Non-parametric identification of multivariable systems

Non-parametric identification of multivariable systems: a local rational modeling approach with application to a vibration isolation benchmark

Robbert Voorhoeve*,[a], Annemiek van der Maas[a], Tom Oomen[a]

[a] *Eindhoven University of Technology, Department of Mechanical Engineering, Control Systems Technology. P.O. Box 513, 5600 MB Eindhoven, The Netherlands (* e-mail: r.j.voorhoeve@tue.nl).*

Abstract

Frequency response function (FRF) identification is often used as a basis for control systems design and as a starting point for subsequent parametric system identification. The aim of this paper is to develop a multiple-input multiple-output (MIMO) local parametric modeling approach for FRF identification of lightly damped mechanical systems with improved speed and accuracy. The proposed method is based on local rational models, which can efficiently handle the lightly-damped resonant dynamics. A key aspect herein is the freedom in the multivariable rational model parametrizations. Several choices for such multivariable rational model parametrizations are proposed and investigated. For systems with many inputs and outputs the required number of model parameters can rapidly increase, adversely affecting the performance of the local modeling approach. Therefore, low-order model structures are investigated. The structure of these low-order parametrizations leads to an undesired directionality in the identification problem. To address this, an iterative local rational modeling algorithm is proposed. As a special case recently developed SISO algorithms are recovered. The proposed approach is successfully demonstrated on simulations and on an active vibration isolation system benchmark, confirming good performance of the method using significantly less parameters compared with alternative approaches.

Keywords: System identification, non-parametric, frequency response function, local parametric modeling, matrix fraction description

1. Introduction

Accurate and fast identification of multivariable frequency response functions (FRFs) is essential in many applications, including the analysis of mechanical structures, see, e.g., [1–3]. Furthermore, the FRF is often used for controller design and validation [4], and as an intermediate step towards the identification of a parametric system model [5, 6]. Recent developments in the design and control of mechanical precision systems have led to an increased relevance of the lightly damped resonant dynamics of such systems [7], as well as an increase in the number of sensors and actuators used to control them [8]. Efficient FRF identification is a crucial step in enabling the use of such advanced design and control methods. In this paper, this problem of efficient FRF identification for lightly damped multivariable systems is considered.

The quality of an FRF estimate depends on the identification method used to obtain it. Key aspects include the applied excitation signal, noise mitigation through averaging, and the suppression of transient contributions. In early literature on time series- and spectral analysis, the methods have been developed for transfer function estimation based on measurements using random time series excitations, see, e.g., [9–11]. These methodologies have been further developed in, e.g., [12]. Subsequently, the advantages of periodic excitations have been advocated, see, e.g., [6, 13], including the analysis of nonlinear distortions [14]. In recent years, more advanced non-parametric identification methods have been proposed to further improve the transfer function estimate, e.g., [15–17]. One of these methods is the local polynomial method (LPM) [17–19], which uses least squares estimation in a small local frequency band to obtain improved estimates. A key mechanism through which this improvement is achieved is the explicit estimation and suppression of transient contributions. In [20], this method is further extended towards the use of rational models, known

as the local rational method (LRM). In [20], the LRM is formulated for single-input single-output (SISO) systems and uses a linear cost criterion, in the sense of [21], to estimate the local rational models. The use of these local rational models instead of polynomials has been shown to lead to significant improvements in the estimation quality for systems containing lightly damped resonances, [22–24]. For a recent overview of FRF identification methods applied to a lightly-damped mechanical positioning system, see, e.g., [24].

Although FRF identification has been significantly developed, the application of these advanced methods to multivariable systems, especially those with a large number of inputs and outputs, leads to several aspects that are presently unclear and which are essential for efficient FRF identification. In particular, the LPM can suffer from large interpolation errors around the lightly damped resonances [22, 23]. The LRM, on the other hand, has been successfully applied for lightly damped SISO systems [20, 22], but, as will be shown in the present paper, the multivariable extension allows substantial design freedom in the model parametrization.

The main contribution of this paper is a framework for local rational modeling of multivariable systems with high input-output dimensionality and lightly-damped complex dynamics. The focus herein lies on the extension of the existing LRM to a multi-input multi-output (MIMO) approach and the involved parametrization issues. One of the main parametrization issues for such a MIMO LRM approach is the number of parameters that are used. It is important to use a low number of parameters since this leads to a smaller minimal windows size, which in general leads to a smaller interpolation bias [6, Section 7.2.2.3]. Therefore, parsimonious parametrizations are investigated. Of particular interest herein are the directionality aspects associated with such multivariable system parametrizations. Iterative algorithms are proposed to mitigate the possible negative effects of this directionality. Finally, the advantages of the considered methods are demonstrated on relevant simulation examples as well as experimental data from a recently proposed benchmark system.

The outline of this paper is as follows. In Section 2, the local parametric modeling approach is introduced and the advantages of this approach for the identification of mechanical systems are explained. In Section 3, the problem of finding a suitable parametrization for the MIMO LRM is considered and a number of parametrizations are proposed and analyzed. In Section 4, iterative methods for solving the output error LRM problem are considered. In Section 5, the results of the simulation study are shown. In Section 6, the results based on experimental data from a recent system identification benchmark system are presented. In Section 7, the conclusions of this paper are presented as well as an outlook on ongoing research.

2. Problem formulation

In this section, the considered problem of local rational modeling for multivariable systems is formulated. First, the core idea of local parametric modeling for FRF identification is introduced. Second, a practically relevant example is presented showcasing the superior transient suppression properties of the local parametric modeling approach. Third, an example is presented which shows the advantage of rational parametrizations for the identification of lightly damped systems. Fourth, the bias and variance aspects of the LRM estimator are discussed. Last, the challenges concerning multivariable local rational parametrizations are explained, which are addressed in the remaining sections of this paper.

2.1. Local parametric modeling for FRF identification

Consider the discrete time signal $u(n),\ n = 0, \ldots, N-1$. The discrete Fourier transform of $u(n)$ is defined as

$$U(k) = \frac{1}{\sqrt{N}} \sum_{n=0}^{N-1} u(n) e^{-j2\pi nk/N} \,. \tag{1}$$

When the signal $u(n)$ is applied as input to a linear time invariant system G_0 with additive output noise $v(n)$, as in Figure 1, the resulting output in the frequency domain equals

$$Y(k) = G_0(\Omega_k)U(k) + T(\Omega_k) + V(k) \,, \tag{2}$$

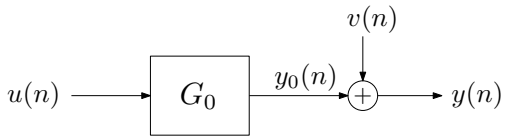


Figure 1: LTI discrete time system in an output error setup.

where $T(\Omega_k)$ represents the transient contribution and $V(k)$ represents the noise contribution. The argument Ω_k denotes the generalized frequency variable evaluated at DFT-bin k, which, when formulated in, e.g., the Laplace domain becomes $\Omega_k = j\omega_k$ and in the Z-domain $\Omega_k = e^{j\omega_k T_s}$.

The main idea of the local modeling approach, e.g., as in [17], is to identify a model, with validity over only a small frequency range, which can be used to provide a non-parametric estimate of the FRF and the transient at the central point k. To achieve this, a small frequency window around DFT-bin k is considered, denoted by the variable $r \in \mathbb{Z}$, which can range from $-n_W$ to n_W, i.e.,

$$Y(k+r) = G(\Omega_{k+r})U(k+r) + T(\Omega_{k+r}) + V(k+r)\,. \tag{3}$$

Next, both the plant, $G(\Omega_{k+r})$, and the transient contribution, $T(\Omega_{k+r})$, which are assumed to be smooth functions of the frequency, are parametrized. For instance, using the polynomial parametrization

$$G(\Omega_{k+r}) = G(\Omega_k) + \sum_{s=1}^{R} g_s(k) r^s\,, \tag{4}$$

$$T(\Omega_{k+r}) = T(\Omega_k) + \sum_{s=1}^{R} t_s(k) r^s\,. \tag{5}$$

Using this parametrization (3) is rewritten as

$$Y(k+r) = \Theta(k)K(k+r) + V(k+r), \tag{6}$$

with

$$\begin{aligned}
\Theta(k) &= \begin{bmatrix} \Theta_G(k) & \Theta_T(k) \end{bmatrix}, \\
\Theta_G(k) &= \begin{bmatrix} G(\Omega_k) & g_1(k) & g_2(k) & \dots & g_R(k) \end{bmatrix}, \\
\Theta_T(k) &= \begin{bmatrix} T(\Omega_k) & t_1(k) & t_2(k) & \dots & t_R(k) \end{bmatrix}, \qquad (7) \\
K(k+r) &= \begin{bmatrix} K_1(r) \otimes U(k+r) \\ K_1(r) \end{bmatrix}, \qquad K_1(r) = \begin{bmatrix} 1 & r & \dots & r^R \end{bmatrix}^T. \qquad (8)
\end{aligned}$$

Finally, the parameters of the local model are determined by solving the linear least squares problem

$$\hat{\Theta}(k) = \underset{\Theta(k)}{\arg\min} \sum_{r=-n_W}^{n_W} \|Y(k+r) - \Theta(k)K(k+r)\|_2^2 = Y_n(k)K_n(k)^+\,, \tag{9}$$

with $X_n(k) = \begin{bmatrix} X(k-n_W) & X(k-n_W+1) & \cdots & X(k+n_W) \end{bmatrix}$, and with $A^+ = A^H(AA^H)^{-1}$ the right Moore-Penrose pseudoinverse. Performing this least squares estimation for each DFT-bin, k, and evaluating the local models at the center frequency $r = 0$, yields a non-parametric estimate, $G(\Omega_k)$, for the FRF. For the parametrization described here this evaluation is trivial, since for $r = 0$ only the zeroth order polynomial term remains, which is why $G(\Omega_k)$ directly appears in this parametrization, see (4). It should be noted that

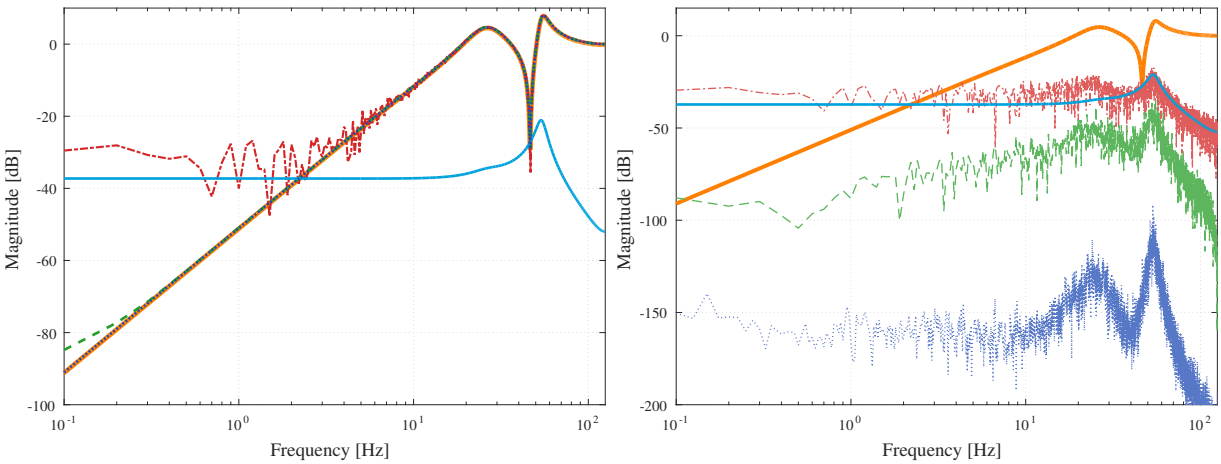


Figure 2: Simulation example showing the improved transient mitigation of the local parametric modeling approach. Both figures show the simulated system (orange solid) and the transient contribution (cyan solid). The left figure shows: First, the spectral analysis (SA) estimate with a rectangular window (red dash-dotted), which provides no transient mitigation and a poor estimate. Second, the SA estimate with a Hanning window (green dashed), which provides better transient mitigation but is still visibly perturbed by the transient at the first few frequency points. Last, the LPM estimate (dark-blue dotted), which provides an accurate estimate through enhanced transient mitigation. These observations are confirmed in the right figure, which shows the estimation errors of the same FRF estimates, where the estimation errors for the LPM are several orders of magnitude smaller than those of the two SA approaches.

this least squares estimation can be subject to certain weighting factors, as in, e.g., [25], which is true for any estimation problem considered in this paper. However, for clarity, such weighting factors have been omitted in the present paper.

It has been shown, e.g, in [17, 22, 24], that the local polynomial method leads to an improved FRF estimate, especially in cases where there is a significant transient contribution. This is most relevant when the available measurement time is limited or when there are lightly damped resonances. Another case where the transient contribution can significantly affect the quality of the FRF estimates is in closed-loop identification, as will be shown in an example in the following section.

2.2. Motivating example for local parametric modeling in closed-loop identification

For many systems, including precision motion systems, identification experiments need to be performed in closed-loop, since feedback is required to ensure stable and safe operation. Closed-loop FRF identification is commonly performed using the indirect approach, as explained in, e.g., [6, Section 2.6.4]. An essential part of this indirect approach is the accurate identification of the closed-loop sensitivity function. This sensitivity function often has low gain at low frequencies [26]. This often causes problems in accurately identifying the sensitivity at low frequencies, as can be seen in, e.g., [27, Fig. 7 and 8]. In the following example, it is shown that this poor low-frequency estimation performance is caused by a transient contribution, and it is shown how local parametric modeling mitigates this.

Figure 2 shows the results of a simulated identification experiment for a typical sensitivity function. For details on the simulation example considered here, see Appendix A. In Figure 2, a comparison is shown between the FRF estimate obtained using the LPM, i.e. through (4)-(7) with $R = 2$ and $n_W = 4$, and the estimates obtained using standard spectral analysis methods. Two spectral analysis estimates are shown, one where a rectangular window is applied and one where a Hanning window is utilized to mitigate the transient contribution. The window size for the spectral analysis methods is 10 s, or 2500 samples and an overlap of 50% is used when the Hanning window is applied. This means a total of 4 windows are used in the estimate with the rectangular window and 7 half-overlapping windows in the Hanning window estimate. In this simulation example, there are no noise sources perturbing the identification signals. Therefore, the observed estimation errors can be solely attributed to transient contributions, windowing effects, and

interpolation bias. The spectrum of the transient contribution as estimated by the LPM, i.e. by (5), is also shown in Figure 2. This spectrum is scaled with the inverse input power to enable a direct comparison between the magnitudes of the transient and the estimated FRFs.

In Figure 2, it can be seen that in the low-frequency region the transient contribution is significantly larger than the contribution of the linear system. Consequently, the spectral analysis method with the rectangular window has a large error in this region, as the transient contribution is not mitigated. The spectral analysis method which uses a Hanning window provides more effective transient mitigation and therefore remains relatively accurate for a larger frequency range. The LPM estimate clearly outperforms both of the spectral analysis estimates and remains effective even at the lowest frequencies due to the effective mitigation of the transient contribution.

This example clearly shows that the local parametric modeling approach provides effective transient mitigation, thereby significantly enhancing the estimation accuracy in cases where this transient contribution is significant. In the next example, the advantages of rational parametrizations are highlighted for cases where lightly damped resonances are present in the FRF.

2.3. Motivating example for local rational parametrizations

In this subsection, a comparison is made between the polynomial parametrization, given by equations (4)–(7), and a rational parametrization. This comparison between the LPM and the LRM is based on a noise-free simulation example, similar to the example in Section 2.2, containing lightly damped resonant dynamics. For further details on this simulation example, see Appendix A.

In the SISO formulation of the LRM, the system and transient are parametrized as

$$G(\Omega_{k+r}) = \frac{n(\Omega_{k+r})}{d(\Omega_{k+r})}\,, \qquad T(\Omega_{k+r}) = \frac{m(\Omega_{k+r})}{d(\Omega_{k+r})}\,, \tag{10}$$

$$n(\Omega_{k+r}) = G(\Omega_k) + \sum_{s=1}^{R} n_s(k) r^s\,, \tag{11}$$

$$m(\Omega_{k+r}) = T(\Omega_k) + \sum_{s=1}^{R} m_s(k) r^s\,, \tag{12}$$

$$d(\Omega_{k+r}) = 1 + \sum_{s=1}^{R} d_s(k) r^s\,. \tag{13}$$

Substituting these expressions in (3) leads to a formulation that is nonlinear in the parameters $\Theta(k)$, i.e., it can not be written in the same form as (6). To be able to solve this problem in a linear least squares sense, the resulting equation is multiplied with the denominator polynomial. This leads to

$$d(\Omega_{k+r})Y(k+r) = n(\Omega_{k+r})U(k+r) + m(\Omega_{k+r}) + \tilde{V}(k+r)\,, \tag{14}$$

which is linear in the parameters, and where $\tilde{V}(k+r) = d(\Omega_{k+r})V(k+r)$ is treated as an unknown noise term. The parameters of the local rational model are determined by solving the linear least squares problem,

$$\hat{\Theta}(k) = \arg\min_{\Theta} \sum_{r=-n_W}^{n_W} \left| d(\Omega_{k+r})Y(k+r) - n(\Omega_{k+r})U(k+r) - m(\Omega_{k+r}) \right|^2\,. \tag{15}$$

Performing this least squares estimation for each frequency point and evaluating the local model at $r = 0$, again yields a non-parametric estimate of the FRF $G_0(\Omega_k)$. This approach is known as the local rational method (LRM).

In Figure 3, an example is shown of a response containing lightly damped resonances. In this figure, it can clearly be seen that the LPM estimate, with $R = 2$ and $n_W = 4$, depicted in green, provides a poor estimate of the FRF around the resonance peaks. In this example, the frequency resolution around the resonances is clearly insufficient for an accurate polynomial interpolation in the LPM. It is asserted in [28]

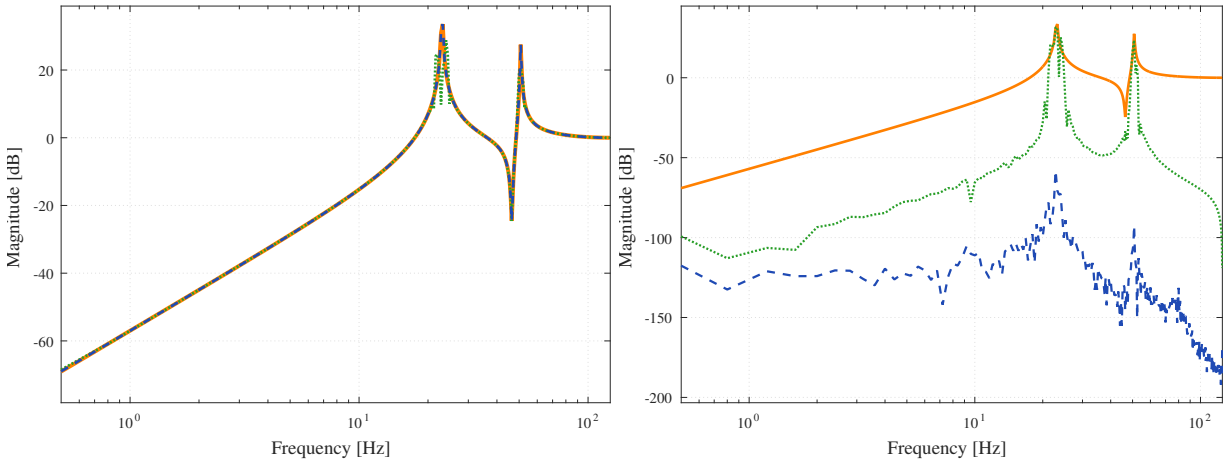


Figure 3: Simulation example showing the reduced interpolation errors of the local rational model parametrization for systems with lightly-damped resonant dynamics. Both figures show the simulated system (orange solid). The left figure shows the LPM estimate (green dotted), for which large errors are visible around the resonance peaks, and the LRM estimate (dark-blue dashed), which remains accurate around the resonances. These observations are confirmed in the right figure, which shows the estimation errors of the LPM (green dotted) and LRM (dark-blue dashed) estimates, here the estimation errors for the LRM are several orders of magnitude smaller than those of the LPM, especially around the resonance peaks.

that the minimum required measurement time for the LPM should be such that there are at least seven frequency points in the 3 dB bandwidth of the resonance. The 3 dB bandwidth for the first resonance of this example is $B_{3dB} = 0.08$ Hz and the frequency resolution is 0.4 Hz, meaning the resolution, and therefore the measurement time, should be increased by at least a factor 30 before the interpolation errors of the LPM are sufficiently bounded.

The LRM, with $R = 2$ and $n_W = 5$, depicted in blue in Figure 3, does accurately model the local plant and transient behavior around the resonances in this example, leading to a smaller interpolation error and therefore an improved FRF estimate. This is clearly observed in Figure 3, where the LRM significantly outperforms the LPM, especially around the resonances. Since lightly damped resonances are common in mechanical systems, see, e.g., [29], the LRM is typically much better suited than the LPM to provide fast and accurate FRF identification for this class of systems.

2.4. Bias and variance aspects of the LRM estimator

The accuracy of the LRM estimator is influenced by several aspects. First, the output noise v leads to a variance error on the estimate, which should be quantified and mitigated if possible. Second, there are various aspects that can lead to a bias in the LRM estimator. In this section, several of these bias and variance aspects are explored.

- Two sources of bias for both the local parametric estimates are the interpolation error $O_{\text{int}\,G}$, and the residual system leakage errors $O_{\text{leak}\,G}$, which both decrease as the local window size decreases [18, Sections 3.3]. This is one of the main reasons to try and keep the window size as small as possible and to use low-order parsimonious parametrizations when possible.

- In general, the variance on the local parametric estimates decreases with increasing window size, as this increases the amount of noisy data points used to estimate a fixed number of parameters, decreasing the variance on those parameters. This essentially provides a way to trade-off bias and variance errors by increasing the window size when the variance errors are dominant and decreasing it when the bias errors dominate [6, Section 7.2.6].

- The variance on the LPM estimate can be estimated from the least squares residuals as is detailed in [18, Sections 3.1 & 5.1]. For the LRM this same method is not directly applicable due to the noisy

output data that is present in the regression matrix K, which apart from complicating the variance estimation also leads to a bias in the LRM estimate [22].

- In [22, Section 3.3], it is shown that this bias term is inherently linked to the SNR of the output signal in the frequency domain.
- The analysis in [22, Section 3.4], based on extensive Monte-Carlo simulations with different SNRs, suggests that for systems with a good SNR ($>$ 20 dB) this bias term is small compared to, e.g, the variance errors and the interpolation and leakage errors for the LPM and ETFE.
- It is a topic of ongoing research to further assess and quantify the bias errors for the LRM and to provide an accurate estimation of the variance on the LRM estimate [30].

In this paper, the focus is on the extension of the existing LRM to a MIMO approach and the involved parametrization issues. Therefore, to focus on the influence of the parametrization on the quality of the LRM estimate, systems with a good SNR ($>$ 20 dB) are considered and the bias and variance aspects related to the SNR are not further examined in detail here. This assumption of good SNR is also often satisfied for the class of mechanical systems.

To provide an estimate for the covariance of the LRM estimate, the biasing effect of the noise in the regression matrix K is neglected in which case the derivation and expressions for the estimated covariance becomes similar to the expressions for the LPM case as presented in, e.g., [18, Sections 3.1 & 5.1]. Under the assumption of a high SNR ($>$ 20 dB) this is a reasonable estimate. For further details on this covariance estimate for the specific parametrizations proposed in this paper see Appendix C. To asses the overall quality of the LRM estimates in this paper, including all bias and variance aspects, estimation errors with respect to reference models are used where possible.

2.5. Parametrization challenges for multivariable systems with large input/output dimensions

In Section 2.3, the advantages of rational parametrizations for local parametric modeling are established. Since lightly damped resonances are often encountered in mechanical systems, there is a strong preference for the LRM over the LPM for these systems. However, these systems often have multiple inputs and outputs, and the LRM has so far only been formulated for single-input single-output systems.

The LPM straightforwardly extends to multivariable systems, in fact, the formulation (2)–(7) is directly applicable for multivariable systems where $Y(k) \in \mathbb{C}^{n_y \times 1}$ and $U(k) \in \mathbb{C}^{n_u \times 1}$. In this MIMO LPM formulation, the polynomial coefficients in Θ_G and Θ_T are coefficient matrices of size $n_y \times n_u$ and $n_y \times 1$ respectively. This bring the total amount of parameters in the LPM parametrization to:

$$n_{\theta,\,\mathrm{LPM}} = n_y(n_u + 1)(R + 1), \tag{16}$$

which for systems with many inputs and outputs becomes very large, especially when the polynomial order, R, required to model the local behavior becomes large. Such a large amount of parameters can be problematic as it increases the required size of the local modeling window. In turn, this can lead to larger interpolations errors [6, Section 7.2.2.3].

Because of the increased flexibility offered by a rational parametrization, a multivariable LRM can potentially be used to accurately model the local system and transient behavior with a reduced number of parameters. The parametrization challenge herein is to obtain a multivariable rational parametrization that is suitable for local parametric modeling and uses a limited number of parameters. In the following section, several multivariable parametrizations for the local rational approach are proposed and compared.

3. Multivariable LRM parametrizations

In this section, parametrizations for multivariable local rational modeling are investigated. The specific requirements for parametrizations for local modeling are first analyzed. Next, several parametrizations are proposed and subsequently compared.

3.1. Parametrizations for local modeling

The general form for multivariable rational parametrizations used in this paper is the left matrix fraction description (LMFD):

$$G(\Omega_{k+r}) = D(\Omega_{k+r})^{-1}\, N(\Omega_{k+r})\,, \tag{17}$$

$$T(\Omega_{k+r}) = D(\Omega_{k+r})^{-1}\, M(\Omega_{k+r})\,. \tag{18}$$

Assuming this general LMFD formulation, the remaining question is how to parametrize $D(\Omega_{k+r})$, $N(\Omega_{k+r})$, and $M(\Omega_{k+r})$.

Although substantial literature exists on MFD parametrizations for system identification, e.g., [31–38], the global identification problem considered in these references is significantly different from the local identification problem considered here. This distinction can be seen in multiple aspects, which are investigated in detail, next.

In the global identification problem, one of the main challenges concerning the parametrization, is its ability to describe all systems of a given complexity or order. This is important because, in parametric identification, it is often required that the true system is contained in the model set. In the local parametric modeling approaches, this is not a requirement, as the model only needs to approximate the local system behavior. Therefore, a single generic parametrization is sufficient to parametrize the model for local identification, since such a generic parametrization can be used to approximate any system of a given order, as is explained in, e.g., [32].

In system identification, model complexity in dynamic models is often related to the McMillan degree. For MFDs, this McMillan degree is equal to the order of the determinant of the denominator matrix D. For local parametric modeling, however, this McMillan degree of the parametrized model is of secondary importance. It is more important that a parametrization uses the smallest number of parameters to accurately describe the local system behavior. Since a small amount of parameters directly translates into a smaller minimal window size, generally reducing the interpolation bias [6, Section 7.2.2.3].

Lastly, to obtain an identifiable parametrization, constraints need to be added to satisfy the uniqueness condition required for identifiability, see [32, 39]. In global identification it is common to constrain the highest polynomial order of the D matrix, see, e.g. [33, 40, 41], as this effectively constrains the McMillan degree and enforces (strict-)properness of the parametrized models. For local rational modeling, however, it is preferred to constrain the zeroth degree of the the D matrix. This has the following advantages. First, the plant estimate for the LRM is given by $D(\Omega_{k+0})^{-1}N(\Omega_{k+0}) = D_0^{-1}N_0$; constraining the D_0 matrix guarantees its invertibility. Second, having no constraints on the highest degree matrices puts no constraint on the order of the local model. This can be beneficial when some local frequency windows contain no significant dynamics and can therefore be best approximated with a low order model, while in other windows the additional freedom provided by the higher order terms in the parametrization is required to describe the local behavior.

In the remainder of this section, several parametrizations are proposed. For ease of presentation and to avoid over-complicating the comparisons, only bi-proper parametrizations are considered, i.e., parametrizations which have the same order for the numerator as the denominator. After defining them, the parametrizations are evaluated and compared, taking into account the described distinctive characteristics of the local identification problem. Here, one of the main properties to consider is the number of parameters in the respective parametrizations.

3.2. Common denominator

The first parametrization considered in this paper for the multivariable local rational method is the common denominator parametrization. One advantage of using a common denominator for all elements in the transfer function is that it significantly limits the number of parameters introduced to model the denominator $D(\Omega_{k+r})$. In the terms of the general LMFD formulation of (17)–(18), this parametrization is defined as follows.

Parametrization 1 (Common Denominator (CD))**.** In the common denominator parametrization, the polynomial matrices $D(\Omega_{k+r})$, $N(\Omega_{k+r})$ and $M(\Omega_{k+r})$ are defined as

$$D(\Omega_{k+r}) = d_c(\Omega_{k+r})\ I_{n_u}\,, \tag{19}$$

$$d_c(\Omega_{k+r}) = 1 + \sum_{s=1}^{R} d_{c,s}(k) r^s\,, \qquad d_{c,s}(k) \in \mathbb{C}^{1\times 1}\,, \tag{20}$$

$$N(\Omega_{k+r}) = G(\Omega_k) + \sum_{s=1}^{R} N_s(k) r^s\,, \qquad G(\Omega_k),\, N_s(k) \in \mathbb{C}^{n_y\times n_u}\,, \tag{21}$$

$$M(\Omega_{k+r}) = T(\Omega_k) + \sum_{s=1}^{R} M_s(k) r^s\,, \qquad T(\Omega_k),\, M_s(k) \in \mathbb{C}^{n_y\times 1}\,. \tag{22}$$

△

Note that the △-symbol is used here to clearly denote the end of a parametrization. The choice for a common denominator for all outputs introduces a coupling between the local models for all inputs and outputs, making this an inherently multivariable parametrization. The different transfer function elements for a multivariable system indeed often contain common pole dynamics, justifying the common denominator formulation.

3.3. MISO parametrization

In this next parametrization, the denominator matrix $D(\Omega_{k+r})$ has the same diagonal structure as in the common denominator parametrization. However, here each diagonal element is modeled as a separate polynomial. This yields the following parametrization for $D(\Omega_{k+r})$.

Parametrization 2 (MISO)**.** In the MISO parametrization, the polynomial matrix $D(\Omega_{k+r})$ is defined as,

$$D(\Omega_{k+r}) = \operatorname{diag}\left(d_1(\Omega_{k+r}), \ldots, d_{n_y}(\Omega_{k+r})\right)\,, \tag{23}$$

$$d_i(\Omega_{k+r}) = 1 + \sum_{s=1}^{R} d_{i,s}(k) r^s\,, \qquad d_{i,s}(k) \in \mathbb{C}^{1\times 1}\,. \tag{24}$$

The matrices, $N(\Omega_{k+r})$ and $M(\Omega_{k+r})$ are defined as in the common denominator parametrization, i.e., by (21) and (22). △

Using this parametrization it is possible to separately consider the local parametric fit for each individual output, making this a MISO parametrization. Note that the local polynomial method can also be separated as such and is therefore also a MISO parametrization.

3.4. Full MFD parametrization

In the following parametrization, $D(\Omega_{k+r})$ is modeled as a full matrix polynomial, i.e., each element of the $D(\Omega_{k+r})$ is modeled with separate polynomials. This yields the following parametrization for $D(\Omega_{k+r})$.

Parametrization 3 (MFD full)**.** In the full MFD parametrization, the polynomial matrix $D(\Omega_{k+r})$ is defined as,

$$D(\Omega_{k+r}) = I_{n_y} + \sum_{s=1}^{R} D_s(k) r^s\,, \qquad D_s(k) \in \mathbb{C}^{n_y\times n_y}\,. \tag{25}$$

The matrices, $N(\Omega_{k+r})$ and $M(\Omega_{k+r})$ are again defined by (21) and (22). △

This choice of parametrization is clearly richer than the previous two parametrizations. However, it also uses significantly more parameters to model $D(\Omega_{k+r})$. This parametrization introduces a coupling between the different outputs through the off-diagonal elements, making it an inherently MIMO parametrization. This parametrization is known in system identification literature as the full polynomial form, see, e.g., [42, Chapter 6].

3.5. Parsimonious MFD parametrization

The complexity of the parametrizations proposed in the previous sections is fully determined by the degree of the matrix polynomials, R. The McMillan degree for all these parametrizations, which is the degree of the determinant of $D(\Omega_{k+r})$, is equal to $n_y \cdot R$ for the MISO and full MFD parametrizations and equal to $\min(n_y, n_u) \cdot R$ for the common denominator parametrization. The minimal increment with which the degree of these parametrizations can be increased is therefore n_y or $\min(n_y, n_u)$, which can be large for systems with many outputs and outputs. Such large increments in model order can often lead to situations of under-modeling or over-modeling. This is undesirable for local parametric modeling, therefore a parametrization is desired for which the complexity increases in smaller increments.

In, e.g., [31–34], MFD parametrizations have been proposed for which the McMillan degree can be equal to any given integer. These parametrizations are characterized by a certain degree structure for the matrices D and N, which is defined using a set of indices, $\eta_1, \ldots, \eta_{n_y}$. Here, η_i corresponds to the degree of the i-th row of the D-matrix. To parametrize a system of a certain order, a number of different degree structures are possible, therefore a choice for this degree structure is necessary to obtain a single suitable parametrization. In this paper it is proposed to use a *quasi-constant* degree structure, meaning the degrees should not differ by more than one, with the highest degrees on the first rows. This degree structure is given by

$$\eta_i = \begin{cases} \left\lceil \dfrac{n_x}{n_y} \right\rceil & i = 1, \ldots, l \\ \left\lceil \dfrac{n_x}{n_y} \right\rceil - 1 & i = l+1, \ldots, n_y \end{cases}, \tag{26}$$

with $l = n_y - (n_y \lceil n_x/n_y \rceil - n_x)$, and where $\lceil x \rceil$ denotes the ceiling function, i.e., rounding up to the nearest integer. The corresponding degree structure for $D(\Omega_{k+r})$ and $N(\Omega_{k+r})$ is

$$\deg D(\Omega_{k+r}) = \begin{bmatrix} \delta \\ \delta - 1 \end{bmatrix} \begin{matrix} \updownarrow l \\ \updownarrow (n_y - l) \end{matrix}, \qquad \deg N(\Omega_{k+r}) = \begin{bmatrix} \delta \\ \delta - 1 \end{bmatrix} \begin{matrix} \updownarrow l \\ \updownarrow (n_y - l) \end{matrix}, \tag{27}$$

with $\delta = \lceil n_x/n_y \rceil$. This corresponds to the fully-parametrized left matrix fraction descriptions (F-LMFD) as described in [43], which is an over-parametrized description. Constraints need to be added to address this over-parametrization.

In global identification, the highest degree matrix polynomials are often constrained, which corresponds to constraining the denominator polynomial to be monic in the SISO case. This constraint on the highest degree matrix polynomials takes the following form,

$$\tilde{D}_{ch} = I_{n_y}, \quad \text{and} \quad \tilde{D}_{rh} = \begin{bmatrix} I_l & 0 \\ \tilde{D}_{\delta-1,21} & I_{(n_y-l)} \end{bmatrix}, \tag{28}$$

where $\tilde{D}_{ch}$ is the matrix containing the polynomial coefficients with the highest column degree and $\tilde{D}_{rh}$ is the matrix containing the polynomial coefficients with the highest row degree. The resulting parametrization is often used in global identification and is also called the generic observable canonical LMFD parametrization which is described in, e.g., [32–34]. This parametrization is generic in the sense that it can be used to approximate all proper LTI systems of the given order up to arbitrary precision, see [32]. Here this parametrization is defined as follows.

Parametrization 4 (MFD high-degree constraint)**.** Using the *quasi-constant* degree structure of (26) and the high-degree constraint of (28), the $D(\Omega_{k+r})$, $N(\Omega_{k+r})$ and $M(\Omega_{k+r})$ matrices are defined as

$$\tilde{D}(\Omega_{k+r}) = \sum_{s=0}^{\delta-2} \tilde{D}_s(k) r^s + \begin{bmatrix} \tilde{D}_{\delta-1,11}(k) & 0 \\ \tilde{D}_{\delta-1,21}(k) & I_{(n_y-l)} \end{bmatrix} r^{\delta-1} + \begin{bmatrix} I_l & 0 \\ 0 & 0 \end{bmatrix} r^{\delta}, \tag{29}$$

$$\tilde{N}(\Omega_{k+r}) = \sum_{s=0}^{\delta-1} \tilde{N}_s(k) r^s + \begin{bmatrix} \tilde{N}_{\delta,1}(k) \\ 0 \end{bmatrix} r^{\delta}, \tag{30}$$

$$M(\Omega_{k+r}) = \sum_{s=0}^{\delta-1} M_s(k) r^s + \begin{bmatrix} M_{\delta,1}(k) \\ 0 \end{bmatrix} r^\delta \,, \tag{31}$$

with

$$\begin{aligned}
&\tilde{D}_{\delta-1,11}(k) \in \mathbb{C}^{l\times l} \,, \quad \tilde{D}_{\delta-1,21}(k) \in \mathbb{C}^{(n_y-l)\times l} \,, \quad \tilde{D}_s(k) \in \mathbb{C}^{n_y\times n_y} \,, \\
&\tilde{N}_s(k) \in \mathbb{C}^{n_y\times n_u} \,, \quad \tilde{N}_{\delta,1}(k) \in \mathbb{C}^{l\times n_u} \,, \\
&\delta = \lceil n_x/n_y \rceil \,, \quad l = n_y - (n_y\delta - n_x) \,.
\end{aligned}$$

△

For the LRM, it is preferred to constrain the zeroth degree matrix polynomials instead of the highest degree, as is described in Section 3.1. The MFD parametrization proposed in this paper therefore has a slightly different structure than the one used in literature [32–34] and defined by (29)–(31). The constraint on the zeroth degree matrix for this parametrization takes the form

$$D_0 = \begin{bmatrix} I_l & 0 \\ D_{0,21} & I_{(n_y-l)} \end{bmatrix} . \tag{32}$$

With this alternative constraint, the definition of the proposed MFD parametrization is as follows.

Parametrization 5 (MFD low-order constraint)**.** In the proposed MFD parametrization for local rational modeling, the polynomial matrices $D(\Omega_{k+r})$, $N(\Omega_{k+r})$ and $M(\Omega_{k+r})$ are defined as,

$$D(\Omega_{k+r}) = \begin{bmatrix} I_l & 0 \\ D_{0,21}(k) & I_{(n_y-l)} \end{bmatrix} + \sum_{s=1}^{\delta-1} D_s(k) r^s + \begin{bmatrix} D_{\delta,11}(k) & 0 \\ 0 & 0 \end{bmatrix} r^\delta \,, \tag{33}$$

$$N(\Omega_{k+r}) = \sum_{s=0}^{\delta-1} N_s(k) r^s + \begin{bmatrix} N_{\delta,1}(k) \\ 0 \end{bmatrix} r^\delta \,, \tag{34}$$

$$M(\Omega_{k+r}) = \sum_{s=0}^{\delta-1} M_s(k) r^s + \begin{bmatrix} M_{\delta,1}(k) \\ 0 \end{bmatrix} r^\delta \,, \tag{35}$$

with

$$\begin{aligned}
&D_{0,21}(k) \in \mathbb{C}^{(n_y-l)\times l} \,, \quad D_s(k) \in \mathbb{C}^{n_y\times n_y} \,, \quad D_{\delta,11}(k) \in \mathbb{C}^{l\times l} \,, \\
&N_s(k) \in \mathbb{C}^{n_y\times n_u} \,, \quad N_{\delta,1}(k) \in \mathbb{C}^{l\times n_u} \,, \quad M_s(k) \in \mathbb{C}^{n_y\times 1} \,, \quad M_{\delta,1}(k) \in \mathbb{C}^{l\times 1} \,, \\
&\delta = \lceil n_x/n_y \rceil \,, \quad l = n_y - (n_y\delta - n_x) \,.
\end{aligned}$$

△

Note that in this parametrization $N_0(k)$ and $M_0(k)$ are not equal to $G(\Omega_k)$ and $T(\Omega_k)$, since for this parametrization $D_0(k)$ is not, in general, equal to the identity matrix, so here $G(\Omega_k) = D_0(k)^{-1}N_0(k)$ and $T(\Omega_k) = D_0(k)^{-1}M_0(k)$. The equivalence between parametrization 4 and 5 is established by the following theorem.

Theorem 1. *A bijection exists almost everywhere, i.e. generically, between parametrizations 4 and 5. More specifically, let $\tilde{G}(\Omega_{k+r}) = \tilde{D}(\Omega_{k+r})^{-1}\tilde{N}(\Omega_{k+r})$ be parametrized with parametrization 4 of order n_x, then as long as $\tilde{D}_{0,22}(k)$ and its Schur complement, $\tilde{D}_0(k)/\tilde{D}_{0,22}(k) = \tilde{D}_{0,11}(k) - \tilde{D}_{0,12}(k)\tilde{D}_{0,22}(k)^{-1}\tilde{D}_{0,21}(k)$ are invertible, there is a $G(\Omega_{k+r}) = D(\Omega_{k+r})^{-1}N(\Omega_{k+r})$ parametrized with parametrization 5 of the same order n_x, such that $G(\Omega_{k+r}) = \tilde{G}(\Omega_{k+r})$. Conversely, let $G(\Omega_{k+r})$ be parametrized with parametrization 5 of order n_x, then as long as $D_{\delta,11}(k)$ and $D_{\delta-1,22}(k)$ are invertible, there is a $\tilde{G}(\Omega_{k+r})$ parametrized with parametrization 4 of the same order n_x, such that $G(\Omega_{k+r}) = \tilde{G}(\Omega_{k+r})$.*

Table 1: Number of parameters for the proposed parametrizations.

	Number of parameters, n_θ	Example: $n_u = 4$, $n_y = 8$
LPM	$(n_u+1)n_y + (n_u n_y + n_y)R$	$R = \{1, 2, 3\} \rightarrow n_\theta = \{80, 120, 160\}$
CD	$(n_u+1)n_y + (n_u n_y + n_y + 1)R$	$R = \{1, 2, 3\} \rightarrow n_\theta = \{81, 122, 163\}$
MISO	$(n_u+1)n_y + (n_u n_y + n_y + n_y)R$	$R = \{1, 2, 3\} \rightarrow n_\theta = \{88, 136, 184\}$
MFD full	$(n_u+1)n_y + (n_u n_y + n_y + n_y^2)R$	$R = \{1, 2, 3\} \rightarrow n_\theta = \{144, 248, 352\}$
MFD	$(n_u+1)n_y + (n_u + 1 + n_y)n_x$	$n_x = \{1, 2, 3\} \rightarrow n_\theta = \{53, 66, 79\}$

See Appendix B, for a proof of Theorem 1.

Since parametrization 4 is known to be generic in the sense that it can be used to approximate any system of order n_x up to arbitrary precision, Theorem 1 shows that this property also holds for parametrization 5. The invertibility conditions on $\tilde{D}_{0,22}(k)$ and $\tilde{D}_0(k)/\tilde{D}_{0,22}(k)$ are also satisfied generically, i.e., any system for which these conditions are not satisfied can be approached arbitrarily closely by a system for which they are satisfied, see [32]. To show this, consider $\tilde{D}_0(k)$ for which the invertibility conditions are not satisfied, and $\breve{D}_0(k)$ for which they are satisfied, then for any $\varepsilon > 0$

$$\tilde{D}_{0,n}(k) = \tilde{D}_0(k) + \varepsilon \breve{D}_0(k)\,, \tag{36}$$

does satisfy the invertibility conditions. In other words, an arbitrarily small perturbation of $\tilde{D}_0(k)$ suffices to negate any problematic singularities.

As described in Section 3.1, there are significant advantages to constraining the low-order polynomials instead of the high-order polynomials in the context of the LRM. In particular, consider again that the FRF estimate computed by the LRM is given by

$$G(\Omega_k) = D(\Omega_{k+0})^{-1} N(\Omega_{k+0})\,, \tag{37}$$

where $D(\Omega_{k+0})$ is equal to an arbitrary matrix of estimated parameters when using parametrization 4, and equal to (32), when using parametrization 5, which is easy to invert. Not having to invert an arbitrary matrix of uncertain parameters to obtain the FRF estimate, is a large advantage in the context of the LRM, both computationally and in terms of bias and variance on the estimate. Therefore, in the remainder of this paper, only the MFD parametrization with the low-order constraint, i.e., parametrization 5, is considered and it is simply referred to as the MFD parametrization.

3.6. Evaluation and comparison of parametrizations

As is argued in Section 3.1, the most relevant measure for model complexity in the context of LRM is the number of parameters in the parametrization. This is important because the number of parameters directly translates into a smaller minimal window size, thereby improving the potential for a favorable bias/variance trade-off [6, Section 7.2.2.3]. In Table 1, the number of parameters for the parametrizations proposed in the previous sections are shown, including an example for a system with 4 inputs and 8 outputs. This table clearly shows that the proposed MFD parametrization enables the formulation of a MIMO LRM problem using significantly fewer parameters compared with the alternatives. This is especially true when a system with a large number of inputs and outputs is considered.

An important aspect of the MFD parametrization is its structured nature. The effects of this structure on the solution mechanisms of the LRM need to be addressed. The multivariable LRM can be realized by solving the following linear least squares problem for all k,

$$\hat{\theta}_{\text{lin}}(k) = \arg\min_\theta \sum_{r=-n_W}^{n_W} \|D(\Omega_{k+r}, \theta)Y(k+r) - N(\Omega_{k+r}, \theta)U(k+r) - M(\Omega_{k+r}, \theta)\|_2^2\,. \tag{38}$$

The MFD parametrization enables the use of a lower order system model and fewer parameters by introducing structure in the denominator and numerator matrix polynomials. Due to this structure, the problem can no longer be expressed as in (6) and the standard solution, as in (9), as well as the covariance estimation as presented in, e.g., [18, Sections 3.1 & 5.1], are no longer applicable and need to be modified as is detailed in Appendix C.

The structure in (38) for parametrization 5 also leads to a certain directionality or asymmetry in the linearized identification problem (38). More specifically a distinction between high-order and low-order outputs is made due to the degree structure of the D matrix in this parametrization given by (33). As shown in Section 3.5, in its rational MFD form, i.e. $D(\Omega_{k+r})^{-1}N(\Omega_{k+r})$, parametrization 5 can be used to describe any system of a given order, which means that there is no inherent directionality in this description. However, when considering the linearized terms $D(\Omega_{k+r})Y(k+r)$ and $N(\Omega_{k+r})U(k+r)$, it becomes clear that the column structure of D creates a distinction between the manner in which the first, high-order, outputs are treated in contrast to the remaining, low-order, outputs. This is undesirable since there is typically no preference from a system identification perspective. In the next section, this issue is addressed through iterative algorithms. In particular, these iterative algorithms aim to minimize the original nonlinear least squares problem

$$\hat{\theta}_{\mathrm{nl}}(k) = \arg\min_{\theta} \sum_{r=-n_W}^{n_W} \|Y(k+r) - G(\Omega_{k+r},\theta)U(k+r) - T(\Omega_{k+r},\theta)\|_2^2 , \tag{39}$$

with $G(\Omega_{k+r})$ and $T(\Omega_{k+r})$ given by (17) and (18).

4. Iterative solution methods

The linearization used in the standard non-iterative LRM solution, (38), is known in system identification as the Levy Method [21]. In global system identification this method often leads to biased results due to the additional weighting of the output error, i.e. $V(k)$ in (2), with $D(\Omega_k)$, which due to its polynomial character typically over-emphasizes high-frequency noise contributions. This is especially relevant when a large frequency range is considered and when a high polynomial order for $D(\Omega_k)$ is used. This problem is presumably less severe for the LRM due to the limited frequency window and relatively low polynomial orders that are considered. Because of this, the gain that can be achieved by utilizing iterative solution methods aimed at solving the original output error problem, (39), is expected to be limited for the SISO LRM. This is confirmed in [22, Section 3.4], where an iterative version for the SISO LRM is investigated and it is concluded that the iterative algorithms overall lead to larger errors.

For a multivariable LRM approach, however, the size of the frequency window and the polynomial order can increase significantly. Furthermore, depending on the multivariable parametrization, the multiplication with $D(\Omega_k)$ not only constitutes a frequency dependent weighting, but can also weight different directions differently from others, as is explained in Section 3.6. For the proposed MFD parametrization this effect is the most obvious because of the structure of $D(\Omega_k)$ in this parametrization. Therefore, in the context of the multivariable LRM, it is prudent to reconsider the iterative methods aimed at solving the output error problem (39).

4.1. Iterative algorithms

Iterative algorithms that can be used to achieve this are well-known in the field of global system identification. One of these algorithms is the Sanathanan-Koerner (SK) algorithm [44]. In this algorithm, the additional weighting of the cost function with $D(\Omega_k)$ is compensated by iteratively re-weighting the cost function with an estimate of its inverse, $\hat{D}(\Omega_k)^{-1}$. This estimate is obtained from the solution of the previous iteration. This algorithm is defined as follows.

Algorithm 1 (Sanathanan-Koerner (SK))**.** The algorithm is initialized by an initial parameter vector

$$\hat{\theta}_{\mathrm{SK}}^{\langle 0\rangle}(k) = \theta_{\mathrm{SK,\,init}} , \tag{40}$$

subsequently for a number of iterations, or until convergence is reached, the parameters are updated by

$$\hat{\theta}^{\langle i+1\rangle}_{\mathrm{SK}}(k) = \arg\min_{\theta} \sum_{r=-n_W}^{n_W} \left\| D(\Omega_{k+r}, \hat{\theta}^{\langle i\rangle}_{\mathrm{SK}}(k))^{-1} \left(D(\Omega_{k+r}, \theta) Y(k+r) - N(\Omega_{k+r}, \theta) U(k+r) - M(\Omega_{k+r}, \theta) \right) \right\|_2^2 . \tag{41}$$

This algorithm has been shown to provide good results in parametric identification and fits well within the existing non-iterative LRM framework since the first SK iteration with $D(\Omega_{k+r}, \theta_{\mathrm{SK,\,init}}) = I$, is the same as the standard LRM. However, this algorithm does not necessarily converge, and when convergence is reached it does not converge to a local minimum of (39), see [45].

Another possible approach is a gradient-based approach, such as the Gauss-Newton or Levenberg-Marquart algorithms. This can be used either as an alternative or complementary to the SK algorithm. These gradient-based algorithms enable monotonic convergence to a local minimum of the non-linear cost function, yielding favorable results when an initial estimate of sufficient quality is available. Such an initial estimate can, for example, be obtained from the non-iterative, linearized LRM or from several iterations of the SK algorithm. The Gauss-Newton algorithm is defined as follows.

Algorithm 2 (Gauss-Newton (GN))**.** The algorithm is initialized by an initial parameter vector

$$\hat{\theta}^{\langle 0\rangle}_{\mathrm{GN}}(k) = \theta_{\mathrm{GN,\,init}} , \tag{42}$$

subsequently for a number of iterations, or until convergence is reached, the parameters are updated by

$$\hat{\theta}^{\langle i+1\rangle}_{\mathrm{GN}}(k) = \hat{\theta}^{\langle i\rangle}_{\mathrm{GN}}(k) + \Delta_\theta(k, \hat{\theta}^{\langle i\rangle}_{\mathrm{GN}}(k)) , \tag{43}$$

$$\Delta_\theta(k, \hat{\theta}) = \arg\min_{\Delta} \sum_{r=-n_W}^{n_W} \left\| e(k+r, \hat{\theta}) + J_r(k+r, \hat{\theta}) \Delta \right\|_2^2 , \tag{44}$$

$$e(k+r, \hat{\theta}) = Y(k+r) - G(\Omega_{k+r}, \hat{\theta}) U(k+r) - T(\Omega_{k+r}, \hat{\theta}) , \tag{45}$$

$$J_r(k+r, \hat{\theta}) = \left. \frac{\partial r(k+r, \theta)}{\partial \theta} \right|_{\theta=\hat{\theta}} . \tag{46}$$

The Levenberg-Marquardt algorithm can be interpreted as a damped version of the Gauss-Newton algorithm where the absolute size of Δ in (44) is also considered in the cost function, weighted by a damping factor, λ. How to determine an appropriate value for this damping factor is explained in, e.g., [46].

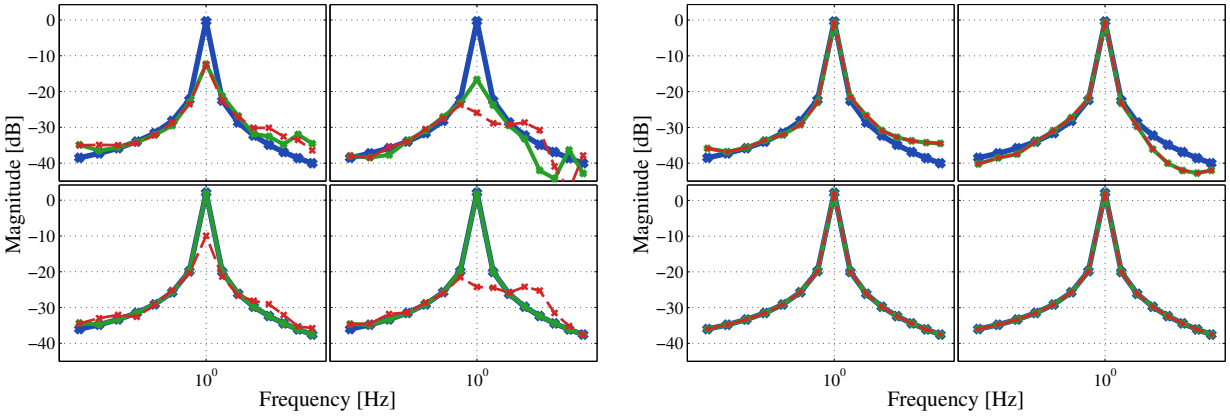
4.2. Numerical example

The following example aims to illustrate the asymmetry in the proposed MFD parametrization, as described in Section 3.6, and to show the potential of the proposed iterative approaches to mitigate this problem and thereby significantly improve the LRM results. In this example, a second-order, 2×2 system is considered, of the form

$$P_{\mathrm{ex}}(s) = \begin{bmatrix} a \\ b \end{bmatrix} \frac{1}{s^2 + 2\zeta\omega_n s + \omega_n^2} \begin{bmatrix} c & d \end{bmatrix} , \tag{47}$$

with $[a,\, b,\, c,\, d,\, \omega_n,\, \zeta] = [0.15,\, 0.2,\, 1,\, 1,\, 2\pi,\, 2.0 \cdot 10^{-3}]$. This system is simulated for $T_{\mathrm{sim}} = 40$ s with a sampling frequency of $f_s = 100$ Hz, and it is excited with zero-mean unit-variance Gaussian white noise. The first output of this system is contaminated with significant white measurement noise yielding a poor signal to noise ratio (SNR) of -2 dB, while the second output is almost noiseless with an SNR of 55 dB.

Using the MFD parametrization, as described in Section 3.5, with $n_x = 1$, this first, low-SNR, output is used as the high-order output. This means the high-order parameters in $N_{\delta,1}(k)$, $M_{\delta,1}(k)$ and $D_{\delta,11}(k)$ are all determined based on this low-SNR output yielding poor estimates. This results in a poor LRM estimate as can be seen from the red line in Figure 4a. Applying iterations in this case leads to significantly improved



(a) bad SNR on the high-order output

(b) good SNR on the high-order output

Figure 4: Simulation example illustrating the need for iterative solution methods. The simulated system (blue), the model identified using the non-iterative LRM (red) and the model identified using the iterative LRM (green) are shown for the cases where, (a) the identification is performed using the bad SNR output as the high-order output, and (b) the identification is performed using the good SNR output as the high-order output.

results, as can seen from the green line in Figure 4a, because as a result of the iterations the estimation of the high-order parameters also uses the information from the good-SNR output.

In Figure 4b, the MFD parametrization is modified such that the good-SNR output is used as the high-order output. In this case, the non-iterative LRM estimate is already accurate. Applying iterations therefore yields no further improvement, and hence the iterative LRM estimate (green line) overlaps with the non-iterative estimate (red line).

The iterative solutions (green lines) in Figures 4a and 4b do not correspond to one another. Ideally, the iterative methods should have converged to the same solution, since the same data is used and the respective parametrizations can be used to describe the same set of systems. A likely reason for this is that the iterative LRM algorithm tends to converge to a local minimum of the cost function, which is not equal to the global minimum. In this example, the noisy nature of the first output makes it likely that such local minima exist. Also, due to the low-SNR of the first output, there is little information in the data which can be used to identify the corresponding FRF elements. In light of this, the large apparent errors in the estimation of these FRF elements are not as remarkable.

5. Simulation results

In this section, the different parametrizations proposed in Section 3 are analyzed on a simulation example. This simulation example enables a clear comparison between the parametrizations due to the precise control over the simulation environment and the fact that the true model is known.

5.1. System description

In this simulation example, a 4×4 MIMO system is considered which is based on the wafer stage setup considered in [25]. The system is a model of a moving square plate with actuators and sensors on the four corners. Only the direction perpendicular to the plate is considered. This means the considered system has three rigid-body modes, one translational mode and two rotational modes. In addition to these rigid body dynamics, a number of lightly damped non rigid-body resonances are present in the system. A Bode magnitude diagram of the system is shown in Figure 5. For further details on this simulation example, see Appendix D.

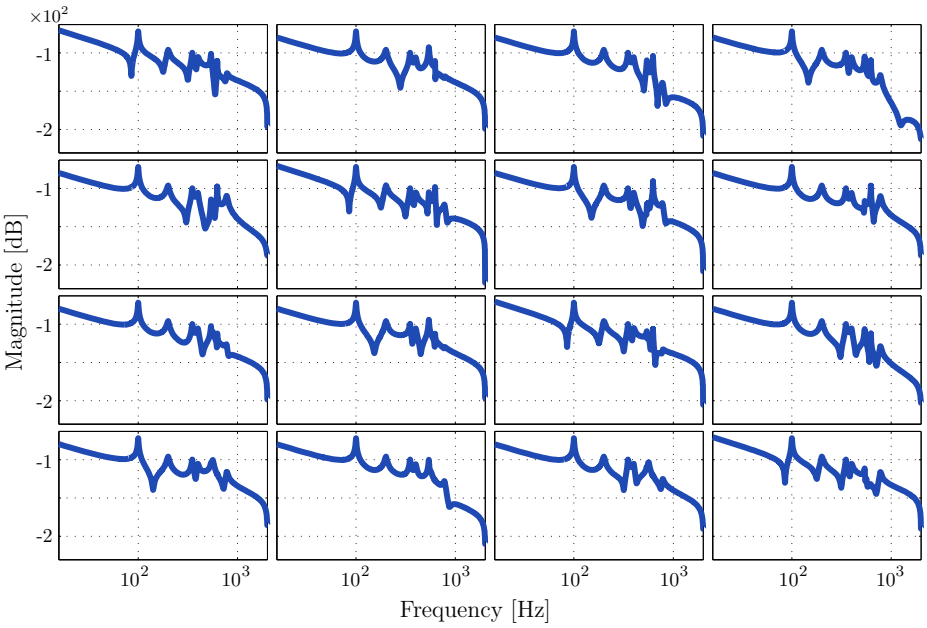


Figure 5: Bode magnitude diagram of the 4×4 simulated system.

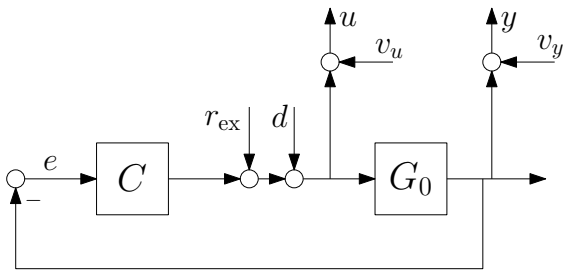


Figure 6: Closed-loop identification scheme.

5.2. Methods

A feedback controller is designed to stabilize the rigid body dynamics of the system. This controller is designed through rigid-body decoupling and decentralized loop-shaping design. The closed-loop bandwidth for the individual control loops, defined as the crossover frequency of the open-loop gain, is designed to be 20 Hz.

The system is simulated in a closed-loop setup for 2 seconds with a sampling frequency of $f_s = 4000$ Hz. The simulation setup is shown in Figure 6. The system is excited by four independently generated Gaussian white noise signals with zero mean and a variance of 1. These excitation signals, r_{ex}, are applied at the system input, additive to the control signals. To simulate process noise, additional Gaussian white noise signals, d, are added to the system at the input with a magnitude of 2% of the excitation signal, i.e., an SNR of 34 dB. To simulate measurement noise, additional white noise signals, v_u and v_y, are added to the simulated system outputs, with an SNR of 60 dB.

From the excitation signals, r_{ex}, and the total simulated inputs, u, and outputs, y, the FRF of the system, G_0, can be determined through the indirect method. In this indirect method the FRF estimate of G_0 is obtained as the quotient of the FRFs of the transfer functions from $r_{\text{ex}} \rightarrow u$ and $r_{\text{ex}} \rightarrow y$, see [6,

Section 7.2.7], i.e.,

$$\hat{G}_{yu}(\Omega_k) = \hat{G}_{yr_{\text{ex}}}(\Omega_k)\hat{G}_{ur_{\text{ex}}}(\Omega_k)^{-1}. \tag{48}$$

Note that both $\hat{G}_{yr_{\text{ex}}}(\Omega_k)$ and $\hat{G}_{ur_{\text{ex}}}(\Omega_k)$ are essentially measured in a traditional, open-loop, setting with a noise-less input r_{ex} and a noisy output y or u, so these FRFs can be estimated using any method that works in this traditional setting, such as the LPM and LRM.

The non-parametric identification is performed using all of the considered methods, i.e., spectral analysis (SA), the LPM and the proposed MIMO LRM methods. Each of these methods is set up to obtain a similar level of noise mitigation. For the local modeling approaches, the level of noise mitigation is determined by the degrees of freedom of the residual in the least squares estimation step [6, Section 7.2.2.2]. This number of degrees of freedom is equal to,

$$\text{dof}_{\text{LPRM}} = \frac{n_{\text{eq}} - n_\theta}{n_y}, \tag{49}$$

where n_{eq} is the number of equations in the estimation problem and n_θ is the number of parameters.

For the simulation results presented in this section, the minimum number of degrees of freedom for all local modeling approaches is set to be equal to the amount of outputs of the simulated closed loop system, i.e.,

$$\text{dof}_{\min} = n_z = n_u + n_y = 8. \tag{50}$$

The number of neighboring frequencies, n_W, to consider in the local models is then selected to be equal to the lowest number for which this constraint is met.

For the spectral analysis methods, a comparable level of noise mitigation is obtained by ensuring the number of considered measurement windows is such that the residual for the spectral analysis method also has $\text{dof}_{\min} = n_z$ degrees of freedom. The degrees of freedom for the spectral analysis residual is equal to,

$$\text{dof}_{\text{SA}} = n_{\text{win}} - n_u, \tag{51}$$

i.e., the number of measurement windows n_{win}, minus the number of inputs [6, Section 7.2.3]. So here, $n_{\text{win}} = 4 + 8 = 12$. The spectral analysis (SA) method considered here uses Hanning windows with 50% overlap to estimate the cross-spectral densities of the signals r_{ex} and u, i.e., $\hat{S}_{U_W R_{\text{ex},W}}$, and of r_{ex} and y, i.e., $\hat{S}_{Y_W R_{\text{ex},W}}$. From these cross-spectral densities, the FRF of the plant G is estimated by

$$\hat{G}_{yu,\text{SA}}(\Omega_k) = \hat{S}_{Y_W R_{\text{ex},W}}(k)\hat{S}_{U_W R_{\text{ex},W}}(k)^{-1}. \tag{52}$$

For details on the spectral analysis approach see, e.g., [6, Section 7.2.3].

The methods described here have all been implemented in MATLAB®. This includes an iterative LRM version which uses the proposed MFD parametrization and where a user-specified number of SK-iterations are used to initialize the LM-algorithm which then runs until convergence or until a user-specified maximum number of iterations is reached.

The criterion used to compare the performance of the different methods is the square root of a sample maximum likelihood cost function, given by

$$V = \left| \frac{1}{n(K_{\text{est}})} \sum_{k \in K_{\text{est}}} e^H_{\text{vec}(\hat{G})}(\Omega_k)\; C^{-1}_{\text{vec}(G_0)}(\Omega_k)\; e_{\text{vec}(\hat{G})}(\Omega_k) \right|^{1/2}, \tag{53}$$

here K_{est} is the set of frequency bins k for which the estimated FRF $\hat{G}(\Omega_k)$ exists and $n(K_{\text{est}})$ is the cardinality of this set. Additionally, in (53)

$$e_{\text{vec}(\hat{G})}(\Omega_k) = \text{vec}(\hat{G}(\Omega_k) - G_0(\Omega_k)), \tag{54}$$

and $G_0(\Omega_k)$ is the true simulated system evaluated at Ω_k, and $\hat{G}(\Omega_k)$ is the estimated FRF. The covariance matrix $C_{\text{vec}(G_0)}(\Omega_k)$ is calculated using the known noise characteristics and transfer functions of the simulated system, details on the calculation of this covariance matrix are provided in Appendix E.

Table 2: Cost function values for different methods applied to the simulated system.

	SA	LPM-1	CD-1	MISO-1	MFD-3	MFD full-1 (= MFD-8)
Cost	19	30	2.7	4.7	2.2	0.59
N_{par}	-	80	81	88	79	144
n_W	-	9	9	9	9	13

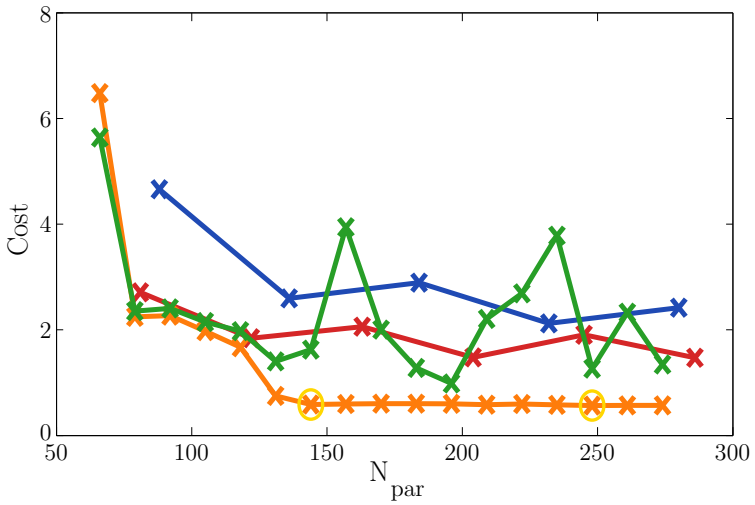


Figure 7: Comparison of different parametrizations. The maximum likelihood cost, as described in Section 6.2, is shown as a function of the number of parameters. The LRM results for the non-iterative MFD (orange), iterative MFD (green), MISO (blue) and Common Denominator (red) parametrizations are shown. The yellow circles denote the MFD orders which coincide with the full MFD parametrization. The considered orders range from $n_x = 2$–18 for the MFD parametrization, $R = 1$–5 for MISO and $R = 1$–6 for the CD parametrization.

5.3. Results

The resulting cost function values for the applied methods are shown in Table 2. For the MFD parametrization the third and eighth order versions are shown, for all other parametrization the $R = 1$ case is shown. Note that full-MFD parametrization with $R = 1$ coincides with the MFD with $n_x = 8$. The cost function values in Table 2 clearly show that for such resonant systems with short data records, the LPM and spectral analysis methods perform worse than the proposed MIMO-LRM methods. Of the LRM methods, the full-MFD parametrization obtains the lowest cost function value. However, it also uses the highest number of parameters and also the largest window size in which to perform the local fits.

To further analyze the proposed LRM parametrizations, the data has been processed using varying orders for all parametrizations. The resulting cost function values are shown in Figure 7, where they are plotted against the number of parameters in the parametrizations.

From Figure 7, it is concluded that all of the proposed parametrizations lead to viable approaches for the multivariable LRM, as the differences between them are small compared with the difference between these LRM methods and alternatives such as the LPM and the spectral analysis method. Also, in this figure one can clearly observe the smaller increments between subsequent orders of the MFD parametrization compared with the other parametrizations, showing the increased flexibility of the proposed MFD parametrization. Furthermore, it can be observed that for the considered data-set, the MISO parametrization performs the worst and the non-iterative MFD parametrization performs the best for a similar number of parameters. Figure 7 also shows that, for this example, the iterative method leads to an increase of the overall cost function, except for the lowest order used. A possible reason for this is that in the considered example

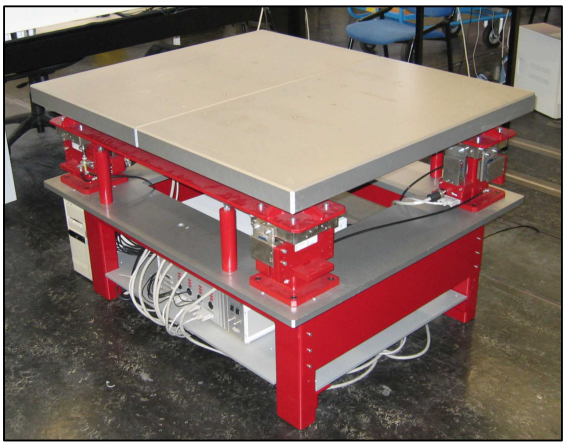


Figure 8: Benchmark active vibration isolation system (AVIS).

the directionality issues, as described in Section 4, are not as relevant. Also, it is likely that the iterative methods still suffer from similar convergence and over-fitting issues as are observed in [22].

6. Benchmark results

In this section experimental data from an active vibration isolation system (AVIS), shown in Figure 8, is used to showcase the advantages of the proposed methods on a real industrial system.

6.1. Benchmark description

The considered experimental setup was recently proposed as an identification benchmark [47], and an extensive set of measurement data for this system is freely available, see [48]. This system is well-suited as an identification benchmark because of its industrial relevance, with a relatively large number of inputs, $n_u = 8$, and outputs, $n_y = 6$, and the high-order resonant dynamics that can be observed in a large frequency range.

This benchmark poses a challenge for both parametric and non-parametric system identification. The challenge of this benchmark for parametric identification lies mainly in the numerical aspects and the ability of the identification scheme to reliably cope with the large system orders and the large amount of data. The challenge for non-parametric identification mainly lies in the efficiency of the identification methods. Even though a very large data set is available for this system, it is the challenge to use as few data as possible and still obtain an accurate non-parametric estimate of the plant dynamics.

6.2. Methods

The data used in this paper is from an open-loop experiment with random non-periodic excitation signals (noise experiment 1 from [48]). In total this measurement is 300 seconds long. Here it is investigated how well the proposed approaches perform with much shorter data records.

The cost function that is used to compare the performance of the different methods is again a sample maximum likelihood cost function, as given by (53). In this case, the reference model, $G_0(\Omega_k)$, and covariance matrix, $C_{\mathrm{vec}(G_0)}(\Omega_k)$, are taken as the spectral analysis estimate obtained from a validation data set using the full 300 seconds of available measurement data. The validation data comes from an experiment performed using a different realization of the excitation signal on the same experimental setup and under the same experimental conditions (noise experiment 2 from [48]). Linear interpolation to the sparsest frequency grid is used for all methods to be able to compare them on equal footing. The methods compared here are a non-iterative LRM approach, using the full MFD parametrization of order 1, i.e. the MFD of 6^{th} order, and the spectral analysis approach, using Hanning windows, with 50% overlap.

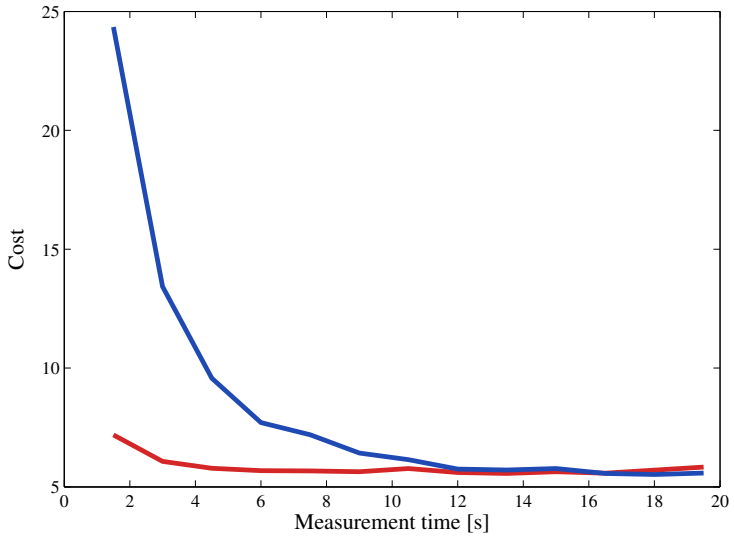


Figure 9: Overall performance of the spectral analysis method versus the LRM for the AVIS data. The maximum likelihood cost, as described in Section 6.2, is shown as a function of the measurement time used to obtain the FRF estimates.

6.3. Results

The main result of this investigation is shown in Figure 9, in this figure the cost for the spectral analysis and the full MFD LRM method are shown versus the used measurement time in seconds. This figure clearly shows that the LRM is much more efficient for short data records than the spectral analysis method. Using the LRM it is possible to obtain a very accurate non-parametric estimate of the FRF using just 4 seconds of measurement time whereas with the spectral analysis method more than 10 seconds of data is needed to obtain an estimate of comparable quality. Also note that with the spectral analysis method, averaging steps are needed to reduce the variance on the estimate. These averaging steps also reduce the frequency resolution of the estimated FRF. For the LRM the frequency resolution is not reduced, so the LRM provides an estimate on a much denser frequency grid than the spectral analysis method using the same amount of measurement time. To expand on these results, some examples are shown which highlight the differences between the spectral analysis and LRM estimates.

Figure 10 shows the full 6×8 bode magnitude plot of the reference model and the LRM result using just 1.5 seconds of measurement time, including the standard deviations of both these estimates. On a whole these two FRF estimates are in good agreement. In the low-frequency range, below 6 Hz, the estimates do differ but here both estimates are unreliable due to the fact that the geophone sensors of the system do not perform as well in this frequency range leading to a poor signal to noise ratio, this is reflected by the high standard deviations for both estimates in this frequency region. Additionally, both estimates have relatively few estimated points in this range as the resolution for the LRM using 1.5s of data is 2/3 Hz and the resolution for the SA estimate is 0.2 Hz. The good agreement of both estimate in mid-higher frequency range shows that for such a large scale industrial system the LRM is indeed capable of efficiently estimating the FRF.

Figure 11a shows a the Bode magnitude plot for the $G_{(1,1)}$ element of the full identified model G, estimated using the shortest data record of 1.5 seconds. This figure shows the reference model as well as the estimates obtained using the LRM and the spectral analysis methods as well as the standard deviations of all these estimates. For this short data record, there is a clear difference in the quality of the LRM and spectral analysis estimates, with the LRM being superior to the spectral analysis method. This difference is especially pronounced around the resonances, as can be seen in the zoomed-in plot in Figure 11b. It is clear from this figures that around the resonance peaks the spectral analysis method provides a poor estimate of the FRF whereas the LRM method already yields relatively accurate results. For many applications, such as

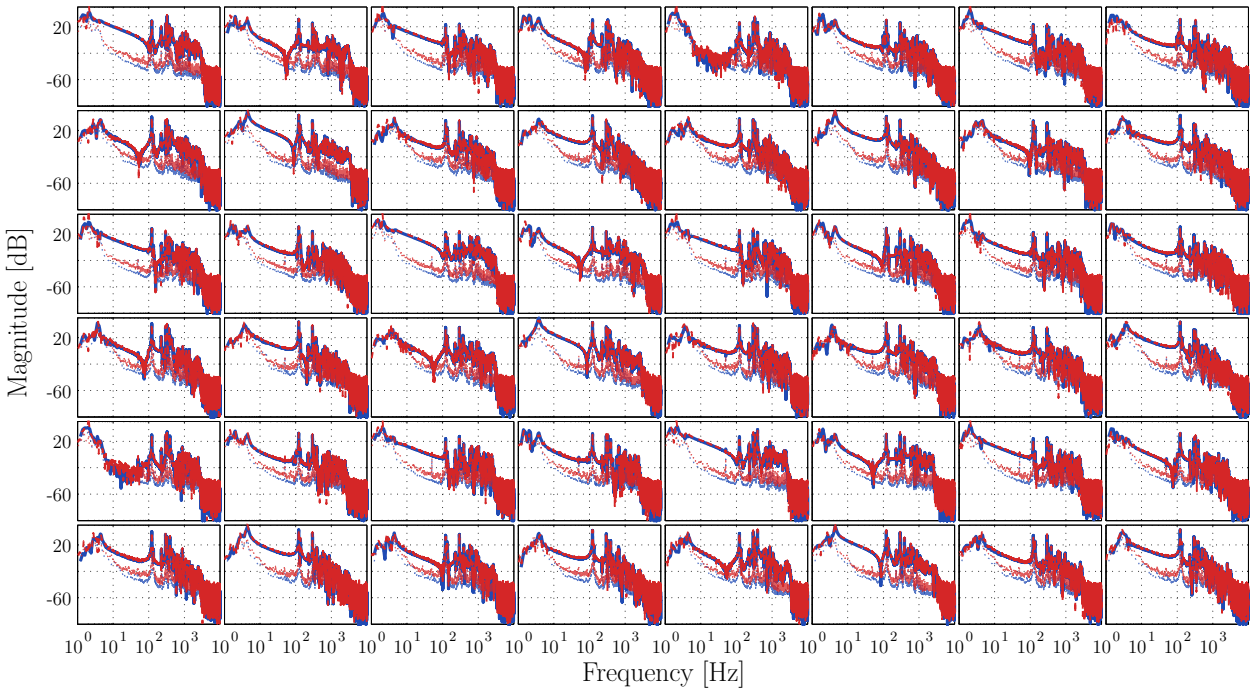
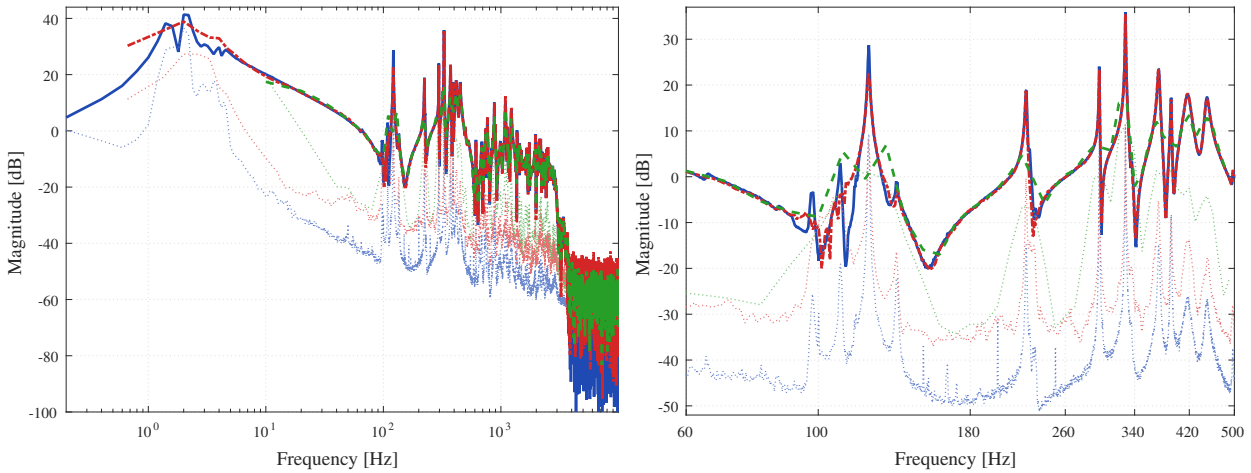
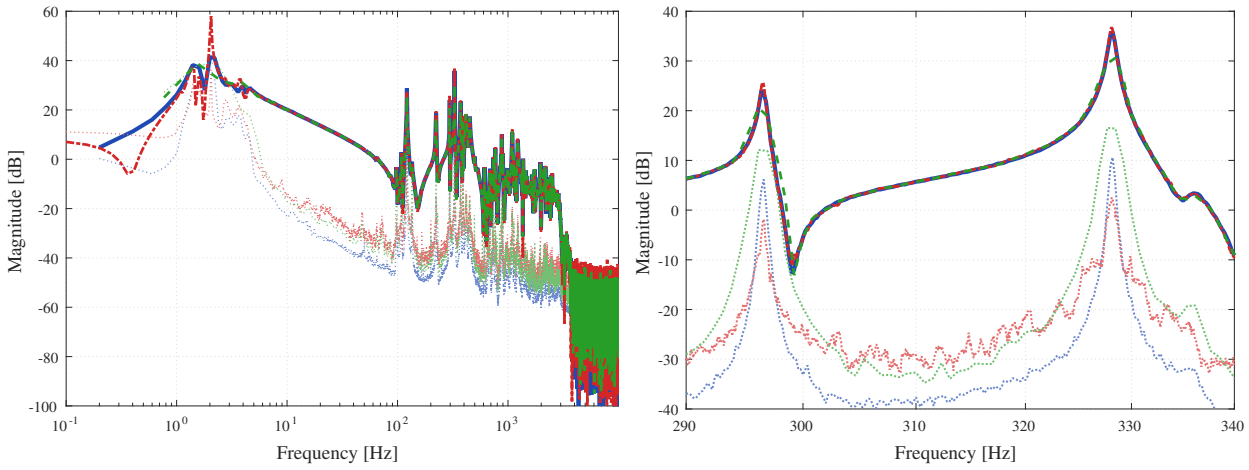


Figure 10: Full 6×8 Bode magnitude plot of the AVIS system. The reference model (blue solid) and the LRM estimate (red dashed) and their respective estimated standard deviations (same colors, thin-dotted) using only 1.5 seconds of measurement data, are shown. This figure shows that the multivariable LRM approach enables the efficient identification of the full 6×8 FRF from a short data record of just 1.5 seconds with good accuracy.



(a) Full view (b) Zoomed-in view

Figure 11: Bode magnitude plot of the $G_{(1,1)}$ element of the multivariable model G, showing the reference model (blue solid), the spectral analysis estimate (green dashed) and the LRM estimate (red dash-dotted) and their respective estimated standard deviations (same colors, thin-dotted) using 1.5 seconds of measurement data. For this short data record, the SA method provides a poor estimate, especially around the resonances, where clearly a higher frequency resolution and longer experiment time are required to obtain an adequate FRF estimate. The proposed multivariable LRM estimate already provides an accurate estimate for this short data record of 1.5 seconds, showing the efficiency and the potential of the proposed methods.



(a) Full view (b) Zoomed-in view

Figure 12: Bode magnitude plot of the $G_{(1,1)}$ element of the multivariable model G, showing the reference model (blue solid), the spectral analysis result (green dashed) and the LRM result (red dash-dotted) and their respective estimated standard deviations (same colors, thin-dotted) for 19.5 seconds of measurement data. In the full view (a), no clear difference in the quality of the SA and LRM estimates is observed. In the zoomed-in view, a difference in estimation quality can be observed around the resonance peaks due to the superior frequency resolution of the LRM estimate.

control and modal analysis, it is especially important to obtain an accurate non-parametric estimate around the resonances so this is a significant result. Around 100 Hz there are still significant discrepancies between the LRM estimate and the reference model, this can be explained from the limited frequency resolution of the LRM estimate when using only 1.5 s of measurement data as well as the large amount of relatively small resonances around this frequency range. When more measurement time is used the resolution of the LRM estimate increases and these discrepancies mostly disappear.

In Figure 12a the Bode magnitude plot is shown for the results with the data record of 19.5 seconds. In this figure no clear overall difference in the quality of the SA and LRM estimates of the FRF is observed. However, when zooming in on the resonance peaks as is done in Figure 12b, a difference in estimation quality can still be seen. In fact, at the resonance frequencies the LRM estimate using 19.5 seconds of measurement data is of a similar quality as the reference model that is obtained using 300 seconds of measurement data. A possible reason why this superior estimation quality of the LRM around the resonances is not reflected in the cost function values, as shown in Figure 9, is that in the calculation of the cost function the resolution of the LRM estimate is down-sampled to the resolution of the SA estimate. When considering the full resolution that is obtained using the LRM, this difference in estimation quality is much clearer.

7. Conclusions and outlook

The goal of the present paper is to develop methods for accurate and fast frequency response function identification for multivariable systems with lightly-damped complex dynamics. The local rational modeling approach that is explored and extended in this paper is a powerful method to achieve this goal.

In this paper, first, the potential of the local parametric modeling approach for FRF identification of mechanical systems is illustrated through a closed-loop identification example which shows the importance of improved transient suppression for such applications. Also, the advantages of using a rational parametrization are illustrated for cases where the system contains lightly damped resonant dynamics.

As the key contribution, this paper points out the freedom in the parametrization of the multivariable LRM and investigates a number of possible parametrizations. These parametrizations can all be represented as left matrix fraction descriptions (LMFDs) with varying structures of the denominator matrix. From

simulations of a representative motion system, it can be concluded that all the proposed parametrizations are viable, yielding improved results when compared with classical methods. Furthermore, from an analysis based on experimental data from a recent system identification benchmark, it can be concluded that the proposed multivariable LRM methods are significantly more efficient than classical spectral analysis methods.

The most parsimonious and flexible of the proposed parametrizations, is also the most structured. As a consequence of its structure, this parametrization introduces a certain directionality when an equation error criterion in the sense of Levy [21] is considered, as is often done in the local modeling approach. More specifically, a distinction can be seen between the way certain high-order outputs are handled as opposed to the remaining low-order outputs. This directionality is generally undesirable. It can be mitigated by applying iterative approaches aimed at solving the output-error LRM problem. Such iterative methods have been implemented and investigated, showing some promising results but ultimately still yielding larger overall estimation errors than the equivalent non-iterative versions of the LRM. These increased errors can most likely be attributed to over-fitting and convergence issues of the iterative approaches, as has also been observed in [22]. Additional research is required to investigate when and how iterative methods can be applied to benefit the overall accuracy of the LRM.

A potential avenue to improve the performance of iterative methods could be to incorporate an appropriate method for order selection to alleviate the problems with over-fitting. Also, for the investigated iterative methods, the definition of the cost function is an essential step. Therefore, another potential avenue for improvement of the proposed iterative LRM methods, is to further explore the freedom in the definition of this cost function, e.g., by applying appropriate weighting specifically aimed at obtaining a better non-parametric estimate of the FRF. Further ongoing research on the local rational method involves a detailed investigation of the bias and variance on the obtained FRF estimates for the proposed parametrizations.

Acknowledgment

The authors would like to thank Robin de Rozario, and Maarten Steinbuch from the Eindhoven University of technology, and Dieter Verbeke, Egon Geerardyn, and Johan Schoukens from the Vrije Universiteit Brussel for their contributions to this paper and the ongoing research collaboration on this topic. Furthermore, the authors gratefully acknowledge one of the anonymous reviewers for pointing out an important step regarding the influence of noise in Appendix C which has improved the result. This research is supported by the TU/e Impuls program and ASML research mechatronics as well as the research program VENI with project number 13073, which is (partly) financed by the Netherlands Organization for Scientific Research (NWO).

References

[1] P. Verboven, E. Parloo, B. Cauberghe, P. Guillaume, Improved modal parameter estimation for lowly damped systems using non-parametric exponential windowing techniques, Mechanical Systems and Signal Processing 19 (4) (2005) 675–699.
[2] L. Hermans, H. van der Auweraer, Modal testing and analysis of structures under operational conditions: Industrial applications, Mechanical Systems and Signal Processing 13 (2) (1999) 193–216.
[3] J. Noël, M. Schoukens, F-16 aircraft benchmark based on ground vibration test data, in: 2017 Workshop on Nonlinear System Identification Benchmarks, April 24-26, 2017.
[4] M. Steinbuch, M. Norg, Advanced motion control: An industrial perspective, European Journal of Control 4 (4) (1998) 278–293.
[5] E. Geerardyn, M. L. D. Lumori, J. Lataire, FRF smoothing to improve initial estimates for transfer function identification, IEEE Transactions on Instrumentation and Measurement 64 (10) (2015) 2838–2847.
[6] R. Pintelon, J. Schoukens, System identification: a frequency domain approach, 2nd Edition, Wiley-IEEE press, Hoboken, NJ, USA, 2012.
[7] T. Oomen, R. van Herpen, S. Quist, M. van de Wal, O. Bosgra, M. Steinbuch, Connecting system identification and robust control for next-generation motion control of a wafer stage, IEEE Transactions on Control Systems Technology 22 (1) (2014) 102–118.
[8] R. van Herpen, T. Oomen, E. Kikken, M. Van De Wal, W. Aangenent, M. Steinbuch, Exploiting additional actuators and sensors for nano-positioning robust motion control, Proceedings of the American Control Conference 24 (6) (2014) 984–990.
[9] G. Jenkins, D. Watts, Spectral analysis and its applications, Holden-Day series in time series analysis, Holden-Day, 1969.
[10] G. Box, G. Jenkins, Time series analysis: forecasting and control, Holden-Day series in time series analysis, Holden-Day, 1976.

[11] J. Bendat, A. Piersol, Random Data: Analysis and Measurement Procedures, 4th Edition, Wiley Series in Probability and Statistics, Wiley, 2011.
[12] L. Ljung, On the estimation of transfer functions, Automatica 21 (6) (1985) 677–696.
[13] R. Pintelon, J. Schoukens, Measurement of frequency response functions using periodic excitations, corrupted by correlated input/output errors, IEEE Transactions on Instrumentation and Measurement 50 (6) (2001) 1753–1760.
[14] J. Schoukens, R. Pintelon, T. Dobrowiecki, Y. Rolain, Identification of linear systems with nonlinear distortions, Automatica 41 (3) (2005) 491–504.
[15] P. Hägg, J. Schoukens, M. Gevers, H. Hjalmarsson, The transient impulse response modeling method for non-parametric system identification, Automatica 68 (2016) 314–328.
[16] J. Lataire, T. Chen, Transfer function and transient estimation by gaussian process regression in the frequency domain, Automatica 72 (2016) 217–229.
[17] J. Schoukens, G. Vandersteen, K. Barbe, R. Pintelon, Nonparametric preprocessing in system identification: a powerful tool, European Journal of Control 15 (3-4) (2009) 260–274.
[18] R. Pintelon, J. Schoukens, G. Vandersteen, K. Barb, Estimation of nonparametric noise and FRF models for multivariable systems–part I: Theory, Mechanical Systems and Signal Processing 24 (3) (2010) 573–595.
[19] R. Pintelon, G. Vandersteen, J. Schoukens, Y. Rolain, Improved (non-)parametric identification of dynamic systems excited by periodic signals – the multivariate case, Mechanical Systems and Signal Processing 25 (8) (2011) 2892–2922.
[20] T. McKelvey, G. Guérin, Non-parametric frequency response estimation using a local rational model, IFAC Proceedings Volumes 45 (16) (2012) 49–54, 16th IFAC Symposium on System Identification.
[21] E. C. Levy, Complex-curve fitting, IRE Transactions on Automatic Control AC-4 (1) (1959) 37–43.
[22] E. Geerardyn, Development of user-friendly system identification techniques, Ph.D. thesis, Vrije Universiteit Brussel (2016).
[23] E. Geerardyn, T. Oomen, J. Schoukens, Enhancing $\mathcal{H}_\infty$ norm estimation using local LPM/LRM modeling: Applied to an AVIS, IFAC Proceedings Volumes 47 (3) (2014) 10856–10861, 19th IFAC World Congress.
[24] R. van der Maas, A. van der Maas, J. Dries, B. de Jager, Efficient nonparametric identification for high-precision motion systems: A practical comparison based on a medical x-ray system, Control Engineering Practice 56 (2016) 75–85.
[25] R. Voorhoeve, R. de Rozario, T. Oomen, Identification for motion control: Incorporating constraints and numerical considerations, in: 2016 IEEE American Control Conference, 2016, pp. 6209–6214.
[26] J. Doyle, G. Stein, Multivariable feedback design: Concepts for a classical/modern synthesis, IEEE Transactions on Automatic Control 26 (1) (1981) 4–16.
[27] T. Oomen, M. van de Wal, O. Bosgra, Design framework for high-performance optimal sampled-data control with application to a wafer stage, International Journal of Control 80 (2007) 919–934.
[28] J. Schoukens, G. Vandersteen, R. Pintelon, Z. Emedi, Y. Rolain, Bounding the polynomial approximation errors of frequency response functions, IEEE Transactions on Instrumentation and Measurement 62 (5) (2013) 1346–1353.
[29] R. Pintelon, Y. Rolain, MIMO lowly damped system: Free-free vibrations of an aluminium tooling plate, IFAC-PapersOnLine 48 (28) (2015) 1466–1470.
[30] D. Verbeke, J. Schoukens, Local parametric modeling based on rational approximation, in: 36th Benelux Meeting on Systems and Control, 2017, p. 143.
[31] T. Kailath, Linear systems, Prentice Hall information and system sciences series, Prentice-Hall, Englewood Cliffs, 1980.
[32] K. Glover, J. C. Willems, Parametrizations of linear dynamical systems: Canonical forms and identifiability, IEEE Transactions on Automatic Control 19 (6) (1974) 640–646.
[33] J. Vayssettes, G. Mercère, P. Vacher, R. De Callafon, Frequency-domain identification of aircraft structural modes from short-duration flight tests, International Journal of Control 87 (7) (2014) 1352–1372.
[34] M. de Mathelin, M. Bodson, Canonical vs pseudo-canonical forms for the structural and parametric identification of multivariable systems, Proceedings of the European Control Conference, Grenoble, France 1 (1991) 1282–1287.
[35] G. Corrêa, K. Glover, Pseudo-canonical forms, identifiable parametrizations and simple parameter estimation for linear multivariable systems: Input-output models, Automatica 20 (4) (1984) 429–442.
[36] A. Van Overbeek, L. Ljung, On-line structure selection for multivariable state-space models, Automatica 18 (5) (1982) 529–543.
[37] M. Gevers, V. Wertz, Uniquely identifiable state-space and arma parametrizations for multivariable linear systems, Automatica 20 (3) (1984) 333–347.
[38] R. P. Guidorzi, Invariants and canonical forms for systems structural and parametric identification, Automatica 17 (1) (1981) 117–133.
[39] B. D. Moor, M. Gevers, G. C. Goodwin, L_2-overbiased, L_2-underbiased and L_2-unbiased estimation of transfer functions, Automatica 30 (5) (1994) 893–898.
[40] B. Cauberghe, Applied frequency-domain system identification in the field of experimental and operational modal analysis, Ph.D. thesis, Vrije Universiteit Brussel (2004).
[41] R. van Herpen, T. Oomen, M. Steinbuch, Optimally conditioned instrumental variable approach for frequency-domain system identification, Automatica 50 (9) (2014) 2281–2293.
[42] T. Söderström, P. Stoica, System Identification, Prentice Hall, Hemel Hempstead, UK, 1989.
[43] J. Vayssettes, G. Mercère, O. Prot, New developments for matrix fraction descriptions: A fully-parametrised approach, Automatica 66 (2016) 15 – 24.
[44] C. K. Sanathanan, J. Koerner, Transfer function synthesis as a ratio of two complex polynomials, IEEE Transactions on Automatic Control 8 (1) (1963) 56–58.
[45] A. H. Whitfield, Asymptotic behaviour of transfer function sythesis methods, International Journal of Control 45 (3)

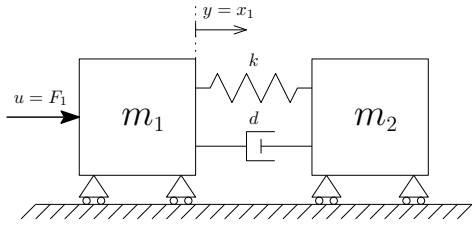


Figure A.1: 2-mass sprig damper system

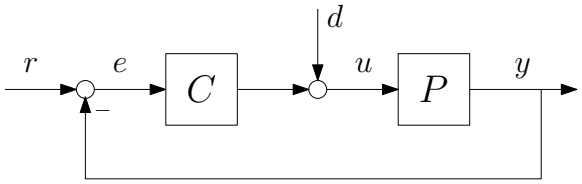


Figure A.2: Feedback scheme

(1987) 1083–1092.

[46] L. Knockaert, F. Ferranti, T. Dhaene, Vector fitting vs. Levenberg-Marquardt: Some experiments, in: IEEE Workshop on Signal Propagation on Interconnects, 2009, pp. 1–4.

[47] R. Voorhoeve, A. van Rietschoten, E. Geerardyn, T. Oomen, Identification of high-tech motion systems: An active vibration isolation benchmark, in: 17th IFAC Symposium on System Identification, Beijing, China, 2015, pp. 1250–1255.

[48] R. Voorhoeve, A. van der Maas, System identification (SYSID) benchmark for an active vibration isolation system (AVIS), 4TU-datacenter: `https://data.4tu.nl/` (December 2016). doi:10.4121/uuid:494e738d-e2aa-49e4-b076-ac96d3a142e8.

Appendix A. Details of the simulation examples from Section 2

The simulation examples of Section 2 are based on a 2-mass-spring-damper system with a single unconstrained rigid-body degree of freedom, as depicted in Figure A.1. In these examples the input, which is the actuation force on the first mass, i.e., F_1, and the output, which is the position of the first mass, are collocated yielding the following equation for the system,

$$P_{\text{col}} = c\frac{m_2 s^2 + ds + k}{m_1 m_2 s^4 + (m_1 + m_2)ds^3 + (m_1 + m_2)ks^2}, \tag{A.1}$$

with parameters $[c,\, k,\, d,\, m_1,\, m_2] = [5.6 \cdot 10^3,\, 5.4 \cdot 10^4,\, 2.0,\, 1.1,\, 0.9]$. In both examples, the system is controlled in feedback as shown in Figure A.2.

In the example of Section 2.2, the controller is tuned such that the bandwidth of the controlled system is equal to 25 Hz. Here, the bandwidth is taken as the point where the loop gain $L = PC$, first crosses the 0 dB threshold from above, i.e., when the loop gain enters the unit circle. Furthermore, in this example, the controller is tuned such that the peak of the sensitivity function, $S = (1 + L)^{-1}$, is below 8 dB. To achieve this, the controller is designed using a lead filter, with a zero frequency of 10 Hz and a pole frequency of 40 Hz, and a gain of 5.

In the example of Section 2.3, a different controller is designed. For this example the goal is to show the difference between the LPM and LRM in estimation quality around lightly damped resonances. Therefore, the controller is tuned such that the sensitivity function had a large, lightly damped, resonance peak. To achieve this, the controller in this example is simply a gain of 10.

In both examples the sample frequency is $f_s = 250$ Hz and the simulation time is $T_{\text{sim},2.2} = 40$ s for the example of Section 2.2 and $T_{\text{sim},2.3} = 2.5$ s for the example of Section 2.3. As excitation signals, d, one period of full-grid random phase multisines are used, i.e. a signal with a constant amplitude for all frequencies in the DFT grid and with a random phase. All discretizations are performed using the Tustin method without prewarping.

Appendix B. Proof of Theorem 1

The bijection between parametrizations 4 and 5 can be shown by the existence of a unimodular transformation which transforms either formulation into the other, i.e.,

$$\tilde{D} = QD, \quad \tilde{N} = QN, \quad G = \tilde{D}^{-1}\tilde{N} = D^{-1}Q^{-1}QN = D^{-1}N\,, \tag{B.1}$$

with $\det(Q) = c \in \mathbb{C} \setminus \{0\}$.

Since the transformation involves a pre-multiplication with Q, the column-degree structure of all polynomial matrices is maintained. To maintain the row-degree structure of the $N(\Omega_{k+r})$ and $M(\Omega_{k+r})$ matrices in both parametrizations, the unimodular matrix, Q, must have the following upper diagonal block structure,

$$Q = \begin{bmatrix} Q_{11} & Q_{12} \\ 0 & Q_{22} \end{bmatrix} . \tag{B.2}$$

This guarantees that the structure of the highest degree matrix of $N(\Omega_{k+r})$ is maintained, i.e.,

$$QN_\delta = \begin{bmatrix} Q_{11} & Q_{12} \\ 0 & Q_{22} \end{bmatrix} \begin{bmatrix} N_{\delta,1} \\ 0 \end{bmatrix} = \begin{bmatrix} \tilde{N}_{\delta,1} \\ 0 \end{bmatrix} = \tilde{N}_\delta \, . \tag{B.3}$$

First, consider the case of transforming a system parametrized with parametrization 4 to the equivalent system parametrized with parametrization 5. From the zeroth order matrix polynomial of $\tilde{D}(\Omega_{k+r})$ one can obtain a formulation for the unimodular transformation by taking,

$$\begin{aligned} \tilde{D}_0 &= \begin{bmatrix} \tilde{D}_{0,11} & \tilde{D}_{0,12} \\ \tilde{D}_{0,21} & \tilde{D}_{0,22} \end{bmatrix} = \begin{bmatrix} I & \tilde{D}_{0,12}\tilde{D}_{0,22}^{-1} \\ 0 & I \end{bmatrix} \begin{bmatrix} \tilde{D}_0/\tilde{D}_{0,22} & 0 \\ 0 & \tilde{D}_{0,22} \end{bmatrix} \begin{bmatrix} I & 0 \\ \tilde{D}_{0,22}^{-1}\tilde{D}_{0,21} & I \end{bmatrix} \\ &= \begin{bmatrix} \tilde{D}_0/\tilde{D}_{0,22} & \tilde{D}_{0,12} \\ 0 & \tilde{D}_{0,22} \end{bmatrix} D_0 = QD_0 \, , \end{aligned} \tag{B.4}$$

where $\tilde{D}_0/\tilde{D}_{0,22}$ is the Schur complement of $\tilde{D}_{0,22}$ in $\tilde{D}_0$, i.e.,

$$\tilde{D}_0/\tilde{D}_{0,22} = \tilde{D}_{0,11} - \tilde{D}_{0,12}\tilde{D}_{0,22}^{-1}\tilde{D}_{0,21} \, ,$$

In this case,

$$Q^{-1} = \begin{bmatrix} \tilde{D}_0/\tilde{D}_{0,22} & \tilde{D}_{0,12} \\ 0 & \tilde{D}_{0,22} \end{bmatrix}^{-1} = \begin{bmatrix} (\tilde{D}_0/\tilde{D}_{0,22})^{-1} & -(\tilde{D}_0/\tilde{D}_{0,22})^{-1}\,\tilde{D}_{0,12}\,\tilde{D}_{0,22}^{-1} \\ 0 & \tilde{D}_{0,22}^{-1} \end{bmatrix} , \tag{B.5}$$

can be used to transform parametrization 4 to 5, under the condition that $\tilde{D}_{0,22}$ and $\tilde{D}_0/\tilde{D}_{0,22}$ are invertible.

Next, consider the case of transforming a system parametrized with parametrization 5 to the equivalent system parametrized with parametrization 4. Here, the matrices of highest column degree are considered, yielding

$$D_{ch} = \begin{bmatrix} D_{\delta,11} & D_{\delta-1,12} \\ 0 & D_{\delta-1,22} \end{bmatrix} = Q^{-1}\tilde{D}_{ch} = Q^{-1}I = Q^{-1} \, . \tag{B.6}$$

In this case,

$$Q = \begin{bmatrix} D_{\delta,11} & D_{\delta-1,12} \\ 0 & D_{\delta-1,22} \end{bmatrix}^{-1} = \begin{bmatrix} D_{\delta,11}^{-1} & -D_{\delta,11}^{-1}\,D_{\delta-1,12}\,D_{\delta-1,22}^{-1} \\ 0 & D_{\delta-1,22}^{-1} \end{bmatrix} , \tag{B.7}$$

can be used to transform parametrization 5 to 4, under the condition that $D_{\delta,11}$ and $D_{\delta-1,22}$ are invertible.

Appendix C. LRM solution and covariance estimation for parametrization 5

Due to the structure in parametrization 5, rewriting the LRM problem as in (6) yields

$$Y(r) = \Theta K(r) + \tilde{V}(r) \, , \tag{C.1}$$

$$K(r) = \begin{bmatrix} K_1(r) \otimes U(r) \\ K_1(r) \\ Y_{1:l}(r) \\ K_2(r) \otimes Y(r) \\ Y_{1:l}(r) r^{\delta} \end{bmatrix}, \qquad K_1(r) = \begin{bmatrix} 1 \\ r \\ \vdots \\ r^{\delta} \end{bmatrix}, \qquad K_2(r) = \begin{bmatrix} r \\ \vdots \\ r^{\delta-1} \end{bmatrix},$$

$$\Theta = \begin{bmatrix} \Theta_N & \Theta_T & \Theta_D \end{bmatrix}, \qquad \Theta_N = \begin{bmatrix} \Theta_{N_{0:\delta-1}} \begin{bmatrix} \Theta_{N_{\delta,1}} \\ 0 \end{bmatrix} \end{bmatrix}, \qquad \Theta_T = \begin{bmatrix} \Theta_{T_{0:\delta-1}} \begin{bmatrix} \Theta_{T_{\delta,1}} \\ 0 \end{bmatrix} \end{bmatrix},$$

$$\Theta_D = \begin{bmatrix} \begin{bmatrix} 0 \\ \Theta_{D_{0,21}} \end{bmatrix} & \Theta_{D_{1:\delta-1}} & \begin{bmatrix} \Theta_{D_{\delta,11}} \\ 0 \end{bmatrix} \end{bmatrix}, \tag{C.2}$$

where the frequency argument k has been omitted for notational clarity and where $\tilde{V}(r) = D(r)V(r)$ is treated as an unknown noise source. This problem can not be solved as in (9) due to the elements of Θ that are equal to zero. In vectorized form these parameter constraints can be included by eliminating the columns of the regression matrix K corresponding to the parameters θ that are equal to zero. Therefore the LRM problem for this parametrization can be written as

$$Y(r) = \mathcal{K}(r)\theta + \tilde{V}(r), \tag{C.3}$$

where $\mathcal{K}(r) = I_{n_u} \otimes K(r)$ and $\theta = \text{vec}(\Theta)$ where the trivial rows (zero-elements) of θ and the corresponding columns of $\mathcal{K}(r)$ have been removed. The least squares solution to (C.3), as defined in (38), is given by,

$$\hat{\theta} = \mathcal{K}_n^{+} \mathcal{Y}_n, \tag{C.4}$$

where $\mathcal{K}_n = \begin{bmatrix} \mathcal{K}(-n_W)^T & \mathcal{K}(-n_W+1)^T & \cdots & \mathcal{K}(n_W)^T \end{bmatrix}^T$ and $\mathcal{Y}_n = \begin{bmatrix} Y(-n_W)^T & \cdots & Y(n_W)^T \end{bmatrix}^T$ and with $A^{+} = (A^H A)^{-1} A^H$ the left Moore-Penrose pseudoinverse.

To estimate the covariance of the LRM estimate with parametrization 5, consider the LRM estimate as given by (37) with D_0 as in (32), i.e.,

$$\hat{G} = \hat{D}_0^{-1} \hat{N}_0 = \begin{bmatrix} I_l & 0 \\ -\hat{D}_{0,21} & I_{n_y-l} \end{bmatrix} \begin{bmatrix} \hat{N}_{0,1} \\ \hat{N}_{0,2} \end{bmatrix}. \tag{C.5}$$

Defining $\Delta\hat{X} = (\hat{X} - X)$ and assuming $\mathbb{E}(\hat{G}) = G$, the covariance of the estimate is given by

$$\text{Cov}(\text{vec}(\hat{G})) = \mathbb{E}(\text{vec}(\Delta\hat{G})\text{vec}(\Delta\hat{G})^H), \tag{C.6}$$

with

$$\Delta\hat{G} = \begin{bmatrix} I_l & 0 \\ -D_{0,21} & I_{n_y-l} \end{bmatrix} \Delta\hat{N}_0 - \begin{bmatrix} 0_{l \times n_y - l} \\ I_{n_y-l} \end{bmatrix} \Delta\hat{D}_{0,21} \left(N_{0,1} + \Delta\hat{N}_{0,1} \right). \tag{C.7}$$

Using a first order approximation this yields,

$$\text{vec}(\Delta\hat{G}) \approx \text{vec}\left(\begin{bmatrix} I_l & 0 \\ -D_{0,21} & I_{n_y-l} \end{bmatrix} \Delta\hat{N}_0 \right) - \text{vec}\left(\begin{bmatrix} 0_{l \times n_y - l} \\ I_{n_y-l} \end{bmatrix} \Delta\hat{D}_{0,21} N_{0,1} \right). \tag{C.8}$$

Using $\text{vec}(AXB) = \left(B^T \otimes A\right) \text{vec}(X)$, the two terms on the right-hand side of (C.8) can be rewritten as

$$\text{vec}\left(\begin{bmatrix} I_l & 0 \\ -D_{0,21} & I_{n_y-l} \end{bmatrix} \Delta\hat{N}_0 \right) = \left[\left(I_{n_u} \otimes \begin{bmatrix} I_l & 0 \\ -D_{0,21} & I_{n_y-l} \end{bmatrix} \right) \quad 0_{n_y n_u \times n_\theta - n_y n_u} \right] \Delta\hat{\theta} \tag{C.9}$$

$$\text{vec}\left(\begin{bmatrix} 0_{l \times n_y - l} \\ I_{n_y-l} \end{bmatrix} \Delta\hat{D}_{0,21} N_{0,1} \right) = \left[0_{n_y n_u \times n_{\theta,N,T}} \quad \left(N_{0,1}^T \otimes \begin{bmatrix} 0_{l \times n_y - l} \\ I_{n_y-l} \end{bmatrix} \right) \quad 0_{n_y n_u \times n_{\theta,D} - l(n_y-l)} \right] \Delta\hat{\theta}, \tag{C.10}$$

where $n_{\theta,N,T}$ is the total number of parameters, θ, used to describe the numerator, $\hat{N}$, and transient term, $\hat{T}$, and $n_{\theta,D}$ is the number of parameters used to describe the denominator, $\hat{D}$. Next, it follows from (C.3) and (C.4) that,

$$\hat{\theta} = \mathcal{K}_n^+ \mathcal{Y}_n = \theta + \mathcal{K}_n^+ \tilde{\mathcal{V}}_n \,, \tag{C.11}$$

with $\tilde{\mathcal{V}}_n = \begin{bmatrix} \tilde{V}(-n_W)^T & \cdots & \tilde{V}(n_W)^T \end{bmatrix}^T$, and, hence,

$$\Delta\hat{\theta} = \mathcal{K}_n^+ \tilde{\mathcal{V}}_n \,. \tag{C.12}$$

The effect of noise in the regression matrix, $\mathcal{K}_n$, only leads to second-order noise effects in $\Delta\hat{\theta}$ due to the multiplication with $\tilde{\mathcal{V}}_n$ in (C.12), and can therefore be neglected in the first-order approximation made here, which holds for the considered high-SNR cases.

This leads to the following expression for $\Delta\hat{G}$,

$$\mathrm{vec}(\Delta\hat{G}) \approx \left[\left(I_{n_u} \otimes \begin{bmatrix} I_l & 0 \\ -D_{0,21} & I_{n_y-l} \end{bmatrix} \right) \quad 0 \quad -\left(N_{0,1}^T \otimes \begin{bmatrix} 0 \\ I_{n_y-l} \end{bmatrix} \right) \quad 0 \right] \mathcal{K}_n^+ \tilde{\mathcal{V}}_n \,. \tag{C.13}$$

The covariance on the LRM estimate is then given by

$$\mathrm{Cov}(\mathrm{vec}(\hat{G})) \approx S \, \mathrm{Cov}(\tilde{\mathcal{V}}_n) \, S^H \,, \tag{C.14}$$

where

$$S = \left[\left(I_{n_u} \otimes \begin{bmatrix} I_l & 0 \\ -D_{0,21} & I_{n_y-l} \end{bmatrix} \right) \quad 0 \quad -\left(N_{0,1}^T \otimes \begin{bmatrix} 0 \\ I_{n_y-l} \end{bmatrix} \right) \quad 0 \right] \mathcal{K}_n^+ \,. \tag{C.15}$$

Assuming the noise is uncorrelated over the frequencies and the true noise covariance matrix is equal in the considered frequency window, as is also done in e.g. [6, Section 7.2.2.2], it holds that

$$\mathrm{Cov}(\tilde{\mathcal{V}}_n) = I_{2n_W+1} \otimes C_{\tilde{V}} \,. \tag{C.16}$$

An estimate of the noise covariance matrix can be obtained from the least squares residual as shown in, e.g., [6, Section 7.2.2.2], i.e.,

$$\hat{C}_{\tilde{V}} = \frac{1}{q} \sum_{r=-n_W}^{n_W} \hat{V}(r) \hat{V}(r)^H \,, \tag{C.17}$$

where

$$\hat{V}(r) = \hat{Y}(r) - \mathcal{K}(r)\hat{\theta} \,, \tag{C.18}$$

and where q is equal to the appropriate degrees of freedom of the residual $\hat{V}$, e.g., $q = \mathrm{dof}_{\mathrm{LPRM}}$ as defined in (49). The estimated covariance matrix of the LRM is then given by,

$$\mathrm{Cov}(\mathrm{vec}(\hat{G})) \approx \hat{S} \, (I_{2n_W+1} \otimes \hat{C}_V) \, \hat{S}^H \,, \tag{C.19}$$

with $\hat{S}$ given by (C.15) but where $D_{0,21}$ and $N_{0,1}$ have been replaced with their respective estimates.

Appendix D. Details of the simulation example from Section 5

In this simulation example, the plant is defined as:

$$G = G_{\mathrm{RB}} + G_{\mathrm{nRB},1} + G_{\mathrm{nRB},2} \,. \tag{D.1}$$

with

$$G_{\mathrm{RB}} = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & -1 \\ 1 & -1 & 1 \\ 1 & -1 & -1 \end{bmatrix} \frac{1}{s^2} \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & -1 & -1 \\ 1 & -1 & 1 & -1 \end{bmatrix} ,$$

$$G_{\mathrm{nRB},1} = \begin{bmatrix} -1 & 1 & 1 & 1 \\ 1 & 1 & 1 & -1 \\ 1 & 1 & -1 & 1 \\ -1 & 1 & -1 & -1 \end{bmatrix} \operatorname{diag}\left(\frac{1}{s^2 + 2\beta_{1,i}\omega_{1,i}s + \omega_{1,i}^2} \right) \begin{bmatrix} -1 & 1 & 1 & -1 \\ 1 & 1 & 1 & 1 \\ 1 & 1 & -1 & -1 \\ 1 & -1 & 1 & -1 \end{bmatrix},$$

with $\beta_1 = \begin{bmatrix} 0.005 & 0.02 & 0.01 & 0.015 \end{bmatrix}$ and $\omega_1 = 2\pi \begin{bmatrix} 100 & 200 & 350 & 400 \end{bmatrix}$, and

$$G_{\mathrm{nRB},2} = \begin{bmatrix} L \end{bmatrix} \operatorname{diag}\left(\frac{1}{s^2 + 2\beta_{2,i}\omega_{2,i}s + \omega_{2,i}^2} \right) \begin{bmatrix} R \end{bmatrix},$$

with 4 additional modes, and where the elements of L and R as well as $\beta_{2,i}$ and $\omega_{2,i}$ are determined randomly as $l_{ij} \sim \mathcal{N}(0,\, 1)$, $r_{ij} \sim \mathcal{N}(0,\, 1)$, and $\beta_{2,i} \sim \mathcal{U}(0, 0.03)$, $\omega_{2,i} \sim \mathcal{U}(1000\pi, 2000\pi)$ (MATLAB® 2014a, random number generator: Mersenne Twister, seed 197; order of generations $\beta_{2,i}$, $\omega_{2,i}$, L, and R). Discretization is performed using the Tustin method without prewarping.

Appendix E. Covariance calculation for simulation results

The covariance matrix, $C_{\mathrm{vec}(G_0)}$, which is used in Section 5 to define the maximum likelihood cost function, (53), is derived from the known covariance matrices of the applied input signals as follows. First consider the input and output signals as defined in Figure 6, then take

$$z = \begin{bmatrix} y \\ u \end{bmatrix} = G_{rz}(r_{\mathrm{ex}} + d\,) + \begin{bmatrix} v_y \\ v_u \end{bmatrix}. \tag{E.1}$$

The covariance matrix of the measurement of z then becomes,

$$C_z = C_v + G_{rz} C_d G_{rz}^H, \tag{E.2}$$

with

$$C_v = 10^{-6} \operatorname{diag}(\mathrm{rms}(\mathrm{z}_0)^2), \quad C_d = 4 \cdot 10^{-4} I_{n_u}, \tag{E.3}$$

where the factors 10^{-6} and $4 \cdot 10^{-4}$ reflect the signal to noise ratios of 60 dB and 34 dB. Next, using [6, equation (7-42)], the following is obtained

$$C_{\mathrm{vec}(G_{rz})} = S_{RR}^{-1} \otimes C_z, \tag{E.4}$$

note that in this expression the use of any windowing and averaging is disregarded, i.e., this is the covariance matrix that would be obtained without any windowing and averaging. In this derivation, the auto-power spectrum of the excitation signal is approximated by its expected value, i.e.,

$$S_{RR} \approx \mathbb{E}(S_{RR}) = I_{n_u}. \tag{E.5}$$

Finally, the covariance matrix, $C_{\mathrm{vec}(G_0)}$, can be derived using [6, equation (7-50)],

$$C_{\mathrm{vec}(G_0)} = (G_{ru}^{-T} \otimes [I_{n_y} \;\; -G_0]) C_{\mathrm{vec}(G_{rz})} (G_{ru}^{-T} \otimes [I_{n_y} \;\; -G_0])^H. \tag{E.6}$$