Integer Program where the vector \mathbf{x} consists of integers. There are several commercial packages for solving integer nonlinear program such as [Mathworks \(2013\)](#) or [Holmström, Göran, and Edvall \(2010\)](#). \square

4. Simulation example

In this section, we present two simulation examples to illustrate the procedures for constructing confidence regions developed in the previous sections. We consider the same first-order system as described by (1) and (3) in the preview example in Section 1.1 and construct simultaneous confidence regions for a two-frequency and a ten-frequency case.

4.1. Two-frequency case

In order to construct confidence regions at $\Omega_1 = 1$ and $\Omega_2 = 2$ rad/s (in this case the baseline frequency corresponds to $\Omega_0 = \Omega_1 = 1$ rad/s), we first determine the sampling time period $T = 0.0262$ s using (23) with $i_{\text{max}} = 2$, $P = 3$, and $S = 30$. The number of the samples within one period of the baseline sinusoid is $N_0 = 240$. By choosing $n = 4$, the total number of samples to be used for the construction of the confidence regions is $n \cdot N_0 = 4 \times 240 = 960$.

By applying the following input signal to the system

$$u(t) = \begin{cases} \cos(\Omega_1 t) + \cos(\Omega_2 t), & t \geq 0 \\ 0, & t < 0 \end{cases}$$

and gathering the output measurements $\{y(kT)\}$, we construct confidence regions for the frequency responses at the two frequencies. In order to avoid the transient phase, we discard the first $\ell = 150$ data points, and then collect 960 samples of input–output data. The total number of data points is hence $N_1 = 1110$. The sampled noise $v(kT)$ is a zero-mean Gaussian white noise sequence

with variance of 0.16². This information about the noise is given for completeness of description but it is unknown to the user except for the fact that $v(kT)$ is an independent sequence with symmetric distribution around zero.

The parameter vector is $\theta^0 = [a_1^0, b_1^0, a_2^0, b_2^0]^T$ with $a_i^0 = \text{Re}\{G^0(j\Omega_i)\}$ and $b_i^0 = \text{Im}\{G^0(j\Omega_i)\}$. The parameters bounding the tail are $M_g = 3$ and $\rho = 1.7$. The prediction and prediction error are given by

$$\hat{y}_k(\theta) = \sum_{m=1}^2 [a_m \cos(\Omega_m kT) - b_m \sin(\Omega_m kT)], \\ \varepsilon_k(\theta) = y(kT) - \hat{y}_k(\theta), \text{ for } k = 151, \dots, 1110,$$

and we calculate

$$f_{r,k}^a(\theta) = \varepsilon_k(\theta) \cdot \cos(\Omega_r kT), \quad f_{r,k}^b(\theta) = \varepsilon_k(\theta) \cdot \sin(\Omega_r kT)$$

for $r = 1, 2$ and $k = 151, \dots, 1110$.

In order to construct separate confidence regions for the parameters, we generate decoupling binary strings by following the steps in (P3') of Section 3. We generate $n = 4$ index sets $\mathbf{K}_{1,p}$ for $p = 1, \dots, 4$ according to (25) with $S = 30$. Then, we generate $\mathbf{K}_{j,p}$ for $j = 1, \dots, 8$ and $p = 1, \dots, 4$ as in (26). \mathbf{J}_p is constructed as

$$\mathbf{J}_p = \{\mathbf{K}_{1,p}, \mathbf{K}_{2,p}, \dots, \mathbf{K}_{8,p}\}, \quad \text{for } p = 1, \dots, 4$$

and finally we construct

$$\mathcal{J}_1 = \{\mathbf{J}_1, \mathbf{J}_2, \mathbf{J}_3, \mathbf{J}_4\}.$$

By repeating this procedure 798 times and adding the null set \mathcal{J}_0 , we obtain the final set \mathcal{J} in (29) with $M = 800$ and the corresponding binary strings $\{h_0; h_1; \dots; h_{M-1}\}$ using (30).

[Fig. 6](#) illustrates the generation of decoupling time indices: if a time index k_0 is randomly chosen in the first segment, then 7 additional time indices are chosen in the remaining 7 segments separated by $S = 30$ samples from each other. It can be observed that these eight time indices satisfy the requirement (24) for the four frequencies $\Omega_0, 2\Omega_0, 3\Omega_0, 4\Omega_0$.

Using the generated binary strings we calculate the scaled empirical correlation functions for $r = 1, 2$ and $i = 0, \dots, 799$

$$C_{r,i}^a(\theta) = \sum_{k=151}^{1110} h_{i,k} f_{r,k}^a(\theta), \quad C_{r,i}^b(\theta) = \sum_{k=151}^{1110} h_{i,k} f_{r,k}^b(\theta).$$

The confidence region Θ_r^a is constructed by discarding those values of a_r for which at most four empirical correlation functions satisfy $C_{r,i}^a(\theta) - \Gamma_{r,i}^a < 0$ or $C_{r,i}^a(\theta) + \Gamma_{r,i}^a > 0$. The construction for Θ_r^b is similar. Then, following [Theorem 2](#) with $L = 2$ and $q = 5$, θ^0 belongs to the simultaneous region $\hat{\Theta} = \bigcap_{r=1}^2 (\Theta_r^a \cap \Theta_r^b)$ with probability at least $1 - 2 \cdot 2 \cdot 2 \cdot 5/800 = 0.95$.

These results are shown in [Figs. 7 and 8](#) where the blank areas are the confidence regions at each frequency and the true values are marked with \star . The regions where at most four $C_{r,i}^a(\theta) - \Gamma_{r,i}^a$ functions were negative are marked with \square , and the regions where at most four $C_{r,i}^a(\theta) + \Gamma_{r,i}^a$ were positive are marked with \circ . Likewise \times and $+$ represents the regions where at most four values of $C_{r,i}^b(\theta) - \Gamma_{r,i}^b$ and $C_{r,i}^b(\theta) + \Gamma_{r,i}^b$ were negative and positive, respectively. As each function only depends on one parameter, the confidence regions are rectangular, and each step in the construction excludes a particular region of the parameter space.

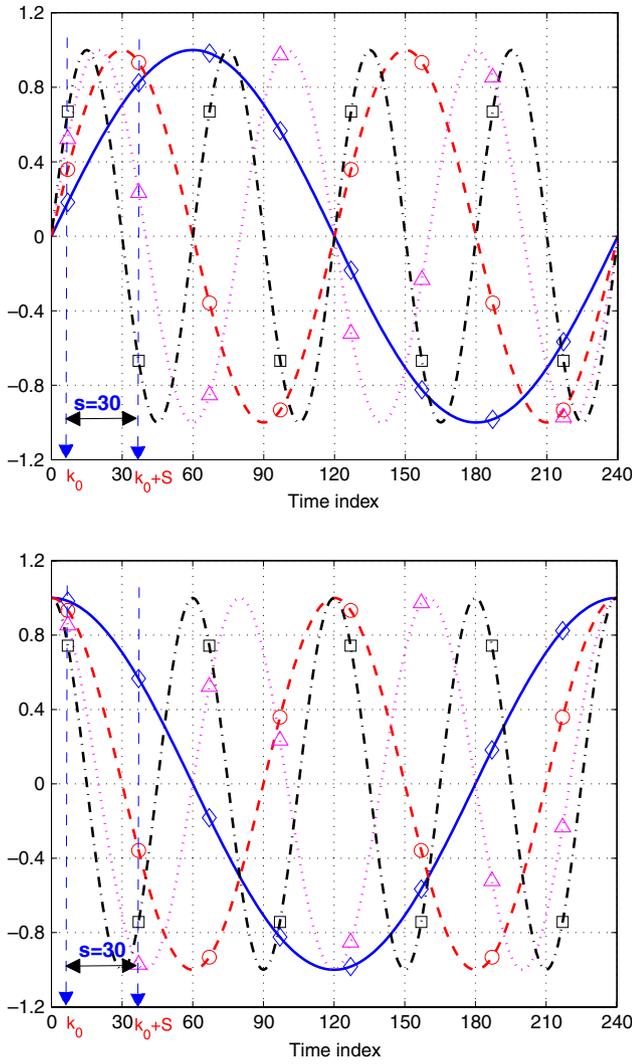


Fig. 6. Generation of a set of decoupling time indices: (Left) Solid: $\sin(\Omega_0 T)$, Dashed: $\sin(2\Omega_0 T)$, Dotted: $\sin(3\Omega_0 T)$, Dash-Dot: $\sin(4\Omega_0 T)$, (Right) Solid: $\cos(\Omega_0 T)$, Dashed: $\cos(2\Omega_0 T)$, Dotted: $\cos(3\Omega_0 T)$, Dash-Dot: $\cos(4\Omega_0 T)$.

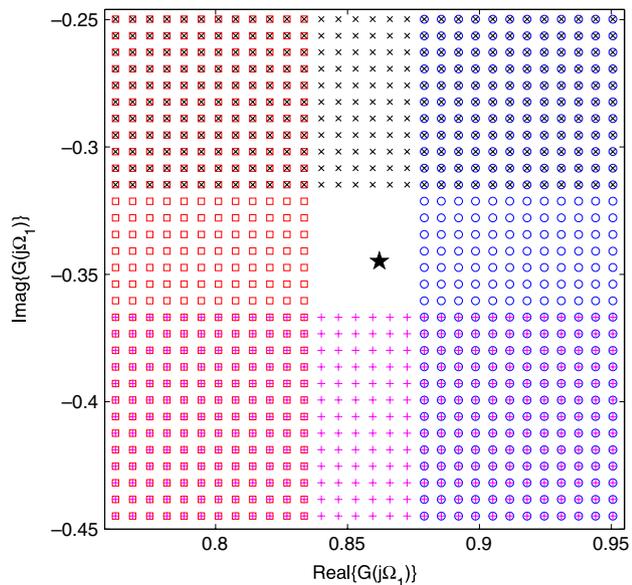


Fig. 7. Confidence region for $G^0(j\Omega_1)$.

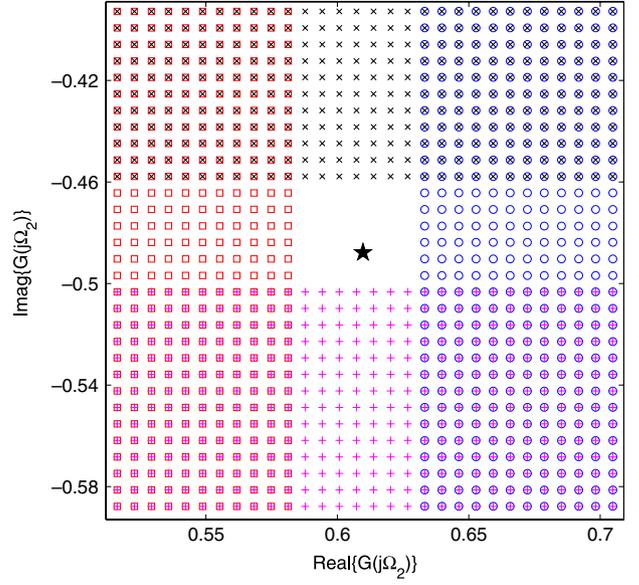


Fig. 8. Confidence region for $G^0(j\Omega_2)$.

4.2. Ten-frequency case

Consider the same system as in the previous subsection. Our task is now to construct a simultaneous confidence region with 95% probability for the frequency response at the ten frequencies

$$\Omega = \{0.1, 0.2, 0.4, 0.6, 0.8, 1, 2, 4, 6, 8\} \text{ rad/s} \quad (33)$$

(the baseline frequency is $\Omega_0 = 0.1$ rad/s). The available experiment time $T_{\text{exp}} = 300$ seconds. Here we apply a Schroeder-phased multi-sine input (Bayard, 1993) with the ten frequencies. The amplitude and phases are given by

$$A_m = \sqrt{2/L}, \quad \psi_m = 2\pi \sum_{r=1}^m rA_r^2/2 = \frac{2\pi}{L} m(m+1) \quad (34)$$

with $L = 10$ for $m = 1, 2, \dots, 10$.

4.2.1. CASE 1: freely selectable sampling period

Here $P = 8$, and choosing $S = 4$ gives the sampling period $T = 0.0614$ s from (23) and the number of the samples within one period of the baseline sinusoid is $N_0 = 4 \cdot 2^8 = 1024$ samples.

Out of approximately $N_{\text{exp}} = T_{\text{exp}}/T = 4889$ samples available, after waiting $\ell = 793$ samples, we gather 4096 samples which corresponds to $n = 4$ periods of the baseline sinusoid, and calculate 4000 scaled empirical correlation functions after generating decoupling binary strings.

Figs. 9 and 10 show the constructed simultaneous confidence region (converted using (13)) with probability at least $1 - 2 \cdot 10 \cdot 2 \cdot 5/4000 = 0.95$ with $L = 10$, $M = 4000$, and $q = 5$. In this example the sampled noise sequence $v(kT)$ is a white noise sequence uniformly distributed on $[-0.25, 0.25]$.

4.2.2. CASE 2: a priori fixed sampling period

For the same system and the task considered in CASE 1, suppose that the sampling period is now fixed as $T_{\text{fix}} = 1/16$ s = 0.0625 s for the same experiment time $T_{\text{exp}} = 300$ s so that the total number of samples available is now $N_{\text{exp}} = T_{\text{exp}}/T_{\text{fix}} = 4800$ samples.

We would like to find a set of frequency points, $\hat{\Omega}$, which allows decoupling strings with the fixed sampling period T_{fix} . To do this, we solve the integer program (32). To see the effect of the weighting factor α_m , $m = 1, \dots, 10$, we consider two different sets of weighting factors: one is the unity-weighting set consisting of all

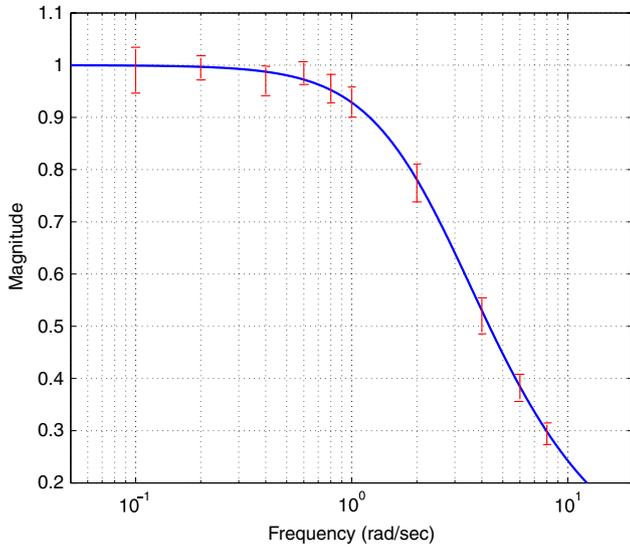


Fig. 9. True frequency response and simultaneous 95% confidence region: magnitude plot.

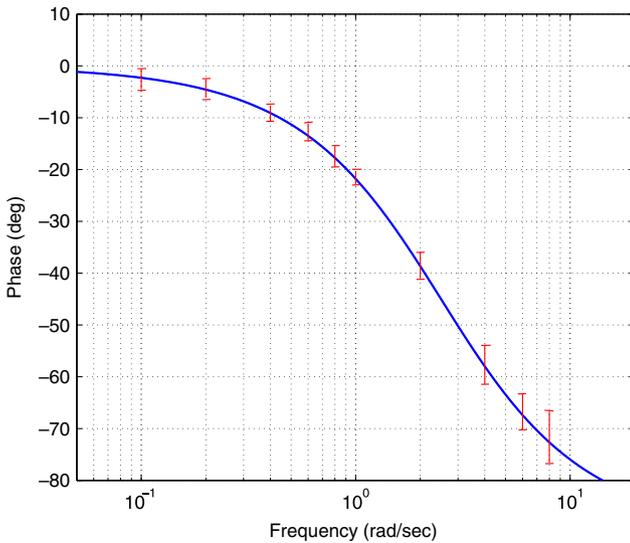


Fig. 10. True frequency response and simultaneous 95% confidence region: phase plot.

ones $\alpha_m^{(1)} = 1$, $m = 1, \dots, 10$, and the other is the exponential weighting set whose weights decrease with frequency given by $\alpha_m^{(2)} = (\Omega_1/\Omega_m)^2$. For both cases, we used the fixed parameters $S_\ell = 4$ and $S_u = 10$.

The results are shown in Table 1. Compared to the unity-weighting case, the exponential weighting set resulted in more balanced frequencies for the ten frequencies. Furthermore, the exponential weighting set yielded $\hat{S} = 4$ and $\hat{P} = 8$ and thus the number of samples in one approximate baseline sinusoid is $\hat{S} \cdot 2^{\hat{P}} = 1024$ samples. Hence we may use up to 4 cycles (=4096 samples) of the approximate baseline sinusoid. However for the unity-weighting case we can use only 3072 samples corresponding to 2 cycles of the approximate baseline sinusoid. By applying the procedure (P3') in Section 3 we can construct simultaneous confidence regions for the approximate frequency points given in Table 1.

5. Conclusion

In this paper, we have extended the LSCR algorithm introduced in Campi and Weyer (2005) to the problem of constructing confi-

dence regions for the frequency response at multiple frequencies using a finite number of input–output data points. No information about the tail of the impulse response can be obtained from a finite number of data points, and hence *a priori* information has been used to bound the effects of the tail. Three theorems have been established describing the probabilistic properties of the constructed confidence regions: Theorem 1 shows that the constructed confidence sets contain the true real and imaginary parts of the frequency response at a single frequency, while Theorem 2 extends Theorem 1 to cover the frequency response at multiple frequencies. Theorem 3 proved that the constructed confidence set eventually converges to the true frequency response as the number of data points increases. In order to reduce the computations required for implementing the general algorithm for multiple frequencies, we developed a fast numerical method with decoupling binary strings. The developed algorithm was demonstrated with good results in two simulation examples with multi-sine inputs.

Acknowledgments

The research of S. Ko and E. Weyer was partly supported by the Australian Research Council under the Discovery Grant Scheme, Project DOP0558579 and the National Research Foundation of Korea (NRF) grant funded by the Korean Government (MSIP) (NRF-2013M1A3A3A02042434). The research of M.C. Campi was supported by MIUR under the project ‘‘Identification and Adaptive Control for Industrial Systems’’.

Appendix A. Proof for Theorem 1

For the proof of the main theorem, we first state the following propositions from Campi et al. (2009) and Campi and Weyer (2010).

Proposition 1. Let \mathbf{H} be a stochastic $M \times N$ matrix with elements $h_{i,k}$, $i = 0, \dots, M-1$, $k = 1, \dots, N$, constructed according to point (P3) of the algorithm in Section 2.2, and further let $\boldsymbol{\xi} \triangleq [\xi_1, \dots, \xi_N]^T$ be a vector independent of \mathbf{H} of mutually independent random variables symmetrically distributed around $\mathbf{0}$. Given an $\bar{i} \in [0, M-1]$, let $\mathbf{H}_{\bar{i}}$ be the $M \times N$ matrix whose rows are all equal to the \bar{i} th row of \mathbf{H} . Then, $\mathbf{H}\boldsymbol{\xi}$ and $(\mathbf{H} - \mathbf{H}_{\bar{i}})\boldsymbol{\xi}$ have the same M -dimensional distribution provided that the \bar{i} th element of $(\mathbf{H} - \mathbf{H}_{\bar{i}})\boldsymbol{\xi}$ (which is 0) is repositioned as the first entry of the vector.

The next proposition proves that the elements of the vector $\mathbf{H}\boldsymbol{\xi}$ exhibit a precise ordering property. Through a simple modification of the proof for Proposition 4 in Campi et al. (2009) this can be easily proved.

Proposition 2. Let \mathbf{H} and $\boldsymbol{\xi}$ be as in Proposition 1, and in addition assume that $\boldsymbol{\xi}$ admits a density. Then, the random vector $\mathbf{H}\boldsymbol{\xi}$ has the following property: each element of the vector $\mathbf{H}\boldsymbol{\xi}$ has the same probability $1/M$ to be in the j th position (i.e. there are exactly $j-1$ other elements in $\mathbf{H}\boldsymbol{\xi}$ smaller than the variable under consideration) and this holds for any choice of j between 1 to M .

Now consider the following events (with the notation $\xi_k \triangleq v(kT) \cos \varphi_r(kT)$)

$$\mathbf{A} = \left\{ \sum_{k=\ell+1}^{N_1} h_{i,k} \xi_k < 0 \text{ for at most } q-1 \text{ selections of } i \right\} \\ \cup \left\{ \sum_{k=\ell+1}^{N_1} h_{i,k} \xi_k > 0 \text{ for at most } q-1 \text{ selections of } i \right\}$$

Table 1
Approximate frequency points for different weighting factors using the Integer Programming.

m	Ω_m	$\alpha_m^{(1)}$	$\hat{\Omega}_m^{(1)}$	$\mathbf{x}_{\text{opt}}^{(1)}$	$\alpha_m^{(2)}$	$\hat{\Omega}_m^{(2)}$	$\mathbf{x}_{\text{opt}}^{(2)}$
1	0.1	1	0.0654	$\hat{S} = 6, \hat{P} = 8$	1.00000	0.0982	$\hat{S} = 4, \hat{P} = 8$
2	0.2	1	0.1963	$\hat{i}_2 = 3$	0.25000	0.1963	$\hat{i}_2 = 2$
3	0.4	1	0.3927	$\hat{i}_3 = 6$	0.06250	0.3927	$\hat{i}_3 = 4$
4	0.6	1	0.5890	$\hat{i}_4 = 9$	0.02778	0.5890	$\hat{i}_4 = 6$
5	0.8	1	0.7854	$\hat{i}_5 = 12$	0.01563	0.7854	$\hat{i}_5 = 8$
6	1.0	1	0.9817	$\hat{i}_6 = 15$	0.01000	0.9817	$\hat{i}_6 = 10$
7	2.0	1	2.0289	$\hat{i}_7 = 31$	0.00250	1.9635	$\hat{i}_7 = 20$
8	4.0	1	3.9924	$\hat{i}_8 = 61$	0.00063	4.0252	$\hat{i}_8 = 41$
9	6.0	1	6.0214	$\hat{i}_9 = 92$	0.00028	5.9887	$\hat{i}_9 = 61$
10	8.0	1	7.9849	$\hat{i}_{10} = 122$	0.00016	7.9522	$\hat{i}_{10} = 81$

and

$$\mathbf{B} = \left\{ C_{r,i}^a(\theta^0) \Gamma_{r,i}^a < 0 \text{ for at most } q-1 \text{ selections of } i \right\} \\ \cup \left\{ C_{r,i}^a(\theta^0) + \Gamma_{r,i}^a > 0 \text{ for at most } q-1 \text{ selections of } i \right\}. \quad (\text{A.1})$$

Then, in view of Proposition 2 we have

$$\Pr(\mathbf{A}) = \frac{2q}{M} \quad (\text{A.2})$$

since an element, say the j th, of the vector $\mathbf{H}\xi$ is $\sum_{k=\ell+1}^{N_1} h_{i,k} \xi_k$. From (21), we have

$$C_{r,i}^a(\theta^0) - \Gamma_{r,i}^a \leq \sum_{k=\ell+1}^{N_1} h_{i,k} \xi_k \leq C_{r,i}^a(\theta^0) + \Gamma_{r,i}^a. \quad (\text{A.3})$$

If there are at most $q-1$ functions such that $C_{r,i}^a(\theta^0) - \Gamma_{r,i}^a < 0$, then from (A.3), there are also at most $q-1$ functions such that $\sum_{k=\ell+1}^{N_1} h_{i,k} \xi_k < 0$. Similarly, if there are at most $q-1$ functions such that $C_{r,i}^a(\theta^0) + \Gamma_{r,i}^a > 0$, then there are also at most $q-1$ functions such that $\sum_{k=\ell+1}^{N_1} h_{i,k} \xi_k > 0$. Hence, we have $\mathbf{B} \subset \mathbf{A}$ and

$$\Pr(\mathbf{B}) \leq \Pr(\mathbf{A}). \quad (\text{A.4})$$

Suppose that we have extracted a probabilistic outcome β from \mathbf{B} . Then, from (A.1), either $C_{r,i}^a(\theta^0) - \Gamma_{r,i}^a < 0$ for at most $q-1$ selections of i or $C_{r,i}^a(\theta^0) + \Gamma_{r,i}^a > 0$ for at most $q-1$ selections of i , so that $\theta^0 \notin \Theta_r^a$ (recall the construction of Θ_r^a). Vice versa, if $\beta \notin \mathbf{B}$, then $C_{r,i}^a(\theta^0) - \Gamma_{r,i}^a < 0$ for at least q selections of i and $C_{r,i}^a(\theta^0) + \Gamma_{r,i}^a > 0$ for at least q selections of i , yielding $\theta^0 \in \Theta_r^a$. Using (A.2) and (A.4), the conclusion is drawn that $\Pr(\theta^0 \in \Theta_r^a) \geq 1 - 2q/M$. $\Pr(\theta^0 \in \Theta_r^b) \geq 1 - 2q/M$ can be proved similarly. \square

Appendix B. Proof for Theorem 3

Before proving the convergence, we need the following lemmas.

Lemma 1. For $\Omega T \neq 0$, as $N \rightarrow \infty$, we have

- (1) $\frac{1}{N} \sum_{k=1}^N \cos(\Omega T k) \rightarrow 0$ and $\frac{1}{N} \sum_{k=1}^N \sin(\Omega T k) \rightarrow 0$,
- (2) $\frac{1}{N} \sum_{k=1}^N (h_{i,k} - 0.5) \cos(\Omega T k) \rightarrow 0$ and $\frac{1}{N} \sum_{k=1}^N (h_{i,k} - 0.5) \sin(\Omega T k) \rightarrow 0$ with probability 1,
- (3) $\frac{1}{N} \sum_{k=1}^N h_{i,k} \cos(\Omega T k) \rightarrow 0$ and $\frac{1}{N} \sum_{k=1}^N h_{i,k} \sin(\Omega T k) \rightarrow 0$ with probability 1,
- (4) $\frac{1}{N} \sum_{k=1}^N h_{i,k} \gamma(kT) |\cos \varphi_r(kT)| \rightarrow 0$ and $\frac{1}{N} \sum_{k=1}^N h_{i,k} \gamma(kT) |\sin \varphi_r(kT)| \rightarrow 0$ with probability 1.

Proof. (1) As

$$\left| \frac{1}{N} \sum_{k=1}^N e^{j\Omega T k} \right| = \left| \frac{1}{N} \frac{e^{j\Omega T} - e^{j\Omega T(N+1)}}{1 - e^{j\Omega T}} \right| \\ \leq \frac{1}{N} \frac{|e^{j\Omega T}| + |e^{j\Omega T(N+1)}|}{|1 - e^{j\Omega T}|} \\ \leq \frac{1}{N} \frac{2}{|1 - e^{j\Omega T}|} \rightarrow 0 \text{ as } N \rightarrow \infty,$$

$$\lim_{N \rightarrow \infty} \sum_{k=1}^N \frac{e^{j\Omega T k}}{N} = \lim_{N \rightarrow \infty} \sum_{k=1}^N \frac{\cos \Omega T k + j \sin \Omega T k}{N} = 0.$$

Hence (1) is proved.

(2) Since $E\{(h_{i,k} - 0.5) \cos(\Omega T k)\} = 0$ and $E\{(h_{i,k} - 0.5)^2 \cos^2(\Omega T k)\} \leq 0.25$, (2) follows from Kolmogorov's strong law of large numbers, see Appendix C.

(3) This result can be obtained directly using (1) and (2).

(4) This follows from (14) and

$$\frac{1}{N} \sum_{k=1}^N h_{i,k} e^{-\rho k T} |\cos \varphi_r(kT)| \leq \frac{1}{N} \sum_{k=1}^N e^{-\rho k T} \\ = \frac{1}{N} \frac{e^{-\rho T} - e^{-\rho(N+1)T}}{1 - e^{-\rho T}} \rightarrow 0 \text{ as } N \rightarrow \infty. \quad \square$$

Lemma 2. Let $\tilde{a}_r = a_r^0 - a_r$, $\tilde{b}_r = b_r^0 - b_r$, and $N = N_1 - \ell$. Then as $N \rightarrow \infty$

$$\frac{1}{N} (C_{r,i}^a(\theta) \pm \Gamma_{r,i}^a) \rightarrow \frac{A_r \tilde{a}_r}{4}, \\ \frac{1}{N} (C_{r,i}^b(\theta) \pm \Gamma_{r,i}^b) \rightarrow -\frac{A_r \tilde{b}_r}{4}, \text{ with probability 1} \quad (\text{B.1})$$

for all $i = 0, 1, \dots, M-1$ and $r = 1, 2, \dots, L$.

Proof. From (20) and (19), we have

$$\frac{1}{N} (C_{r,i}^a(\theta) \pm \Gamma_{r,i}^a) = A_r \tilde{a}_r \frac{1}{N} \sum_{k=\ell+1}^{N_1} h_{i,k} \cos^2 \varphi_r(kT) \\ - A_r \tilde{b}_r \frac{1}{N} \sum_{k=\ell+1}^{N_1} h_{i,k} \cos \varphi_r(kT) \sin \varphi_r(kT) \\ + \sum_{m=1(m \neq r)}^L A_m \left\{ \tilde{a}_m \left[\frac{1}{N} \sum_{k=\ell+1}^{N_1} h_{i,k} \cos \varphi_m(kT) \cos \varphi_r(kT) \right] \right. \\ \left. - \tilde{b}_m \left[\frac{1}{N} \sum_{k=\ell+1}^{N_1} h_{i,k} \sin \varphi_m(kT) \cos \varphi_r(kT) \right] \right\} \\ + \frac{1}{N} \sum_{k=\ell+1}^{N_1} h_{i,k} \sum_{m=1}^L A_m \tilde{y}_m(kT) \cos \varphi_r(kT)$$

$$\begin{aligned}
& + \frac{1}{N} \sum_{k=\ell+1}^{N_1} h_{i,k} v(kT) \cos \varphi_r(kT) \\
& \pm A \frac{1}{N} \sum_{k=\ell+1}^{N_1} h_{i,k} \gamma(kT) |\cos \varphi_r(kT)|. \tag{B.2}
\end{aligned}$$

We show convergence of each term of (B.2):

◇ 1st term: Using trigonometric formulae, the first term becomes

$$0.5A_r \tilde{a}_r \frac{1}{N} \sum_{k=\ell+1}^{N_1} h_{i,k} + 0.5A_r \tilde{a}_r \frac{1}{N} \sum_{k=\ell+1}^{N_1} h_{i,k} \cos(2\Omega_r kT + 2\psi_r)$$

and as $N \rightarrow \infty$ the first term converges to $A_r \tilde{a}_r / 4$ and the second term goes to 0 w.p. 1, using (3) of Lemma 1.

◇ 2nd and 3rd terms: Using trigonometric formulae and (3) of Lemma 1, all terms go to zero w.p. 1 as $N \rightarrow \infty$.

◇ 4th and 6th terms: The magnitude of the 4th term is bounded by the 6th term by (14) and (19). Since the 6th term converges to zero w.p. 1 from (4) of Lemma 1, the 4th term also goes to zero w.p. 1 as $N \rightarrow \infty$.

◇ 5th term: Follows from Kolmogorov's strong law of large numbers (See Appendix C).

The convergence of $\frac{1}{N} (C_{r,i}^b(\theta) \pm \Gamma_{r,i}^b)$ to $-A_r \tilde{b}_r / 4$ can be proved similarly. □

Now we return to the proof of Theorem 3. It follows from Lemma 2 that when $a_r \neq a_r^0$ and $b_r \neq b_r^0$, for all $i = 0, 1, \dots, M-1$, the empirical correlation functions $\frac{1}{N} (C_{r,i}^a(\theta) \pm \Gamma_{r,i}^a)$ and $\frac{1}{N} (C_{r,i}^b(\theta) \pm \Gamma_{r,i}^b)$ will with probability 1 have the same signs as $A_r \tilde{a}_r / 4$ and $-A_r \tilde{b}_r / 4$, respectively, for all $N > \bar{N}$ for a sufficiently large (realization dependent) value of \bar{N} , and hence θ will be excluded from the confidence set $\hat{\Theta}_N$. This completes the proof. □

Appendix C. Strong law of large numbers (Kolmogorov)

Suppose Y_1, Y_2, \dots , are independent random variables with $E[Y_i] = 0$, and that $\sum_{k=1}^{\infty} E\{Y_k^2\}/k^2 < \infty$, then $\frac{1}{N} \sum_{k=1}^N Y_k \rightarrow 0$ with probability 1 as $N \rightarrow \infty$. See e.g. Shiryaev (1995) Chapter IV.3 for a proof.

Appendix D. Proof for the generation of decoupling strings

It is sufficient to prove that (24) is 0 when summed over an arbitrary index set \mathbf{J}_p (27), so we consider only \mathbf{J}_1 . Here we prove the results for the sine function in (24). The same approach can be used for the cosine function.

A chosen time index k in the first segment generates another $2^p - 1$ indices in the remaining segments, according to (26). Let say they are

$$\begin{aligned}
\{n_1 = k, n_2 = k + S, n_3 = k + 2S, \dots, \\
n_{2^p} = k + (2^p - 1)S\}. \tag{D.1}
\end{aligned}$$

(D.1) can be written as a union of 2^{p-1} pairs $\{k_j, k'_j\}$ such that the sum of the sine functions evaluated over each pair is zero, and thus we have

$$\sum_{j=1}^{2^p} \sin(\Omega_m T n_j) = 0 \quad \text{for all } m \in \{1, \dots, 2i_{\max}\}. \tag{D.2}$$

Before providing the proof, we present a simple example describing the idea.

Example 1. Consider the case with $\Omega_0 = 1$ rad/s and $i_{\max} = 6$. Here $P = 4$ and $T = 0.0131$ s for $S = 30$. One period of the baseline sinusoid is divided into $2^P = 16$ segments. Suppose that a time index $k = 10$ is chosen in the first segment. We evaluate $\sin(\Omega_m T n_j)$ at the time indices in (D.1) with $k = 10$ for the three frequencies $\Omega_1 = 3\Omega_0$, $\Omega_2 = 4\Omega_0$, and $\Omega_3 = 6\Omega_0$. They are shown in Table D.1.

We can observe that the sum of each column in the table is zero: The index set of (D.1) is divided into two subsets \mathcal{A} and \mathcal{B} , and the sum over the indices in \mathcal{A} has the negative value of the sum over the indices in \mathcal{B}

$$\sum_{j \in \mathcal{A}} \sin(\Omega_m T n_j) = - \sum_{j \in \mathcal{B}} \sin(\Omega_m T n_j). \tag{D.3}$$

(D.3) holds for $\Omega_m = 3\Omega_0$ with

$$\begin{aligned}
\mathcal{A} &= \{1, 2, 3, 4, 5, 6, 7, 8\}, \\
\mathcal{B} &= \{9, 10, 11, 12, 13, 14, 15, 16\},
\end{aligned}$$

for $\Omega_m = 4\Omega_0$ with

$$\begin{aligned}
\mathcal{A} &= \{1, 2, 5, 6, 9, 10, 13, 14\}, \\
\mathcal{B} &= \{3, 4, 7, 8, 11, 12, 15, 16\},
\end{aligned}$$

and for $\Omega_m = 6\Omega_0$ with

$$\begin{aligned}
\mathcal{A} &= \{1, 2, 3, 4, 9, 10, 11, 12\}, \\
\mathcal{B} &= \{5, 6, 7, 8, 13, 14, 15, 16\}. \quad \square
\end{aligned}$$

In the above example the frequencies are (i) an odd multiple of Ω_0 ($3\Omega_0$), (ii) a power of 2 times Ω_0 ($4\Omega_0$), and (iii) an even (but not a power of 2) multiple of Ω_0 ($6\Omega_0$). We now prove (D.2) in general for these three subsets of the frequencies. We only show the results for the sines, the cosines follow by the same argument. We use the notation $\omega_0 \triangleq \Omega_0 T$.

(1) For odd-multiple frequencies $i_m \Omega_0$ with $i_m = 2l + 1$, $l = 0, 1, \dots$: We divide 2^p sample indices into two subsets

$$\begin{aligned}
\mathcal{A} &= \{1, 2, 3, \dots, 2^{p-1}\}, \\
\mathcal{B} &= \{j + d : \forall j \in \mathcal{A}\} \quad \text{with } d \triangleq 2^{p-1}.
\end{aligned}$$

Now we show that each index in \mathcal{A} has a canceling counterpart in \mathcal{B} .

Select $j \in \mathcal{A}$ and then $j + d \in \mathcal{B}$. Note that $n_j = k + (j - 1)S$ and $n_{j+d} = k + (j + d - 1)S$ for some k . Observe that

$$\begin{aligned}
z_j &= \sin[(2l + 1)\omega_0 n_j] \\
&= \sin[(2l + 1)\omega_0 (k + (j - 1)S)] \\
&= \sin\left[(2l + 1)\omega_0 k + 2\pi \frac{(j - 1)(2l + 1)}{2^p}\right],
\end{aligned}$$

and

$$\begin{aligned}
z_{j+d} &= \sin[(2l + 1)\omega_0 (k + (j - 1 + 2^{p-1})S)] \\
&= \sin\left[(2l + 1)\omega_0 k + 2\pi \frac{(j - 1)(2l + 1)}{2^p} + (2l + 1)\pi\right] \\
&= -\sin\left[(2l + 1)\omega_0 k + 2\pi \frac{(j - 1)(2l + 1)}{2^p}\right] = -z_j.
\end{aligned}$$

Here $S = 2\pi / (2^p \omega_0)$ was used. Therefore (D.2) is satisfied for these frequencies.

Table D.1Evaluation of the sine and cosine functions at different time indices n_j .

j	$\sin(3\Omega_0 T n_j)$	$\sin(4\Omega_0 T n_j)$	$\sin(6\Omega_0 T n_j)$	$\cos(3\Omega_0 T n_j)$	$\cos(4\Omega_0 T n_j)$	$\cos(6\Omega_0 T n_j)$
1	+0.3827	+0.5000	+0.7071	+0.9239	+0.8660	+0.7071
2	+1.0000	+0.8660	0	0	-0.5000	-1.0000
3	+0.3827	-0.5000	-0.7071	-0.9239	-0.8660	+0.7071
4	-0.7071	-0.8660	+1.0000	-0.7071	+0.5000	0
5	-0.9239	+0.5000	-0.7071	+0.3827	+0.8660	-0.7071
6	0	+0.8660	0	+1.0000	-0.5000	+1.0000
7	+0.9239	-0.5000	+0.7071	+0.3827	-0.8660	-0.7071
8	+0.7071	-0.8660	-1.0000	-0.7071	+0.5000	0
9	-0.3827	+0.5000	+0.7071	-0.9239	+0.8660	+0.7071
10	-1.0000	+0.8660	0	0	-0.5000	-1.0000
11	-0.3827	-0.5000	-0.7071	+0.9239	-0.8660	+0.7071
12	+0.7071	-0.8660	+1.0000	+0.7071	+0.5000	0
13	+0.9239	+0.5000	-0.7071	-0.3827	+0.8660	-0.7071
14	0	+0.8660	0	-1.0000	-0.5000	+1.0000
15	-0.9239	-0.5000	+0.7071	-0.3827	-0.8660	-0.7071
16	-0.7071	-0.8660	-1.0000	+0.7071	+0.5000	0

- (2) For the frequencies $i_m \Omega_0$ with $i_m = 2^l$, $l = 1, \dots, P-1$: We divide 2^P sample indices into two subsets

$$\mathcal{A} = \left\{ \left[(2r-2)d+1, (2r-2)d+2, \dots, (2r-2)d+d \right], \right. \\ \left. r = 1, \dots, 2^l \right\}, \\ \mathcal{B} = \left\{ j+d : \forall j \in \mathcal{A} \right\} \quad \text{with } d \triangleq 2^{P-(l+1)}.$$

Select $j \in \mathcal{A}$ and then $j+d \in \mathcal{B}$. Observe that

$$z_j = \sin[2^l \omega_0 n_j] = \sin[2^l \omega_0 (k + (j-1)S)] \\ = \sin \left[2^l \omega_0 k + 2\pi \frac{(j-1)2^l}{2^P} \right]$$

and

$$z_{j+d} = \sin[2^l \omega_0 (k + (j-1)S + 2^{P-(l+1)}S)] \\ = \sin \left[2^l \omega_0 k + 2\pi \frac{(j-1)2^l}{2^P} + 2\pi \frac{2^{P-(l+1)}2^l}{2^P} \right] \\ = \sin \left[2^l \omega_0 k + 2\pi \frac{(j-1)2^l}{2^P} + \pi \right] \\ = -z_j.$$

This shows that (D.2) is satisfied for these frequencies.

- (3) For the other even-multiple frequencies $i_m \Omega_0$ with $i_m = 2l$: the allowable form of $2l$ is $2l = (2r_1 - 1)2^{q_1}$ for some positive integers r_1 and q_1 . Otherwise $2l$ becomes a power of two. Note that $2l = (2r_1 - 1)2^{q_1} \leq i_{\max} \cdot 2 \cdot i_{\max} < 2^P$, $2 < 2r_1 - 1$, and so $2^{q_1} < 2^{P-1}$. For this case, we set

$$d \triangleq \frac{2^{P-1}}{\gcd(2^{P-1}, 2l)} = \frac{2^{P-1}}{2^{q_1}} = 2^{P-1-q_1},$$

where $\gcd(A, B)$ denotes the greatest common divisor of A and B , and divide 2^P sample indices into two subsets

$$\mathcal{A} = \left\{ \left[(2r-2)d+1, (2r-2)d+2, \dots, (2r-2)d+d \right], \right. \\ \left. r = 1, \dots, \gcd(2^{P-1}, 2l) \right\}, \\ \mathcal{B} = \left\{ j+d : \forall j \in \mathcal{A} \right\}.$$

Select $j \in \mathcal{A}$ and then $j+d \in \mathcal{B}$. Observe that

$$z_j = \sin[2l \omega_0 n_j] = \sin[2l \omega_0 (k + (j-1)S)] \\ = \sin \left[2l \omega_0 k + 2\pi \frac{2l(j-1)}{2^P} \right] \\ \text{and } z_{j+d} = \sin \left[2l \omega_0 k + 2\pi \frac{2l(j-1)}{2^P} + d \frac{2l}{2^{P-1}} \pi \right].$$

Since

$$d \frac{2l}{2^{P-1}} = \frac{2^l}{\gcd(2^{P-1}, 2l)} \\ = \frac{(2r_1 - 1)2^{q_1}}{\gcd\{2^{P-1}, (2r_1 - 1)2^{q_1}\}} = 2r_1 - 1$$

which is an odd number, we have $z_{j+d} = -z_j$.

Therefore (D.2) is satisfied for these frequencies.

This completes the proof. \square

References

- Bayard, D. S. (1993). Statistical plant set estimation using Schroeder-phased multisinusoidal input design. *Applied Mathematics and Computation*, 58, 169–198.
- Campi, M. C., & Weyer, E. (2005). Guaranteed non-asymptotic confidence regions in system identification. *Automatica*, 41(10), 1751–1764.
- Campi, M. C., & Weyer, E. (2006). Identification with finitely many data points: the LSCR approach. In *Proceedings of the 14th IFAC symposium on system identification* (pp. 46–64). Available at <http://www.ing.unibs.it/~campi/LSCRwebsite>.
- Campi, M. C., & Weyer, E. (2010). Non-asymptotic confidence sets for the parameters of linear transfer functions. *IEEE Transactions on Automatic Control*, 55(12), 2708–2720.
- Campi, M. C., Ko, S., & Weyer, E. (2009). Non-asymptotic confidence regions for model parameters in the presence of unmodelled dynamics. *Automatica*, 45(10), 2175–2186.
- de Vries, D. K., & Van den Hof, P. M. J. (1995). Quantification of uncertainty in transfer function estimation: a mixed probabilistic-worst-case approach. *Automatica*, 31(4), 543–557.
- den Dekker, A. J., Bombois, X., & Van den Hof, P. M. J. (2008). Finite sample confidence regions for parameters in prediction error identification using output error models. In *Proceedings of the 17th IFAC World Congress*.
- Garatti, S., Campi, M. C., & Bittanti, S. (2004). Assessing the quality of identified models through the asymptotic theory - When is the result reliable? *Automatica*, 40(8), 1319–1332.
- Goodwin, G. C., Gevers, M., & Ninnies, B. (1992). Quantifying the error in estimated transfer functions with application to model order selection. *IEEE Transactions on Automatic Control*, 37(7), 913–928.
- Hjalmarsson, H., & Ninnies, B. (2006). Least-squares estimation of a class of frequency functions: a finite sample variance expression. *Automatica*, 42(4), 589–600.
- Holmström, K., Görán, A. O., & Edvall, M. M. (2010). *User's guide for TOMLAB7*. Tomlab Optimization Inc., <http://tomopt.com/docs/TOMLAB.pdf>.
- Ko, S., Weyer, E., & Campi, M. C. (2007). Non-asymptotic uncertainty assessment of frequency responses using the LSCR approach. In *Proceedings of the international conference on control, automation and systems 2007*.
- Ko, S., Weyer, E., & Campi, M. C. (2008). Non-asymptotic model quality assessment of transfer functions at multiple frequency points. In *Proceedings of the 17th IFAC World Congress*.

Ljung, Lennart (1999). *System identification: theory for the user*. Upper Saddle River, NJ: Prentice Hall.

Mathworks, (2013). *Global optimization toolbox: user's guide*.

Pintelon, R., & Schoukens, J. (2001). *System identification: A frequency domain approach*. New York, USA: IEEE press.

Shiryayev, A. N. (1995). *Probability*. New York, USA: Springer-Verlag.



Sangho Ko is an associate professor at the School of Aerospace and Mechanical Engineering, Korea Aerospace University, South Korea. He received his Ph.D. degree from the Department of Mechanical and Aerospace Engineering of the University of California, San Diego (UCSD) in 2005. From 2005 to 2006, he was a post-doctoral scholar in UCSD to research system identification techniques. From March 2006 to February 2008 he was a research fellow in the Department of Electrical and Electronic Engineering of the University of Melbourne, Australia, where he worked on finite sample quality assessment of system identification.

From 1992 to 1999, he was with Samsung Aerospace Industries, Ltd., Kyungnam, Korea, where he was a research and development engineer for digital flight control systems. His research interests include: state estimation and control, system identification, flight control, propulsion system control.



Erik Weyer is a professor in the Department of Electrical and Electronic Engineering at the University of Melbourne. He received the Siv. Ing. degree in 1988 and the Ph.D. in 1993, both from the Norwegian Institute of Technology, Trondheim, Norway. From 1994 to 1996 he was a Research Fellow at the University of Queensland, and since 1997 he has been with the Department of Electrical and Electronic Engineering at the University of Melbourne. He has held visiting positions at the University of Brescia, Italy, the Technical University of Vienna, Austria, and Politecnico di Milano, Italy. From 2010 to 2012 he was an associate editor

of IEEE Transactions of Automatic Control, and he is currently an associate editor of Automatica. His research interests are in the areas of system identification and control, with particular emphasis on finite sample properties of system identification methods, and modeling and control of irrigation channels and rivers. He was a co-recipient of the IEEE CSS Control System Technology Award in 2014.



Marco Claudio Campi is Professor of Automatic Control at the University of Brescia, Italy.

In 1988, he received the Doctor degree in electronic engineering from the Politecnico di Milano, Milano, Italy. From 1988 to 1989, he was a Lecturer at the Department of Electrical Engineering of the Politecnico di Milano. From 1989 to 1992, he was a Research Fellow at the Centro di Teoria dei Sistemi of the National Research Council (CNR) in Milano and, in 1992, he joined the University of Brescia, Brescia, Italy. He has held visiting and teaching appointments at the Australian National University, Canberra, Australia; the University of Illinois at Urbana-Champaign, USA; the Centre for Artificial Intelligence and Robotics, Bangalore, India; the University of Melbourne, Australia; the Kyoto University, Japan.

Marco Campi is the chair of the *Technical Committee IFAC on Modeling, Identification and Signal Processing (MISP)*. He has been in various capacities on the Editorial Board of *Automatica*, *Systems and Control Letters* and the *European Journal of Control*. Marco Campi is a recipient of the "Giorgio Quazza" prize, and, in 2008, he received the IEEE CSS George S. Axelby outstanding paper award for the article *The Scenario Approach to Robust Control Design*. He has delivered plenary and semi-plenary addresses at major conferences including SYSID, MTNS, and CDC, and has been a distinguished lecturer of the Control Systems Society. Marco Campi is a Fellow of IEEE, a member of IFAC, and a member of SIDRA.

The research interests of Marco Campi include: system identification, stochastic systems, randomized methods, adaptive and data-based control, robust optimization, and learning theory.