

## Frequency domain weighted nonlinear least squares estimation of parameter-varying differential equations

Goos, Jan; Lataire, John; Louarroudi, Ebrahim; Pintelon, Rik

*Published in:*  
Automatica

*DOI:*  
[10.1016/j.automatica.2016.09.031](https://doi.org/10.1016/j.automatica.2016.09.031)

*Publication date:*  
2017

*Document Version:*  
Submitted manuscript

[Link to publication](#)

### *Citation for published version (APA):*

Goos, J., Lataire, J., Louarroudi, E., & Pintelon, R. (2017). Frequency domain weighted nonlinear least squares estimation of parameter-varying differential equations. *Automatica*, 75(1), 191-199.  
<https://doi.org/10.1016/j.automatica.2016.09.031>

### **Copyright**

No part of this publication may be reproduced or transmitted in any form, without the prior written permission of the author(s) or other rights holders to whom publication rights have been transferred, unless permitted by a license attached to the publication (a Creative Commons license or other), or unless exceptions to copyright law apply.

### **Take down policy**

If you believe that this document infringes your copyright or other rights, please contact [openaccess@vub.be](mailto:openaccess@vub.be), with details of the nature of the infringement. We will investigate the claim and if justified, we will take the appropriate steps.

# Weighted nonlinear least squares estimation of linear differential equations with parameter-varying coefficients. <sup>★</sup>

Jan Goos, John Lataire, Ebrahim Louarroudi, Rik Pintelon

*ELEC, Vrije Universiteit Brussel  
Pleinlaan 2, 1050 Brussels, Belgium*

---

## Abstract

This paper presents a frequency domain identification technique for estimating of Linear Parameter-Varying (LPV) differential equations. In a band-limited setting, it is shown that the time derivatives of the input and output signals can be computed exactly in the frequency domain, even for non-periodic inputs and parameter variations. The method operates in an errors-in-variables framework (noisy input and output), but the scheduling signal is assumed to be known. Under these conditions, the proposed estimator is proven to be consistent.

*Key words:* Linear Parameter-Varying systems; Identification methods

---

## 1 Introduction

A good example of an LPV system is a construction crane, which is basically a pendulum of varying length  $l(t)$ . The cable length directly influences the poles of the system, thereby determining its eigenfrequency. In the linear parameter varying (LPV) framework [16,18], we call variables like  $l(t)$  scheduling parameters, and denote them as  $p(t)$ . The dynamic relation between input  $u(t)$  and output  $y(t)$  is still linear, but it depends on the (continuously varying) scheduling parameter.

There are two main classes of LPV identification techniques: local and global approaches. In the local LPV framework [2,3], a nonlinear or parameter-varying model is linearized at different operating points. The result is a set of LTI models, which are then interpolated over the operating range. Here, we opt to directly estimate an LPV model, from a single, global experiment, where the scheduling parameter  $p(t)$  varies during the measurement, covering its entire operating range. Amongst other advantages, a global modeling approach captures transient dynamics, when the plant shifts from one operating point to another. Additionally, the rate of change of the scheduling parameter can be directly accounted for.

In this paper we study parameter-varying differential equations. During the past decade, a lot of research has been dedicated to the identification of LPV input-output equations, but mostly in discrete-time [9,19]. For an overview, see [8]. In [10], the first steps are taken towards direct identification of an LPV differential equation, using an instrumental variables approach. However, the input signals is assumed to be known, and the time derivatives are approximated using filtering operations. The main contribution of this paper is the frequency domain identification of continuous-time LPV systems from input-output data without approximating the time derivatives, and where both  $u(t)$  and  $y(t)$  can be corrupted by colored noise. The proposed identification algorithm is based on the Linear Time-Varying (LTV) identification algorithm described in [7]. A first key difference is that the time-varying coefficients are now replaced by functions of the scheduling parameter  $p(t)$ .

$$\sum_{n=0}^{N_a} a_n(p(t)) \frac{d^n y_0(t)}{dt^n} = \sum_{n=0}^{N_b} b_n(p(t)) \frac{d^n u_0(t)}{dt^n} \quad (1)$$

where the subscript of  $x_0$  denotes a noiseless quantity. The coefficients of the LPV Input-Output (IO) model (1) are approximated by linear combinations of known/chosen basis functions in  $p(t)$ , viz.

$$\begin{bmatrix} a_n(p(t)) \\ b_n(p(t)) \end{bmatrix} = \sum_{i=0}^{N_p} \begin{bmatrix} a_{[n,i]} \\ b_{[n,i]} \end{bmatrix} \phi_i(p(t)) \quad (2)$$

---

<sup>★</sup> This paper was not presented at any IFAC meeting.

Email addresses: jan.goos@vub.ac.be (Jan Goos), john.lataire@vub.ac.be (John Lataire), ebrahim.louarroudi@vub.ac.be (Ebrahim Louarroudi), rik.pintelon@vub.ac.be (Rik Pintelon).

The results in this paper also hold if the scheduling parameter is multivariable. As an alternative to (2), Support Vector Machines (SVMs) [11] or Gaussian Processes (GPs) can be used to model the coefficient functions.

A second key difference with [7], is that the full covariance matrix is used to weigh the residual errors, to ensure consistency. The consistency and correctness of the proposed Linear Parameter-Varying Input-Output (IO) estimator is proven, and illustrated on a simulation example. The results hold for arbitrary non-steady-state, non-periodic data. Even though the identification problem is considered in the frequency domain, the input  $u(t)$  and scheduling  $p(t)$  do not have to be periodic.

## 2 The sampled LPV differential equation in the frequency domain

The Fourier transform of (1) will be computed from the measured time domain signals, which are sampled uniformly at a sample frequency  $f_s = 1/T_s$  ( $T_s$  is the sample time). A total of  $N$  samples of each signal is acquired.

**Definition 1** The Discrete Fourier Transform (DFT), at the angular frequencies  $\omega_k = 2\pi k f_0 \forall k \in [0, N-1]$ , with  $f_0 = f_s/N$ , is defined as

$$\text{DFT} \{x(nT_s)\} = X(k) = \sum_{n=0}^{N-1} x(nT_s) e^{-j\omega_k n T_s} \quad (3)$$

**Definition 2** Similarly, the inverse Discrete Fourier Transform (iDFT) is defined as

$$\text{iDFT} \{X(k)\} = \frac{1}{N} \sum_{k=0}^{N-1} X(k) e^{j\omega_k n T_s}. \quad (4)$$

We denote the DFTs of the input  $u(t)$  and the output  $y(t)$  with  $U(k)$  and  $Y(k)$  respectively.

**Assumption 3** Band-limited excitation: the Fourier transforms of the true input and scheduling signals are zero beyond the Nyquist frequency:  $|U_0(j\omega_k)| = 0$  and  $|P_0(j\omega_k)| = 0$  for  $k f_0 \geq f_{nyq} = f_s/2$ . Furthermore, the basis functions are band-limited w.r.t  $f_{nyq}$ .

The continuous time signals are windowed, because only the time frame  $[0, T]$  is considered. If a rectangular window  $w(t)$  is used, the Fourier transform becomes

$$\begin{aligned} \mathcal{F} \left\{ w(t) \frac{d^n x(t)}{dt^n} \right\} \Big|_{j\omega_k} &= (j\omega_k)^n X(j\omega_k) \\ &+ \underbrace{\sum_{r=0}^{n-1} (j\omega_k)^{n-1-r} \left( x^{(r)}(T_-) - x^{(r)}(0_+) \right)}_{=T_x^n(j\omega_k)} \end{aligned} \quad (5)$$

where  $x^{(r)}$  is the  $r^{\text{th}}$  time derivative of  $x$ ,  $X(j\omega_k)$  is the Fourier transform of  $x(t)$  and  $\omega_k = 2\pi k f_0$  is the angular frequency  $\forall k \in [0, N-1]$ . Equation (5) is proven in Appendix 5.B of [14] using integration by parts. Note that the difference between the initial and end conditions of the signal determine the polynomial  $T_x^n(j\omega_k)$ . By including the transient term  $T_x^n(j\omega_k)$  in (5), the derivatives of arbitrary signals can be represented exactly in the frequency domain.

### 2.1 Frequency domain model

Taking the affine approximation (2) into account, the DFT of the sampled and windowed (1) equals

$$\begin{aligned} &\sum_{n=0}^{N_a} \sum_{i=0}^{N_p} a_{[n,i]} \Phi_i\{p\} * [(j\omega_k)^n Y(k) + T_y^n(j\omega_k)] \\ &= \sum_{n=0}^{N_b} \sum_{i=0}^{N_p} b_{[n,i]} \Phi_i\{p\} * [(j\omega_k)^n U(k) + T_u^n(j\omega_k)] \end{aligned} \quad (6)$$

where  $\Phi_i\{p\} = \text{DFT} \{\phi_i(p(t))\}$  are the DFT basis functions of the scheduling parameter, and  $*$  represents the circular convolution product.

**Assumption 4** Weierstrass approximation theorem: the basis functions  $\phi(p(t))$  can be approximated by a polynomial in  $t$  of degree  $m$  in the finite interval  $t \in [0, T]$ .

**Assumption 5** The basis functions  $\phi(p(t))$  are periodic in the time window  $T$ , and can therefore be represented exactly by their Fourier series over the interval  $t \in [0, T]$ .

**Theorem 6** Assumptions 4 and 5 are alternative. If either one of them holds, the convolution of the basis functions and the polynomials  $\Phi_i\{p\} * T_y^n(j\omega_k)$  and  $\Phi_i\{p\} * T_u^n(j\omega_k)$  are also polynomials in  $j\omega_k$ . For the proof, we refer to Appendix A.

**Corollary 7** The polynomials  $T_y^n(j\omega_k)$  and  $T_u^n(j\omega_k)$  can be extracted from the summation in (6), and grouped into one transient polynomial  $T_{uy}^{N_T}(j\omega_k) = \sum_{i=0}^{N_T} \gamma_i (j\omega_k)^i$ , of order  $N_T = \max\{N_a, N_b\} - 1$ .

$$\begin{aligned} T_{uy}^{N_T}(j\omega_k) &= - \sum_{n=0}^{N_a} \sum_{i=0}^{N_p} a_{[n,i]} \Phi_i\{p\} * T_y^n(j\omega_k) \\ &+ \sum_{n=0}^{N_b} \sum_{i=0}^{N_p} b_{[n,i]} \Phi_i\{p\} * T_u^n(j\omega_k) \end{aligned} \quad (7)$$

Computing derivatives of the signals is exact in the frequency domain. On the other hand, the computationally complex convolution can be avoided by multiplying the signals in the time domain. This results in the following computationally efficient frequency domain model:

$$\begin{aligned}
& \sum_{n=0}^{N_a} \sum_{i=0}^{N_p} a_{[n,i]} \text{DFT} \{ \phi_i(p(t)) \text{iDFT} \{ (j\omega_k)^n Y(k) \} \} \\
&= \sum_{n=0}^{N_b} \sum_{i=0}^{N_p} b_{[n,i]} \text{DFT} \{ \phi_i(p(t)) \text{iDFT} \{ (j\omega_k)^n U(k) \} \} \\
&+ \sum_{i=0}^{N_T} \gamma_i (j\omega_k)^i
\end{aligned} \tag{8}$$

where  $T_{uy}^{N_T}(j\omega_k)$  is defined in (7). Note that the proposed model (8) is linear in the parameters  $a_{[n,i]}$  and  $b_{[n,i]}$ .

## 2.2 Aliasing in the LPV output spectrum

It is well-known that the steady state response of an LTI system to a single sine wave  $\sin(\omega_u t)$ , is a sine wave at the same frequency, but with a changed magnitude and phase [12,14]. In the output spectrum of an LPV system, energy appears at frequencies that were not directly excited. These extra spectral components, or *harmonic frequencies*, appear at integer multiples of the basic scheduling frequency  $\omega_p$ , around the excited frequencies  $\omega_u$ :  $\omega_u + k\omega_p$ . In theory, these *skirt-like* shapes, shown in Figure 1, extend infinitely far. In fact, Appendix B shows that they have a hyperbolic decay. In practice, the skirts will eventually disappear below the noise level, possibly beyond the Nyquist frequency  $f_{\text{nyq}} = f_s/2$ . In the sampled frequency domain, this frequency content will fold back and cause *aliasing*. Note however, that the harmonic components decay smoothly over the frequencies, as seen on the right of Figure 1.

**Assumption 8** *The aliasing contributions of the harmonics can be approximated by a polynomial in  $j\omega_k$ .*

**Remark 9** *In case of aliasing, the equality in Corollary 7 becomes an inequality:  $N_T \geq \max\{N_a, N_b\} - 1$ .*

Even more, windowing a non-periodic signal results in leakage terms [13], which fold back over the Nyquist frequency. If the highest excited frequency is too close to  $f_{\text{nyq}}$ , then this leakage can cause a significant aliasing effect, but it varies smoothly over the frequency. Again, by increasing the degree  $N_T$  of the additional polynomial  $T_{uy}^{N_T}(j\omega_k)$ , this aliasing term can be captured.

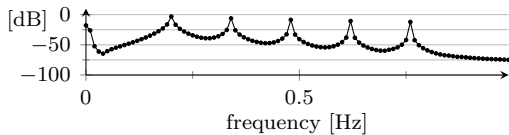


Figure 1. Output spectrum (•) from a sparsely excited LPV model. The parameter variation creates spectral content around the excited frequencies. These skirts theoretically extend beyond the Nyquist frequency, resulting in aliasing.

## 2.3 Noise assumptions

**Assumption 10** *The input and output signals  $u(t)$  and  $y(t)$  are noisy observations:  $x(t) = x_0(t) + v_x(t)$  with  $x_0(t)$  the true unknown value and  $v_x(t)$  stationary (band-limited) filtered white noise with finite  $m^{\text{th}}$  order moments. Furthermore, let the noisy part  $v_x(t)$  be uncorrelated with the true signal  $x_0(t)$ .*

**Corollary 11** *From Assumption 10, it holds asymptotically ( $N \rightarrow \infty$ ) that the noise  $V_U(k)$  and  $V_Y(k)$  on the input and output spectra  $U(k)$  and  $Y(k)$  is circular complex, normally distributed and uncorrelated over the frequency:*

$$C_U(k,l) = \mathbb{E} \{ V_U(k), \overline{V_U(l)} \} = \delta_{k,l} \sigma_U^2(k) \tag{9}$$

$$C_Y(k,l) = \mathbb{E} \{ V_Y(k), \overline{V_Y(l)} \} = \delta_{k,l} \sigma_Y^2(k) \tag{10}$$

$$C_{UY}(k,l) = \mathbb{E} \{ V_U(k), \overline{V_Y(l)} \} = \delta_{k,l} \sigma_{UY}^2(k) \tag{11}$$

where  $\delta_{k,l} = 1$  only if  $k = l$  and is zero otherwise, and  $\bar{x}$  denotes the complex conjugate of  $x$ .

In some cases, the measurement covariances  $\sigma_U^2(k)$  and  $\sigma_Y^2(k)$  are known on beforehand. Otherwise, they have to be estimated from the data. A non-parametric noise model in the frequency domain can be estimated easily if multiple periods of the input and scheduling are observed [15]. The extension to arbitrary inputs and smooth parameter variations is more involved [6].

**Assumption 12** *The scheduling parameters  $p(t)$  are known exactly.*

## 3 LPV model estimation

In this section we propose an estimator for the model parameters  $a_{[n,i]}$  and  $b_{[n,i]}$  in the differential equation (8). If the measured input and output signals are noisy, the equation cannot hold exactly. A “good” model minimizes the residual error between both sides of the equation. Let us first recapitulate why the differential equation is treated in the frequency domain.

- (1) In the band-limited setting of Assumption 3, the time derivatives can be computed exactly.
- (2) It is straightforward to select only the frequency band of interest.
- (3) It is possible to obtain a non-parametric noise model, which can be used to weigh the residuals in the cost function.

### 3.1 Equation error

The column vector  $\theta$  stacks all the model parameters  $a_{[n,i]}$ ,  $b_{[n,i]}$  and  $\gamma_i$  vertically:

$$\theta = \begin{bmatrix} a_{[0,0]}, a_{[0,1]}, \dots, a_{[n,i]}, \dots, a_{[N_a, N_p]}, \dots \\ b_{[0,0]}, b_{[0,1]}, \dots, b_{[n,i]}, \dots, b_{[N_b, N_p]}, \dots \\ \gamma_0, \gamma_1, \dots, \gamma_{N_T} \end{bmatrix}^T \quad (12)$$

Now, for a given model with parameters  $\theta$  and data  $z$ , the equation error  $e(\theta, z)$  is defined at each frequency as the difference between the left and right side of the differential equation in the frequency domain. If a set of basis functions  $\phi(p(t))$  is chosen to model the varying coefficients, (8) can be written as

$$\mathbf{K} \theta = e(\theta, z) \approx 0 \quad (13)$$

The regressor matrix  $\mathbf{K} \in \mathbb{C}^{[N \times (N_a + N_b + 2)(N_p + 1) + N_T + 1]}$  gathers the convolved terms  $\mathbf{K}_Y$  and  $\mathbf{K}_U$ , complemented with the monomials from the transient term.

$$\mathbf{K} = \begin{bmatrix} \mathbf{K}_Y & -\mathbf{K}_U & -\boldsymbol{\Omega} \end{bmatrix} \quad (14)$$

$$\begin{aligned} \mathbf{K}_Y(:, nN_p + i + 1) &= \text{DFT} \{ \phi_i(p(t)) \text{iDFT} \{ (j\omega_k)^n Y(k) \} \} \\ \mathbf{K}_U(:, nN_p + i + 1) &= \text{DFT} \{ \phi_i(p(t)) \text{iDFT} \{ (j\omega_k)^n U(k) \} \} \\ \boldsymbol{\Omega}(k, :) &= \begin{bmatrix} 1 & (j\omega_k) & \dots & (j\omega_k)^{N_T} \end{bmatrix} \end{aligned} \quad (15)$$

The indices follow the notation of the frequency domain differential equation (8):  $i \in [0, N_p]$ ,  $n \in [0, N_a]$  in  $\mathbf{K}_Y$  and  $n \in [0, N_b]$  in  $\mathbf{K}_U$ . Each row  $k$  in  $\mathbf{K}$  corresponds to a frequency  $\omega_k = 2\pi k f_0 \forall k \in [0, N - 1]$ .

### 3.2 Weighted Nonlinear Least Squares Estimator

The Weighted Nonlinear Least Squares Estimator (WNLSE) minimizes the magnitude of the squared complex residuals in (13), weighted with its covariance matrix  $C_e(\theta)$ . The latter can be computed from the input-output noise covariances, as illustrated in Appendix C.1.

**Remark 13** *The noise on the input and output is assumed to be circular complex, normally distributed and uncorrelated over the frequency (Corollary 11). In this case, the equation error is also normally distributed, but no longer circular complex. Additionally, the time-variation introduces a correlation over the frequency.*

To cope with Remark 13, the complex equation error is decomposed in its real and imaginary part, and the corresponding covariance matrix  $C_{\text{Re}}(\theta)$  is computed, as discussed in Appendix C.2.

$$X_{\text{re}} = \begin{bmatrix} \text{Re} \{X\} \\ \text{Im} \{X\} \end{bmatrix}, \quad X_{\text{Re}} = \begin{bmatrix} \text{Re} \{X\} & -\text{Im} \{X\} \\ \text{Im} \{X\} & \text{Re} \{X\} \end{bmatrix} \quad (16)$$

The Weighted Nonlinear Least Squares (WNLS) cost function is now defined as

$$V_{\text{WNLS}}(\theta, z) = e_{\text{re}}(\theta, z)^T C_{\text{Re}}^{-1}(\theta) e_{\text{re}}(\theta, z) \quad (17)$$

$$\hat{\theta}_{\text{WNLS}} = \underset{\theta}{\text{argmin}} V_{\text{WNLS}}(\theta, z) \quad (18)$$

The equation error covariance matrix  $C_{\text{Re}}(\theta)$  is symmetric and positive definite, which allows a decomposition  $C_{\text{Re}}(\theta)^{-1} = W(\theta)^T W(\theta)$ , as explained in Appendix C.3. The WNLS cost function (17) can thus be rewritten as

$$\begin{aligned} V_{\text{WNLS}}(\theta, z) &= e_{\text{re}}(\theta, z)^T W(\theta)^T W(\theta) e_{\text{re}}(\theta, z) \\ &= \varepsilon(\theta, z)^T \varepsilon(\theta, z) \end{aligned} \quad (19)$$

where  $\varepsilon(\theta, z)$  denotes the weighted equation error, decomposed in its real and imaginary components. **In the frequency domain representation (13), it is easy to consider only a specific frequency band, by selecting only the appropriate rows in the regressor matrix  $\mathbf{K}$ . However, unlike in the LTI case, the computation of the equation error requires all frequencies where the spectrum might contain signal energy.**

### 3.3 Nonlinear optimization problem

The WNLS cost function (18) is nonlinear in the model parameters. To this end, an iterative Gauss-Newton method is employed, which requires the Jacobian  $J(\theta, z) = \frac{\partial \varepsilon(\theta, z)}{\partial \theta}$  of the weighted equation error (19) to the model coefficients  $\theta$ , given the data  $z$ . At each iteration  $i$ , a correction term  $\Delta\theta^{(i)}$  is given by solving the overdetermined set of equations

$$J(\theta^{(i-1)}, z) \Delta\theta^{(i)} = -\varepsilon(\theta^{(i-1)}, z) \quad (20)$$

in least square sense. For details, we refer to Chapter 9.4 in [14]. Solving the iterative minimization step (20) requires an assumption on the system, the input and the scheduling trajectory.

**Assumption 14** *Persistency of excitation and scheduling condition: the rank of the Jacobian matrix  $J$  should be  $(N_a + N_b + 2)(N_p + 1) + N_T$ , which is the number of effective parameters.*

Because the identification problem is nonlinear in the parameters, a good initial estimate is needed. **For example, in [7], the Total Least Squares estimator [20] is used for initialization.** To expand the convergence region of the Newton-Gauss method, a Levenberg-Marquard [4] method is used in the parameter update step (20).

### 3.4 Statistical properties of the WNLSE

Because of the weighting with the covariance of the equation error, the Weighted Nonlinear Least Squares Estimator (18) is consistent, which is proven in this section

for an increasing number of independent experiments  $N_{\text{exp}} \rightarrow \infty$ . This consistency definition differs from the usual LTI interpretation, because of the conflict with Assumption 4, stating that the measurement of an arbitrary (but smooth) parameter-varying system must happen over a finite time interval  $t \in [0, T]$ .

**Assumption 15** *The true model is in the model class (1), covering the dynamical order  $n_x = \max\{N_a, N_b\}$  of the system, and the proposed basis functions  $\phi_i(p(t))$ .*

**Assumption 16** *The cost function  $V_{\text{WNLS}}(\theta, Z)$  in (17) and its derivative are continuous in a compact set around the true model parameters  $\theta_0$ .*

**Assumption 17** *Multiple experiments are available, where each experiment has similar experimental conditions, satisfying Assumption 14.*

- (1) *the RMS of each input lies in a specified range.*
- (2) *the excited frequency band of the input is similar for all experiments*
- (3) *the scheduling trajectory is limited between a given minimal and maximal value.*

In this framework, it becomes possible to add some measurements, where the scheduling parameter is fixed to a constant value, effectively adding an LTI experiment. Whether or not this is practically feasible, depends on the application at hand.

**Theorem 18** *Under Assumptions 3, 10 ( $m = 4$ ), 12, 14, 15, 16 and 17, the WNLS estimator (18) is strongly consistent (for  $N_{\text{exp}} \rightarrow \infty$ ). If in addition Assumption 10 ( $m = \infty$ ) holds, then (18) is asymptotically normally distributed.*

Each independent experiment has different initial conditions, which are not consistently estimated. However, following Theorem 18, it follows that this does not affect the consistency of  $a_{[n,i]}$  and  $b_{[n,i]}$ . The practical limit of the WNLS estimator (18) is determined by the inversion of the equation error covariance matrix  $C_e(\theta)^{-1} \in \mathbb{C}^{N \times N}$ , suggesting the use of short data records. However, under Assumptions 4 and 5,  $C_e(\theta)$  is band-dominant (and Hermitian positive definite), and the inverse can still be computed for moderately sized measurements.

As discussed in Appendix E.3, in the periodic case, it is also possible to prove consistency for an increasing measurement time. Additional periods allows to average out the noise, increasing the accuracy of the estimate.

## 4 Simulation results

To demonstrate the properties of the proposed identification algorithm, let us study a fourth order simulation example. The LPV differential equation is given by

$$\begin{aligned} & y^{(4)} + 0.220y^{(3)} + (33.561 + 11.310p + p^2)y^{(2)} \\ & + (0.955 + 0.226p + 0.020p^2)y^{(1)} \\ & + (50.494 + 17.8591p + 50.494p^2)y \\ & = u^{(2)} + 0.020u^{(1)} + (5.565 + p)u \end{aligned} \quad (21)$$

Both the input and output dynamics are dependent on  $p$  and  $p^2$ . Figure 2 shows some frozen frequency response functions (FRFs), where the scheduling signal is kept constant  $p(t) = p$ . The system has two resonance peaks, and one antiresonance. The second resonance moves to a higher frequency as  $p$  is increased.

A single experiment was used to identify the coefficients of the given model structure (21). The sampling frequency is  $f_s = 3$  Hz, and  $N = 2000$  samples are taken. Therefore, the frequency resolution is  $f_0 = 0.0015$  Hz, and the time window  $T \simeq 666.66$  s. The experiment is global, in the sense that the scheduling parameter is varying during the experiment, as shown in Figure 2.

$$p(t) = \frac{3}{5}\pi \cos(2\pi f_0 t) + \frac{2}{5}\pi \sin(4\pi f_0 t) \quad (22)$$

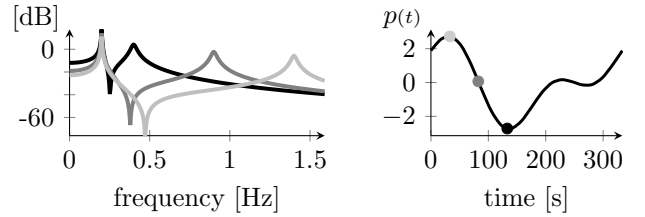


Figure 2. (left) Frozen LTI frequency response functions for low, mid and high values of the scheduling signal. (right) Scheduling trajectory in the identification experiment.

Figure 3 shows the input and output signals in the time domain. The model (21) is excited with a sum of random phase sines. Only half of the available frequency band ( $f_{\text{nyq}} = f_s/2 = 1.5$  Hz) is excited, as illustrated in Figure 4. Recall from Section 3.2 that a frequency band of interest can be selected in the WNLS cost function (17). Also, the further the excited frequencies are from the Nyquist frequency  $f_{\text{nyq}}$ , the smoother the contribution of the hyperbolic skirts due to the aliasing effect becomes, as discussed in Section 2.2.

In this simulation setup, the model structure is assumed to be known. The dynamic orders (1) are  $N_a = 3$ ,  $N_b = 2$ , the polynomial dependency on the scheduling (2) is of degree  $N_p = 2$ . As stated in Assumption 10, both the input and output measurements are disturbed by additive stationary filtered white noise, as shown in Figure 4. The signal-to-noise ratio of the measured input and output RMS is 10 dB. The covariances are estimated from the periodic, but non-steady-state data [15].

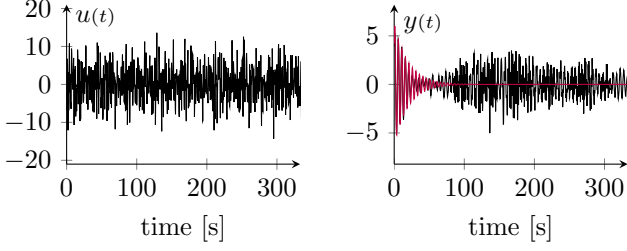


Figure 3. Input  $u(t)$  and output  $y(t)$  of the identification experiment. From the difference (—) between a second and the first period of the output  $y(t)$ , we can see that the system is not in steady state. Therefore, a transient term will be needed in the frequency domain representation (8).

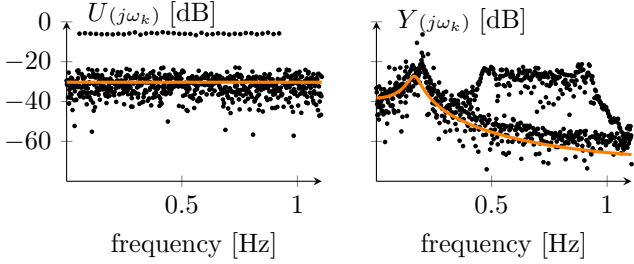


Figure 4. Input  $U(j\omega_k)$  and output  $Y(j\omega_k)$  spectra (•) of the identification experiment. Only a limited frequency band is excited at the input  $U(j\omega_k)$ , where the system dynamics are important. Transient and alias effects are present at the output  $Y(j\omega_k)$ , but they are hidden below the noise (—), or mixed with the output signal.

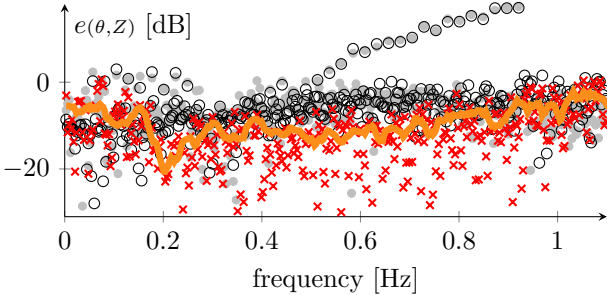


Figure 5. The left (○) and right (●) hand sides of the differential equation (8) lie on top of each other. The equation residuals  $e$  (×), and the predicted standard deviation (—) are shown at each frequency. Since  $e$  (×) is comparable to the expected noise standard deviation (—), no bias can be detected.

The order  $N_T$  of the additional polynomial has to be estimated from even if  $N_a$ ,  $N_b$  and  $N_p$  are known. In practice,  $N_a$ ,  $N_b$  and  $N_p$  should be determined from the data, which is an order selection problem. In the given simulation setup,  $N_T = 5$  captures the transient and alias effect well. A higher degree does not decrease the cost function (17), and a lower degree introduces larger modeling errors  $e(\theta, Z)$ . If the residual equation error  $e$  is normally distributed (Theorem 18), a confidence interval can be constructed using the equation error covari-

ance matrix  $C_e$  computed in Appendix C. If  $e$  does not coincide with the confidence interval, there may be some unmodeled dynamics, or the order of the transient is not high enough. In Figure 5, the 63% confidence bound corresponds well to the residual equation error (36% lies beyond the confidence range), so no bias can be detected, which suggests the estimate is consistent.

#### 4.1 Monte Carlo simulations

To further illustrate the consistency properties of the proposed estimator, two Monte Carlo runs are performed. In each dataset, the same experimental conditions are used, as explained earlier in Section 4. Every simulation has a different noise realization. The results are compared by studying an invariant of the model. In this case, the frozen Frequency Response Function is chosen, which was used in Figure 2 to give an impression of the varying dynamics. Given the different identified frozen FRF<sup>[i]</sup>s, the Mean Squared Error (MSE) can be computed as follows:

$$\text{MSE}(f, t) = \frac{1}{N_{\text{exp}}} \sum_{i=1}^{N_{\text{exp}}} |\text{FRF}^{[i]} - \text{FRF}_0|^2 \quad (23)$$

In the first set of Monte Carlo simulations,  $N_{\text{exp}} = 10$  different realizations are identified independently. In the second Monte Carlo run,  $N_{\text{exp}} = 40$  different realizations are identified, but they are gathered in groups of four simulations. The four corresponding cost functions are combined, by stacking the weighted equation errors (E.11). The following table shows the RMS value of the errors on the resulting frozen FRFs.

|     | $N_{\text{exp}} = 10$ | $N_{\text{exp}} = 40$ (grouped per 4) |
|-----|-----------------------|---------------------------------------|
| RMS | 3.73e-6               | 1.61e-6                               |

Because the amount of data was increased by a factor four, the RMS on the estimated frozen FRFs drops by a factor two (6 dB). This indicates that the  $(N_a + N_b + 2)(N_p + 1) = 24$  dynamic coefficients are estimated consistently, even when  $4(N_T + 1) = 24$  transient coefficients have to be identified simultaneously. The latter estimates are inconsistent. Therefore, the Monte Carlo simulations validate the presented proof of consistency: adding additional measurement records improves the identification of the LPV model parameters.

## 5 Conclusion

In this paper, a consistent estimator for Linear Parameter-Varying differential equations is presented. Both arbitrary and periodic variations of the scheduling parameters are handled. As with most LPV identification tools, the choice in basis functions is important in the model selection step. It is possible to incorporate prior knowledge, like known dynamic effects on the

scheduling signal, in the basis  $\phi(p(t))$ . Consistency is proven for an increasing number of similar (but not identical) experiments. From the proof, it follows that the full covariance matrix of the equation error is needed. The latter depends on the input-output covariances, that can be estimated non-parametrically from the data.

## Acknowledgements

This work was supported in part by the Fund for Scientific Research (FWO-Vlaanderen), by the Flemish Government (Methusalem Fund, METH1), and by the Belgian Federal Government (IUAP VII, DYSCO).

## A Extracting the polynomial transient term

### A.1 The basis functions are polynomials in $t$

Following Assumption 4, the coefficient functions  $a_n(p(t))$  and  $b_n(p(t))$  can be approximated with a polynomial in  $t$ . Multiplying a signal  $x(t)$  by  $t^n$  corresponds to an  $n^{\text{th}}$  order derivative in the frequency domain.

$$\mathcal{F}\{t^n x(t)\} = (-1)^n \frac{d}{d j\omega_k} \mathcal{F}\{x(t)\} \quad (\text{A.1})$$

The  $n^{\text{th}}$  derivative of a polynomial of degree  $N$  remains a polynomial, but with degree  $\max\{N - n, 0\}$ . The total sum of polynomials in (8), therefore, results in a transient polynomial  $T_{yu}^{N_T}$  of degree  $\max\{N_a, N_b\} - 1$ .

### A.2 The basis functions are periodic in $t$

Because the Fourier transform is linear, we only consider the convolution with a single sinusoidal signal. It is well known that such a signal consists of two complex conjugate Dirac impulses in the frequency domain:

$$\mathcal{F}\{a \sin(\omega_k t + \psi)\} = (\alpha + j\beta)\delta(\omega - \omega_k) + (\alpha - j\beta)\delta(\omega + \omega_k).$$

Denote the Laplace variable  $s = j\omega$ . Convolving the Dirac impulses with a transient polynomial  $T_x^n(s)$  yields two shifted polynomials, and the sum of both terms yields a polynomial of order  $n$  with real coefficients. Using the binomial of Newton, the convolution becomes

$$\begin{aligned} & [(\alpha + j\beta)\delta(\omega - \omega_k) + (\alpha - j\beta)\delta(\omega + \omega_k)] * \gamma_n s^n \quad (\text{A.2}) \\ &= (\alpha + j\beta)\gamma_n (s - j\omega_k)^n + (\alpha - j\beta)\gamma_n (s + j\omega_k)^n \\ &= \sum_{r=0}^n \binom{n}{r} \gamma_n \omega_k^r s^{n-r} 2 \operatorname{Re} \left\{ (-j)^r \alpha + j(-j)^r \beta \right\} \end{aligned}$$

For all values of  $r \in \mathbb{N}_+$ , the coefficients are real, and the degree of the resulting polynomial in  $s$  is of degree  $n$ , which concludes the proof.

## B Smooth harmonics at high frequencies

This section proves that the harmonics are smooth functions of the frequency, specifically beyond the Nyquist frequency  $f_s/2$ . Because of the sampling, this alias contribution folds back in the observed frequency range. Since the alias error is smooth over the frequency, it can be modeled by a polynomial in  $j\omega_k$ , justifying Assumption 8. We consider both polynomial and periodic basis functions  $\phi_i(p(t))$ , as discussed in Assumptions 4 and 5. The basic idea is the same for both cases. The response of the parameter-varying system, for a specific trajectory  $p(t)$ , is first rewritten as a general convolution integral [17]

$$y(t) = \int_0^\infty g(t, \tau) u(t) d\tau, \quad (\text{B.1})$$

The convolution model (B.1) is then approximated with a parallel structure, obtained via a specific series expansion of the *Time-Variant Transfer Function* (TV-TF), which is the Fourier transform of the time-variant impulse response.

$$G(j\omega, t) = \int_0^\infty g(t, t - \tau) e^{-j\omega\tau} d\tau \quad (\text{B.2})$$

$$= \sum_{i=0}^\infty G_i(j\omega) \psi_i(t) \quad t \in [0, T] \quad (\text{B.3})$$

with  $\psi_i(t)$  a complete set of basis functions, e.g. polynomials (arbitrary time-variation) or sines and cosines (periodic time-variation). In practice, a finite number  $N_h$  of branches is used, and the output is computed as a sum of time-weighted LTI responses

$$y(t) \simeq \sum_{i=0}^{N_h} \mathcal{F}^{-1} \{G_i(j\omega) U(j\omega)\} \psi_i(t) \quad (\text{B.4})$$

$$= \sum_{i=0}^{N_h} y_i(t) \psi_i(t) \quad (\text{B.5})$$

### B.1 Parallel structure using polynomial basis functions

In the case of (smooth) arbitrary time variations, a polynomial basis is used, and (B.6) corresponds to a truncated Taylor series of the TV-TF (B.2).

$$y(t) = \sum_{i=0}^{N_h} \mathcal{F}^{-1} \{G_i(j\omega) U(j\omega)\} t^i \quad t \in [0, T] \quad (\text{B.6})$$

The Fourier transform of a windowed monomial  $t^r$  yields a sum of hyperbolas in the frequency domain.

$$\mathcal{F}\{t^r w(t)\} \big|_{j\omega_k} = \begin{cases} -\sum_{n=1}^r \frac{1}{(j\omega_k)^n} \frac{r!}{(r-n+1)!} \frac{T^{r-n+1}}{r+1} & \forall k \neq 0 \\ \frac{T^{r+1}}{r+1} & k = 0 \end{cases}$$

Because the input signal is band-limited (Assumption 3),  $G_i(j\omega)U(j\omega)$  will be band-limited as well. The multiplication with the basis functions in (B.6) becomes a convolution in the frequency domain, resulting in hyperbolas, centered around the excited frequencies. For sufficiently large frequencies  $f > f_s/2$ , the contributions of these hyperbolas are all smooth<sup>1</sup>, and can be approximated well by a polynomial in  $j\omega_k$ .

**Assumption 19** *The output of the parameter-varying system can, for the considered trajectory of  $p(t)$ , be approximated by the truncated Taylor series (B.6).*

The sum of a finite amount of shifted and scaled hyperbolas is smooth for sufficiently large frequencies. Therefore, the alias error can be approximated arbitrarily well by a polynomial  $T_{uy}^{N_T}(j\omega_k)$ . Note that the degree of this polynomial can be higher than the degree of the polynomial, required to model the transient behavior in Corollary 7, which leads to Remark 9:  $N_T \geq \max\{N_a, N_b\} - 1$ .

### B.2 Parallel structure using periodic basis functions

For periodic parameter variations, (B.1) is approximated using the Fourier series expansion of the TV-TF (B.2).

$$G_{N_h}(j\omega, t) = \sum_{k=-N_h}^{N_h} G_k(j\omega) e^{j\omega_k t} \quad (\text{B.7})$$

Provided that the time-variant impulse response  $g(t, \tau)$  has a uniform exponential decay and is a smooth function of order  $L$ , it is proven in [1] that the  $G_k(j\omega)$ s are hyperbolas in the frequency domain.

$$G_k(j\omega) = \sum_{r=1}^{n_a} \sum_{n=-N_h}^{N_h} \frac{\alpha_{r,n}}{j\omega - (\lambda_r - j2\pi n f_0)} + D$$

In the frequency domain, the multiplication with the periodic basis functions  $e^{j\omega_k t}$  only shift the transfer functions. For sufficiently large frequencies  $f > f_s/2$ , the finite sum of hyperbolas yields a smooth function. Therefore, the alias error can be approximated arbitrarily well by a polynomial in  $j\omega_k$ .

## C Computing the equation error covariance matrix from input-output noise

### C.1 The complex equation error covariance matrix

The differential equation in the frequency domain (8) can be rewritten as

$$e = \mathbf{A}Y - \mathbf{B}U - T_{uy}^{N_T} \quad (\text{C.1})$$

<sup>1</sup> A finite sum of smooth functions remains smooth.

where  $Y$  and  $U$  are the DFT spectra of the output and input, and  $\mathbf{A}$  and  $\mathbf{B} \in \mathbb{C}^{N \times N}$  are complex matrices:

$$\mathbf{A} = \sum_{n=0}^{N_a} \sum_{i=0}^{N_p} a_{[n,i]} \begin{bmatrix} \text{DFT} & \phi_i(p) & \text{iDFT} & (j\omega_k)^n \end{bmatrix} \quad (\text{C.2})$$

$$\mathbf{B} = \sum_{n=0}^{N_b} \sum_{i=0}^{N_p} b_{[n,i]} \begin{bmatrix} \text{DFT} & \phi_i(p) & \text{iDFT} & (j\omega_k)^n \end{bmatrix} \quad (\text{C.3})$$

The boxes represent matrices of dimension  $N \times N$ . Indeed, the DFT and iDFT operators can be written as multiplication by a square, complex matrix. The matrices containing the basis functions  $\phi(t)$  and the powers  $(j\omega_k)^n$  are diagonal.

Then, the covariance matrix of the equation error  $C_e(\theta) \in \mathbb{C}^{N \times N}$  can be computed as follows:

$$C_e(\theta) = \begin{bmatrix} \mathbf{A} & -\mathbf{B} \end{bmatrix} \begin{bmatrix} C_Y & C_{YU} \\ C_{YU}^H & C_U \end{bmatrix} \begin{bmatrix} \mathbf{A}^H \\ -\mathbf{B}^H \end{bmatrix} \quad (\text{C.4})$$

or

$$C_e(\theta) = \mathbf{A}C_Y\mathbf{A}^H + \mathbf{B}C_U\mathbf{B}^H - 2\text{Herm}\{\mathbf{A}C_{YU}\mathbf{B}^H\}$$

with the Hermitian operator defined as  $\text{Herm}\{X\} = \frac{X+X^H}{2}$ . The transient term does not appear in (C.4), because it does not contain the measured signals. The equation error  $e$ , and therefore  $C_e$ , depends directly on the model parameters  $a_{[n,i]}$  and  $b_{[n,i]}$ . The diagonal of the covariance matrix is interpreted as the variance of the equation error, at each individual frequency  $k f_0$ . The non-diagonal elements of  $C_e$  represent the covariance of the equation error  $e$  over different frequencies. Note that these elements are non-zero, even if the disturbing noise is uncorrelated over the frequency. This is because, in the general case, the matrices  $\mathbf{A}$  and  $\mathbf{B}$  are full. Thus, for time-varying systems, the equation error is correlated over the frequency.

### C.2 The real-imaginary equation error covariance

To cope with the fact that the equation error  $e$  is not circular complex (Remark 13), it is decomposed into its real and imaginary part. This approach ensures that the model parameters are real-valued as well. Using the decomposition (16), the equation error (C.3) becomes

$$e_{\text{re}} = \mathbf{A}_{\text{Re}}Y_{\text{re}} - \mathbf{B}_{\text{Re}}U_{\text{re}} - \mathbf{\Omega}_{\text{Re}}\theta_{\text{poly}} \quad (\text{C.5})$$

The covariance matrix  $C_{\text{Re}}(\theta) \in \mathbb{R}^{2N \times 2N}$  of these real and imaginary parts can be computed as in (C.4).

$$C_{\text{Re}}(\theta) = \frac{1}{2} \begin{bmatrix} \mathbf{A}_{\text{Re}} & -\mathbf{B}_{\text{Re}} \end{bmatrix} \begin{bmatrix} C_{Y_{\text{Re}}} & C_{Y_{\text{Re}}U_{\text{re}}} \\ C_{Y_{\text{Re}}U_{\text{re}}}^H & C_{U_{\text{re}}} \end{bmatrix} \begin{bmatrix} \mathbf{A}_{\text{Re}}^H \\ -\mathbf{B}_{\text{Re}}^H \end{bmatrix}$$

Assumption 10, and subsequently Corollary 11, state that the measurement noise is circular complex, which implies that the real and imaginary components have the same real and diagonal covariance matrix. The cross-correlation matrix  $C_{YU}$  is complex, so it does have a real and imaginary blocks, as in (16). The resulting covariance matrix of the real and imaginary equation error depends on the parameter variation  $p(t)$ , but it is generally block band dominant. In case of a periodic scheduling trajectory,  $C_{\text{Re}}(\theta)$  becomes block band diagonal.

### C.3 Inverting a covariance matrix

In the WNLS cost function (17), the inverse of the covariance matrix is needed. Since the covariance matrices  $C_e(\theta)$  and  $C_{\text{Re}}(\theta)$  are Hermitian and positive definite, they can therefore be factorized by a Cholesky decomposition  $R^H R$ , with  $R$  an upper triangular matrix, which can be inverted efficiently.

When the real and imaginary part of the equation error are separated (Appendix C.2), the resulting covariance matrix is block band dominant. By using a unitary transformation  $T^T T = T T^T = I$ , this  $C_{\text{Re}}$  can be transformed into a band dominant form. Then, the (possibly sparse) band structure remains intact through the Cholesky decomposition. The cost function (19) can then be rewritten as follows

$$\begin{aligned} V_{\text{WNLS}} &= e_{\text{re}}^T C_{\text{Re}}^{-1} e_{\text{re}} = e_{\text{re}}^T T (T^T C_{\text{Re}} T)^{-1} T^T e_{\text{re}} \\ &= e_{\text{re}}^T T (R^T R)^{-1} T^T e_{\text{re}} = \underbrace{e_{\text{re}}^T T}_{\varepsilon^T} \underbrace{T^T R^{-1} T^T}_{W} e_{\text{re}} \\ &= \varepsilon^T \varepsilon \end{aligned} \quad (\text{C.6})$$

## D Removing the additional polynomial from the equation error

Recall that the model parameter vector  $\theta$  contains the coefficients  $a_{[n,i]}$  and  $b_{[n,i]}$  of the original differential equation (1) with affine dependence (2), as well as the coefficients of a polynomial in  $j\omega_k$ . The first partition is denoted with  $\theta_{\text{dyn}}$ , and the latter with  $\theta_{\text{poly}}$ . The contribution of the polynomial on the equation error in (13) is then given by  $-\Omega \theta_{\text{poly}}$ , where  $\Omega$  is the polynomial regressor from (15), stacking the powers of  $j\omega_k$ . Using the variable projection method [5], the effect of a polynomial in  $j\omega_k$ , like  $T_{uy}^{N_T}(j\omega_k)$  in (8) can be projected away. It is then possible to compute the weighted residuals as

$$\begin{aligned} W_{\text{Re}} e_{\text{re}} &= W_{\text{Re}} (\mathbf{A}_{\text{Re}} Y_{\text{re}} - \mathbf{B}_{\text{Re}} U_{\text{re}} - \Omega_{\text{Re}} \theta_{\text{poly}}) \\ &= \left[ I - W_{\text{Re}} \Omega_{\text{Re}} [\Omega_{\text{Re}}^T C_{\text{Re}}^{-1} \Omega_{\text{Re}}]^{-1} \Omega_{\text{Re}}^T W_{\text{Re}}^T \right] \times \\ &\quad W_{\text{Re}} [\mathbf{A}_{\text{Re}} Y_{\text{re}} - \mathbf{B}_{\text{Re}} U_{\text{re}}] \end{aligned} \quad (\text{D.1})$$

$$= \Pi r \quad (\text{D.2})$$

$\Pi$  is a symmetrical idempotent matrix ( $\Pi \Pi = \Pi$ ), and  $r$  represents the weighted equation error from which the polynomial influence has been removed. The WNLS cost function (C.6) can now be rewritten as

$$V_{\text{WNLS}} = r^T \Pi r \quad (\text{D.3})$$

## E Proof of the strong consistency of the WNLS estimator

Appendix D showed that the polynomial can be removed from the equation error, resulting in a new WNLS cost function (D.3). Now, Appendix E.1 proves that this new WNLS cost function is minimal in the true parameters. Next, Appendix E.2 shows that it converges uniformly with probability 1 to the expected value. These are sufficient conditions for strong consistency.

### E.1 The WNLS cost function without the additional polynomial is minimal in the true parameters

Since the expected value of the cost function is a real number, observe the trace of the result:

$$\mathbb{E} \{V_{\text{WNLS}}\} = \mathbb{E} \{ \text{trace} [r^T \Pi r] \} \quad (\text{E.1})$$

$$= \text{trace} [\Pi \mathbb{E} \{r r^T\}] \quad (\text{E.2})$$

The residual  $r$  can be decomposed into a deterministic part  $r_0$ , and a noisy part  $\Delta r$ .

$$r(\theta_{\text{dyn}}, Z) = r_0(\theta_{\text{dyn}}, Z) + \Delta r(\theta_{\text{dyn}}, Z) \quad (\text{E.3})$$

If we substitute these residuals in the projected WNLS cost function (E.2), we find the expected value

$$\mathbb{E} \{r r^T\} = \mathbb{E} \{r_0 r_0^T\} + \mathbb{E} \{\Delta r \Delta r^T\} \quad (\text{E.4})$$

$$+ \mathbb{E} \{r_0 \Delta r^T\} + \mathbb{E} \{\Delta r r_0^T\} \quad (\text{E.5})$$

$$\begin{aligned} &= r_0 r_0^T + W C_{\text{Re}} W^T \\ &= r_0 r_0^T + W [W^T W]^{-1} W^T \\ &= r_0 r_0^T + I_N \end{aligned} \quad (\text{E.6})$$

The first term in (E.4) is minimal only in the true model parameters  $\theta_0$ . Following Assumption 10, the cross-terms disappear. We end up with an  $N \times N$  identity matrix  $I_N$ , but only if we use the full covariance matrix  $C_{\text{Re}}$  in the weighting. Substituting the residual in the trace of the cost function (E.2), yields

$$\mathbb{E} \{V_{\text{WNLS}}\} = \text{trace} (\Pi r_0 r_0^T) + \text{trace}(\Pi) \quad (\text{E.7})$$

$$= r_0^T \Pi r_0 + \{2N - (N_T + 1)\} \quad (\text{E.8})$$

By removing the polynomial in  $j\omega_k$ , less model parameters are used, and thus the trace of the symmetrical

idempotent projection matrix  $\Pi$  becomes  $2N - (N_T + 1)$ , the effective number of parameters. Now, in the absence of model errors (Assumption 15), it can be found that

$$\begin{aligned} r_0 &= W_{\text{Re}} [\mathbf{A}_{\text{Re}} Y_{\text{re}} - \mathbf{B}_{\text{Re}} U_{\text{re}}] \\ &= W_{\text{Re}} \mathbf{\Omega}_{\text{Re}} \theta_{\text{poly}} \end{aligned} \quad (\text{E.9})$$

Combining the latter with (D.1) and (D.2) leads to the following result:

$$\Pi r_0 = \Pi W_{\text{Re}} \mathbf{\Omega}_{\text{Re}} \theta_{\text{poly}} = 0 \quad (\text{E.10})$$

The projected residuals are zero in the true parameters  $\theta_0$ , which means that the expected value of the cost function (E.2) is minimal in  $\theta_0$ .

### E.2 Convergence of the WNLS cost function

Suppose multiple ( $N_{\text{exp}}$ ) independent experiments are conducted. Then, by the strong law of large numbers, the WNLS cost function  $V_{\text{WNLS}}(\theta)$  in (17) converges ( $N_{\text{exp}} \rightarrow \infty$ ) uniformly (in  $\theta$ ) with probability 1 to the expected value  $\mathbb{E}\{V_{\text{WNLS}}(\theta)\}$ . The input signal  $u(t)$  and the scheduling trajectory  $p(t)$  can change over the experiments, resulting in different regressors  $\mathbf{K}_m$  for each experiment. The model structure (13) remains the same, and the model coefficients  $\theta$  (12) (containing the model coefficients  $a_{[n,i]}$  and  $b_{[n,i]}$ ) are identical. The multiple experiments can be described by one big cost function, that is built from smaller, single experiment blocks.

$$\mathbf{K}_{\text{exp}} \theta = e_{\text{exp}} \quad (\text{E.11})$$

$$\begin{bmatrix} \mathbf{K}_1 \\ \vdots \\ \mathbf{K}_{N_{\text{exp}}} \end{bmatrix} \theta = \begin{bmatrix} e_1 \\ \vdots \\ e_{N_{\text{exp}}} \end{bmatrix} \quad (\text{E.12})$$

The regressor matrix  $\mathbf{K}_{\text{exp}}$  grows linearly with the amount of experiments. It is not required that the measurements have the same amount of samples  $N$ .

Note that the equation error  $e$  is correlated over the frequencies for a single experiment via the covariance matrix  $C_e(\theta)$  in (C.4), but it is uncorrelated over the experiments. The covariance matrix of the combined equation errors  $\mathbf{C}_e$  then becomes block diagonal.

$$\mathbf{C}_e = \begin{bmatrix} \boxed{C_{e[1]}} & 0 & \dots & 0 \\ 0 & \boxed{C_{e[2]}} & \ddots & \vdots \\ \dots & \ddots & \ddots & 0 \\ 0 & 0 & \dots & \boxed{C_{e[N_{\text{exp}}]}} \end{bmatrix} \quad (\text{E.13})$$

The equation error of the combined experiments is cumulant mixing of order  $\infty$ , and the results of Chapter 17 in [14] apply; the variance of the cost functions converges to zero. This result can be interpreted as follows: several independent measurements give more information about the system, reducing the influence of the noise.

### E.3 Consistency for periodic parameter variations

In the case of periodic parameter-variation, an additional interpretation for consistency is possible. Because the Fourier series expansion in (B.7) is truncated, there is a possibly large, but finite correlation length of the weighted equation error  $\varepsilon$  over the frequency. This means that  $\varepsilon$  is cumulant mixing of order  $\infty$ , and the results of Chapter 17 in [14] apply. It is easy to see that, in the periodic case, a longer measurement  $N \rightarrow \infty$  gives more information about the system, by allowing to average out the influence of the noise.

## References

- [1] Matthew S. Allen, M. W. Sracic, S. Chauhan, and M. H. Hansen. Output-only modal analysis of linear time-periodic systems with application to wind turbine simulation data. *Mechanical Systems and Signal Processing*, 25(4):1174–1191, 2011.
- [2] F. Bruzelius and C. Breitholtz. Gain scheduling via affine linear parameter-varying systems and  $H_\infty$  synthesis. In *IEEE Conference on Decision and Control*, pages 2386–2391, 2001.
- [3] J. De Caigny, J. F. Camino, and J. Swevers. Interpolation-based modeling of MIMO LPV systems. *IEEE Transactions on Control Systems Technology*, 19(1):46–63, January 2011.
- [4] R. Fletcher. *Practical methods of optimization (2nd ed.)*. Wiley-Interscience, New York, USA, 2000.
- [5] G. Golub and V. Pereyra. The differentiation of pseudoinverses and nonlinear least squares problems whose variables separate. *SIAM Journal on Numerical Analysis*, 10:413–432, 1973.
- [6] J. Lataire and R. Pintelon. Estimating a non-parametric, colored noise model for linear, slowly time-varying systems. *IEEE Trans. Instrum. and Meas.*, 58(5):1535–1545, May 2009.
- [7] J. Lataire and R. Pintelon. Frequency domain weighted nonlinear least squares estimation of continuous-time, time-varying systems. *IET Control Theory & Applications*, 5(7):923–933, 2011.
- [8] V. Laurain, M. Gilson, H. Garnier, and R. Tóth. *Linear Parameter-Varying System Identification: New Developments and Trends*, chapter 5, pages 95–132. World Scientific Publishing, Singapore, 2011.
- [9] V. Laurain, M. Gilson, R. Tóth, and H. Garnier. Refined instrumental variable methods for identification of lpv box-jenkins models. *Automatica*, 46:959–967, 2010.
- [10] V. Laurain, R. Tóth, M. Gilson, and H. Garnier. Direct identification of continuous-time linear parameter-varying input/output models. *IET Control Theory and Applications*, 5(7):878–888, 2010.

- [11] V. Laurain, R. Tóth, W.X. Zheng, and M. Gilson. Nonparametric identification of LPV models under general noise conditions: an LS-SVM based approach. In *Proc. of the 16th IFAC Symposium on System Identification*, pages 1761–1766, Brussels, Belgium, 2012.
- [12] L. Ljung. *System Identification : Theory for the User*. Prentice Hall, Upper Saddle River, NJ, USA, 1999.
- [13] R. Pintelon and J. Schoukens. Identification of continuous-time systems using arbitrary signals. *Automatica*, 33(5):991–994, 1997.
- [14] R. Pintelon and J. Schoukens. *System Identification: A Frequency Domain Approach (2nd ed.)*. Wiley-IEEE Press, 2012.
- [15] R. Pintelon, G. Vandersteen, J. Schoukens, and Y. Rolain. Improved (non-) parametric identification of dynamic systems excited by periodic signals - the multivariate case. *Mechanical Systems and Signal Processing*, 25(8):2892–2922, 2011.
- [16] W. J. Rugh and J. S. Shamma. Research on gain scheduling. *Automatica*, 36(10):1401–1425, 2000.
- [17] H. Sandberg, E. Möllerstedt, and Bernhardsson. Frequency-domain analysis of linear time-periodic systems. *IEEE Transactions on Automatic Control*, 50(12):1971–1983, 2005.
- [18] R. Tóth. *Modeling and Identification of Linear Parameter-Varying Systems*. Springer, Heidelberg, Germany, 2010.
- [19] R. Tóth, V. Laurain, M. Gilson, and H. Garnier. Instrumental variable scheme for closed-loop LPV model identification. *Automatica*, 48:2314–2320, 2012.
- [20] S. Van Huffel and J. Vandewalle. *The Total Least Squares Problem: Computational Aspects and Analysis*. Frontiers in Applied Mathematics. Society for Industrial and Applied Mathematics, 1991.