Nonparametric data-driven modeling of linear systems: estimating the frequency response and impulse response function

Nonparametric Data-Driven Modeling of Linear Systems. Estimating the frequency resonse and impulse response function.
Schoukens, Johan; Godfrey, Keith; Schoukens, Maarten
IEEE CONTROL SYSTEMS MAGAZINE Volume: 38 Issue: 4 Pages: 49-88 Published: AUG 2018

Nonparametric Data Driven Modeling of Linear Systems

Estimating the Frequency Response and Impulse Response Function

Johan Schoukens, Keith Godfrey, and Maarten Schoukens — March 14, 2018

The aim of this article is to give a tutorial overview of frequency response function (FRF) or impulse response (IR) function measurements of linear dynamic systems. These nonparametric system identification methods provide a first view on the dynamics of a system. Three main messages will be passed.

The first message is to *replace the classic FRF measurement techniques from the 1960's, based on spectral analysis methods, by the more advanced recently developed algorithms* that are presented in this article. User guidelines will be given to select the best among all these methods according to four specific users' situations: i) measurements with a high or low signal-to-noise ratio (SNR), ii) systems with smooth or fast varying transfer functions as a function of the frequency, iii) batch or real time processing, and iv) low or high computational cost.

The next message is to *store the reference signal together with the data*. This will be very useful whenever there are closed-loops in the system to be tested, including interactions between the generator and the setup.

The last message is to *use periodic excitations whenever it is possible*. Periodic excitations give access to a full nonparametric noise model, even under closed-loop experimental conditions. Combing periodic signals with the advanced methods presented in this article, gives access to high quality FRF-measurements, while the measurement time is reduced by eliminating disturbing transient effects.

Depending upon the situation, it might be possible to reduce the measurement time or the measurement error with a factor 2 to 100. The complete palette from the simple classical methods to the advanced methods that were recently developed will be covered. The article provides a deep understanding of the problems related to FRF and IR measurements. These insights are used as a basis to understand better the classical methods, and to explain how these methods can be improved to get better FRF-IR measurements than the classical methods that are very popular and still dominate the field today. This leads to a completely new class of IR-FRF estimators.

From the classic time and frequency-domain methods of the 1950's to todays advanced new processing techniques

The overview of the classic time and frequency-domain methods, developed in the 1950's and 1960's, will be complemented by an introduction to the more powerful methods that were developed in the last decade. The measurement time can be significantly reduced using these new techniques, at a cost of increased computational demands. Because the available computer power has grown by several orders of magnitude since the late 1960's, it is clear that there is no

reason any more to stick to the initial choices that were made more than 50 years ago, keeping the practical restrictions of that period in mind. The computing time was the dominant constraint that was driving the research efforts at those days, confining the algorithms to be simple [1]–[4] (See Section Smoothing the FRF using the classical spectral estimation methods). However, with today's possibilities, more complex algorithms are possible. This will either allow the measurement time to be reduced, or the quality of the measurements to be improved. It is the goal of this article to make these new algorithms accessible for a wide group of measurement- and control engineers.

Frequency response function and impulse response function measurements

The goal is to obtain the IR or FRF of the system in Figure 1, together with a confidence bound, starting from discrete time measurements of the input $u(kT_s)$ and output $y(kT_s)$, $k = 1, 2, \ldots, N$, with T_s the sample period that is the inverse of the sample frequency f_s.

The FRF measurement evaluates $G(\omega)$ at a discrete set of frequencies $f_k = kf_0 = kf_s/N$, while the IR measurement returns an estimate for $g(t)$ at a discrete time grid kT_s.

Both measurements give equivalent information. The FRF is the Fourier transform of the IR. Depending upon the needs, some information can be more easily accessed in the time or in the frequency domain. A dominant resonance can be more easily analyzed in the frequency domain, the presence of a dead time will be more easily observed in the time domain. Often the choice between one of both domains is set by the user's needs, experiences, and personal preferences.

For didactic reasons, the major part of the article is focused on single-input single-output (SISO) systems. In "FRF Measurements For MIMO Systems" an introduction to the multiple-input multiple-output (MIMO) is given, paying special attention to the design of the experiment [5]–[9]. Recent results on FRF measurements using some of the methods presented in this article are discussed in [10].

The discussion considers both continuous-time (CT) and discrete-time (DT) systems. The actual nature of the system does not critically influence the presented algorithms provided that the sampling frequency f_s is sufficiently high.

Time-domain methods to estimate the IR: basic ideas

Consider the system in Figure 1 with

$$y(t) = \int_{-\infty}^{\infty} g(\tau)u(t-\tau)d\tau = g(t) * u(t), \tag{1}$$

where $*$ denotes the convolution. A first class of methods estimates the impulse response $g(t)$ starting from the measured input and output signals by solving the deconvolution problem in (1). The equivalent description based on the crosscorrelation $R_{yu}(\tau) = E\{y(t+\tau)u(t)\}$ and autocorrelation function $R_{uu}(\tau) = E\{u(t+\tau)u(t)\}$ turns out to be very useful to simplify the computations [4], [11], [12]

$$R_{yu}(\tau) = g(t) * R_{uu}(\tau). \tag{2}$$

White random noise excitations reduce the autocorrelation of the input to $R_{uu}(\tau) = \sigma_u^2 \delta(\tau)$, and the crosscorrelation becomes

$$R_{yu}(\tau) = \sigma_u^2 g(t), \tag{3}$$

allowing the impulse response to be measured directly, without making an explicit deconvolution. In the 1960's, pseudo random binary sequences (PRBS) [13], [14] were used to replace the white noise excitations, resulting in a lower uncertainty on the estimated impulse response for a given measurement time. These methods will be discussed in the Section Time-Domain Approach.

Nowadays, using the increased computer power, it is possible to estimate the IR directly, even for arbitrary excitations. Combining the experimental data with prior user information like exponential decay and smoothness of the impulse response function reduces the uncertainty even more [15]. All these aspects will be discussed in detail in the Section Variance reduction by combining data and prior knowledge: regularization.

Frequency-domain methods to estimate the FRF: basic ideas

An alternative approach to get around the deconvolution problem in (1) or (2) to estimate $g(t)$ is to transform the equation (1) to the frequency domain. Define $U(k), Y(k)$ as the discrete Fourier transform (DFT) of the measured input and output u, y with k the frequency index (see "The Discrete Fourier Transform"). Neglecting the finite length measurement effects, the following relation holds [16]

$$Y(k) = G(k)U(k). \tag{4}$$

It is very tempting to estimate the FRF by a direct division $Y(k)/U(k)$, but this works well only if $U(k)$ is not becoming very small or equal to zero (see Section FRF Measurements Using Periodic Excitations). This can be realized by using well-designed signals (see "Design Of Excitation Signals"), but in general it is better to average the data over multiple subrecords before making the division. To do so, the Fourier transform of (1), is replaced by the Fourier transform of (2), leading to the crossspectrum $S_{yu} = F(R_{yu})$ and autospectrum $S_{uu} = F(R_{uu})$, with $S = F(R)$ the Fourier transform of R. Equation (2) becomes

$$G(k) = S_{yu}(k)/S_{uu}(k). \tag{5}$$

This method became the standard approach in the 1960's and is still used today in all dynamic signal analyzers [1]–[4], [11], [17], [18] (see Section Smoothing the FRF using the classical spectral estimation methods, and Section Time and frequency-domain interpretation of windows).

Recently, a complete new class of spectral methods has been developed that provide superior results over (5) [19], [20], again at a cost of higher computational demands (see Section Improved FRF MeasurementsUusing Local Parametric Methods). With the resources that are nowadays available, there is no reason why these new methods should not be the default choice.

Parametric and nonparametric models

Parametric models are described by a finite number of parameters, and this number does not depend upon the data length. A typical example is a transfer function model for a linear dynamic system, where the number of parameters is set by the number of poles and zeros. In nonparametric models, the number of parameters grows with the data length. A typical example is the FRF of a system. It will be shown that the frequency resolution is inversely proportional

to the data length, and hence the number of frequencies where an FRF measurement is made grows proportional with the data length. There is no sharp definition for both classes of models. IR-measurements can be assigned to both classes. The length of the impulse response that can be estimated from a given data set grows with the length of the data (a nonparametric model characteristic). However, it depends only logarithmically on the data length so that it becomes almost constant for longer data records (a parametric model characteristic). Usually, it makes no sense to estimate the impulse response for more than a few time constants of the system.

This article is completely focused on nonparametric identification. There is an extensive literature on the topic of parametric identification, such as [16], [21]–[23].

Experiment design: choice of the excitation signal

In both the time and the frequency-domain approach, the choice of the excitation signal has an impact on the methods to be used [24], [25]. Also the variance of the measured impulse or frequency response function strongly depends on this choice. Initially, only 'simple' excitations like steps, impulses, or sines could be generated. In the late 1950's and in the 1960's, more advanced periodic binary excitations could be generated using simple hardware [14], [26]–[28]. These signals simultaneously excite multiple frequencies with the same power, See "Design Of Excitation Signals", and Figure S10. This was a major step forward to reduce the required measurement time.

In the late 1970's and early 1980's, arbitrary random generators became available. From then on it became possible to generate directly advanced signals that were designed and computed on a digital computer. It opened many possibilities to increase the signal-to-noise-ratio (SNR) of the measurements, and to reduce the impact of nonlinear distortions on the FRF measurement [16], [29], [30]. In "Design Of Excitation Signals" a detailed discussion of these signals, including the relevant design parameters and user choices, advantages and disadvantages are discussed.

Outline

The outline is organized with the following major sections. First the measurement setup and noise assumptions are presented. The FRF methods are split along the use of periodic excitations and random excitations. These sections cover both the classical and the recently developed methods. Also the time-domain methods are organized along the same line. The authors have added a small historical overview of the methods that were used before the early 1960's. This gives an idea of the elegance of these methods that solved the IR and FRF measurement problems using the tools that were available at those times. In the sidebars of the paper, some specific background knowledge and advanced applications are discussed. At many places in the article, user guidelines are included that give many practical tips and highlight some useful attention points for the interested user.

Measurement Setup

Each data driven modeling process should start with a careful inspection of the measurement setup. A short discussion with the plant operators or the measurement team can save a lot of time in the remaining part of the modeling effort. It is very important to know what preprocessing is applied to the data. Were filters turned on? Is a drift removal applied to the raw

data? Are the measurements made around a given set point, or are the mean values included in the raw data? How are outliers and missing data handled? Without being aware of these actions, a lot of time and effort can be wasted by trying to include their effects in the model.

Although the setup in Figure 3 covers already many interesting situations, it still doesn't cover the full reality. Often the plant to be modeled is a part of a larger complex network with many interacting loops. Under these conditions it is not obvious at all if it is possible to isolate the sub-system of interest from the rest of the plant using the available measurements. In [31], [32], and succeeding work, a detailed analysis about the minimum required measurement conditions is made to assure that indeed the IR/FRF of the actual subsystem is measured.

Intersample assumptions

All the data processing in this article starts from discrete-time measurements $u(kT_s), y(kT_s)$, sampled at a sampling frequency $f_s = 1/T_s$. No information is available how the continuous-time signals $u(t), y(t)$ vary in between the measured samples. For that reason assumptions are needed, and the measurement setup should match the intersample assumption well. The two most popular intersample assumptions [16], [33] are shown in Figure 2: the zero-order-hold (ZOH) and the band-limited assumption (BL). A detailed discussion of both intersample assumptions (the facts and their appreciation) is given in Sections 13.2 and 13.3 of [16].

ZOH setup

The ZOH setup puts a condition on the excitation that is assumed to remain constant in between the samples. In this setup, the IR/FRF are estimated between the discrete-time reference signal in the memory of the generator, and the sampled output. The intersample behavior and the actuator characteristic are an intrinsic part of the model: if the intersample behavior changes, also the corresponding model will change. The ZOH-assumption is very popular in digital control. In that case the sampling frequency f_s is commonly chosen 10 times larger than the frequency band of interest. A discrete-time model gives an exact description of the continuous-time system.

Band-limited setup

The BL set-up assumes that above f_{max} there is no power in the signals: $U(|f| > f_{max}) = 0$. The continuous-time signals are filtered by well-tuned anti-alias filters (cut-off frequency below $f_s/2$)), before they are sampled. These should eliminate the signal power above half the sampling frequency $f_s/2$ to a user specified level to keep the alias errors under control: the high frequency ($f > f_s/2$) content of the measured signals is folded down in the frequency band of interest and act there as a disturbance. For that reason it is strongly advised to use always anti-alias filters in the measurement set-up. Outside the digital control world, the BL-set-up is the standard choice for discrete-time measurements.

Measurement setup - notations

The general band-limited setup that is given in Figure 3 will be considered as the standard setup. The alternative simplified zero-order-hold 'ZOH setup' is obtained, as a special case of the BL-setup, by using the exactly known discrete-time generator sequence $r_d(k)$ in the memory

of the generator as the 'measured' input, and by removing the anti-alias filters.

The system is excited by the generator output signal $r(t)$ that is applied to the plant using the actuator. The generator signal is disturbed by generator noise $n_g(t)$, the output of the system is disturbed by the process noise $n_p(t)$. The continuous-time input and output signals are first low-pass filtered by the anti-alias filters. The measurement noise on the filtered input and output is respectively $m_u(t)$ and $m_y(t)$. These signals are sampled at a sampling rate $f_s = 1/T_s$. Eventually, the discrete-time measurements $u(kT_s), y(kT_s), k = 1, ..., N$ will be used as the raw data from which the IR and FRF estimates will be obtained.

From here on, for notational simplicity, the difference between continuous-time and discrete-time signals will be no longer explicitly indicated. In the remaining part of the article, the sampled signals $u(kT_s), y(kT_s)$ will be denoted as $u(k), y(k)$.

The discrete Fourier transform (DFT) of the measurements $u(k), y(k), k = 1, ..., N$ is calculated using the fast Fourier transform (FFT) algorithm [34], and denoted as $U(k), Y(k)$. The frequency index k indicates the frequency kf_s/N. Because of the presence of the anti-alias filters, this setup is called the band-limited setup.

Noise assumptions

The measured time-domain signals are

$$\begin{aligned} u(k) &= u_0(k) + n_u(k), \\ y(k) &= y_0(k) + n_y(k), \end{aligned} \tag{6}$$

where u_0, y_0 are the disturbance free signals, and $n_u(k), n_y(k)$ model the combined disturbance contributions on respectively the input and output. The noise sequences n_u, n_y are assumed to be stationary sequences that can be mutually dependent (the input and output noise can be related). Without any loss of generality, they are assumed to be generated as filtered white noise, e.g. $n_y = h_y * e_y$ where e_y is a discrete-time white noise source. Although this simplified discrete-time description of continuous-time stochastic signals is nowadays very common in the system identification community, it is not obvious to make this conceptual step. A profound theoretical foundation for this simplified representation is given in [16], [35]

Assumption 1: In the time domain, the disturbing noise is $n = h * e$, where the white noise $e \sim N(0, \sigma_e^2)$. Different noise sources can be mutually correlated.

Remark: the Gaussian assumption is not needed for many of the results. The actual distribution becomes important to obtain quantified uncertainty bounds in the time domain for very small data sets. Asymptotically, the impact of the distributions on the uncertainty bounds disappears, and a Gaussian setting can be used.

Similar assumptions are made in the frequency domain

$$\begin{aligned} U(k) &= U_0(k) + N_U(k), \\ Y(k) &= Y_0(k) + N_Y(k), \end{aligned} \tag{7}$$

where U_0, Y_0 are the disturbance free signals, and $N_U(k), N_Y(k)$ model the combined disturbance contributions on respectively the input and output at frequency k.

Assumption 2: In the frequency domain, the disturbing noise $N_U(k), N_Y(k)$, for all frequencies k, are complex circular normally distributed [16], with the following properties

$$\begin{aligned} &E\{N_U(k)\} = 0, E\{N_Y(k)\} = 0, \\ &E\{|N_U(k)|^2\} = \sigma_U^2(k), E\{|N_Y(k)|^2\} = \sigma_U^2(k), \\ &E\{N_Y(k)\overline{N}_U(k)\} = \sigma_{YU}^2(k), E\{N_Y(k)N_U(k)\} = 0. \end{aligned} \tag{8}$$

In the last expression, $\bar{x}$ denotes the complex conjugate of x.

Assumption 3: The noise $N_U(k), N_Y(k)$ is independent of $U(l), Y(l), \forall k \neq l$.

The role of the generator noise in Figure 3 differs depending on the selected processing approach, and it affects also the definition of u_0, y_0. (i) In the *periodic framework*, where the excitation is assumed to be periodic, $u_0(t), y_0(t)$ are the signals that are solely due to the reference excitation $r(t)$. In that case, using the notation of Figure 3, $u_1(t) = u_0(t)$, and $y_1(t) = y_0(t)$ if all noise sources n_g, n_p, m_u, m_y are put equal to zero. The generator noise will act as a nonperiodic disturbance that does not affect the uncertainty of the estimates (see Special case: generator noise only in the Section on FRF measurements). (ii) In the *nonperiodic framework*, using for example random excitations, $u_0(t), y_0(t)$ are the signals that are solely due to the reference excitation $r(t)$ and the generator noise $n_g(t)$. In that case $u_1(t) = u_0(t)$, and $y_1(t) = y_0(t)$ if the noise sources n_p, m_u, m_y are put equal to zero. The generator noise $n_g(t)$ can be considered to be a part of the excitation that is out of user control.

User guidelines

The experimental setup has a significant impact on the final quality. Making small modifications can simplify and improve the processing of the data and the quality of the results. For that reason, the user needs to pay enough attention to the following aspects:

- *Anti-alias filters*: Verify if anti-alias filters are present, and make a proper selection of the cut-off frequency.
- *Synchronization*: Make sure that the data acquisition and data generation channels are well synchronized. Lack of synchronization can jeopardize the quality of the data. Nonparametric periodic post processing heavily relies on a perfect synchronization. If the clocks of the data acquisition and the generator are not well synchronized, it is still possible to resample the signals using advanced signal processing algorithms that estimate the clock mismatch, and estimate the corrected discrete Fourier spectra [36].
- *Preprocessing*: Check with the operators what manipulations are applied to the data: drift/trend removal, missing data handling, removal of outliers, and prefiltering.
- *Reference signal*: Store the reference signal together with the measured data. Especially for closed-loop measurements, it becomes much easier to avoid systematic errors if the reference signal is available.
- *Signal assumption*: Make sure that the measurement setup is in agreement with the ZOH (anti-alias filters switched off) or BL (anti-alias filters switched on) intersample assumption.

FRF Measurements Using Periodic Excitations

In this section, the FRF measurement is analyzed in full detail using periodic excitations. This allows the analysis to be simplified a lot, while it is still a very important case from practical point of view. Initially, the FRF was measured frequency per frequency using a sine excitation [37]. By moving to periodic excitations that excite multiple frequencies at once, it was possible to reduce the measurement time significantly [38]. Such periodic signals are standard available in all commercial dynamic signal analyzers. Their use was strongly stimulated by the introduction of the Fourier transform in the electrical engineering field [39]. The development of efficient numerical procedures to calculate the Fourier transform was the start of a new era [40]–[43]. Extending the study to include random excitations will be done in the next section. This will lead to the classic FRF measurement methods that dominate the field since the 1960's.

Consider the SISO system in Figure 1. The frequency-domain analysis is started from the measured input and output DFT spectra (7). In order to simplify the expressions, the frequency index k will be omitted. Only when multiple frequencies are combined or when it is not clear from the context what frequency is used, the frequency index will be added.

Stochastic analysis of periodic excitations

The input $u(t)$ and output $y(t)$ are measured over multiple periods, and broken in subrecords of one period each, for example $u^{[l]}, l = 1, ..., P$, as shown in Figure 4. In the next step, the mean value, and the (co-)variance is calculated as a function of the frequency by analyzing the variations of the periodic input- and output signals over the measurements of the repeated periods. While the disturbing noise n_u, n_y varies from one period to another, the noiseless periodic signals u_0, y_0 do not so. This results eventually in the following simple procedure: For each subrecord $u^{[l]}, y^{[l]}$, corresponding to a period, the discrete Fourier transform $U^{[l]}, Y^{[l]}$ is calculated using the fast Fourier transform (FFT) algorithm. If an integer number of periods is measured under steady-state conditions (no transients present), there will be no leakage in the results. The sample mean and noise (co-)variances at frequency k are then given by

$$\hat{U}(k) = \frac{1}{P}\sum_{l=1}^{P} U^{[l]}(k) \quad \hat{Y}(k) = \frac{1}{P}\sum_{l=1}^{P} Y^{[l]}(k), \tag{9}$$

and

$$\begin{aligned} \hat{\sigma}_U^2(k) &= \tfrac{1}{P-1}\textstyle\sum_{l=1}^{P} |U^{[l]}(k) - \hat{U}(k)|^2, \\ \hat{\sigma}_Y^2(k) &= \tfrac{1}{P-1}\textstyle\sum_{l=1}^{P} |Y^{[l]}(k) - \hat{Y}(k)|^2, \\ \hat{\sigma}_{YU}^2(k) &= \tfrac{1}{P-1}\textstyle\sum_{l=1}^{P} (Y(k) - \hat{Y}(k))(\bar{U}(k) - \bar{\hat{U}}(k)). \end{aligned} \tag{10}$$

Remark that the variance of the estimated mean values $\hat{U}(k), \hat{Y}(k)$ is given respectively by $\sigma_U^2(k)/P, \sigma_Y^2(k)/P$. Adding together all this information in one figure results in a full nonparametric analysis of the SNR of the measurements. Remark that no interaction with the user is needed during the processing. This makes the method well suited to be implemented in standard measurement procedures. The FRF estimate is then $\hat{G} = \hat{Y}/\hat{U}$. Its properties are studied in Section Smoothing FRF measurements for periodic excitations.

Example: the flexible robot arm

As an example, experimental data measured under the BL-assumption on a one degree-of-freedom flexible robot arm (see Figure 5 and [44] for a detailed description) are processed using (9) and (10). A periodic multisine excitation [45], [46] is used (see also "Design Of Excitation Signals"). The sample mean $\hat{U}, \hat{Y}$ and standard deviation $\hat{\sigma}_U, \hat{\sigma}_Y$ are plotted in Figs. 6(a), and 6(b), respectively. Observe that the uncertainty on the mean values will be 10 dB below the actual shown noise levels because 10 periods are averaged (see "Impact Of Averaging On The Variance Of The FRF Estimate"). These results show that already during the measurement process, an excellent impression of the quality of the measurement is obtained, and this without any user interaction. In this case, the SNR is about 20 dB before averaging, except for the input around $f/f_s = 0.015$ corresponding to the first resonance frequency of the robot arm. The drop in the SNR is due to the impedance mismatch between the motor and the structure, creating a power drop at the input, resulting in a low SNR at the input (about 0 dB before averaging). This explains also why it is often very difficult to make a precise damping measurement of lowly damped mechanical structures.

The correlation ρ measures the linear relation between two noise sources

$$\rho = \frac{\sigma_{YU}^2}{\sigma_U \sigma_Y}, \tag{11}$$

and is shown in Figure 6(c). In this case, the high correlation (values close to 1) indicates that the noise on the force and accelerator measurements are highly correlated. This can be either due to generator noise (which is not so in this setup), or to the presence of a dominant noise source in a closed-loop. The latter is the case in this setup, the output of the motor is strongly loaded by the input of the mechanical structure, leading to an internal feedback mechanism. The measured FRF that is shown in Figure 6(d) is obtained using the methods discussed in the next section. Observe that at the first resonance and anti-resonance, the correlation is much lower. This is due to the fact that at those frequencies either the input or the output drops to very low values, as shown in Figs. 6(a) and 6(b), and then the measurement noise will become dominant. The latter is not correlated with the noise in the loop.

User guidelines

- *Periodic inputs*: Use periodic inputs whenever it is possible.
- *Number of repeated measurements*: The (co-)variance estimates improve with the number of repeated periods. However, in practice a small number of repetitions ($P \geq 4$) will do. If the variances are used in a parametric estimate, then $P \geq 4$ is enough to guarantee consistency of the estimates. $P \geq 7$ guarantees the existence of the covariance matrix of the estimated parameters, and their limiting distribution [16], [47].
- *Overlapping subrecords*: all these results can be generalized to measurements with only two repeated periods using overlapping subrecords [48].
- *Co-variance information*: It is strongly advised to calculate also the co-variances because these give valuable information about the presence of generator noise or feedback loops.

Smoothing FRF measurements for periodic excitations

Smoothing: average over the periods

The most simple FRF estimate $\hat{G}$ is the empirical transfer function estimate (ETFE) [21] starting from the sample mean $\hat{U}, \hat{Y}$ (9)

$$\begin{aligned} \hat{G} &= \hat{Y}/\hat{U}, \\ &= \frac{Y_0 + N_{\hat{Y}}}{U_0 + N_{\hat{U}}}, \\ &= G_0 \frac{1 + N_{\hat{Y}}/Y_0}{1 + N_{\hat{U}}/U_0}, \end{aligned} \tag{12}$$

with $G_0 = Y_0/U_0$. Remark that this is a 'local' nonparametric method, the estimates $\hat{G}(k)$ at frequency k make no use of information at other frequencies $k \neq l$. By repeating this calculation at all frequencies of interest, all the estimates are retrieved.

Bias expression

Making the Taylor series expansion of (12) up to degree 2, and under the noise Assumption 3, the bias is

$$\begin{aligned} E\{\hat{G}\} - G_0 &= -G_0 exp(-|U_0|^2/\sigma_{\hat{U}}^2)(1 - \rho\frac{U_0/\sigma_{\hat{U}}}{Y_0/\sigma_{\hat{Y}}}), \\ &= -G_0 exp(-P|U_0|^2/\sigma_U^2)(1 - \rho\frac{U_0/\sigma_U}{Y_0/\sigma_Y}), \end{aligned} \tag{13}$$

see [49] for $\rho = 0$, and [50] for the correlated noise situation. Observe that even for signal-to-noise (SNR) values as low as 10 dB, the relative bias $|b/G_0|$ is still below 10^{-4}. From (13) it follows that the bias on the FRF decreases exponentially with the number of averaged periods P. Observe that the order of the operations is important: first the spectra should be averaged, and next the division can be calculated. If the order is reversed, there will be no bias reduction! This is a general rule of thumb: the data should be first averaged before nonlinear operations like divisions or multiplications or applied. Failing to do so makes a method more prone to systematic errors. In case of non-synchronized measurements, it becomes difficult to apply an averaging before the division. In that case, nonlinear averaging methods can be used to reduce the bias [49], [51]. This comes with a (small) loss in efficiency. For that reason, synchronized measurements are still to be preferred whenever it is possible.

Practical variance expression: study of the distribution of $\hat{G}$

The study of the variance of $\hat{G}$ is more tedious. In [52], [53] it is shown that the variance of (12) does not exist due to the presence of outliers when the denominator comes close to zero. Nevertheless, it is still possible to provide exact uncertainty intervals on the FRF measurement starting from the probability density distribution (PDF) of $\hat{G}$. The exact expression is known and studied in detail in [52], [54]. A detailed discussion of these results is beyond the scope of this article, instead the practical conclusions of the full analysis are discussed here.

Distribution: Under the Gaussian noise Assumption 2, and for a high SNR of the input

signal ($>$ 40dB), the FRF $\hat{G}$ is approximately Gaussian distributed because in (12) the division $1/(1+N_Y/\hat{U}) \approx 1 - N_Y/\hat{U}$. For lower SNR of the input, the linear approximation no longer holds, and the higher order terms in the approximation will affect the distribution.

Uncertainty bounds: It is still possible to calculate a 'practical' variance expression of $\hat{G}$ that can be used to generate uncertainty bounds assuming a complex Gaussian distribution for $\hat{G}$. The higher the input SNR, the tighter the uncertainty interval can be, for example a tight 50% uncertainty interval can be obtained for a SNR that is larger than 10 to 15 dB, and a tight 95% uncertainty bound requires a SNR that is better than 20 dB. The variance expression is obtained from the first order Taylor series expansion of (12), under Assumption 3 [16]

$$\begin{aligned}\sigma^2_{\hat{G}} &= |G_0|^2\left(\frac{\sigma^2_{\hat{Y}}}{|Y_0^2|} + \frac{\sigma^2_{\hat{U}}}{|U_0^2|} - 2Re\left(\frac{\sigma^2_{\hat{Y}\hat{U}}}{Y_0\bar{U}_0}\right)\right), \\ &= \frac{1}{P}|G_0|^2\left(\frac{\sigma^2_{Y}}{|Y_0^2|} + \frac{\sigma^2_{U}}{|U_0^2|} - 2Re\left(\frac{\sigma^2_{YU}}{Y_0\bar{U}_0}\right)\right).\end{aligned} \tag{14}$$

The variance is inversely proportional to the squared SNR of the measurements, and it drops in $1/P$. In practice, the variance is estimated by replacing U_0, Y_0 by $\hat{U}, \hat{Y}$ (9), and the (co-)variances by the sample (co-)variances (10).

This result will guide the excitation design in "Design Of Excitation Signals", good excitation signals should maximize the SNR at all frequencies of interest. Doubling the SNR allows the measurement time to be reduced by a factor of four, so a good excitation design (increasing $|U_0|$), and a good measurement setup design (decreasing the variances) reduces the measurement time significantly.

Special case: generator noise only

If only generator noise n_g is present in Figure 3, $\sigma^2_U = \sigma^2_{N_g}$, $\sigma^2_Y = |G|^2\sigma^2_{N_g}$, and $\sigma^2_{YU} = G\sigma^2_{N_g}$. Substituting these results in (14) shows that $\sigma^2_{\hat{G}} = 0$. In the periodic framework, the generator noise does not contribute to the variance on the FRF measurement. The same conclusion holds true in the general excitation framework. In that case the generator noise adds directly to the measured input signal, and acts as an input that is not under user control.

Example: the flexible robot arm: FRF measurement

The analysis (12) and (14) is applied to the flexible robot arm data that were discussed before. The FRF estimate is calculated starting from the averaged input and output data, and the variance is calculated starting from the sample variances. The results are shown in Figure 6 (d). A smooth FRF estimate is obtained, but it is also clearly visible that the uncertainty around the resonance (of the first mode) is very high. This is due to the poor SNR at the input in that frequency band, indicating once more that low damping values are very difficult to measure with high precision.

User guidelines

- Use periodic excitations whenever it is possible. Check the synchronization between the generator and the data acquisition channels. Use advanced signal processing methods to

remove synchronization errors [36].

- *Nonparametric noise analysis*: Make a nonparametric noise analysis to make a first quality check of the measurements. Identify the dominant noise source (input or output noise). This information might be useful to refine the experiment design, or to improve the measurement setup.
- *Mutual correlation*: Check the cross correlation between the input and output noise. It indicates if one or two independent noise sources are present in the setup. It can also be an indication for strong closed-loop effects. These should be handled with care (see "Measuring The FRF Under Closed-Loop Conditions").
- *Steady-state conditions*: Check if the measurements are made under steady-state conditions or not. The initial transient is estimated by subtracting the last period from the first period. If transients are present, more advanced methods should be used (see Section on Improved FRF measurements using local parametric methods).
- *Averaging*: Average the input and output DFT spectra before making the division to reduce bias effects on the FRF estimate, see (12).

FRF Measurements Using Random Excitations

In this section, the discussion of the FRF measurement is repeated for random excitations. In Section "Smoothing FRF Measurements For Random Excitations Using The Empirical Transfer Function Estimate (ETFE)", the basic problems are introduced: why do noiseless data result in the poor measurements as shown in Figure 7? How can these errors be removed using smoothing methods? What variance/bias tradeoff is made when smoothing is applied? Next the classical methods that were developed during the 1960's to answer these questions are discussed in Section "Smoothing The FRF Using The Classical Spectral Estimation Methods". The recent insights in the structured nature of leakage errors (See "Models For Dynamic Systems: Finite Length Effects") will be used to give a deeper understanding of these well-established methods.

Smoothing FRF measurements for random excitations using the Empirical Transfer Function Estimate (ETFE)

The poor results that are shown in Figure S6 were retrieved by applying the ETFE (12) with $P = 1$ (the data are not split in subrecords) on simulations obtained with a random noise excitation. So no smoothing was applied before the division is made. Figure 7 plots the same results as shown in Figure S6, but now together with the amplitude $|U|$ of the DFT of the actual realization of the input that was used for this measurement. The latter is a random variable too, having dips at some frequencies [4]. It is mainly at those frequencies that $\hat{G}$ becomes very prone to disturbances, resulting in the very scattered nature on the left side of the figure. In this example, the errors are completely due to leakage errors as will be explained later (See "Models For Dynamic Systems: Finite Length Effects") since there was no disturbing noise. Disturbing noise has a very similar effect, for that reason no distinction is made in the further discussion in this section. A detailed analysis of the disturbing noise effects is postponed to Section "Smoothing The FRF Using The Classical Spectral Estimation Methods".

In the absence of input noise, the ETFE can be shown to be asymptotically unbiased for N growing to ∞ if the expected value is calculated with respect to the output noise, conditioned on a given realization of the random input [21]. However, its variance does not decrease to

zero [21], even if there is no input noise present. To reduce the variance, a proper smoothing procedure is needed. The simple averaging procedure used in the previous section works well for synchronized measurements of periodic data, but fails for random noise excitations. For periodic excitations, with P measured periods, the sample mean converges to the true value, for example

$$\lim_{P\to\infty} \frac{1}{P}\sum_{l=1}^{P} U^{[l]} = U_0. \tag{15}$$

For random noise excitations, that have no periodic nature, it is not possible to split the record over the successive periods. Instead, the original data record is split in P subrecords. In this case, the sample mean $\hat{U}(k), \hat{Y}(k), k \neq 0$, converges to zero (e.g. $E\{\hat{U}\} = 0$) at the same rate as its standard deviation of the output noise does, so that no increase in SNR would be obtained. Moreover, the estimate (12) degenerates to $0/0$, and hence an alternative averaging procedure is needed. Two approaches are discussed, either smoothing over neighboring frequencies, or smoothing over successive realizations (subrecords) of the data.

Smoothing the ETFE over neighboring frequencies

If the frequency resolution f_s/N of the FRF measurement is small enough, it can be safely assumed that G does not vary much in the frequency interval $B = [k-n, k+n]$, and hence a smoothed estimate for $\hat{\hat{G}}(k)$ can be obtained by calculating a properly weighted average [21], [55], [56]

$$\begin{aligned}\hat{\hat{G}}(k) &= \frac{\sum_B \hat{G}(m)/\sigma_G^2(m)}{\sum_B 1/\sigma_G^2(m)},\\ &= \frac{\sum_B \hat{G}(m)|U(m)|^2}{\sum_B |U(m)|^2},\\ &= \frac{\sum_B Y(m)\bar{U}(m)}{\sum_B |U(m)|^2}.\end{aligned} \tag{16}$$

The second equality holds true if the output noise dominates (lower SNR at the output than at the input) such that

$$\sigma^2_{\hat{G}}(m) = |G_0|^2 \frac{\sigma^2_{\hat{Y}}(m)}{|Y_0^2(m)|} = \frac{\sigma^2_{\hat{Y}}(m)}{|U_0^2(m)|}, \tag{17}$$

and the variance $\sigma^2_Y(m)$ remains constant in the interval B.

The major disadvantage of the ETFE smoothing technique is the creation of bias errors, mainly at the resonances and anti-resonances, because the estimate $\hat{G}(k)$ is averaged over the frequency interval B. This introduces errors that are proportional to the second derivative of G of order $O(d^2G(f)/df^2)$. For that reason, the width of the smoothing window B should be well tuned. Choosing it too small results in a large variance, while selecting it too large results in a large bias. Optimal tuning methods have been proposed in [57]. First an initial estimate of the noise variance $\sigma^2_Y(m)$ is made, and next this information is used to tune the bandwidth B in (16). The proposed method makes use of local polynomial approximations to estimate the noise variance. This idea will be later repeated and extended in the Section Improved FRF measurements using local parametric methods.

Observe that in (16) the measurements are again averaged before the division is made. However, in the denominator, $|U|^2$ is averaged, and so there is still a nonlinear operation applied before the averaging. This leads to biased estimates if the input measurement is disturbed by noise.

The discussion of the bias and variance effects of the disturbing noise is postponed till the Sections Bias analysis and Variance analysis.

Smoothing the ETFE over successive realizations

An alternative approach to average the ETFE is to split the original data record into $P = 2n+1$ subrecords as was done for the periodic excitation solution above. A new estimate is defined by averaging over the subrecords, instead of averaging over the neighbouring frequencies

$$\hat{G}(k) = \frac{\sum_{l=1}^{P} Y(k)^{[l]} \bar{U}(k)^{[l]}}{\sum_{l=1}^{P} |U(k)^{[l]}|^2}. \tag{18}$$

Apparently, smoothing over the frequency interval B is avoided, but the frequency resolution of the subrecords is P times smaller than that of the original record, and equals exactly the width of the interval B. Hence, similar bias errors will appear for (18) as for (16) (see also Section Bias and variance analysis of leakage, and Table 2).

The scaled numerator and denominator in (18) can also be interpreted as estimates of the cross spectrum S_{YU} and auto spectrum S_{UU} between u, y, so that (18) can be rewritten as

$$\hat{G}(k) = \frac{\hat{S}_{YU}(k)}{\hat{S}_{UU}(k)}, \tag{19}$$

which links the ETFE method directly to the classical FRF estimation methods of the 1960's [1], [2], [11], [17], [58], [59] that will be discussed in detail in the next section. The stochastic properties are also analyzed in the next section.

In Figure 8, the data from Figure S6 are processed using (19). The original data record of 4096 samples is split in 16 subrecords of 256 samples each, and next (18) is calculated. Averaging over 16 subrecords reduced the presence of the dips in S_{UU} (see Figure 8 (b)) with respect to $|U|$ significantly, resulting in a smoother FRF estimate as shown in Figure 8(a), compared to the previous results using a single record, that were shown in Figure 7. This came at a cost of a reduced frequency resolution, the FRF is measured at 128 frequencies instead of 2048 frequencies in (8).

User guideline

The spectral resolution is $\Delta_f = BP/N$, with B the smoothing width, P the number of subrecords or repeated experiments, and N the total data length. The noise reduction is of $O(1/\sqrt{BP})$. The number of subrecords P, and the smoothing bandwidth B can be interchanged without affecting the effective frequency resolution or the uncertainty of the result if the disturbing noise is dominating. For a high SNR, the leakage errors will dominate, and in that case it is better to increase B and keep P as small as possible, because for a given length N, the impact of the leakage errors grows with $\sqrt{P}$.

Smoothing the FRF using the classical spectral estimation methods

For a long time, a FRF measurement was considered as the division of two spectra [1], [2], [4], [11], [18], see (12), (19). Hence, first the individual spectra U, Y had to be accurately estimated, instead of finding directly a model between these two signals. This turns out to be a more demanding approach that is more sensitive to leakage errors as explained below. Measuring the spectrum $X(k), k = 0, ..., N/2$ of a random signal is a tedious job. Due to its random nature, it is not straightforward to get a high quality measurement of the spectrum starting from a finite length measurement $x(t), t = 1, ..., N$. The DFT (S2) $X(k)$ of $x(t)$ is not a consistent estimate because its variance does not decrease to zero for a growing length N. The main reason for that problem is that also the spectral resolution is growing with N, so that no 'averaging' is taking place, as is shown in Figure 9. For that reason, it is necessary to introduce explicitly an averaging step in the estimation procedure to get a reliable spectral estimate. This can be done either over neighboring frequencies or over multiple realizations (see [11] and references therein), similar to the discussion for the ETFE in the previous section. Eventually, the latter approach became the dominant one, especially after the introduction in the engineering world of the FFT [34] to calculate the DFT [60]. The main motivation was a faster computation, and a reduction of the core storage that was needed to process the shorter records [2].

These different approaches to experimental spectral analysis were also reflected in different proposals for FRF estimators using noise excitations. The standard procedure, known as Welch's method [2], or the weighted-overlapped segment averaging (WOSA) procedure, splits the original long data record $u(t), y(t)$ in P shorter (overlapping) subrecords $u^{[l]}, y^{[l]}, l = 1, ..., P$ of length N, and calculate for each of these the DFT $U^{[l]}, Y^{[l]}$, see also Figure 11. The auto spectrum S_{UU} and cross spectrum S_{YU} are estimated using the expressions

$$\begin{aligned} \hat{S}_{UU}(k) &= \frac{1}{P}\sum_{l=1}^{P} |U(k)^{[l]}|^2, \\ \hat{S}_{YU}(k) &= \frac{1}{P}\sum_{l=1}^{P} Y(k)^{[l]}\bar{U}(k)^{[l]}. \end{aligned} \tag{20}$$

Remark that these are equal to the numerator and denominator in (18), which clearly links both methods. In Figure 10, it is illustrated that averaging over P realizations using these expressions reduces the variability. The dips disappear even more rapidly than would be expected on the basis of the $1/\sqrt{P}$ rule (see "Impact Of Averaging On The Variance Of The FRF Estimate"), especially for low values of P.

The FRF $\hat{G}(k)$ is then obtained using (19). It is essential in this approach to realize that $\hat{G}(k)$ is obtained as the division of two estimated spectra that are prone to errors due to the finite data length effects, called leakage errors, and due to the disturbing noise. The errors in Figure 8 are completely due to the leakage, there was no disturbing noise added to the data.

The majority of efforts in the earlier times were directed towards a better understanding and tuning of the leakage errors using well adopted time windows [4], aiming to reshape the errors to become less harmful. The impact of both the leakage and noise errors will be analyzed in more detail in this section because their combined effect determines the stochastic properties of $\hat{G}_(k)$.

Remark that the alias errors are not considered anymore, it is assumed that the anti-alias filters and the sampling frequency are properly matched, so that these errors can be neglected.

Time and frequency-domain interpretation of windows

Calculated DFT spectra are prone to leakage errors due to finite length effects [34]. A subrecord is obtained from an infinite long signal by multiplying the original signal with a 'selection window' $w(t)$ that is equal to zero outside the selected time interval. For example, as shown in Figure 11, the l^{th} subrecord, $u^{[l]}(t) = w^{[l]}(t)u(t)$. Multiplication in the time domain becomes a convolution in the frequency domain, so that

$$U^{[l]}(k) = W(k) * U^{[l]}(k). \tag{21}$$

Many windows have been discussed in the literature, each being optimized for a given application, e.g. spectral resolution, detecting small spurious frequency components, amplitude measurements [61]. In this article, the Rectangular, Hanning, and Diff window will be considered, but instead of studying their impact on the spectral estimates U, Y, as is classically done [3], the focus will be directly on the impact of windowing on the relation (S13)

$$Y(k) = G(k)U + T_G(k) \tag{22}$$

following the approach in [19], [62], [63]. All the windows are defined on the half open interval $x \in [0, 1[$, for $x = [0, 1, ...N-1]/N]$ (see Table 1, outside this interval $w(x) = 0$). The definition is given in the time domain, the impact of the window on the DFT spectrum is given in the last column. The Diff and the Half-sine window are shown to have very similar properties for FRF measurements [63].

Observe that the Diff window $X_w(k+1/2) = [X(k+1)-X(k)]$ acts as a difference centered around $k+1/2$, where the latter frequency index points to the frequency $f = (k+0.5)f_s/N$. The Hanning window $X_w(k) = \frac{1}{4}[-X(k-1) + 2X(k) - X(k+1)]$ is a double difference centered around k. From this observation it can be easily understood that windows push down spectra that are smooth in function of the frequency, like it is the case for the transient $T_G(k)$ in (S13). The first term in this expression $G(k)U(k)$ will be 'rough' for random noise excitations and hence not be reduced by the windowing operation. This is also illustrated in Figure 12 where the amplitude spectrum of the 'rough' signal and the smooth transient term is plotted using a rectangular and a Hanning window. This shows that windowing can decrease the leakage errors since these are directly coupled to the transient term T_G.

Understanding the impact of leakage on FRF measurements

Windowing will not remove all the errors, the tapered output will not be identical to the response of the system on the tapered input, as was already shown in Figure S5. The precise impact on the FRF measurement is studied in this section. To keep the focus completely on the leakage errors, it is assumed that the disturbing noise is zero, so the errors in the FRF measurement will be solely due to leakage effects.

Assumption 4: No disturbing noise

$$\begin{aligned} U(k) &= U_0(k), \\ Y(k) &= Y_0(k). \end{aligned} \tag{23}$$

The impact of the Hanning window on the FRF measurement is obtained by replacing $Y = GU + T_G$ at the three frequencies $k-1, k, k+1$ in $Y_w(k) = \frac{1}{4}[-Y(k-1)+2Y(k)-Y(k+1)]$. The smooth functions G ($O(N^0)$), and T_G ($O(N^{-1/2})$) can be approximated by [62]

$$\begin{aligned} G(k \pm 1) &= G(k) \pm \triangle_G + O(N^{-2}), \\ T_G(k \pm 1) &= T_G(k) \pm \triangle_{T_G} + O(N^{-5/2}), \end{aligned} \tag{24}$$

with $\triangle_G = O(N^{-1})$. Eventually this results in

$$Y_w(k) = G(k)U_w(k) + E_G(k) + E_{T_G}(k). \tag{25}$$

with

$$\begin{aligned} E_G(k) &= -\triangle_G(U(k+1) + U(k-1)) = O(N^{-1}), \\ E_{T_G} &= O(N^{-5/2}). \end{aligned} \tag{26}$$

The transient error E_{T_G} is what remains from the transient after the double differentiation. $E_G(k) = O(N^{-1})$ is a new error that is due to interpolation of G over the left and right frequency in the Hanning window. Observe that it became the dominating error. So it can be concluded that the Hanning window replaces the leakage error that is of $O(N^{-1/2})$ by a smaller $O(N^{-1})$ interpolation error. These results are tabled in Table 2 [63], and compared to those of the Rectangular and Diff (Half-sine) window. The results in this table are normalized by the time constants of the system [64]. To do so, the dominant time constant τ_f in the frequency band of interest is selected, and N is replaced by N/τ_f. For a resonating system, the dominant time constant is set by the damping of the actual resonance that is studied (see "Characterizing A Resonance By Its 3 dB Bandwidth").

Since the rectangular window makes no interpolation, it follows that only the leakage error will be present

$$Y_R(k) = G(k)U_R(k) + 0 + E_{T_G}. \tag{27}$$

For the Diff window that has a width of one bin, the same expression as for the Hanning window holds. The latter has a width of two bins, and so the interpolation error for the Diff window will be smaller, although it is also $O((N/\tau_f)^{-1})$ as it was for the Hanning window. This is still larger than the leakage error of the Diff window, which is $E_{T_G} = O((N/\tau_f)^{-3/2})$, because only one difference is made.

Remark: In modal analysis (analysis of vibrating mechanical structures) [18], exponentially decaying windows are also used [65]. These add an artificial but known damping to data that can be compensated for when physical parameters are extracted from the FRF data.

Bias and variance analysis of FRF leakage errors

For random excitations, the leakage error is itself a random variable. In "Basic Concepts For FRF Measurements: Models For Dynamic Systems" it is explained that leakage errors are created by transient effects at the output of the system, and these have a smooth spectrum $T_G(k)$. The leakage error is then given by $T_G(k)/U(k)$, and since $U(k)$ is a random variable, the leakage error will be too. This error is conditioned on the random input, and is characterized by its mean value (bias errors) and variance [4].

Bias errors: Averaging over successive realizations of the input reduces this error, but it

will not decrease towards zero for the number of averages P growing to ∞. Remember that the leakage error is due to the beginning and end transient. These depend linearly on the random input signal, and hence the correlation $E\{T_{T_G}U\}$ between the input and the transient will be different from zero, leading to systematic errors (bias) in the FRF measurement [66]. For the rectangular window, the bias error is shown to be $O((N/\tau_f)^{-1})$ [62], [63]. For the Hanning and the Diff (Half-sine) windows, the bias error is mainly due to the interpolation errors and drops to $|G^{(2)}(k)O((N/\tau_f)^{-2})$. The bias error of the Diff window is slightly smaller than that of the Hanning window because the width of the Diff window is half that of the Hanning window.

Variance errors: In the classical literature [2], [4] the attention was focused on the bias contribution of leakage, and little effort was spent on the study of leakage induced variance although under good SNR conditions the latter becomes the dominating error. The variance error is respectively of $O(P^{-1}(N/\tau_f)^{-1})$ for the Rectangular window, and of $O(P^{-1}(N/\tau_f)^{-2})$ for the Diff (Half-sine) and Hanning window.

Overlapping windows: While it is not possible to reduce the bias, the variance can be further reduced by allowing for an overlap of R samples, when shifting the window with length N over the long record [63], [67]. This increases the calculation time because more subrecords need to be processed, but the variance is further reduced. Because in most applications, the measurement time is a more important concern than the calculation effort, the overlapping strategy is advised as the standard procedure in the current day practice. Because the overlapping subrecords get more and more correlated for an increasing overlap, the gain saturates. In [63], a very detailed study is made. It is shown under very loose conditions that the optimal choice among all windows with a bounded derivative is the Diff (Half-sine) window with an overlap of $1 - R/N = 2/3$ (the window is shifted each time with $N/3$). This results in a further reduction of the leakage induced variance with more than a factor 3.5 (half of this at the zero and half the sample frequency). This result will also extend to the disturbing noise sensitivity, but in that case the gain is about a factor 2.

Variance analysis of the FRF measurements in the presence of disturbing noise

The variance of $\hat{G}$ (19) due to the disturbing input and output noise N_U, N_Y can again be retrieved by linearizing the expression with respect to the noise, similar to the approach in (12), but now applied to the estimates $\hat{S}_{YU}(k)$ and $\hat{S}_{UU}(k)$. This results eventually in

$$\sigma^2_{\hat{G}} = \frac{1}{P}|G_0|^2 \left(\frac{\sigma^2_Y}{\hat{S}_{YY}(k)} + \frac{\sigma^2_U}{\hat{S}_{UU}(k)} - 2Re(\frac{\sigma^2_{YU}}{\hat{S}_{YU}(k)}) \right). \tag{28}$$

For P growing to ∞, the estimated auto and cross power spectra can be replaced by their exact value, and then it is clear that $\sigma_{\hat{G}}$ drops again in $1/\sqrt{P}$. However, for smaller values of P, this approximation is no longer valid due to the presence of dips as illustrated in Figure 10, and then the original expression (28) should be used. The additional loss is plotted in Figure 13 [16], showing that for small values of P the excess loss comes close to 5 for $P = 1$. Using overlapping subrecords, the value of P is artificially increased by about a factor 2, as explained in the previous section.

A very popular alternative for the variance expression (28) is [4], [18]

$$\sigma^2_{\hat{G}(k)} = \frac{1}{P}|G_0(k)|^2 \frac{1-\gamma^2(k)}{\gamma^2(k)}, \tag{29}$$

with

$$\gamma^2(k) = \frac{|S_{YU}(k)|^2}{S_{UU}(k)S_{YY}(k)}. \tag{30}$$

In practice, the coherence γ^2 is estimated by replacing the theoretical values in (30) by their measured values. The coherence (30) has exactly the same interpretation as the correlation (11) for periodic excitations [4]. The coherence $\gamma^2 = 1$ for undisturbed measurements (no leakage, no noise, no nonlinear distortions), pointing to perfectly linearly related data. It drops to zero due to noise disturbances, unmeasured inputs, and for data that are not linearly related [4]. It is a very popular measure for the quality of the data, and is plotted in many commercial signal analyzers [3], [4]. Compared to the variance analysis for periodic data (10), the coherence provides less information because it makes no split between input and output noise, and their mutual correlation. This can be seen in the analysis of the flexible robot arm data in Figure 6, where the coherence (correlation) is plotted in Figure 6(c). If only this information is available, the insight in the origin of the dips in the coherence would be lacking (drop in SNR of the input).

Bias analysis of the FRF: Impact of noise disturbances on the reference input, the H_1, H_2 methods

The most popular FRF estimate $\hat{G}$ is (19). However, this method fails completely in the presence of input noise. Consider

$$\lim_{P\to\infty} \hat{G}(k) = G(k) \frac{1 + \frac{\sigma^2_{YU}}{\hat{S}_{YU}(k)}}{1 + \frac{\sigma^2_{UU}}{\hat{S}_{UU}(k)}}. \tag{31}$$

Assume for simplicity that the noise is not correlated with the input u and the output y, then it follows that (31) equals

$$\lim_{P\to\infty} \hat{G}(k) = G(k) \frac{1}{1 + \frac{\sigma^2_{UU}}{\hat{S}_{UU}(k)}}, \tag{32}$$

showing that (19) underestimates the true value of the FRF in the presence of noise on the input signal [4], [16]. In the mechanical community, this estimator is called the H_1 estimator.

There is a simple trick to get around this problem provided that the SNR of the output is (very) high. In that case the H_2 method can be used [68]

$$\hat{G}(k) = \frac{\hat{S}_{YY}(k)}{\hat{S}_{YU}(k)}. \tag{33}$$

Observe that the squared input $|U|^2$ is no longer averaged, and so the source of the bias in (19) is eliminated. As a rule of thumb, the reference should always be selected to be that signal with the highest SNR, resulting in the H_1 or H_2 method [4], [18]. Commercial dynamic signal analyzers use the H_1 method. Switching the input and output cables turns it into the H_2 method by inverting the estimated FRF $1/\hat{G}$.

User guidelines

- *Periodic or random excitation*: The first advice remains to use periodic excitations whenever it is possible. If this is not possible for technical reasons, the most recent and powerful FRF measurement techniques that are presented in the next section are advised to be used. If for some reason the classical methods that are presented in this section should be used, the following advices can help to get the best results within the classical framework.
- *Bias errors*: Check the experimental setup, and verify the SNR of the input signal. If it is well above 40 dB (1% noise floor), the relative bias of the H_1 will be below 10^{-4}. For a lower SNR, it should be checked that the bias is not too large for the application in mind. If the bias is too large, it might be an option to switch to the H_2 method if the output has a higher SNR than the input.
- *Variance errors*: It is necessary to average over a sufficiently large number ($P > 16$) of subrecords to keep the additional variance loss small. Use overlapping subrecords (66% overlap) to reduce the variance even more.
- *Leakage errors 1*: Even for large SNR levels, the FRF estimate can still be poor due to leakage errors. Use the Half-sine (Diff) window in combination with averaging over subrecords to reduce the error. Keep the subrecords as long as possible because the leakage errors (bias, variance) drop in $1/N^2$. Keep in mind that leakage creates bias and variance errors.
- *Leakage errors 2*: The authors strongly advise use of the methods of the next section because these eliminate the leakage errors almost completely.

Improved FRF Measurements Using Local Parametric Methods

Leakage errors are considered for a long time to be random errors. Only recently it became clear that there is a lot of structure in these errors [66], [67], [69]–[72] as discussed in "Models For Dynamic Systems: Finite Length Effets", leading to a new family of FRF methods [19], [20], [70], [73], [74]. These recently developed methods do not target anymore a precise estimation of the cross and the auto spectrum as the classical methods in the previous section do, but focus directly on the estimation of the relation (S13), $Y(k) = G(k)U(k) + T_G(k)$, between the DFT spectra. That eliminates almost completely the leakage error, so that only the noise errors remain important. The main idea to estimate $G(k)$ is to consider a narrow frequency interval $B = [k - n, k + n]$ around the frequency of interest k. In that interval the parametric models $G = B_G/A_G, T_G = I/A_G$, (S15) can be represented by low order approximations, so that at each frequency k a reduced system identification problem is solved. These methods will be shown to have a better leakage reduction at a higher computational cost, while the noise sensitivity remains very similar to the classical windowing methods of the previous section.

A short discussion and comparison of the properties of the three local parametric models will be given. The mathematical details, proofs, and more extensive comparisons are discussed in [20], [64], [75], [76].

The use of a local parametric model seems to conflict with the concept of nonparametric methods. However, compared to a full parametric approach, it turns out that identifying a local parametric model is much simpler and requires no user interaction. This is because the model order will be fixed, and the (nonlinear) optimization problem is easily solved due to the very

simple nature of the local approximation problem. Moreover, the number of local models to be identified grows with the number of data. For all these reasons, the local parametric methods still belong to the class of nonparametric methods.

The local system identification problem

A system identification problem is defined by three players: the data, the model, and the cost function. In this case the data are given by (7)

$$\begin{aligned} U(k) &= U_0(k) + N_U(k), \\ Y(k) &= Y_0(k) + N_Y(k), \end{aligned} \tag{34}$$

in the interval $B = [k-n, k+n]$. The local parametric model that is valid in the interval B is

$$Y_0(k) = G(k, \theta_G)U_0(k) + T_G(k, \theta_{T_G}). \tag{35}$$

Both the transfer function and transient term in this model are estimated by minimizing the errors $E(k)$ on the frequency interval $B = [k-n, k+n]$. The width $2n$ of this interval is called the local bandwidth of the method.

$$E(k) = Y(k) - G(k)U(k) - T_G(k), \tag{36}$$

using a weighted least-squares cost function

$$V(k) = \sum_B W(l)|E(l)|^2 \tag{37}$$

that is minimized with respect to θ_G and θ_{T_G}. For a sufficiently small frequency interval B, the (co-)variances of the noise N_U, N_Y can be assumed to be constant, so that these do not affect the optimal choice for the weighting $W(l)$ in (37). Instead, the combined choice of the weighting and the model structure will be used to manipulate the complexity of the identification problem.

Iterative Local Rational Method (ILRM): $W(l) = 1$, $G = B_G/A_G, T_G = I/A_G$,

This setting results in

$$V_{ILRM}(k) = \sum_B |Y(l) - \frac{B_G(l)}{A_G(l)}U(l) - \frac{I(l)}{A_G(l)}|^2. \tag{38}$$

The presence of the denominator $A_G(l)$ leads to a nonlinear optimization problem that should be solved iteratively, which is affordable nowadays.

Properties: From the three proposed methods, the ILRM seems to be the most natural choice. This method suffers from a higher noise sensitivity than the other two methods. If the order is not well tuned, excess poles and zeros can create very large narrow spikes due to very closely spaced poles and zeros. This can only be avoided by a dedicated model tuning at every frequency, which reduces the robustness of this method even more. These effects are illustrated in Section 3.4 of [75].

Local Polynomial Method (LPM): $W(l) = 1$, $G = B_G$, $T_G = I$.

The nonlinear optimization problem is turned into a linear least-squares problem by setting

the denominator equal to one $A_G(l) = 1$, which leads directly to the local polynomial method [19], [77],

$$V_{LPM}(k) = \sum_B |Y(l) - B_G(l)U(l) - I(l)|^2. \tag{39}$$

The minimizer of the cost function is found by solving a linear set of equations, hence no iterative procedure is needed any more.

The pole/zero cancellation problems of the ILRM method are completely eliminated by putting the denominator equal to one. This results in a local polynomial approximation of the transfer function G and the transient T_G. This simplified approach turns out to be very attractive. Besides reducing the optimization problem to a linear one, it also makes the model selection problem less critical. The order R of both polynomials, B_G and I, can be put equal to $R = 2$ (or $R = 4$) with good results. Under these conditions, the disturbing noise induced variance of the LPM is 1.74 dB below that of the classical Hanning method if both methods are tuned to the same frequency resolution.

It is clear that the polynomials can only approximate a rational form in the finite frequency window B. The approximation errors are studied in full detail in [64]. The main conclusion is that it is most efficient to choose R to be even. For that choice, the leakage error E_{LPM} is bounded by

$$E_{LPM} = O((B/B_{3dB})^{R+1}) = O((NP)^{-(R+1)}), \tag{40}$$

with B_{3dB} the 3 dB bandwidth of the resonance under study. Compared to the windowing methods that had errors of $O((N)^{-1})$, a huge gain is made in the reduction of the leakage errors. From (40) it follows also that the local bandwidth $2n$ of the interval B should be chosen as small as possible, e.g. more than two times smaller than B_{3dB}, but at the same time it should contain enough frequencies to estimate the $2(R + 1)$ complex coefficients in the two polynomials, so that $n \geq R + 1$. This leads for $R = 2$ and $n = 3$ to at least 7 frequencies in the interval B.

The last result can also be translated in the minimum record length that is needed [64]. Using the relations in "Characterizing A Resonance By Its 3 dB Bandwidth", $B \leq B_{3dB}/2$, and that $\tau = 1/(\zeta\omega_n) = 2/B_{3dB}$, it is found that the frequency resolution of the measurement should be better than $B_{3dB}/(2R + 2)$, and the record length

$$T_{meas} \geq (R + 1)\pi\tau. \tag{41}$$

Local (Linear) Rational Method (LRM): $W(l) = |A(l)|^2$, $G = B_G/A_G, T_G = I/A_G$,

An alternative to linearize the cost function (38) is to put the weighting $W(l) = |A_G(l)|^2$, resulting in another linear least-squares problem [78], [79]

$$V_{LRM}(k) = \sum_B |A_G(l)Y(l) - B_G(l)U(l) - I(l)|^2. \tag{42}$$

Properites: The LPM was the start of a new era in FRF measurements. It was the first method that was proposed and studied in full detail to remove the leakage errors almost completely. However, for systems with low damping, as often happens in advanced mechanical applications, it can be hard to meet the constraint that $B \leq B_{3dB}/2$ or alternatively that

$T_{meas} \geq (R+1)\pi\tau$. In that case, the LRM approach [79] can still solve the problem because it still identifies a rational model that can deal with these lowly damped poles. For $R = 2$ (a 2^{nd} order model), the error of the LRM is $O((B/D)^4)$, with D the shortest distance to the neighboring poles and zeros.

The LRM combines the advantages of the LPM (linear-in-the-parameters) and the IRLM (a rational model). However, for low SNR of the input, a relative bias of $O(\sigma_Y^2/S_{Y_0Y_0})$ appears [75]. It turns out that in most practical problems this is not a real issue, and for that reason the authors advise the LRM method as the default choice among the discussed local parametric models.

General remarks

History: The LPM was the first local parametric method that was proposed in the literature [19], [20], and a detailed discussion of its properties for SISO and MIMO can be found in [20], [77]. In 2012, the LRM method [79] was proposed as an attractive alternative that can better deal with lowly damped systems as often appear in vibrating mechanical structures. The ILRM method is studied in detail in [75]. It is the most expensive method since an iterative algorithm is needed, and it will turn out to be most sensitive to noise among the three proposals.

Alternative Methods: Alternative parametric approaches are discussed in [76], [80]. The first alternative results in a 'global' method that links all frequencies to each other, leading to large sets of equations to be solved. The second method is a Bayesian approach that makes an intrinsic trade-off between variance and bias (see "Bias And Variance Trade-Off Of Estimators"). These aspects are discussed later in this article in more detail.

Automatic tuning of the local bandwidth B: In [57] an automatic local bandwidth tuning algorithm is proposed, starting from a local polynomial model. These ideas can be transferred to the local parametric methods. A first attempt to do so is presented in [81]. At a cost of additional calculations, a lower RMS error can be obtained.

Missing input and output data: If data get lost due to sensor failure, overloads, and/or data transmission errors, special actions are needed to take care for these errors. Instead of making new measurements, advanced signal processing methods can be used. The missing data are then estimated together with the FRF and its variance. If the reference signal is available, missing data in the input and the output can be restored. If that is not the case, the methods assume that only output data are missing [82], [83].

Illustration of the leakage rejection of the Hanning, LPM, and LRM methods

The classical window method (Hanning) and the local parametric methods LPM/LRM are illustrated on a system with two resonances using noise free data (no disturbing noise added) so that the effect of the leakage errors is clearly visible (see Figure 14 for more details) . For the Hanning method in Figure 14(a), subrecords are used with a length of $N = 256$ (red) and $N = 1024$ (pink) samples, both with an overlap of $R = \frac{2}{3} \times N$. Observe that the errors of the Hanning method become very large, especially around the second resonance (30% or more for the short subrecord length). Using longer subrecords reduces this error, but this comes at a cost of fewer averages (larger risks for spiky errors due to dips in the input spectrum $\hat{S}_{UU}$). In Figure

14(b) the results of the LPM (pink) and LRM (red) are shown. For both methods, the errors are an order of magnitude smaller than those of the Hanning method. Around the second resonance, the error of the LPM increases sharply because the local bandwidth B becomes large with respect to the 3_{dB} bandwidth. The LRM method performs still very well under these conditions.

User guidelines: Classical window methods - LPM/LRM methods

- *Random excitation*: For measurements with a high SNR of the input measurements, the LPM/LRM provide results with a comparable quality as those obtained using periodic methods. They outperform the classical windowing methods.
- *Periodic excitation*: For periodic excitations, the LPM/LRM methodology allows the transients to be removed so that there is no need any more to wait for the steady-state regime. This can result in a significant reduction of the measurement time [84].
- *Noise analysis*: High quality noise power spectra estimates are obtained using the LPM/LRM if the input SNR is high. For measurements with comparable SNR at the input and output, it is also possible to make a full noise analysis resulting in high quality (co-)variance estimates, as a function of the frequency, of the input/output noise provided that a good noise-free reference signal is available. If that is not the case, periodic excitations can be a solution to obtain good estimates (see Section Stochastic analysis of periodic excitations).
- *Leakage errors*: The LPM/LRM methods almost completely remove the leakage errors. For that reason these methods are advised as the default choice for FRF measurements whenever random noise excitations are used. There is no reason to continue to use the classical windowing methods, except for very special cases where the memory of the processor would be very limited. The default setting to get a maximum leakage rejection and a minimal bias error is to choose the local bandwidth B in (37) as small as possible (no user interaction is needed).
- *Disturbing noise sensitivity*: With the default settings (local bandwidth as small as possible), the noise sensitivity of the classical windowing methods and the LPM/LRM are comparable. For smooth systems, the local bandwidth can be increased without creating too large bias errors (this is an advanced form of smoothing over the neighboring frequencies). This gives an additional handle to the user to reduce the variance of the results if needed [57].
- *Calculation time*: The calculation demands in real time applications can be strongly reduced by making the majority of the calculations before the measurements are done. An example of a real time implementation on a microprocessor is discussed in [84].

Measuring The Power Spectrum Of The Disturbing Noise On The Output

The major problem in measuring the spectral properties of the disturbing noise is the separation of the true signals U_0, Y_0 and the noise N_U, N_Y. A first possibility is to use periodic signals, as is explained in the Section *Stochastic analysis of periodic excitations*. A full nonparametric analysis of the (co-)variance of the input and output noise could be made, without any interaction of the user [16], [47].

If the periodicity assumption does not hold, an alternative approach is needed. In that case, the noise is separated from the signal, using the estimated FRF, assuming that the input measurement is noise free. This can be generalized to noisy input and output measurements if an exactly known reference signal is available, for example the signal in the memory of the

generator, using the indirect method or the joint-input-output method [85]–[87]. In that case the SISO problem (from the input u to the output y) is replaced by a MIMO problem (from the reference r to the input u and the output y), and eventually it is still possible to obtain the (co-)variance of the input-output noise.

Motivated by the previous discussion, it is assumed in this section that the input is exactly known, while the output is disturbed by noise.

Assumption 5: No disturbing input noise

$$\begin{aligned} U(k) &= U_0(k), \\ Y(k) &= Y_0(k) + N_Y(k). \end{aligned} \tag{43}$$

Under these assumptions, only the variance $\sigma_Y^2(k)$ needs to be measured. This is done simultaneously with the FRF measurement. The classical windowing methods start from the measured coherence, while the new local parametric methods start from the residue analysis.

Coherence based spectral noise analysis

Under Assumption 5, an estimate of the variance is obtained by [4]

$$\begin{aligned} \sigma_Y^2(k) &= \hat{S}_{YY}(k) - \frac{|\hat{S}_{YU}(k)|^2}{\hat{S}_{UU}(k)}, \\ &= \hat{S}_{YY}(k)\frac{1-\gamma^2(k)}{\gamma^2(k)}. \end{aligned} \tag{44}$$

The cross and auto spectra are estimated using the windowing methods as discussed before, and the errors on the power spectrum estimate of the noise will be set by the leakage errors of $\hat{S}_{UU}, \hat{S}_{YY}$ [62], which are of $O(N^{-2})$ for the Hanning window.

Spectral noise analysis using the local parametric methods

The system description (S13) can be extended by a parametric noise model

$$\begin{aligned} Y(k) &= G(k)U(k) + T_G(k) + N_Y(k), \\ &= G(k)U(k) + T_G(k) + H(k)E(k) + T_H(k). \end{aligned} \tag{45}$$

The output noise is modeled as a filtered white noise source with power spectrum $\sigma_e^2|H(k)|^2$, and a transient term $T_H(k)$ that models the leakage effects of the noise shaping system H, completely similar to the role of $T_G(k)$ for the plant model. In practice both transients $T_G(k), T_H(k)$ are added together in the local parametric methods. The power spectrum $\sigma_e^2|H(k)|^2$ is directly estimated as the mean square value of the residuals in the local frequency band B (see previous section) [19], [77], corrected for the degrees of freedom in the least square problem that define the ILRM, LPM, and the RLM. Since these residuals are no longer disturbed by the leakage effect, they have a much higher quality than those directly obtained using the coherence method (44). In this case the errors on $\sigma_Y^2(k)$ are of the same order of magnitude as those of the corresponding local parametric method.

Spectral noise analysis: an example

The output of a lowly damped second-order system is disturbed by filtered white noise. In order to visualize all the important aspects, the noise filter was chosen to have a resonance and an antiresonance. The results are shown in Figure 15 for the coherence method using a Hanning window, and for the LPM. Since the spectral resolution is quite high (many frequencies in the 3 dB bandwidth), the results with the LRM would be very similar. Observe that the frequency resolution for the LPM method is 16 times higher than that of the coherence method. A smoother estimate can be obtained by averaging the results over neighboring frequencies in a sliding window.

Time-Domain Approach

In this section, the direct measurement of the IR is studied. In a second step, it is possible to obtain the FRF by calculating the FFT of the estimated IR. The frequency resolution can be increased by increasing the length of the time record using zero padding . The discussion in this section is under the assumption that the input measurement is not disturbed by noise (Assumption 5). The estimation of the impulse response function in the time domain is studied within a discrete-time setting

$$y(t) = g(t) * u(t) = \sum g(k)u(t-k) = \sum g(t-k)u(k). \tag{46}$$

This restriction does not limit the generality of the discussion and the proposed methods, provided that either a ZOH-input is used, or that the sampling frequency f_s is chosen high enough with respect to the bandwidth of the system as was discussed in the Section Measurement Setup. Without loss of generality, the discussion is restricted to causal systems. The generalization to non-causal systems is straightforward.

Real-life systems have mostly an infinitely long impulse response that will be approximated by finite length models $\tilde{g}(t), t = 0, 1, 2, \ldots, n$

$$\tilde{y}(t) = \sum_{k=0}^{n} \tilde{g}(k)u(t-k). \tag{47}$$

Observe that opposed to the FRF models, it is requested to specify here the model length n. This is seemingly increasing the complexity of this approach, compared to the FRF methods, but in the latter case the frequency resolution f_0 needs to be well selected with respect to the system dynamics as explained in "Basic Concepts For FRF Measurements: Link Between The Data Length And The Frequency Resolution Of The FRF Estimate", leading to a very similar problem.

In general, it is not possible to estimate the individual entries $g(t)$ in (47) independently from each other, all parameters are estimated at once. So, the IR estimation methods are global nonparametric methods. Below it will be shown that for some special choices of the input signal, the problem can be reduced to a local nonparametric problem, where each entry $\tilde{g}(t)$ is estimated individually.

IR estimation, a linear least-squares problem

The underlying idea to all the methods that will be discussed below is to solve the overdetermined set of equation [22]

$$y = K\tilde{g}, \tag{48}$$

with $y = (y(n), y(n+1), \ldots, y(N-1))^T$, $\tilde{g} = (g(0), g(1), \ldots, g(n))^T$, and

$$K = \begin{pmatrix} u(n) & u(n-1) & u(n-2) & \ldots & u(0) \\ u(n+1) & u(n) & u(n-1) & \ldots & u(1) \\ u(n+2) & u(n+1) & u(n) & \ldots & u(2) \\ \ldots & \ldots & \ldots & \ldots & \ldots \\ u(N-1) & u(N-2) & u(N-3) & \ldots & u(N-n-1) \end{pmatrix} \tag{49}$$

by minimizing the following least-squares cost function with respect to $g(k), k = 0, 1, 2, \ldots, n$

$$\begin{aligned} V &= \sum_{t=n}^{N} (y(t) - \tilde{y}(t))^2, \\ &= \sum_{t=n}^{N} (y(t) - \sum_{k=0}^{n} \tilde{g}(k) u(t-k))^2. \end{aligned} \tag{50}$$

The sum (50) starts at $t = n$ to avoid that unknown past values of the input appear in (49). Alternatively, the unknown past values could be estimated too, but that would double the number of unknown parameters. Moreover, the product of the unknown parameters $\tilde{g}(k) \times u_{t<0})$ would turn the linear least-squares problem into a nonlinear one, and iterative methods would be needed to solve it. Observe that these choices are the time-domain equivalent of the leakage handling in the frequency-domain solutions.

No weighting function is added to the cost function, mainly because this would require again prior noise information that is not available.

The minimizer of the cost function V (50) is obtained by solving [22]

$$K^T K \tilde{g} = K^T y, \tag{51}$$

with

$$K^T K = \begin{pmatrix} \hat{R}_{uu}(0) & \hat{R}_{uu}(1) & \hat{R}_{uu}(2) & \ldots & \hat{R}_{uu}(n) \\ \hat{R}_{uu}(1) & \hat{R}_{uu}(0) & \hat{R}_{uu}(1) & \ldots & \hat{R}_{uu}(n-1) \\ \hat{R}_{uu}(2) & \hat{R}_{uu}(1) & \hat{R}_{uu}(0) & \ldots & \hat{R}_{uu}(n-2) \\ \ldots & \ldots & \ldots & \ldots & \ldots \\ \hat{R}_{uu}(n) & \hat{R}_{uu}(n-1) & \hat{R}_{uu}(n-2) & \ldots & \hat{R}_{uu}(0) \end{pmatrix}, \tag{52}$$

and

$$K^T y = \begin{pmatrix} \hat{R}_{yu}(0) \\ \hat{R}_{yu}(1) \\ \hat{R}_{yu}(2) \\ \ldots \\ \hat{R}_{yu}(n) \end{pmatrix}. \tag{53}$$

or

$$\hat{R}_{uu}\tilde{g} = \hat{R}_{yu}, \tag{54}$$

with

$$\begin{aligned} \hat{R}_{uu} &= \frac{1}{N-n} K^T K \in \mathbb{R}^{(n+1)\times(n+1)}, \\ \hat{R}_{yu} &= \frac{1}{N-n} K^T y \in \mathbb{R}^{(n+1)\times 1}. \end{aligned} \tag{55}$$

Observe that the size of the normal equations (54) is significantly smaller than the size of the original set (49) $K \in \mathbb{R}^{(N-n)\times(n+1)}$. Moreover, the matrices can be directly obtained from the sample estimates of the auto and cross-correlation, without any need to form the large matrix K, for example

$$\hat{R}_{yu}(\tau) = \frac{1}{N-n} \sum_{k=n}^{N} y(k)u(k-\tau). \tag{56}$$

In the next sections the solution of (54) will be discussed for different choices of the excitation signal. The latter choice strongly influences the complexity of the solution.

Remark: Instead of using delayed inputs, corresponding to z^{-k} as basis functions in the $z-$domain, it can be advantageous to use more advanced basis functions. A first possibility is to use Laguerre functions [88]. It is possible to inject more knowledge, if the user has some prior idea about the pole locations of the system, using generalized orthogonal functions [89]. The major advantage of these methods is the reduction of the number of unknown parameters to be estimated which results in a reduced uncertainty. Of course, it remains a major challenge to obtain valuable prior knowledge.

IR estimation using impulse excitations

Basic idea

For a long time, impulse (or step) excitations were the only practical possibility to create a wide band excitation. At the same time, (54) is significantly simplified. For a perfect impulse excitation,

$$u(t) = \sigma_u \delta(t), \tag{57}$$

and

$$R_{uu}(\tau) = \sigma_u^2 \delta(t). \tag{58}$$

So $K^T K$ simplifies to

$$K^T K = \sigma_u^2 \begin{pmatrix} 1 & 0 & 0 & \dots & 0 \\ 0 & 1 & 0 & \dots & 0 \\ 0 & 0 & 1 & \dots & 0 \\ \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & 0 & \dots & 1 \end{pmatrix}, \tag{59}$$

and the normal equations (54) reduces to

$$\sigma_u^2 \tilde{g} = \hat{R}_{yu} = (y(0)y(1)y(2)\dots y(n))^T, \tag{60}$$

which give a direct access to the impulse response $\tilde{g}$ in the measured output. For that reason this approach was very popular before 1970 because no computers were needed. A paper recorder, combined with an impulse excitation was enough to have a direct measurement of the impulse response.

Modern use of impulse excitations

Although the computational restrictions are completely removed today, the impulse response excitation is still popular in some application fields.

Mechanical engineering: Mechanical engineers use special instrumented hammer kits [18], [90]. The setup is struck by the hammer, and the impulse response is again directly measured (Figure 16). The major advantage (and the reason that it is still used) is that no complicated actuator connections need to be realized, and the excitation can be easily moved around the whole structure in a few minutes. The instrumented hammer kits measure also the actual applied force so that variations in the excitation level can be compensated for. The duration of the pulse that sets the bandwidth of the excitation is controlled by a proper choice of the hammer tip [18].

Dynamic calibration: During the development of primary standards for dynamic calibration problems, it is not possible to measure the input signal. Special setups have been built to create short impulses with controlled properties [91]–[94]. The main issue is to keep the duration of the impulse short enough compared to the time constant of the sensor that needs to be calibrated.

Biomedicine: In mathematical modeling of biomedical systems, the choice of input is usually very restricted, particularly when the subject is a human being. A single rapidly administered dose, administered either intravenously or orally, is the most commonly used form. It is then assumed that the duration of the administration is rapid compared with the kinetics of the subject, so that the input can be approximated by $Q\delta(t)$, where $\delta(t)$ is a delta function [95], [96]. For intravenous input, this form of input is usually referred to as a bolus dose. The only other form of input commonly used in such applications is a constant continuous infusion, i.e. a step input [95], [96]. The unit step response is then the integral of the unit impulse response; see the Remark below.

Averaging of impulse measurements

The direct impulse response measurement method works well only for measurements with high SNR because there is no averaging in the simple procedure. Increasing the impulse level is often impossible for practical reasons. Under low SNR measurement conditions, the results can be averaged over repeated experiments, but then it is extremely important to align the successive measurements using a good trigger signal or advanced signal processing methods in order to avoid smearing effects in the averaging procedure [97].

Remark: Sometimes it might be easier to measure the step response. This is the integral of the IR [98].

IR estimation using white noise excitation

Not only impulse excitations result in an impulse like autocorrelation, all wide band excitations with a flat amplitude spectrum have a similar behavior. For that reason, white noise

excitations became a popular alternative for the impulse excitation.

Random white noise excitation

For simplicity, only white random noise is considered here. Observe that the actual distribution (for example Gaussian noise, or uniform noise) is not important from a theoretical point of view because the autocorrelation depends only on the second-order properties. Of course, in practice the distribution will set the peak value of the excitation for a given power. Usually it is very desirable to keep this value as low as possible in order not to overload the system, and to improve the SNR. In that case

$$\lim_{N\to\infty} \hat{R}_{uu}(\tau) = E\{\hat{R}_{uu}(\tau)\} = \sigma_u^2 \delta(\tau), \tag{61}$$

and the same expressions as in the previous section are retrieved. However, the major difference with respect to the impulse excitation is that the system is persistently excited during the whole experiment. The variance on the estimate will drop as $O(1/N)$, so that even under low SNR conditions good measurements can be obtained.

For finite length records, the autocorrelation does not reach yet its expected value. It is a random variable by its own, and eventually

$$\hat{R}_{uu}(\tau) = R_{uu}(\tau) + O(\frac{1}{\sqrt{N}}), \tag{62}$$

leading to errors in the estimated impulse response of the same order of magnitude. These can be eliminated by solving again the original normal equations (54), of course at a loss of the simplicity of the method. Nevertheless, it is still advantageous to use white noise excitations, because these are the optimal excitation signal from disturbing noise rejection point of view (See Section Variance analysis of the FRF measurements in the presence of disturbing noise).

Nowadays, the cross-correlation is calculated using digital computers. In the early days, these were not available, and instead analog correlators were built using tape recorders to delay the signals [99], [100] (see "Historical Note On FRF And IR Measurements").

Well-designed deterministic signals

The finite length effects discussed in the previous section are further reduced by replacing the random excitation by a well-designed period excitation. Periodic signals have a periodic autocorrelation function. This creates no problem as long as the period length is longer than the useful length of the impulse response, for example five or ten times the dominating time constant of the system (see "Truncation Error Of Finite Impulse Response Models").

A very popular choice within this class of signals are well-designed binary signals [27], [28], [101], amongst others, the Maximum Length Binary Signals (see "Design Of Deterministic Excitation Signals") that could be generated using dedicated hardware as shown in Figure S10. These signals combine many advantages.

- *Removal of random finite length effects*: The autocorrelation function is (See "Design Of

Deterministic Excitation Signals")

$$\hat{R}_{uu}(\tau) = R_{uu}(\tau) + O(\frac{1}{N}). \tag{63}$$

Observe that the error term $O(\frac{1}{N})$ converges much faster to zero for an increasing data length N than the $O(1/\sqrt{N})$ term in (62) for white noise excitations. In order to fully retrieve these advantages, the clock frequency f_c should be about 2.5 times larger than the highest frequency of interest.

- *Binary correlation*: Because the excitation signal is either 1 or -1, a binary correlator can be used. If only limited computer power is available, or in time critical applications, this is a significant advantage.
- *Periodic signal*: All the advantages of periodic signals as discussed in Section FRF measurements using periodic excitations, are also retrieved here. Amongst others, simple averaging procedures, and direct access to the noise power spectrum can be mentioned.

A disadvantage of the MLBS (and the related sequences) is the high sensitivity to nonlinear distortions, compared to randomized excitations like random noise or random phase multisines [16]. These create large spikes in the estimated impulse response [102]. Averaging over multiple realizations of the MLBS in combination with a median filter can largely remove these spikes [102], but this requires again much longer measurement times. Another alternative is to use randomized periodic ternary signals that do not excite the multiples of the second and third harmonic [103].

The advantages of correlation methods combined with PRBS excitations was early recognized [14], and applied to industrial processes like gas chromatography [26], nuclear power plants [104], and oil refineries [105].

IR estimation using arbitrary excitations

Of course the direct impulse response estimation method (54) can be combined with arbitrary excitations. Calculating the solution of (54) can be speeded up by exploiting the underlying Toeplitz nature of the matrices [106].

Variance analysis of the direct IR method

The covariance matrix C_g is directly available for the linear least-squares problem

$$C_g = (K^T K)^{-1} K^T C_{n_y} K (K^T K)^{-1}. \tag{64}$$

Asymptotic frequency-domain expression

An asymptotic frequency-domain interpretation for the model order $n \to \infty$ of this result can be made [21]

$$\sigma^2_{G_{IR}}(k) \approx \frac{n}{N} \frac{\sigma^2_Y(k)}{S_{UU}(k)}. \tag{65}$$

The subscript IR indicates that this result is obtained via the impulse response estimate. (65) shows that the length of the estimated impulse response should be well balanced: as small as possible to reduce the variance, but long enough to keep the systematic errors (bias) small. From

(65) it follows also that the standard deviation $\sigma_{G_{IR}}$ drops again as $O(1/\sqrt{N})$.

Remark: More general variance expressions for (65) exist that are valid for finite-model orders [107]. The simple expression (65) will be further used because it will lead later to a very user friendly rule of thumb.

Comparison with the spectral analysis methods

In (28) the variance on the FRF $G(k)$ obtained with the spectral methods is given. In the absence of input noise ($\sigma_U = 0, \sigma_{YU} = 0$), this expression simplifies to

$$\sigma^2_{G_{FRF}}(k) = \frac{1}{P}\frac{\sigma^2_Y(k)}{S_{UU}(k)}. \tag{66}$$

The subscript FRF indicates that this result is obtained via a direct FRF measurement. The length of the subrecords is $N_{sub} = N/P$. The additional gain for overlapping subrecords (up to a factor 2) can be added to (66). The additional loss, up to a factor 4^2 (see Figure 13), is not considered for simplicity, and the leakage induced variance is also disregarded.

From (65) and (66), it follows that

$$\frac{\sigma^2_{G_{FRF}}(k)}{\sigma^2_{G_{IR}}(k)} = \frac{1}{2}\frac{\frac{1}{P}}{\frac{n}{N}} = \frac{N_{sub}}{2n}, \tag{67}$$

with $N_{sub} = N/P$ the length of a subrecord in the spectral analysis method, and n the length of the estimated impulse response.

This result is very revealing because it allows comparison of the FRF and the IR based methods, starting from their experiment design parameters. To do so, the choice of N_{sub} and n is directly linked to practical user choices. Using the results in "Basic Concepts For FRF Measurements" the following remarks can be made:

(i) Choice of n: The length of the estimated impulse response should be long enough such that the truncation bias is small as discussed in "Truncation Error Of Finite Impulse Response Models". The relative MSE due to the truncation induced model errors is $e^{-2n/\tau}$. Truncating the impulse response at n equal $2\tau, 5\tau, 7\tau$, results respectively in an RMS error below 10%, 1%, and 0.1%.

(ii) Selecting the subrecord length N_{sub}: The frequency resolution of the spectral method is $\triangle_f = 1/N_{sub}$. In "Characterizing A Resonance By Its 3 dB Bandwidth" it is explained that the required resolution depends on the 3dB bandwidth of the system. For a resonating system, this is $B_{3dB} = 1/\pi\tau$. Hence the relative frequency resolution of G_{FRF} is $\frac{\triangle_f}{B_{3dB}} = \frac{\tau}{2N_{sub}}$, and the number of measured frequencies in the 3dB band is

$$F_{3dB} = \frac{N_{sub}}{\pi\tau}. \tag{68}$$

Plugging both results in (67), shows that for the 1% error level ($n = 5\tau$), the following ratios hold

$$\frac{\sigma^2_{G_{FRF}}(k)}{\sigma^2_{G_{IR}}(k)} = \frac{\pi\tau F_{3dB}}{10\tau} \approx \frac{F_{3dB}}{4}, \tag{69}$$

From this analysis, it can be concluded that for a very low frequency resolution (4 frequencies in the 3dB bandwidth) and 1% RMS bias error, both the G_{FRF} and G_{IR} require about the same experiment length. For more noisy data (higher errors can be tolerated, and so n can be reduced to 2τ) or for higher frequency resolutions, the IR approach becomes more attractive if the calculation time and the size of the computer memory is not an issue.

Variance reduction by combining data and prior knowledge: regularization

The linear least-squares estimate (54) is the maximum likelihood estimate for white Gaussian disturbing noise n_y. Under these conditions, and keeping in mind that the problem is linear-in-the-parameters, it is well known that this estimator reaches the Cramer-Rao lower bound [16], [21], [22], even for finite sample lengths. This means that it is impossible to find another unbiased estimator that will have a smaller uncertainty (covariance matrix) [16], [21], [22]. However, it is also well known that the variance can be further reduced by allowing a (small) bias [15] (see "Bias And Variance Trade-Off Of Estimators"). This observation is at the start of the regularization methods [15]. The cost function (50) is extended with an additional regularization term $\gamma\tilde{g}^T P^{-1}\tilde{g}$

$$\begin{aligned} V &= \sum_{t=n}^{N}(y(t)-\tilde{y}(t))^2 + \gamma\tilde{g}^T P^{-1}\tilde{g}, \\ &= \sum_{t=n}^{N}(y(t)-\sum_{k=0}^{n}\tilde{g}(k)u(t-k))^2 + \gamma\tilde{g}^T P^{-1}\tilde{g}. \end{aligned} \tag{70}$$

The regularization term pulls the estimates towards zero, resulting in a reduced output variance at a cost of an increasing bias b. For some value of γ, the mean squared error $|b|^2 + \sigma^2$ will reach a minimum as shown in Figure S3. From practical experience, it turns out that a large variance reduction can be obtained for a small increase of the bias.

The success of this approach strongly depends on a proper choice of the regularization matrix P. This choice can be guided by putting the regularized cost function (70) in a Bayesian framework. A comprehensive introduction to this approach is given in the survey paper [15]. The Bayesian framework allows the regularization matrix P to be connected to the prior knowledge of the user. For the IR estimation problem, very valuable prior information is stability and smoothness, and both are shortly discussed below.

Stability: If it is known that the unknown system is stable, it follows that the impulse response should decay exponentially to zero for lumped systems [108].

Smoothness: Most impulse responses of physical systems have an intrinsic smoothness. Again this can be expressed by requiring that neighboring points are correlated to each other.

Both aspects can be translated in a proper choice of the regularization matrix P. A popular choice is the TC (Tuned/Correlated) kernel

$$P_{k,j} = \lambda\alpha^{\max(k,j)}. \tag{71}$$

The hyper parameters α, λ are tuned on the data [15]. In [109], it is shown that this regularization can be interpreted as a restriction on the variability of the low-pass filtered impulse response,

that is forced exponentially to zero.

This simple idea turns out to be very powerful, the additional information that is added by the prior knowledge to the data is very significant, and it results in a strong reduction of the uncertainty of the simulated system output at a cost of a (small) bias on the estimated impulse response. Overall the MSE on the output is minimized by these methods. However, it turns out that this bias is focused at the (dominant) resonance as shown in Figure 17. It leads to an underestimation of the damping at those resonances. This is unacceptable in applications where a precise damping estimate is an important design input. This conflicting observation is mainly due to the different objective of regularization methods (objective is to minimize the MSE of the modeled system output) and FRF measurements (objective is to have a reliable FRF measurement at all frequencies of interest).

Comparison of the Hanning, the least-squares, and the regularized least-squares method

In this section, the FRF is estimated using the Hanning method, with an overlap of 2/3. It is compared with the FRF obtained from the estimated IR using the least-squares and the regularized least-squares.

Simulation 1: a resonating system: The system is excited with filtered white noise up to $0.4f_s$. The output is disturbed with white noise so that the global SNR is 6dB (noise power at 25% of the signal power at the output). The RMS error is obtained from 200 repeated simulations. The results are shown and discussed in Figure 17. From these results it follows that the non-regularized LS gives the best in-band results if no (local) bias is tolerated around the dominant resonance frequency. If that is no problem, the regularized least-squares is the best. At the high frequencies, the errors grow very fast because the excitation power drops to zero, resulting in a very low SNR of the measurements. In this band, the regularized method reduces the error by 20dB or more.

Simulation 2: a first order system with delay: The system is excited with filtered white noise up to $0.4f_s$. The output is disturbed with white noise so that the SNR in the passband of the system is 10 dB. The RMS error is obtained from 1000 repeated simulations. The results are shown and discussed in Figure 18. From these results it follows that the regularized LS gives the best results, followed by G_{LS} with the impulse response length fixed to $n = 8\tau = 40$. Tuning the impulse response length by minimizing the error on a validation set increased the in-band error with 2 dB. This shows again that from FRF point of view, it is not obvious how to select the optimal impulse response length. The errors of the Hanning method are 10 times larger while its spectral resolution is reduced by a factor 8 because the original record was split in 8 subrecords. This shows that for systems with slowly varying dynamics, a huge gain can be made by replacing the classical methods with the recently developed tools. In this case, the spectral methods would need 100 times more measurements to come to a similar quality as the (regularized) LS methods.

Remark: The method is not changed to account for the delay. It will be accommodated automatically in the impulse response estimate: $\hat{g}(k) \approx 0, k \leq delay$ at the start of the estimated impulse response.

User guidelines

- *Direct estimation of the impulse response*: This method leads to high quality estimates with a high noise rejection and low bias errors provided that the selected impulse response length is long enough.
- *Selection of the impulse response length*: (i) For systems with fast varying dynamics like resonating systems, classical model complexity selection methods (AIC, BIC, etc.) cannot be used for the purpose of FRF estimation because these methods balance variance versus bias. This bias is concentrated around the dominant resonance frequencies, leading to an over estimation of the damping. (ii) For systems with slowly varying dynamics, the regularized LS methods can further improve the results of the LS method.
- *Tuning the bias error*: The user can set the lowest damping level that should be estimated without bias. On the basis of that result, the length of the impulse response $n = \alpha\tau$ in the least-squares method can be set see "Truncation ErrorOof Finite Impulse Response Models".
- *Regularization*: In the frequency bands with a low SNR, the results can be significantly improved using the regularization approach. This might come with a bias at those frequencies where the dynamics vary rapidly.

Conclusion

This article gives a tutorial overview of IR and FRF estimation methods. The historical approaches from the 1950's and 1960's are discussed, together with the recently developed methods. It is shown that the available computer resources strongly influence the choice of the best suited algorithm to solve the problem. Today, the classical spectral analysis methods are still dominating the field although it is possible to do significantly better with the recently developed local parametric methods or with direct impulse response estimation methods. Using these methods, shorter experiments can be used to obtain a similar or even better quality of the results.

The discussion of the spectral methods is not done along the same lines as the classical text books. The authors rather preferred to give a new interpretation of the properties of these methods starting from the recent insights in the structured nature of leakage errors.

A short historical overview is included, showing that IR and FRF measurements have been done for a long time. There was a strong evolution that was driven by the major changes in the available technology, starting with graphical methods, next using analog correlation methods, to be eventually driven by digital signal processing methods. Too many engineers and scientists are not aware about the new possibilities that are available today, which leads to a waste of efforts and resources. This article aims at filling this gap.

Below, a final set of user guidelines is formulated, directing the user to the best choice keeping the available computing facilities in mind.

Concluding user guidelines

- *Store the reference signal together with the data*: Whenever the external reference signal is available, it is strongly advised to store it together with the data. This will be very useful whenever there are closed-loops in the system to be tested, including interactions between

the generator and the setup.

- *Use periodic excitations whenever it is possible*: Periodic excitations give access to a full nonparametric noise model, even under closed-loop experimental conditions. Combination with the local parametric methods, or with the IR method, allows the user to eliminate transient effects if needed (high computational cost), otherwise a direct division of the output/input spectra can be made (lowest computational cost).
- *Use the direct IR estimation for systems with rapidly varying dynamics and low SNR measurements*: The direct IR method gives the best noise and leakage rejection. Care should be taken that the impulse response length n is well tuned (e.g. 5 to 10 dominant time constants of the system, depending upon the required precision). The calculation effort grows with n as $O(n^3)$ without special numerical implementations.
- *Use the regularized IR estimation for systems with slowly varying dynamics and low SNR measurements*: The regularized IR reduces the mean square error of the direct IR method. Moreover, this method protects intrinsically against over-modeling so that the choice of the impulse response length becomes less critical.
- *Use the local parametric methods for high SNR measurements*: If the SNR of the data is (very) high, the local parametric methods are a very attractive alternative for the direct IR methods. High quality results are produced without needing to solve large sets of equations.
- *Real time applications*: The local polynomial method is well suited for real time applications. For a fixed input signal, all the matrix inverses can be pre-calculated and stored.
- *Windowing methods*: These methods require the least computer resources of all the methods. For FRF measurements, the combination of a sine-window with a 66% overlap is the best combination. These methods have the highest sensitivity to leakage errors. It is advised to use these methods as a last resort, if the calculation time is very critical, and the LPM methods are still too demanding.

Publicly Available Software

Most of the results can be reproduced using (free) public software.

- The book "Mastering System Identification in 100 Exercises" [73] provides an exercise based introduction to system identification, dealing amongst others with the generation and analysis of excitation signals (Chapter 2), and FRF measurements (Chapter 3). The MATLAB® solutions of the exercises are available on the book support site, booksupport.wiley.com. The software to generate the following figures (or similar figures) can be downloaded from this website: Figure 7 and 8 (Execise 37), Figure 9 (Exercise 29), Figure 10 (Exercise 30), Figure 14 and 15 (Exercises 51 and 52).
- The results in Figure 17 and 18 on IR estimation were obtained using the Matlab® system identification toolbox, using the routine "arxRegul". This toolbox has also a routine "etfe" to calculate the emperical transfer function estimate (ETFE).
- The freely available frequency domain identification toolbox FDIDENT can be used to generate the advanced excitation signals that are discussed in this article (Figure S13), and to make the nonparametric noise analysis using periodic excitations (Figure 6)
- Code to generate special (binary and multi-level) signals is available at https://www2.warwick.ac.uk/fac/sci/eng/research/systems/bbsl/signal_design [28].

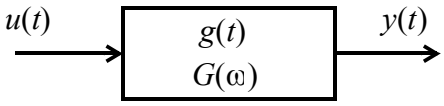

Figure 1. A continuous or discrete-time system is considered with impulse response $g(t)$ and a frequency response function $G(\omega)$. For a continuous-time system $G(\omega)$ is a short hand notation for $G(s = j\omega)$, and for a discrete-time system, $G(\omega)$ stands for $G(z = e^{j\omega})$.

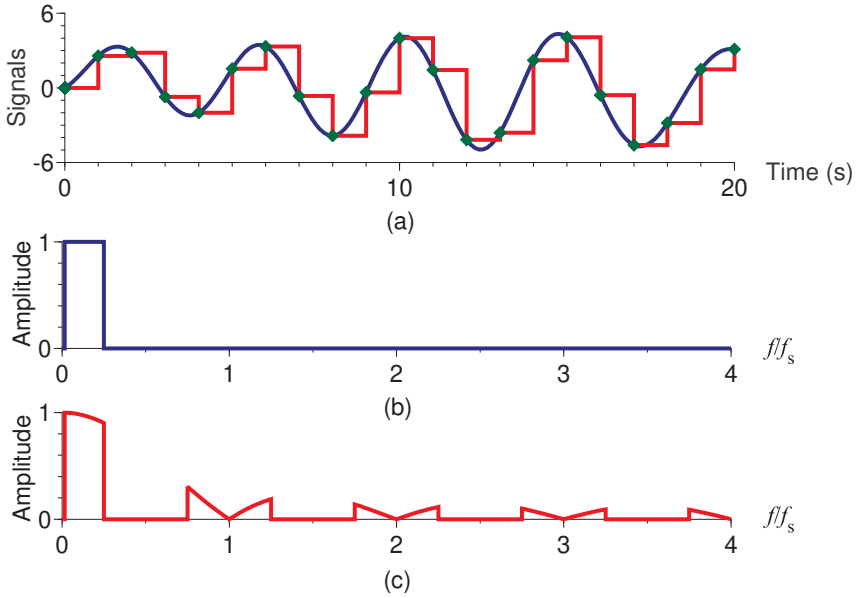

Figure 2. Band-limited and ZOH reconstruction. (a) The time signals, green diamonds: the samples, blue line: BL reconstruction, red line: ZOH reconstruction; (b) amplitude spectrum of the BL limited reconstruction; (c) amplitude spectrum of the ZOH reconstruction. Observe the high frequency repetitions around the multiples of f_s.

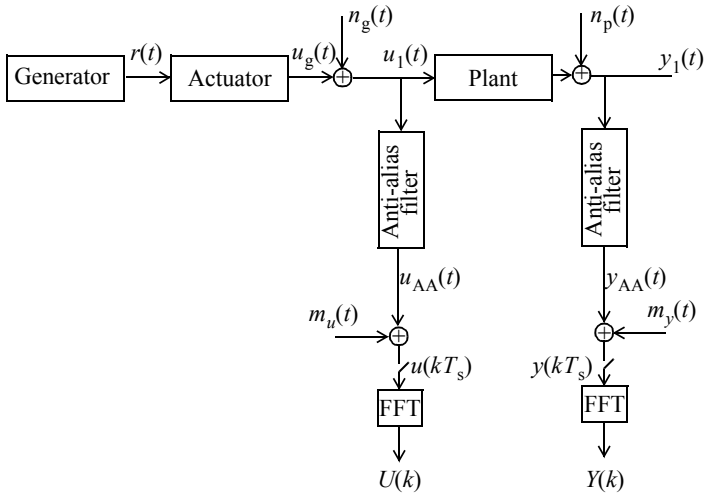

Figure 3. Measurement setup and notations. The generator signal is disturbed by generator noise $n_g(t)$, the output of the system is disturbed by the process noise $n_p(t)$. The measured input and output signals are first low-pass filtered by the anti-alias filters. The measurement noise on the input and output is respectively $m_u(t)$ and $m_y(t)$. These signals are sampled at a sampling rate $f_s = 1/T_s$. The discrete Fourier transform (DFT) of the measurements $u(kT_s), y(kT_s), k = 1, ..., N$ is calculated using the fast Fourier transform (FFT) algorithm, and denoted as $U(k), Y(k)$. The frequency index k indicates the frequency kf_s/N.

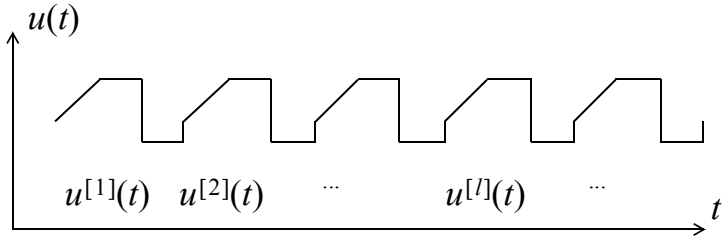

Figure 4. Calculating the sample mean and sample variance of a periodic signal starting from multiple measured periods. The input $u(t)$ and output $y(t)$ are measured over multiple periods, and broken in subrecords of one period each, for example $u^{[l]}, l = 1, ..., P$. The FFT is applied to each subrecord. In the next step, the mean value, and the (co-)variance is calculated as a function of the frequency.

Figure 5. The one degree-of-freedom flexible robot arm of the KULeuven PMA department of Mechanical Engineering. The red mass at the tip of the flexible arm is driven by an electric motor. The input signal is the motor current, the output signal is the acceleration measured at the tip of the robot. A periodic excitation signal is used, the period length is $N = 4096$ with $f_s = 500$ Hz, and $P = 10$ periods are measured in steady-state conditions. Only the odd frequencies $[1, 3, ..., 201]f_0$, with $f_0 = f_s/N = 0.1221$ Hz are excited.

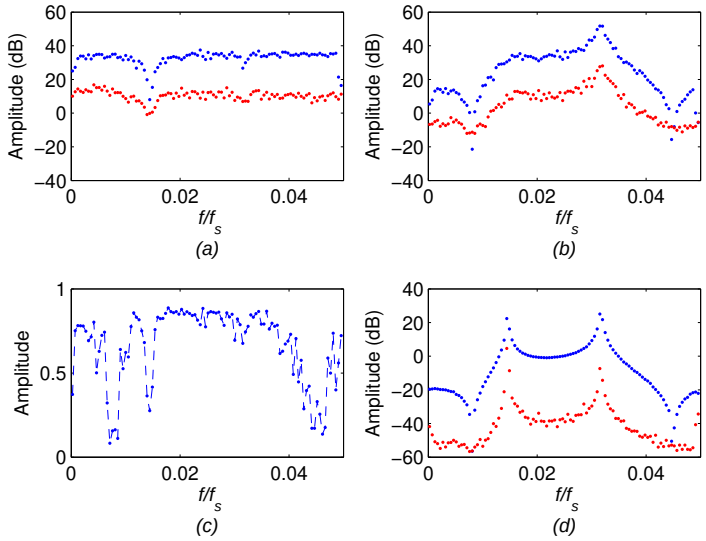

Figure 6. Processing the periodic measurements on the robot arm in Figure 5. (a) DFT spectrum U of the current, (b) DFT spectrum Y of the acceleration, (c) the correlation ρ (11), (d) FRF obtained by (12) $\hat{G} = \hat{Y}/\hat{U}$ (see Section Smoothing FRF measurements for periodic excitations). Blue: the amplitude, Red: the standard deviation.

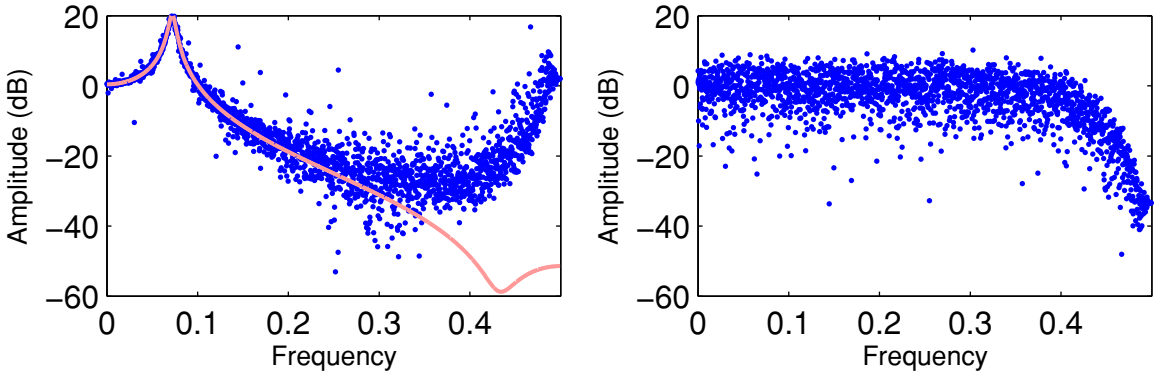

Figure 7. The FRF measurement shown in Figure S6 (left), together with the amplitude $|U|$ of the DFT of the actual realization of the input that was used for this measurement (right). $|U|$ is a random variable too that varies over the frequency. At some frequencies $|U|$ is very small, resulting in large errors on $\hat{G}$

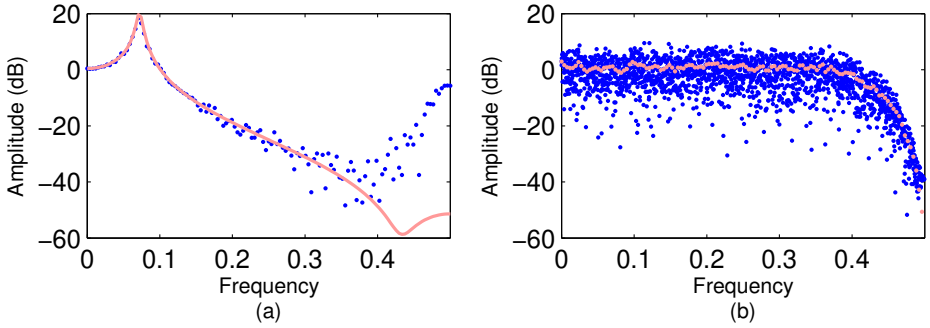

Figure 8. FRF starting from the simulation in Figure S6. The full record is broken in 16 subrecords of length $N = 256$. (a) Pink line: true FRF G, blue dots: FRF measurement $\hat{G}(k) = \hat{S}_{YU}(k)/\hat{S}_{UU}(k)$. (b) Blue dots: $|U|$ for the full record, Pink dots: $\hat{S}_{UU}$. Observe that the frequency resolution of the pink dots is 16 times smaller than that of the blue dots. This is because the length of the subrecords is 16 smaller than that of the full record.

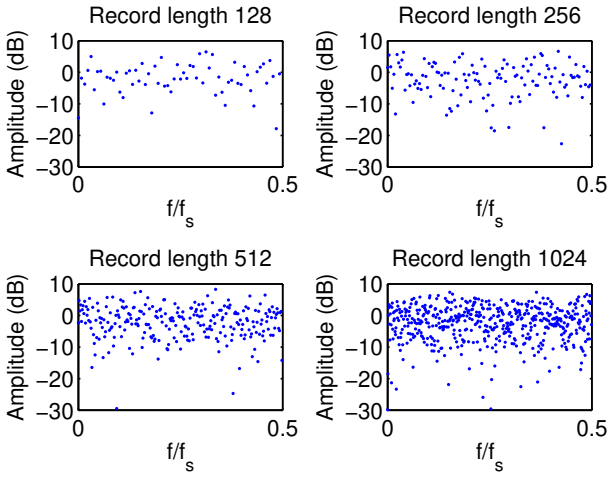

Figure 9. DFT spectrum of white noise for a growing record length. Observe that the spectrum does not converge to a constant as it could be expected for white noise. Longer measuring does increase the number of frequencies on which the spectrum is calculated, but the variability of the measurement does not decrease.

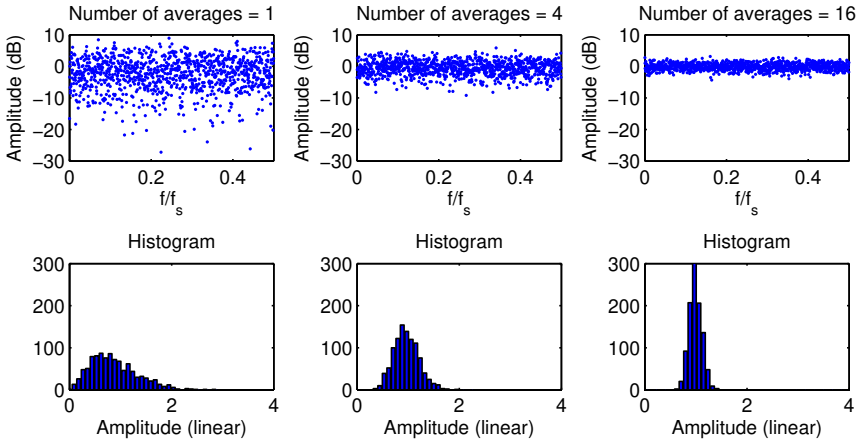

Figure 10. Illustration of the smoothing effect that is obtained by averaging $\hat{S}_{UU}$ over different realizations. The power spectrum is shown in the upper figures for $M = 1, 4, 16$, and $N = 2048$. The lower figures show the corresponding histograms. It can be shown that these follow a χ^2 distribution with $2P$ degrees of freedom [16].

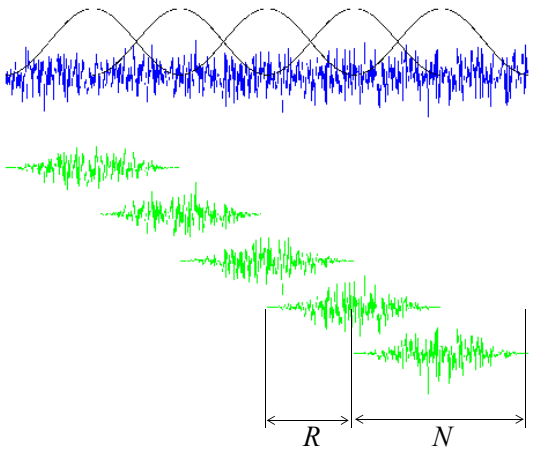

Figure 11. Illustration of the overlapped weighted segments - here with a Hanning window w_H (black lines) and an overlap $1 - R/N = 1/2$ (the window is shifted R samples in each step). Each subrecord is multiplied with the window (shown in green, shifted below the original signal in blue), and the DFT is applied to the product.

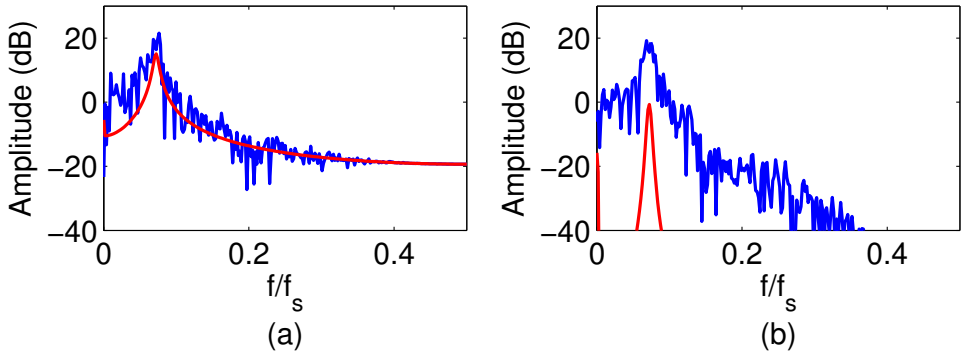

Figure 12. Impact of the Rectangular (a) and Hanning (b) window on the signal term GU (blue) and the transient term T_G (red). Observe that the signal term is 'rough', and the transient term is smooth. The double differentiation of the Hanning window is reducing the smooth transient term, while the amplitude level of the rough signal term remains the same. The system is the same as in Figure S6 and Figure 8, the data length is $N = 512$.

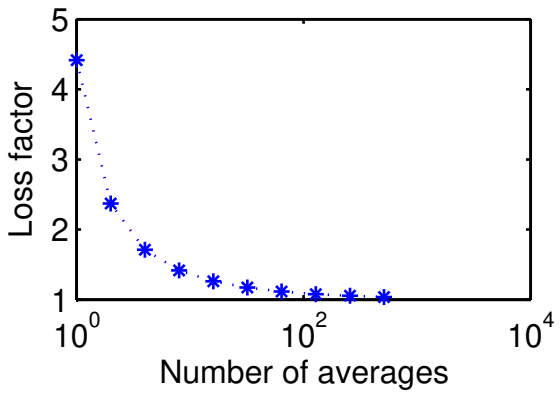

Figure 13. Impact of replacing S_{UU} by its finite sample estimate $\hat{S}_{UU}$ in $\hat{G} = \hat{S}_{YU}/\hat{S}_{UU}$. The estimated power spectrum $\hat{S}_{UU}$ is a random variable that can become very small when the number of averages is small, as shown in Figure 10. These dips create large spikes after the division in $\hat{G}$, leading to an increase of the standard deviation $\sigma_{\hat{G}}$ of $\hat{G}$. The 95% bound of the relative increase of the standard deviation $\sigma_{\hat{G}}$ is plotted.

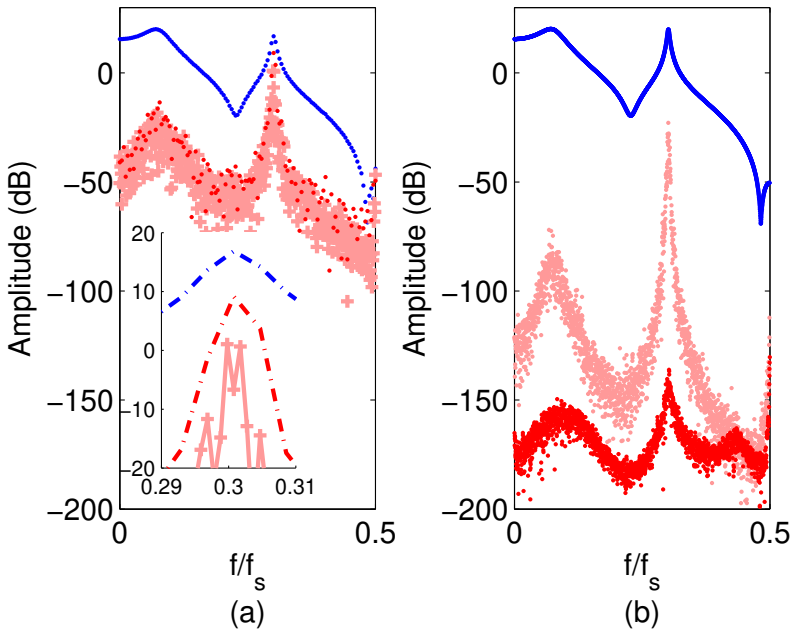

Figure 14. Illustration of FRF measurements using the classical Hanning method and the LPM/LRM on undisturbed data. A system with a highly and a lowly damped resonance is excited with filtered white noise (bandwidth $0.4f_s$). The full record length is 4096 samples. The blue dots are the estimated FRF, the red and pink dots plot the errors. Figure (a) shows the results for the Hanning method that is applied on subrecords with a length of $N = 256$ (red) and $N = 1024$ (pink) samples, both with an overlap of $R = \frac{2}{3} \times N$. In the inset of (a) a zoom around the second resonance is given. Figure (b) shows the results for the LPM (pink) and LRM (red), applied to the full length record. Observe that the errors of the LPM/LRM are an order of magnitude smaller than those of the Hanning method. The LRM method outperforms the LPM method.

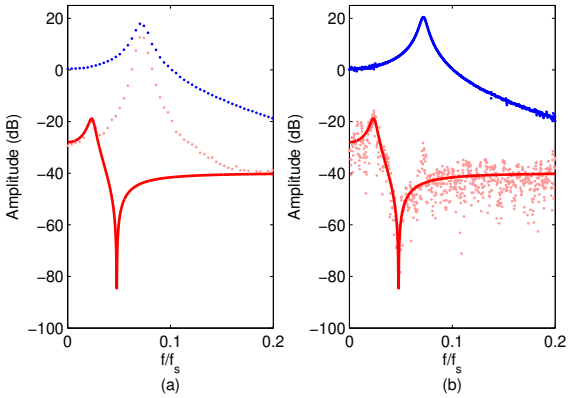

Figure 15. Estimation the output noise variance σ_Y^2 using the coherence method with a Hanning window (a), and the LPM (b). A data record with a total length of 256×16 is split into 16 subrecords for the coherence method, while it is processed as one long record for the LPM approach. The true value of the power spectrum is plotted in red. The blue dots show the estimated FRF, the pink dots show the estimated power spectrum. Observe that for the coherence method the system dynamics are visible in the estimated power spectrum due to the poor separation of the system dynamics and the disturbing output noise. Moreover, the antiresonance is completely lost. This is not the case for the LPM method where the true noise spectrum is well followed by the estimates.

Figure 16. Hammer excitation in a mechanical vibration analysis in the vibration labs of Siemens Industry Software Leuven, Belgium. A force transducer is put in the tip of the hammer, allowing the impact force to be measured. Accelerometers are mounted on the mechanical structure to measure the response of the system.

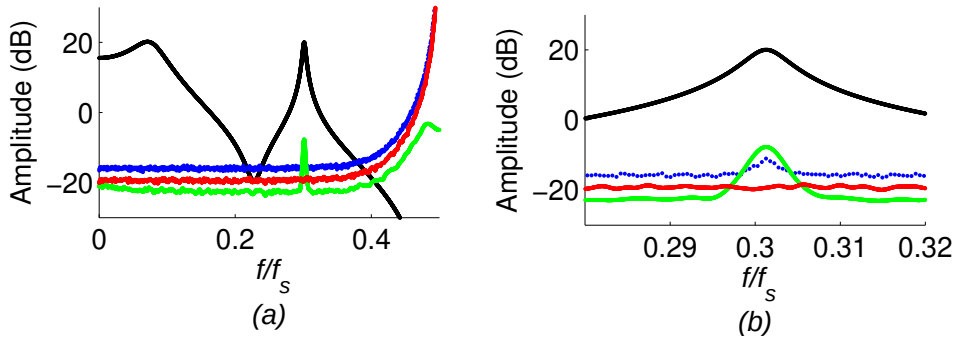

Figure 17. Comparison of three FRF measuring methods. The true FRF (black), and the observed RMS error ($\sqrt{Bias^2 + variance}$) of the Hanning (blue), IR_{LS} (red), and regularized IR_{reg} (green) method are shown. The true system consists of the sum of two resonating systems having respectively a time constant $\tau_1 = 7$ and $\tau_2 = 66$. The Hanning method uses subrecord lengths of $N_{sub} = 2048$, and the total data length is $16 \times N_{sub}$. An overlap of 2/3 is used, resulting in an additional reduction of the variance with a factor 2. For the IR_{LS}, the impulse response length was set to $n = 7 \times \tau_2 = 462$. It can be seen that overall, the RMS error of IR_{reg} is the smallest, especially at the large frequencies where the SNR becomes very low due to the filtered input. This comes with a large bias around the second (dominant) resonance (see also the zoom in Figure (b)). The RMS error of the Hanning method is the largest, and also this method has a significant bias around the second resonance. The IR method shows no bias and has a smaller RMS error than the Hanning method.

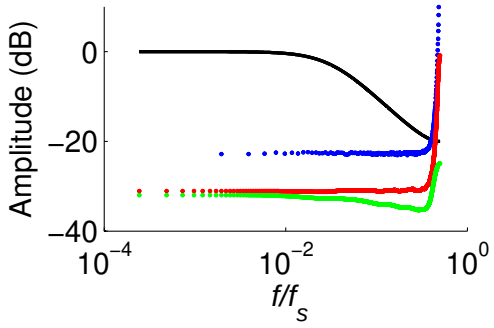

Figure 18. Comparison of three FRF measuring methods. The true FRF (black), and the observed RMS error ($\sqrt{Bias^2 + variance}$) of the Hanning (blue), IR_{LS} (red), and regularized IR_{reg} (green) method are shown. The true system mimics a process. It consists of a first order system with a time constant $\tau = 5$ and a delay of 2τ. The Hanning method uses subrecord lengths of $N_{sub} = 512$, and the total data length is $8 \times N_{sub}$. An overlap of 2/3 is used, resulting in an additional reduction of the variance with a factor 2. For the IR_{LS} and IR_{reg}, the impulse response length was set to $n = 8 \times \tau_2 = 40$. It can be seen that overall, the RMS error of IR_{reg} is the smallest with a gain of 1 dB with respect to IR_{LS} and about 20dB with respect to the Hanning method. At the high frequencies, where the SNR is very low due to filtered input, the gain is even larger. Observe also that the frequency resolution is 8 times higher for the least squares methods than that of the Hanning method.

TABLE 1. In the signal processing field, many windows are defined and studied [61]. Classical windows that are very popular are the Rectangular and Hanning window. More recently the Diff and Half-sine windows were studied in detail [63]. The definition of these windows $w(x)$ is given for $x \in [0, 1[$

Window	$w(x)$	X_w
Rectangular	1	$X_w(k) = X(k)$
Hanning	$0.5(1 - \cos 2\pi x)$	$X_w(k) = \frac{1}{4}[-X(k-1) + 2X(k) - X(k+1)]$
Diff	$e^{j2\pi t} - 1$	$X_w(k + 1/2) = [X(k+1) - X(k)]$
Half-sine	$\sin \pi x$	$F(\sin \pi x)$

TABLE 2. Applying a window in FRF measurements reduces the transient errors E_{T_G} and creates interpolation errors E_G. These errors, together with their bias and variance are given as a function of N: the length of the subrecord, the dominant time constant in the frequency band of interest τ_f, and the second derivative of G with respect to the frequency $G^{(2)}$.

Window	E_{T_G}	E_G	Bias	Variance
Rectangular	$O((N/\tau_f)^{-1/2})$	0	$O((N/\tau_f)^{-1})$	$O(P^{-1}(N/\tau_f)^{-1})$
Hanning	$O((N/\tau_f)^{-5/2})$	$O((N/\tau_f)^{-1})$	$\lvert G^{(2)}(k)\rvert O((N/\tau_f)^{-2})$	$O(P^{-1}(N/\tau_f)^{-2})$
Diff or Half-sine	$O((N/\tau_f)^{-3/2})$	$O((N/\tau_f)^{-1})$	$\lvert G^{(2)}(k)\rvert O((N/\tau_f)^{-2})$	$O(P^{-1}(N/\tau_f)^{-2})$

Sidebar: This is the Sidebar Title

Article Summary

Frequency response function (FRF) and impulse response (IR) measurements give a first view on the dynamics of a system. In the 1960's, FRF measurement methods were developed that were optimized with respect to the available computer power at that time. These by now classic methods are still most popular in engineering curricula and in industry. However, today more advanced algorithms are available that can reduce the measurement time and errors with a factor 2 tot 100 by making better use of the spectacular increased computer power. Because these methods are not well known by the large public, they are not frequently used, leading to a waste of money, resources, and time. The goal of this article is to bridge this gap by providing a deeper insight in the underlying problems of FRF and IR measurements, and to use this better understanding to introduce the recent more powerful methods. Links to publicly available software will be provided. This helps to reconstruct the results shown in this article, and minimizes the effort to use these new algorithms in their daily practice.

Sidebar: This is the Sidebar Title

Random Nature Of FRF Measurements

FRF measurements $\hat{G}(k)$ (often called FRF estimates) that are obtained from finite length measurements differ from the true value $G_0(k)$ by an error term $N_G(k)$, with $\hat{G}(k) = G_0(k) + N_G(k)$ because the estimation process is disturbed by many error sources (see Figure 3). The measured input and output are prone to measurement noise $m_u(t), m_y(t)$; unknown inputs can act on the system to be modeled which leads to process noise $n_p(t)$. Also finite measurement length effects disturb the estimate $\hat{G}(k)$, because the past inputs (before the start of the experiment) and the future outputs (after finishing the measurements) are not always properly included in the calculations.

In Figure S6, a typical FRF measurement is shown. It can be observed that the errors vary rapidly from one frequency to another. Repeating the experiment leads to a new FRF measurement $\hat{G}(k)$ that differs from the previous one because the input changed (for random excitations). Also the measurement and process noise varies from one experiment to the other.

The error term $N_G(k)$ is modeled as a random variable that can be studied using statistical tools. The process of generating each time a new noise sequence is called a 'realization' of the noise.

Sidebar: This is the Sidebar Title

Characterizing The Stochastic Properties Of FRF Estimates

A random variable is fully characterized in amplitude by its probability density function (pdf) [3], [55]. Often it is too difficult to obtain the pdf of $N_G(k)$, and for that reason the pdf is replaced by partial information: its mean value $\mu(k) = E\{N_G(k)\}$ and its variance $\sigma^2 = E\{|N_G(k) - \mu(k)|^2\}$, where $E\{x\}$ stands for the expected value of x [55].

If $\mu(k) \neq 0$, the FRF measurement $\hat{G}(k)$ is prone to a systematic error that cannot be removed by averaging the result over multiple measurements. Such an error is called a bias. Often it is desired to make the bias as small as possible because it can not be (easily) removed any more in the remaining of the data processing. This is further discussed in "Bias And Variance Trade-Off Of Estimators".

The variability of $\hat{G}(k)$ around its expected value is characterized by the variance σ^2. The larger the variance, the wider the spread around the expected value will be. For a Gaussian pdf, the interval $[-1.96\sigma, +1.96\sigma]$ corresponds to the 95% confidence interval.

In practice, the confidence of the measurement $\hat{G}(k)$ is often given by drawing either the $[-1.96\sigma, +1.96\sigma]$ interval around the measured value, or by plotting also $\sigma(k)$ as was done in Figure 6. Observe that these intervals do not account for the presence of a bias (systematic errors).

Sidebar: This is the Sidebar Title

Impact Of Averaging On The Variance Of The FRF Estimate

Averaging an estimate over P independent realizations (loosely speaking this means that one measurement has nothing to do with the other) results in a reduction of the variance σ^2 by a factor P, or a reduction of the standard deviation by $\sqrt{P}$

$$\sigma_{\hat{G}_{aver}} = \sigma_{\hat{G}}/\sqrt{P}. \tag{S1}$$

Averaging can be done by repeating the experiment P times. This looks like a very simple and attractive solution to reduce the standard deviation, but it comes with a price. The measurement time grows proportional with P, while the uncertainty drops only by $\sqrt{P}$. Reducing the standard deviation with a factor 10 increases the measurement time by a factor 100.

Sidebar: This is the Sidebar Title

The Discrete Fourier Transform (DFT)

Consider a discrete time sequence $x(t), t = 0, 1, \ldots, N-1$. The discrete Fourier transform (DFT) and the inverse discrete Fourier transform (IDFT) are then given by

$$X(k) = \frac{1}{\sqrt{N}} \sum_{t=0}^{N-1} x(t) e^{-j2\pi tk/N}, \tag{S2}$$

and

$$x(t) = \frac{1}{\sqrt{N}} \sum_{k=0}^{N-1} X(k) e^{j2\pi kt/N}, \tag{S3}$$

$X(k)$ is the Fourier coefficient of $x(t)$ at frequency $f_k = kf_s/N$, with $f_s = 1/T_s$ the sampling frequency of the discrete time sequence. The scale factor of the DFT can vary from one definition to the other, depending on the purpose. For random excitation the $1/\sqrt{N}$ is popular because it returns an averaged amplitude that is independent of N.

Sidebar: This is the Sidebar Title

Link Between The Data Length And The Frequency Resolution Of The FRF Estimate

The frequency resolution δ_f of a FRF measurement is directly linked to the length of the subrecords N_{sub}

$$\delta_f = f_s/N_{sub} = Pf_s/N = P/T_{meas} \tag{S4}$$

with P the number of subrecords , $N_{sub} = N/P$, f_s the sample frequency, and T_{meas} the total measurement time. The choice of P is a user choice that is set by the balance between the required frequency resolution and the total measurement time. A higher value of P allows for more averages and a lower standard deviation, at a cost of a reduced frequency resolution for a given measurement time.

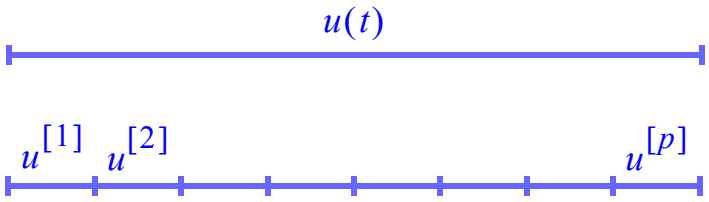

Figure S1. Splitting a long data record in P shorter subrecords reduces the original record length N to a subrecord length N/P. The corresponding frequency resolution drops from f_s/N to f_sP/N.

Sidebar: This is the Sidebar Title

Characterizing A Resonance By Its 3 dB Bandwidth

The transfer function of a system can be written as the sum of first order subsystems (with a real pole) and second order subsystems (with two complex conjugate poles). Lowly damped poles create (sharp) resonances in the FRF, and many FRF methods are most prone to errors around these resonance frequencies as shown in Figure 14.

The frequency resolution δ_f of the FRF measurement should be "small" enough to capture properly the resonance peaks, a too small frequency resolution would underestimate the peak value of the FRF. To make a quantified statement, the 3 dB bandwidth of a resonance can be use. In Figure S2, the 3 dB bandwidth is shown in pink. The 3 dB bandwidth B_{3dB} in rad/s is given by the width of the frequency interval around the peak amplitude of the FRF where $|G(\omega)|_{dB} \geq G_{max} - 3$.

There is a direct link between B_{3dB}, the damping ζ, and the time constant τ of a second order system with resonance frequency ω_n

$$B_{3dB} = 2\zeta\omega_n = 2/\tau. \tag{S5}$$

This equation provides a lot of insight. The frequency resolution of an FRF measurement should be high enough to cover all resonances of interest. For example, the number of frequency points F_{3dB} in the 3dB bandwidth where the FRF is measured is given by

$$F_{3dB} = \frac{N_{sub}}{\pi\tau f_s}, \tag{S6}$$

relating directly the time constant τ of the system to the record length N_{sub}.

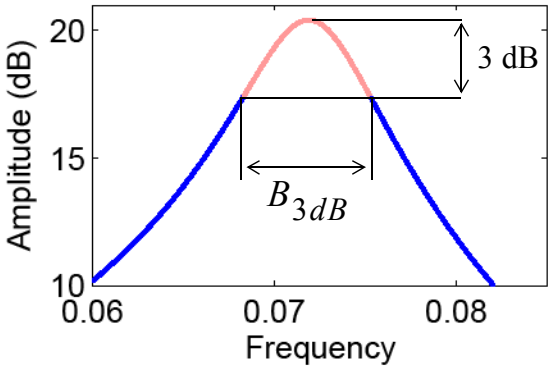

Figure S2. Characterizing a resonating system by its 3 dB bandwidth, indicated in pink.

Sidebar: This is the Sidebar Title

Bias and Variance Trade-Off Of Estimators

An error e can always be written as the sum of its mean value $b = E\{e\}$, called the bias, and the remaining part $v = e - b$ with variance σ^2

$$e = b + v. \tag{S7}$$

The total mean square error is

$$e_{MS} = b^2 + \sigma^2. \tag{S8}$$

Depending on the preference, either the bias b or the mean square error e_{MS} should be as small as possible. It is always possible to scale an unbiased estimator (no bias present) towards zero such that e_{MS} drops. This is illustrated in Figure S3 on the following simple scalar example. Assume that $\hat{\theta}$ is an unbiased estimate of the true parameter $\theta_0 = 1$, with variance $\sigma^2 = 1$. Consider next the scaled estimator $\tilde{\theta} = \lambda\hat{\theta}$. The bias of $\tilde{\theta}$ is $b = (1 - \lambda)$, and the variance $\tilde{\theta}$ is $\sigma^2_{\tilde{\theta}} = \lambda^2$. The MSE becomes

$$e_{MS} = (1 - \lambda)^2 + \lambda^2. \tag{S9}$$

This error, plotted in Figure S3 as a function of the scaling factor λ, shows a minimum at $\lambda = 0.5$.

User guideline Bias and variance trade-off for IR and FRF measurements: Depending on the need of the user, it is preferred to tune the results to have either a low bias or variance.

- *I*f the nonparametric IR or FRF estimates will be used as the input of a parametric modeling step, it is most important that no bias errors are present because these cannot be removed anymore in the later post processing. The parametric estimation step can be considered as an advanced smoothing algorithm that reduces the noise without introducing a bias error if it is well-designed [16], [21], [22].
- *T*o generate initial estimates for a parametric estimation step, a bias can be tolerated [75]. The final estimation should be done on the original not smoothed bias free data.
- *I*f the nonparametric results will be used as such, it is important that the combined bias/variance error is as small as possible so that the mean square error is a better measure of the quality.

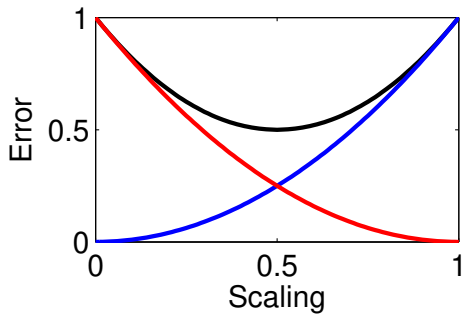

Figure S3. Evolution of the total mean square error as a function of the squared bias and variance error. The bias-variance error trade-off is tuned by scaling an unbiased estimator with a scaling factor between 0 and 1. Black: mean square error, Red: bias error, Blue: variance error

Sidebar: This is the Sidebar Title

Truncation Error Of Finite Impulse Response Models

Most physical stable systems have an impulse response that decays exponentially to zero. In theory, such an impulse response is infinitely long (IIR), in practice it can be arbitrarily well approximated by a finite impulse response (FIR) model. This is very attractive from modeling point of view. The length of the FIR that is needed to keep the truncation error below a given level can be normalized on the time constant of the system. For a first order system, with an impulse response $g(t) = \alpha e^{-t/\tau}$, excited by white noise, and truncated at time T, the relative root mean square truncation error e_{RMS} is given by

$$e_{RMS} = e^{-T/\tau}. \tag{S10}$$

This error is shown in Figure S4 as a function of the truncation time. Truncating the impulse response at T equal to $2\tau, 5\tau, 7\tau$, results respectively in a relative RMS error below 10%, 1%, and 0.1%.

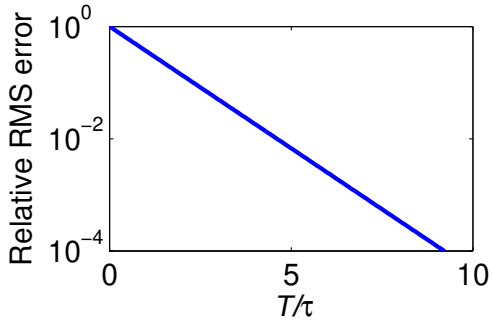

Figure S4. Relative RMS error induced by truncating the impulse response of a first order system with time constant τ.

Sidebar: This is the Sidebar Title

Models For Dynamic Systems: Finite Length Effects

A linear dynamic system is described by (1). When this relation is applied to finite length records, additional effects show up as illustrated in Figure S5 that plots the simulated input and output for a random noise excitation. The time domain record record is split in three subrecords, and for each of these the corresponding output is calculated (blue, green, and red), putting the input equal to zero outside the subrecord. When only the middle subrecord is processed, it is clear that two problems appear: i) The beginning of the output measurement is disturbed by the blue output transient of the first subrecord; ii) The green output-transient of the middle subrecord is missing because it belongs to the third output subrecord.

The effect of both transient terms is that the simple relation $Y(k) = G(k)U(k)$ beween the DFT coefficients $U(k), Y(k)$ of the middle record does no longer hold. An additional term is needed to account for it. Failing to do so creates errors in the IR and FRF estimate, even for noise free measurements, as shown in Figure S6.

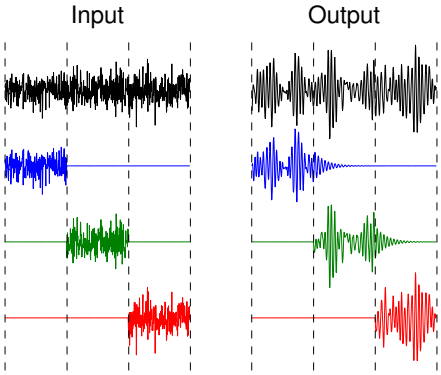

Figure S5. Understanding finite length effects on the relation $Y(k) = G(k)U(k)$ between the DFT coefficients $U(k), Y(k)$ of the middle record for a linear dynamic system. Impact of the beginning (additional blue transient at the start of the middle subrecord) and end effects (missing green transient at the start of the left subrecord)

Adding the transient terms to $Y(k) = G(k)U(k)$ is done below for SISO and MIMO systems, in the time and in the frequency domain. There is a full equivalence between the time and frequency-domain equations.

In the time domain, the finite length effects are known as transient effects that are due to the additional blue transient in Figure S5. In the frequency domain, the finite length effects are called leakage errors. These are due to the combined effect of the blue and green transient. Usually, leakage errors are considered to be random errors, although it is clear from the previous discussion and the mathematical description below that they are highly structured errors. For a long time this was not fully realized, and only recently has this insight been fully explored to develop more powerful nonparametric methods that can remove the finite length effects [19], [20], [66], [67], [69], [71]–[73]. These are discussed in the Section Improved FRF measurements using local parametric methods.

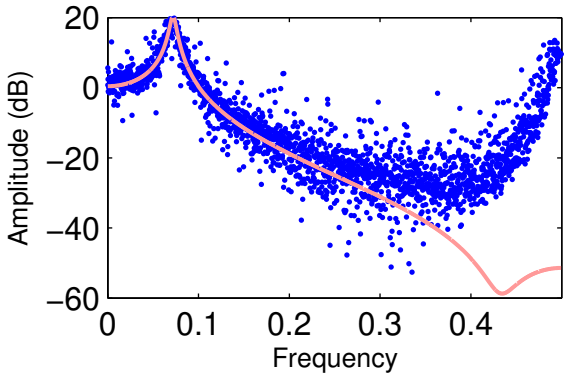

Figure S6. Simulation results on a 2^{nd} order discrete-time system: FRF measurement starting from a finite length record of undisturbed data, using a filtered random noise excitation with a bandwidth of $0.4f_s$. Pink line: true FRF G, blue dots: FRF measurement. Observe that even in the absence of disturbing noise the measurements look very noisy due to the leakage effect. In this case, the response of a second-order system with a time constant of about 8 samples was simulated in $N = 4096$ samples.

SISO systems

Time domain: including the transient effects

Consider the finite length input/output measurement $u_{FL}(t) = u(t), y_{FL}(t) = y(t), t = 1, ..., N$ of the discrete-time SISO system in (1). When equation (1) is applied to $u_{FL}(t), y_{FL}(t)$, putting $u_{FL}(t < 0) = 0$, it needs to be extended with an additional transient term $t_g(t)$ to account for the unmeasured past inputs prior to the start of the experiment

$$y_{FL}(t) = g(t) * u_{FL}(t) + t_g(t). \tag{S11}$$

In a parametric presentation of (S11), the impulse response $g(t)$ and the transient $t_g(t)$ are both modeled by a rational form in the Laplace or z-domain

$$G = B_G/A_G, T_G = I_1/A_G, \tag{S12}$$

with A_G, B_G, I_1, polynomials in s or z.

Frequency domain: modeling the leakage errors

Transforming (S11) to the discrete frequency domain using the discrete Fourier transform (DFT) (S2) gives the following relation between the DFT coefficients $U(k), Y(k)$ [16], [19], [70], [71]

$$Y(k) = G(k)U(k) + T_G(k), k = 0, 1, ..., N/2. \tag{S13}$$

It can be shown that $DFT\{t_g\} \neq T_G$ because the latter depends also on the missing green output transient in subrecord 3 in Figure S5, while this is not the case for $t_g(t)$ [16], [19], [70], [71].

For a random noise or quasi stationary excitation [16], [21]

$$U(k) = O(N^0), T_G(k) = O(N^{-1/2}), \tag{S14}$$

with $O(x)$ an arbitrary function with the property (in probability) $\lim_{x\to 0} |O(x)/x| < \infty$. This shows that the transient effect decreases as $O(N^{-1/2})$ to zero for a growing record length..

The parametric presentation (S15) of (S13) models the impulse response $g(t)$ and the transient $t_g(t)$ again by a rational form in the z-domain that is very similar to (S12)

$$G = B_G/A_G, T_G = I/A_G, \tag{S15}$$

but this time the numerator of the transient term $I = I_1 - I_2$, where the polynomials I_1, I_2 depend respectively on the beginning and end conditions of the finite record. A_G, B_G, I_1, I_2 are polynomials in z as in the time-domain description. Again, these results can be extended to the continuous-time domain by adding a (small) error term that accounts for the alias errors of the sampled transient term [16].

State space interpretation of the leakage errors

An interpretation of these results using an A, B, C, D state space representation is also very revealing [71]. In that case

$$T_G(z) = zC(zI - A)^{-1}(I - A^N)(x_0 - x_p), \tag{S16}$$

with

$$x_p = (I - A^N)^{-1} \sum_{t=0}^{N-1} A^t B u(N - 1 - t), \tag{S17}$$

with x_0 the initial state. These expressions show clearly that the additional term is due to the transient response on $x_0 - x_p$, which is the difference in initial conditions between the true system and the assumed periodic one in the DFT analysis. As mentioned before, these expressions are only valid at the DFT-frequencies kf_s/N.

Observe that the Z-transform of t_g and that of T_G are both given by a rational form of the same order, having the same poles as the plant transfer function $G(z)$. These results extend also to continuous-time systems, provided that the sampling frequency is high enough such that the aliasing error of the sampled transient signals is negligible.

Experimental illustration

The above results show that transient and leakage errors are highly structured. The induced errors have a 'smooth' aspect, in both the time and the frequency domain. This was a key observation for the development of the new generation of nonparametric frequency-domain algorithms. In Figure S7 an experimental illustration of these results is shown.

MIMO systems

All the SISO results can be generalized to MIMO systems with n_u inputs and n_y outputs. These are characterized by the IR matrix (IRM) $g(t) \in \mathbb{R}^{n_y \times n_u}$ in the time domain, and by the frequency response matrix (FRM) $G(k) \in \mathbb{C}^{n_y \times n_u}$ in the frequency domain. The transient term

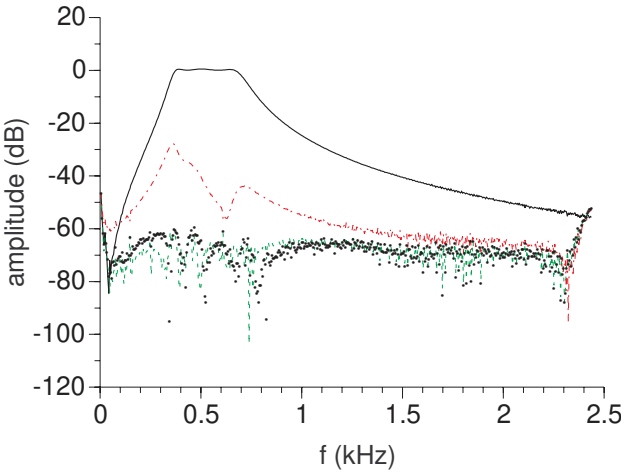

Figure S7. Illustration of the importance of the transient term in (S13). A 6^{th} order bandpass filter is excited with random noise. A reference FRF is measured with a periodic steady-state experiment. Next, the data are used to identify a parametric transfer function model with or without the transient term T_G. Black line: Reference measurement G_{ref} using the steady-state periodic measurements. Red dots: error on the model without T_G estimated on the random noise data. Green dots: error on the model with T_G estimated on the random noise data. Black dots: error of the model that is estimated on the reference data without transient term. Observe that the transient errors are smooth. Including the transient term brings the errors of the full model that is identified on the random data down to the same level as the errors in the model that is identified from the periodic reference data without the transient term.

becomes a vector $t_g \in \mathbb{R}^{n_y \times 1}$, or $T_G \in \mathbb{C}^{n_y \times 1}$.

There are three different choices for the parametric representation of G. The common denominator form

$$G(z) = B(z)/A(z), \tag{S18}$$

with $G \in \mathbb{C}^{n_y \times n_u}$, and A a scalar. Alternatively, the left matrix fraction form

$$G(z) = A(z)^{-1}B(z), \tag{S19}$$

with $A \in \mathbb{C}^{n_y \times n_y}$, and $B \in \mathbb{C}^{n_y \times n_u}$, or the right matrix fraction form

$$G(z) = B(z)A(z)^{-1}, \tag{S20}$$

with $A \in \mathbb{C}^{n_u \times n_u}$, and $B \in \mathbb{C}^{n_y \times n_u}$ can be used.

A detailed discussion on the differences and similarities between these choices is out of the scope of this article, more information can be found in [10].

Concatenating experiments

Using the generalized plant plus transient model structure, it becomes possible to write a model for concatenated data records. At each concatenation point, an additional transient is added to the model [74]. This allows longer records to be obtained starting from short experiments,

resulting in an increased frequency resolution.

Sidebar: This is the Sidebar Title

FRF Measurements For MIMO Systems

All the results for SISO FRF measurements also hold for the measurement of the MIMO frequency response matrix (FRM) at each frequency k

$$\boldsymbol{G}(k) \in \mathbb{C}^{n_y \times n_u}, \tag{S21}$$

with n_u and n_y the number of inputs and output of the system. The major change with respect to the SISO setup is the need to separate the effect of the multiple inputs on a given output. The number of outputs does not add to the complexity of the problem. Consider, without loss of generality, a system with $n_u = 3$ inputs and $n_y = 1$ output

$$Y(k) = \begin{pmatrix} G_1(k) & G_2(k) & G_3(k) \end{pmatrix} \begin{pmatrix} U_1(k) \\ U_2(k) \\ U_3(k) \end{pmatrix}. \tag{S22}$$

The FRM (S22) cannot be identified from one experiment unless constraints are imposed on either the excitation signals U_1, U_2, U_3, or on the smoothness of the system. For uncorrelated inputs, it is possible to consider two unmeasured inputs as a disturbance, but this will increase the uncertainty [4]. Alternatively, three or more experiments can be combined to resolve the problem. Each of these options is briefly discussed below.

FRM measurements using a single experiment

Zippered multisines

A first possibility to separate the different input channels would be to excite one input at a time. This results in a drop of the SNR because the other inputs are set to zero. An alternative implementation of this idea is to realize the separation in the frequency domain by assigning each excitation frequency to only one input. This leads to zippered multisines. The p^{th} input u_p is exciting only the frequencies $k = p + n * n_u$, with $n \in \mathbb{N}$. This choice decouples the inputs in (S22) over the frequency, and at the selected frequencies the SISO methods can be applied. The cost is a reduction of the frequency resolution with a factor n_u. The method requires also a complete control over the excitation signal, input distortions that disturb the zeros in the spectra destroy also the decoupling properties [5].

Local parametric methods

Using the local parametric methods, generalized to MIMO systems, is an alternative to separate the different inputs [20], [77]. If the input is sufficiently rich (for example random phases), the local parameterized models can still be identified from a single experiment by combining the neighboring frequencies in one set of equations. This is a very attractive and robust alternative to the other approaches that are discussed in this section. The properties of this approach remain the same as those discussed in the Section Improved FRF measurements using local parametric methods. A detailed analysis is made in [20]. The more inputs are considered, the larger the local bandwidth B that is used in the local parametric methods will be.

FRM measurements using multiple experiments

In this approach, $n_e \geq n_u$ experiments are made, and the spectra of the input and output at a given frequency k are stored as the columns of the matrices $\boldsymbol{U} \in \mathbb{C}^{n_u \times n_c}$ and $\boldsymbol{Y} \in \mathbb{C}^{n_y \times n_u}$. The FRM at frequency k is then retrieved as the least-squares solution [6]

$$\hat{\boldsymbol{G}} = \boldsymbol{Y}\boldsymbol{U}^H((\boldsymbol{U}^H\boldsymbol{U})^{-1}. \tag{S23}$$

The condition number $\kappa_{\boldsymbol{U}}$ of $(\boldsymbol{U}^H\boldsymbol{U})$ has a direct impact on the covariance matrix of $\hat{\boldsymbol{G}}$. It turns out to grow rapidly with the number of inputs n_u [6]. For (white) noise excitations, the conditioning can be improved by collecting more experiments n_e, but this comes at a cost of an increased measurement time. For that reason alternative experimental procedures were proposed. The first group works for $n_u = 2^n$, and is based on the use of repeated inputs with well selected sign switches that are generated using Hadamard matrices [7], [8]. Later these results were further generalized to an arbitrary number of inputs [6]. Alternatively, well-designed binary signals can be used [9]. All these methods result in a condition number $\kappa_{\boldsymbol{U}} = 1$. This leads to very strong reductions of the uncertainty for systems with a large number of inputs. For example, the standard deviation is reduced with a factor 10 for a system with 10 inputs.

Sidebar: This is the Sidebar Title

Measuring The FRF Under Closed-Loop Conditions

Measuring the FRF of a system under closed-loop conditions (Figure S8) requires special precautions. Depending on the SNR of the measurements, the resulting FRF is the FRF of the feedforward pathway, or the inverse FRF of the feedback pathway, or a combination of both results.

The FRF estimate starting from the cross- and auto-spectrum $\hat{S}_{YU}$, $\hat{S}_{UU}$, is $\hat{G}(k) = \hat{S}_{YU}(k)/\hat{S}_{UU}(k)$ (19). When the measurement is made under feedback conditions (see Figure S8), the output $y(t)$ depends on both the measured input $u(t)$ and the disturbance source $v(t)$. Due to the presence of the feedback loop, the signal u depends also on the disturbance v. As a result, the FRF measurement at frequency k converges to [17]

$$\tilde{G} = \frac{GS_{RR} - \bar{C}S_{VV}}{S_{RR} + |C|^2 S_{VV}}. \tag{S24}$$

This expression reduces to $\tilde{G} = G$, if $S_{VV} = 0$ (r dominates over v), and $\tilde{G} = -1/C$, if $S_{RR} = 0$ (v dominates over r). For mixed SNR, the estimate becomes a mixture of the feedforward and feedback characteristics.

The bias can be eliminated if the external reference signal r is also available [17]. In that case the indirect method can be used [85], [86]

$$\tilde{G} = \frac{G_{yr}}{G_{ur}} = \frac{S_{YR}}{S_{UR}}. \tag{S25}$$

Because r is known exactly and not correlated with v, the bias in (S24) is removed. This approach can also be interpreted within the instrumental variables identification framework [21]–[23].

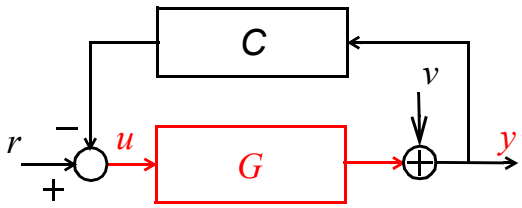

Figure S8. FRF measurement under closed-loop conditions require special care. The FRF of the system G that is captured in a feedback loop is measured starting from the measured input $u(t)$ and output $y(t)$. This leads to a bias because the input u is correlated with the noise v through the feedback path C.

Sidebar: This is the Sidebar Title

Design Of Excitation Signals

Figure S9 highlights some important properties of excitation signals that are discussed in more detail below.

Random excitations: Filtered random noise excitations (a), (d) have a random amplitude spectrum for a given realization. At the dips of the amplitude spectrum, the FRF estimate is most sensitive to disturbing noise (See Section "FRF Measurements Using Random Excitations" and Figure 7.

Periodic excitations: Periodicity is another important aspect. Periodic signals give access to a general and detailed nonparametric noise analysis (see Section FRF measurements using periodic excitations) without any user interaction. The signal (c), (f) has a deterministic amplitude spectrum that can be set by the user. Such signals are not prone to dips in the amplitude spectrum that vary from one realization to the other. This results in a better guaranteed SNR, even for a single realization. Leakage errors need to be avoided/removed by processing an integer number of periods (see Section "FRF Measurements Using Periodic Excitations") or by using the advanced algorithms of Section "Improved FRF Measurements Using Local Parametric Methods".

Deterministic amplitude spectrum: All the signals that are shown in Figure S9 have a random nature, even if the amplitude spectrum can be deterministic. This makes these signals well suited to be used as an excitation signal for FRF measurements in the presence of nonlinear distortions. The random nature is smoothing the impact of the nonlinear distortions so that the FRF still represents the averaged linearized behavior of the system [30].

Special designed excitation signals: Below a number of special signals are discussed: filtered noise, MLBS/PRBS, swept sine, multisines. An extensive discussion of excitation signals can be found in [16], [18], [25], [29].

MLBS

The maximum length binary sequence belongs to the class of pseudo random binary sequences (PRBS). They are deterministic, periodic sequences of length N that switch between one level (e.g., +1) and another level (e.g., -1). The switches can occur only on an equidistant grid at multiples of the clock period T_c, and they are chosen such that the autocorrelation is as spiky as possible so that it mimics a Dirac impulse (see Figure S12) [25], [26]. It is not possible to find a binary sequence with these properties for every arbitrary length N, for example MLBS sequences exist only for length of $2^n - 1$. In [27], an overview is given of binary and near-binary sequences that give the user a large choice of sequence lengths beyond that of the MLBS. Publicly available software to generate these signals is discussed in [28].

The continuous-time sequence is obtained using a ZOH-reconstruction of the discrete-time sequence as shown in Figure S11(a) for a MLBS.

The amplitude spectrum of the discrete sequence is a constant (except for the DC value). The spectrum of the continuous-time sequence rolls off with the $ZOH(f) = sin\pi f/\pi f$

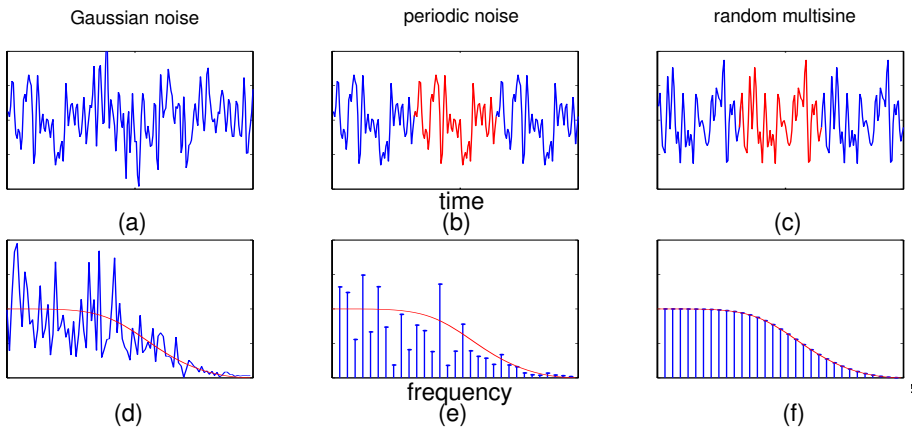

Figure S9. The characteristics of excitation signals are shown in the time and the frequency domain for (a), (d) Gaussian noise; (b), (e) periodically repeated Gaussian noise; and (c), (f) random -phase multisine. In the frequency domain, the amplitude spectrum of the actual realization (blue) and the power spectrum (red) is shown.

characteristic (see Figure S11(d)). This restricts the useful frequency band, because the $ZOH(f)$ has zeros at the multiples of the clock frequency $f_c = 1/T_c$. Moreover, the signal level drops with the frequency which results in a decreasing SNR for higher frequencies.

In the 1960's, dedicated generators were used to generate these signals (See Figure S10 and Figure S14).

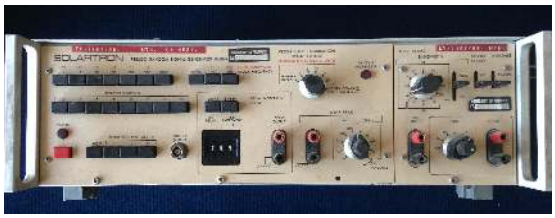

Figure S10. Pseudo random binary signal generator from the 1960's, used for example in laboratories at the School of Engineering, University of Warwick. Dedicated generators to generate MLBS and related signals, like the inverse repeat binary sequence, were built and commercialized. Besides the original signal $u(t)$, also a copy with a user adjustable delay $u(t-\tau)$ was generated. This allowed the correlation $y(t)u(t-\tau)$ to be measured using analog correlators, making a direct measurement of the impulse response $g(\tau)$ possible.

User guidelines

- *Choice of the clock frequency*: Select $f_c = 2.5 f_{max}$ to get a sufficiently flat amplitude spectrum in the frequency band of interest [16].
- *Signal length*: Select the length N of the sequence such that $f_0 = 1/N$ meets the frequency resolution requirement.

- *FFT analysis*: Do not modify the signal by padding the signal with zeros to obtain a period length that is a power of two. Instead, generalized FFT algorithms should be used that can handle arbitrary signal lengths. These are nowadays widely available in commonly used routines. Zero padding will destroy the good spectral properties of the MLBS.
- *Eliminate even nonlinearities*: The major drawback of MLBS or more general PRBS is the sensitivity to nonlinear distortions. These create large spikes in the impulse response estimates [26], [102]. A first possibility to remove these spikes is to calculate the median over multiple randomized realizations. Alternatively, an inverse repeated sequence $[u, -u]$ can be generated. By construction all even nonlinear distortions are eliminated at a cost of a reduction of the frequency resolution with a factor two.
- *Ternary signals*: An increased robustness against nonlinear distortions is obtained using well-designed ternary signals [103]. These signals excite only a set of selected odd frequencies. The remaining unexcited odd frequencies, and the even frequencies can be used to measure the level of the nonlinear distortions [30].

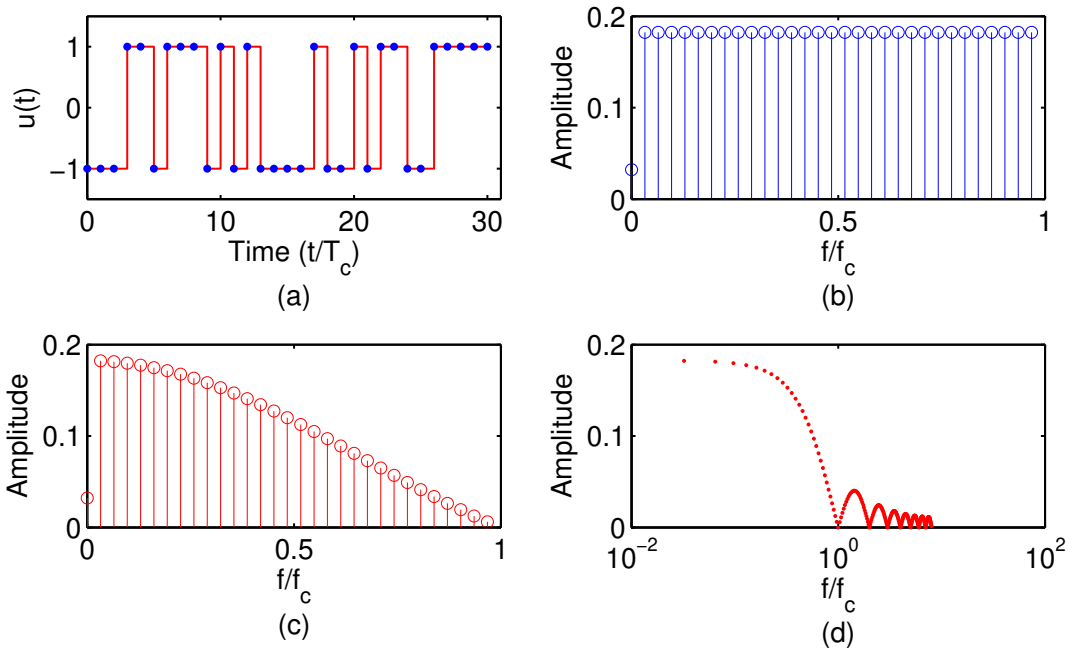

Figure S11. An example of a MLBS with a length $N = 2^5 - 1 = 31$. (a) Discrete-time (blue dots) and continuous-time (red line) MLBS with a clock frequency $f_c = 1/T_c = 1$. The continuous-time sequence is obtained by a ZOH reconstruction. (b) Spectrum of the discrete-time sequence, with $f_s = f_c$. The spectrum of the discrete-time sequence is perfectly flat (except for the DC value). This corresponds to the observation that the discrete correlation function consists within a DC offset of a perfect Dirac as shown in Figure S12. Observe that the spectral resolution is f_c/N which increases with the length of the sequence. The amplitude will drop as $1\sqrt{N}$ because the power is equally spread over N frequencies. Figure (c) and (d) show the amplitude spectrum of the continuous-time MLBS. The spectrum is proportional with $sin(\pi x)/\pi x$ with $x = f/f_c$. The amplitude spectrum has a zero at the multiples of f_c.

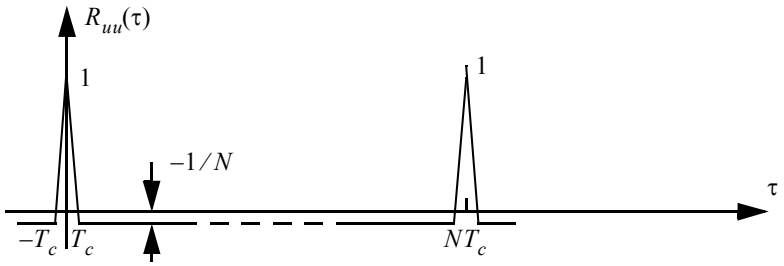

Figure S12. Autocorrelation of a continuous-time MLBS with levels ± 1 and length $N = 2^n - 1$, $n \in \mathbb{N}$. The autocorrelation is periodic with period length N. Observe also the offset of $1/N$. The width of the impulse is set by the clock period T_c. For the discrete-time sequence, the autocorrelation exists only for $\tau = kT_c$. In that case, it becomes a perfect Dirac function plus an offset of $-1/N$.

Some general purpose excitations

In Figure S13 some other general purpose excitation signals are shown, besides the MLBS discussed before. The full details about these signals can be found in [16], [24], [29]. The signals were designed to cover the frequency band [1, 42] Hz in a measurement window of 1 second.

Swept sine or chirp excitation: This is a sine excitation with an instantaneous frequency that is periodically linearly varying between f_1 and f_2

$$u(t) = A \sin((at + b)t), 0 \leq t < T_0, \tag{S26}$$

with $T_0 = 1/f_0$ the period, $a = \pi(k_2 - k_1)f_0^2$, $b = 2\pi k_1 f_0$, $f_1 = k_1 f_0$, $f_2 = k_2 f_0$. Observe in S13(b) that some of the power is put outside the frequency band of interest. Inside the frequency band, a ripple of a few dB is present.

Multisine: A multisine is a sum of harmonically related sines

$$u(t) = \sum_{k=1}^{F} A_k \cos(2\pi k f_0 t + \phi_k). \tag{S27}$$

The user can freely choose the amplitudes A_k and the frequency resolution f_0. The choice of the phases will set the nature of the signal. Random phases between $[0, 2\pi[$ will create a Gaussian like behavior. The phases can also be optimized to minimize the peak value of the signal [45]. In Figure S13 (c,d), the amplitudes were put equal to a constant, and the phases were $\phi_k = -k(k-1)\pi/F$. This is a Schroeder multisine [46] that mimics a swept sine, with a perfectly flat amplitude spectrum in the frequency band of interest, and no power outside this band. This comes at a cost of an increased peak value in the frequency domain. The latter can be reduced by phase optimization algorithms [45]. Multisines are the most flexible class of periodic excitation signals [16], [24].

Filtered white noise: As discussed in the Section Smoothing the FRF using the classical spectral estimation methods, the spectrum of filtered white noise in Figure S13(h) is also a random variable. At some frequencies almost no power will be present, resulting in a low SNR.

For that reason, it is strongly advised to average the FRF-measurements over a number of realizations (subrecords) to get good results. It can also be observed that the time-domain signal in Figure S13(g) has large peak values. This is typical for Gaussian noise excitations signals.

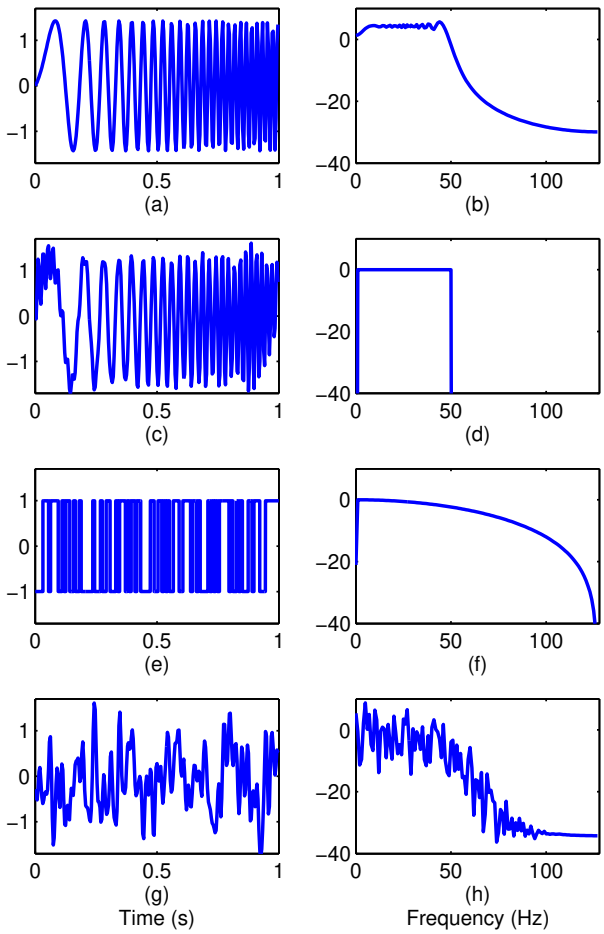

Figure S13. Comparison of general purpose excitation signals in the time and frequency domain. The aim is to excite a frequency band between 1 and 50 Hz, using signals with a length of 256 samples, using a generator clock frequency of 256 Hz. (a) swept sine, (c) Schroeder multisine, (e) MLBS, (g) filtered white noise. For the MLBS, a clock frequency of 127 Hz was used, in order to follow the design rule of the previous section.

Sidebar: This is the Sidebar Title

Historical Note On FRF And IR Measurements

Before the 1950's and 1960's it was not possible to estimate the FRF or IR from experimental data obtained during the normal operation of a system. Specific experiments were made to measure directly the impulse response or the FRF of the system. Using an impulse excitation the impulse response was obtained, while sine excitations were used to measure the FRF at a given frequency [37]. The main reason to stick to these simple methods was the lack of analog to digital converters so that all calculation needed to be done with analog circuits or analog computers. The latter were very tedious to use due to saturation and drift problems. In a more advanced setting for mechanical vibration measurements, a swept sine excitation was used, in combination with a tracking filter that was tuned at the instantaneous sine frequency. The filter reduces (nonlinear) disturbances and noise [18]. In the late 1950's and the start of the 1960's, analog correlators were built to measure the impulse response using wide band random excitations [99], generalizing the possibilities of sine correlators [100]. Later on, these results were significantly improved by using PRBS generators (see Figure S10). These were simple to build, including the generation of arbitrarily delayed copies, using the electronic components that were available at that time. In combination with the analog correlators, it became possible to obtain strongly improved IR estimates in a shorter time.

In a further step, a shift towards digital signal processing was made, the analog correlators were replaced by digital signal processing algorithms. A lot of time and effort was spent at that time to educate the engineering community to understand the required signal theory, dealing mainly with correlation methods. Although the Fourier transform was known among radio engineers [39], it took till the 1960's before it became a practical engineering tool. The FFT was first published in 1965 [60] (see also [41], and [42]), and at the same time digital computers became widely available. This was the start of a new era in FRF measurements, using spectral estimation methods that were efficiently implemented to minimize the size of the required core memory of the computer [1]. The cost to calculate a 2^{10} point Fourier transform dropped from a few dollars to hundredths of a cent, while the calculation time was reduced at the same time from several minutes to tens of milliseconds [42]. It was the start of digital signal processing. These methods became the dominating approach in most fields, and are still the standard tools that are implemented in today's commercial equipment [3], [4], [18]. Only recently, a renewed interest in developing more advanced FRF and IR became visible. These new activities make maximum use of the greatly increased computing power which allowed to shift towards methods that no longer suffer from leakage, and that are less sensitive to disturbing noise.

To give a better idea about the problems that were faced at that time, the measurement of the FRF using a sine excitation, and the measurement of the IR using PRBS signals and analog correlators are discussed in more detail.

Measuring the FRF using sine excitations

This section is based on the methods reported in [37], [40]. Detailed information on the historical methods can be found in [40]. Under steady-state conditions, the output of a linear dynamic system excited by $u(t) = A\sin(2\pi ft + \phi))$ is

$$y(t) = A|G(f)|\sin(2\pi ft + \phi + \angle G(f)). \tag{S28}$$

Graphical method: The input and output is recorded on a plotter. Using visual inspection, the amplitude $|G(f)|$ is obtained as the ratio of the measured amplitudes, while the phase is estimated from the phase difference of both signals. These methods can be refined using Lissajous-plots [37]. For poor SNR, the results become very noise sensitive. For that reason, correlation methods were developed that allow the signals to be averaged over time to improve the SNR.

Correlation method: Two analog correlation methods were used to measure the FRF at a given frequency. The first correlation technique, called *quadrature components method* [37], makes a double correlation with respectively the $\sin$ and $\cos$. This implementation typically requires a correlation time equal to 5 times the dominant time constant of the system [40]. The second correlation method, called the *phase null method*, aligns the correlation signal first with the output signal (phase null) by manually tuning a potentiometer meter, and next a two-step correlation is started. This method requires more time, but results in more precise measurements.

Today, sine based network analyzers are still popular in the microwave field, partly because it is still very expensive to generate arbitrary signals at very high frequencies (today, this is above 20 to 30 GHz), and partly because the SNR is much higher because narrow band filters can be applied.

Measuring the IR using PRBS signals and dedicated generation/correlation hardware

Measuring the FRF using sine excitations is very time consuming, especially at low frequencies, and for systems with long time constants. Only one frequency is probed at a time, and after each frequency change the system should reach steady-state conditions before the next measurement can be made. Using white noise excitations, all dynamics of the system are excited at once, but correlation methods are needed to retrieve the IR from these measurements. In practice, the random white noise excitation is replaced by a periodic PRBS signal (see "Design Of Excitation Signals") that has a deterministic and almost flat amplitude spectrum. Special hardware was developed [13], and combinations of PRBS generators and analog correlators, as shown in Figure S14, were commercialized [38], [99]. A few years later, when FFT's could be implemented on microprocessors, dedicated displays were built that could show the FRF on their screen in amplitude and phase [43]. This became the first generation of dynamic signal analyzers, that were replaced from the mid 1970's by single box instruments, that were based on spectral analysis methods (see Section FRF measurements using random excitations) [58], [59].

References

[1] P.D. Welch. A direct digital method of power spectrum estimation. *IBM Journal of Research and Development*, 5(2):141–156, 1961.

[2] P.D. Welch. The use of fast Fourier transform for estimation of power spectra - a method based on time averaging over short modified periodograms. *IEEE Transactions on Audiio and Electroacoustics*, (2):70–73, 1967.

(a)

(b)

Figure S14. Combination of a pseudo random binary signal generator (a) and correlator (b) from the 1960's (photos provided by Kenneth Kuhn, manager of the HP Memory Project, http://www.hpmemoryproject.org). Besides the original signal $u(t)$, also a copy with a user adjustable delay $u(t-\tau)$ was generated. This allowed the correlation $y(t)u(t-\tau)$ to be measured using the analog correlator in Figure (b), giving a direct measurement of the impulse response. Later on, in the early 1970's, the FFT of the impulse response was calculated and shown on a 'Spectrum Display'. The combination of the generator, correlator, and the spectral display resulted in the first dedicated FRF measurement setups [38], [43], [99].

[3] J.S. Bendat and A.G. Piersol. *Random Data: Analysis and Measurement Procedures*. Wiley, New York, 1971.

[4] J.S. Bendat and A.G. Piersol. *Engineering Applications of Correlations and Spectral Analysis*. Wiley, New York, 1980.

[5] D.E. Rivera, H. Lee, H.D. Mittelmann, and M.W. Braun. Constrained multisine input signals for plant-friendly identification of chemical process systems. *Journal of Process Control*, 19(4):623–635, 2009.

[6] T. Dobrowiecki, J. Schoukens, and P. Guillaume. Optimized excitation signals for MIMO frequency response function measurements. *IEEE Transactions on Instrumentation and Measurement*, 55(6):2072–2079, 2006.

[7] P.A.N. Briggs and K.R. Godfrey. Pseudorandom signals for dynamic analysis of multivariable systems. *Proceedings of the IEE*, 113(7):1259–1267, 1966.

[8] P. Guillaume, R. Pintelon, and J. Schoukens. *Accurate Estimation of Multivariable Frequency Response Functions*. Proceedings of the IFAC96 World Conference, San Fransisco (USA), July 1-5, 1996, Vol. I.,423-428.

[9] A.H. Tan, H.A. Barker, and K.R. Godfrey. Identification of multi-input systems using

simultaneous perturbation by pseudorandom input signals. *IET Control Theory and Applications*, 9(15):2283–2292, 2015.

[10] R. Voorhoeve, A van der Maas, and T. Oomen. Non-parametric identi?cation of multivariable systems: a local rational modeling approach with application to a vibration isolation benchmark. *Mechanical Systems and Signal Processing*, Accepted for publication, 2018.

[11] P.E. Wellstead. Non-parametric methods of system identification. *Automatica*, 17(1):55–69, 1981.

[12] K.R. Godfrey. Correlation methods. *Automatica*, 16(5):527–534, 1980.

[13] A.D.G. Hazlerigg and A.R.M. Noton. Application of crosscorrelating equipment to linear-system identification. *Proceedings IEE*, 112:2385–2400, 1965.

[14] K.R. Godfrey. Theory and application of pseudo-random sequences. *Control*, (6):305 – 308, 1966.

[15] G. Pillonetto, F. Dinuzzo, T. Chen, G. De Nicolao, and L. Ljung. Kernel methods in system identification, machine learning and function estimation: A survey. *Automatica*, 50(3):657–682, 2014.

[16] R. Pintelon and J. Schoukens. *System Identification: A Frequency Domain Approach*. Wiley-IEEE Press, Hoboken, New Jersey, 2nd edition, 2012.

[17] P.E. Wellstead. Reference signals for closed-loop identification. *International Journal of Control*, 26(6):945–962, 1977.

[18] D.J. Ewins. *Modal testing: theory and practice*. Wiley, New York, Princeton, N.J., 1991.

[19] J. Schoukens, G. Vandersteen, K. Barbe, and R. Pintelon. Nonparametric Preprocessing in system identification: a powerful tool. *European Journal of Control*, 15:260–274, 2009.

[20] R. Pintelon, J. Schoukens, G. Vandersteen, and K. Barbe. Estimation of nonparametric noise and FRF models for multivariable systems-Part I: Theory. *Mechanical Systems and Signal Processing*, 24(3):573–595, 2010.

[21] L. Ljung. *System Identification: Theory for the User (second edition)*. Prentice Hall, Upper Saddle River, New Jersey, 1999.

[22] T. Söderström and P. Stoica. *System Identification*. Prentice Hall International (UK) Ltd, Hemel Hempstead, Hertfordshire, 1989.

[23] P.C. Young. *Recursive Estimation and Time-Series Analysis. An Introduction for the Sudent and the Practitioner.* Springer, 2011.

[24] K. Godfrey. Design and application of multifrequency signals. *Computing and Control Engineering Journal*, pages 187–195, 1991.

[25] K.R. Godfrey, editor. *Perturbation signals for system identification*. Prentice Hall, Upper Saddle River, New Jersey, 1993.

[26] K.R. Godfrey and M. Devenish. An experimental investigation of continuous gas chromatography using pseudo-random-binary sequence. *Measurement and Control*, 2(6):228–232, 1969.

[27] A.H. Tan and K.R. Godfrey. The generation of binary and near-binary pseudorandom signals: An overview. *IEEE Transactions on Instrumentation and Measurement*, 51(4):583–588, 2002.

[28] K.R. Godfrey, A.H. Tan, H.A. Barker, and B. Chong. A survey of readily accessible perturbation signals for system identification in the frequency domain. *Control Engineering Practice*, 13(11):1391–1402, 2005.

[29] J. Schoukens, R. Pintelon, E. Vanderouderaa, and J. Renneboog. Survey of excitation signals for FFT based signal analyzers. *IEEE Transactions on Instrumentation and Measurement*, 37(3):342–352, 1988.

[30] J. Schoukens, M. Vaes, and R. Pintelon. Linear system identification in a nonlinear setting: Nonparametric analysis of the nonlinear distortions and their impact on the best linear approximation. *IEEE Control Systems Magazine*, 36:38–69, 2016.

[31] P. M. J. Van den Hof, A. Dankers, P. S. C. Heuberger, and X. Bombois. Identification of dynamic models in complex networks with prediction error methods-Basic methods for consistent module estimates. *Automatica*, 49(10):2994–3006, 2013.

[32] M. Gevers and A. S. Bazanella. Identification in dynamic networks: identifiability and experiment design issues. In *54th IEEE Conference on Decision and Control (CDC), Dec. 15-18, 2015. Osaka, Japan*, pages 4005–4010.

[33] J. Schoukens, R. Pintelon, and H. Van hamme. Identification of linear dynamic systems using piecewise-constant excitations - use, misuse and alternatives. *Automatica*, 30(7):1153–1169, 1994.

[34] E.O. Brigham. *The Fast Fourier Transform*. Prentice-Hall Inc, Englewood Cliffs, New Jersey, 1974.

[35] K.J. Åström. *Introduction to Stochastic Control Theory*. Academic Press, 1970.

[36] J. Schoukens, Y. Rolain, G. Simon, and R. Pintelon. Fully automated spectral analysis of periodic signals. *IEEE Transactions on Instrumentation and Measurement*, 52(4):1021–1024, 2003.

[37] W. Seifert. Kommerzielle frequenzgangmesseinrichtungen. *Regelungstechnik*, 10:350–353, 1962.

[38] G.C. Anderson, B.W. Finnie, and G.T. Roberts. Pseudo-Random and Random Test Signals. *Hewlett-Packard Journal*, 19:2–14, 1967.

[39] M.M. Levy. Fourier transform analysis. *Journal of the British Institute of Radio Engineers*, pages 228–246, 1946.

[40] P.E.A. Cowley. The application of an analog computer to the measurement of process dynamics. *Transactions of the ASME*, 79(4):823–832, 1957.

[41] J.W. Cooley, P.A.W. Lewis, and P.D. Welch. Historical notes on the Fast Fourier Transform. *Proceedings of the IEEE*, 55(2):1675–1677, 1967.

[42] G.D. Bergland. A guided tour of the fast Fourier transform. *IEEE Spectrum*, pages 41–52, 1969.

[43] D.M. Morrison, B.W. Finnie, R.S. Patel, and K.H. Edwards. Versatile display unit extends correlator capability. *Hewlett-Packard Journal*, 24:8–15, 1972.

[44] D.E. Torfs, R. Vuerinckx, J. Swevers, and J. Schoukens. Comparison of two feedforward design methods aiming at accurate trajectory tracking of the end point of a flexible robot arm. *IEEE Transactions on Automatic Control Systems Technology*, 6(1):2–14, 1998.

[45] P. Guillaume, J. Schoukens, R. Pintelon, and I. Kollar. Crest-factor minimization using nonlinear Chebyshev-approximation methods. *IEEE Transactions on Instrumentation and Measurement*, 40(6):982–989, 1991.

[46] M.R. Schroeder. Synthesis of low-peak-factor signals and binary sequences with low autocorrelation. *IEEE Transactions on Information Theory*, 16(1):85–89, 1970.

[47] J. Schoukens, R. Pintelon, G. Vandersteen, and P. Guillaume. Frequency-domain system identification using non-parametric noise models estimated from a small number of data sets. *Automatica*, 33(6):1073–1086, 1997.

[48] K. Barbe, J. Schoukens, and R. Pintelon. Frequency-domain, errors-in-variables estimation of linear dynamic systems using data from overlapping subrecords. *IEEE Transactions on Instrumentation and Measurement*, 57(8):1529–1536, 2008.

[49] P. Guillaume, R. Pintelon, and J. Schoukens. Nonparametric frequency-response function estimators based on nonlinear averaging techniques. *IEEE Transactions on Instrumentation and Measurement*, 41(6):739–746, 1992.

[50] R. Pintelon and J. Schoukens. *System Identification. A Frequency Domain Approach*. IEEE Press, Piscataway, 2001.

[51] P Guillaume. Frequency response measurements of multivariable systems using nonlinear averaging techniques. *IEEE Transactions on Instrumentation and Measurement*, 47(3):796–800, 1998.

[52] R. Pintelon, Y. Rolain, and W. Van Moer. Probability density function for frequency response function measurements using periodic signals. *IEEE Transactions on Instrumentation and Measurement*, 52(1):61–68, 2003.

[53] P.M.T. Broersen. A comparison of transfer-function estimators. *IEEE Transactions on Instrumentation and Measurement*, 44(3):657–661, 1995.

[54] W.P. Heath. The variance of nonparametric errors-in-variables estimates. *IEEE Transaction on Instrumentation and Measurement*, 54(1):228–236, 2005.

[55] A. Papoulis. *Probability, Random Variables, and Stochastic Processes*. McGraw-Hill, 1991.

[56] R.B. Blackman and J.W. Tukey. The measurement of power spectra from the point of view of communication engineering. *Bell Syst. Tech. J.*, 37:183–282, 1958.

[57] A. Stenman, F. Gustafsson, D.E. Rivera, L. Ljung, and T. McKelvey. On adaptive smoothing of empirical transfer function estimates. *Control Engineering Practice*, 8(11):1309–1315, 2000.

[58] P.E. Wellstead. Using digital spectral techniques. *Solartron Instruments*, Report Number 008/83, 1984.

[59] P.E. Wellstead. Theory and statistical accuracy of spectral analysis. *Solartron Instruments*, Report Number 009/83, 1984.

[60] J.W. Cooley and J.W. Tukey. An algorithm for machine calculation of complex Fourier series. *Mathematics of Computation*, 19:297–301, 1965.

[61] F.J. Harris. Use of windows for harmonic-analysis with discrete Fourier-transform. *Proceedings of the IEEE*, 66(1):51–83, 1978.

[62] J. Schoukens, Y. Rolain, and R. Pintelon. Analysis of windowing/leakage effects in frequency response function measurements. *Automatica*, 42(1):27–38, 2006.

[63] J. Antoni and J. Schoukens. A comprehensive study of the bias and variance of frequency-response-function measurements: Optimal window selection and overlapping strategies. *Automatica*, 43(10):1723–1736, 2007.

[64] J. Schoukens, G. Vandersteen, R. Pintelon, E. Zlatko, and Y. Rolain. Bounding the polynomial approximation errors of frequency response functions. *IEEE Transactions on Instrumentation and Measurement*, 62(5, SI):1346–1353, 2013.

[65] P. Verboven, E. Parloo, B. Cauberghe, and P. Guillaume. Improved modal parameter estimation for lowly damped systems using non-parametric exponential windowing techniques. *Mechanical Systems and Signal Processing*, 19(4):675–699, 2005.

[66] J.L. Douce. A note on frequency-response measurement. *IEE Proceedings-Control Theory and Applications*, 133(4):189–190, 1986.

[67] W.D. Widanage, J. L. Douce, and K. R. Godfrey. Effects of overlapping and windowing on frequency response function estimates of systems with random inputs. *IEEE Transactions on Instrumentation and Measurement*, 58(1):214–220, 2009.

[68] P. Cawley. The reduction of bias error in transfer-function estimates using FFT-based analyzers. *Journal of Vibration Acoustics Stress and Reliability in Design - Transactions of the ASME*, 106(1):29–35, 1984.

[69] J.L. Douce and L. Balmer. Statistics of frequency-response estimates. *IEE Proceedings-D Control Theory and Applications*, 137(5):290–296, 1990.

[70] R. Pintelon, J. Schoukens, and G. Vandersteen. Frequency domain system identification using arbitrary signals. *IEEE Transaction on Automatic Control*, 42(12):1717–1720, 1997.

[71] T. McKelvey. Frequency domain identification methods. *Circuits Systems Signal Processing*, 21:39–55, 2002.

[72] J.L. Douce. Improving frequency-response estimates by reduction of end effects. *IEE Proceedings-Control Theory and Applications*, 153(2):247–250, 2006.

[73] J. Schoukens, R. Pintelon, and Y. Rolain. *Mastering System Identification in 100 Exercises*. John Wiley & Sons, Hoboken, New Jersey, 2012.

[74] J. Schoukens, G. Vandersteen, Y. Rolain, and R. Pintelon. Frequency response function measurements using concatenated subrecords with arbitrary length. *IEEE Transaction on Instrumentation and Measurement*, 61(10):2682–2688, 2012.

[75] E. Geerardyn. *Development of User-Friendly System Identification Techniques*. PhD Thesis, Vrije Universiteit Brussel, Brussels, Belgium, 2016.

[76] P. Hagg, J. Schoukens, M. Gevers, and H. Hjalmarsson. The transient impulse response modeling method for non-parametric system identification. *Automatica*, 68:314–328, 2016.

[77] R. Pintelon, J. Schoukens, G. Vandersteen, and K. Barbe. Estimation of nonparametric noise and FRF models for multivariable systems-Part II: Extensions, applications. *Mechanical Systems and Signal Processing*, 24(3):596–616, 2010.

[78] E.C. Levy. Complex-Curve Fitting. *IEEE Transactions on Automatic Control*, 4:37–43, 1959.

[79] T. McKelvey and G. Guerin. Non-parametric frequency response estimation using a local rational model. In *16th IFAC Symposium on System Identification, Brussels, Belgium, July 11-13*, pages 49–54, 2012.

[80] J. Lataire and T. Chen. Transfer function and transient estimation by Gaussian process regression in the frequency domain. *Automatica*, pages 217–229, 2016.

[81] P. Thummala and J. Schoukens. Estimation of the FRF through the improved local bandwidth selection in the local polynomial method. *IEEE Transactions on Instrumentation and Measurement*, 61(10):2833–2843, 2012.

[82] D. Ugryumova, R. Pintelon, and G. Vandersteen. Frequency response function estimation in the presence of missing output data. *IEEE Transactions on Instrumentation and Measurement*, 64(2):541–553, 2015.

[83] D. Ugryumova, R. Pintelon, and G. Vandersteen. Frequency response matrix estimation from missing input-output data. *IEEE Transactions on Instrumentation and Measurement*, 64(11):3124–3136, 2015.

[84] B. Sanchez, J. Schoukens, R. Bragos, and G. Vandersteen. Novel estimation of the electrical bioimpedance using the local polynomial method. Application to in vivo real-time myocardium tissue Impedance characterization during the cardiac cycle. *IEEE Transactions on Biomedical Engineering*, 58(12):3376–3385, 2011.

[85] P.M.J. Van den Hof and R.J.P. Schrama. An indirect method for transfer-function estimation from closed-loop data. *Automatica*, 29(6):1523–1527, 1993.

[86] U. Forssell and L. Ljung. Closed-loop identification revisited. *Automatica*, 35(7):1215–1241, 1999.

[87] B. Ninness and H. Hjalmarsson. On the frequency domain accuracy of closed-loop estimates. *Automatica*, 41(7):1109–1122, 2005.

[88] B. Wahlberg. System-identification using Laguerre models. *IEEE Transaction on Automatic Control*, 36(5):551–562, 1991.

[89] P.S.C. Heuberger, P.M.J. Van den Hof, and B. Wahlberg, editors. *Modelling and Identification with Rational Orthogonal Basis Functions*. Springer, London, 2005.

[90] W.G. Halvorsen and D.L. Brown. Impulse technique for structural frequency-response testing. *Sound and Vibration*, 11(11):8–21, 1977.

[91] V. Wilkens and C. Koch. Amplitude and phase calibration of hydrophones up to 70 MHz using broadband pulse excitation and an optical reference hydrophone. *Journal of the Acoustical Society of America*, 115(6):2892–2903, 2004.

[92] A. Link, A. Taeubner, W. Wabinski, T. Bruns, and C. Elster. Calibration of accelerometers: determination of amplitude and phase response upon shock excitation. *Measurement Science and Technology*, 17(7):1888–1894, 2006.

[93] H. Fueser, S. Eichstaedt, K. Baaske, C. Elster, K. Kuhlmann, R. Judaschke, K. Pierz, and M. Bieler. Optoelectronic time-domain characterization of a 100 GHz sampling oscilloscope. *Measurement Science and Technology*, 23(2), 2012.

[94] M. Kobusch, T. Bruns, L. Klaus, and M. Mueller. The 250 kN primary shock force calibration device at PTB. *Measurement*, 46(5):1757–1761, 2013.

[95] K.R. Godfrey. *Compartmental Models and their Application*. Academic Press, 1983.

[96] J. DiStefano. *Dynamic Systems Biology Modeling and Simulation*. Academic Press, London, 2013.

[97] T.M. Souders, D.R. Flach, C. Hagwood, and G.L. Yang. The effects of timing jitter in sampling systems. *IEEE Transactions on Instrumentation and Measurement*, 39(1):80–85, 1990.

[98] H. Rake. Step response and frequency-response methods. *Automatica*, 16(5):519–526, 1980.

[99] G.C. Anderson and M.A. Perry. A Calibrated Real-Time Correlator/Averager/Probability Analyzer. *Hewlett-Packard Journal*, 19:9–20, 1969.

[100] B. Dwyer. A feedback transfer function analyser. *Control*, 9:676–678, December 1965.

[101] K.R. Godfrey. Introduction to binary signals used in system-identification. In *International Conference on Control 91, Edinburgh, Scotland, March 25-28, Vol. 1*, pages 161–166, 1991.

[102] H.K. Wong, J. Schoukens, and K.R. Godfrey. Structured non-linear noise behaviour and the use of median averaging in non-linear systems with m-sequence inputs. *IET Control Theory and Applications*, 7(7):997–1004, 2013.

[103] A. De Angelis, J. Schoukens, K.R. Godfrey, and P. Carbone. Practical synthesis of ternary sequences for system identification. *IEEE Transaction on Instrumentation and Measurement*, 66:212 – 222, 2017.

[104] K.R. Godfrey. The theory of the correlation method of dynamic analysis and its application to industrial processes and nuclear power plant. *Measurement and Control*, 2(5):T65–72, 1969.

[105] K.R. Godfrey. Dynamic analysis of an oil-refinery unit under normal operating conditions. *Proceedings of the Institution of Electrical Engineers-London*, 116(5):879–888, 1969.

[106] S.L. Marple. Efficient least-squares FIR system-identification. *IEEE Transactions on Acoustics Speech and Signal Processing*, 29(1):62–73, 1981.

[107] B. Ninness and H. Hjalmarsson. Variance error quantifications that are exact for finite-model order. *IEEE Transactions on Automatic Control*, 49:1275–1291, 2004.

[108] A.V. Oppenheim, A.S. Willsky, and S.H. Nawab. *Signals and Systems*. Prentice-Hall, London, 1997.

[109] A. Marconato, M. Schoukens, and J. Schoukens. Filter-based regularisation for impulse response modelling. *IET Control Theory & Applications*, October 2016.

Acknowledgement

The authors like to acknowledge Bart Peeters from LMS International, part of Siemens Product Lifecycle Management for the photo in Figure 16.

This work was supported in part by the Fund for Scientific Research (FWO-Vlaanderen), by the Flemish Government (Methusalem), the Vrije Universiteit Brussel (VUB), and by the ERC advanced grant SNLSID, under contract 320378.

Author Information

Johan Schoukens received both the Master's degree in electrical engineering in 1980, and the PhD degree in engineering sciences in 1985 from the Vrije Universiteit Brussel (VUB), Brussels, Belgium. In 1991 he received the degree of Geaggregeerde voor het Hoger Onderwijs from the VUB, and in 2014 the degree of Doctor of Science from The University of Warwick. From 1981 to 2000, he was a researcher of the Belgian National Fund for Scientific Research (FWO-Vlaanderen) at the Electrical Engineering Department of the VUB, From 2000 to 2018 he was a full-time professor in electrical engineering, since 2018 he is Emiritus Professor at the department INDI of the VUB. From 2009 to 2016, he was Visiting Professor at the Katholieke Universiteit Leuven (Belgium). His main research interests include system identification, signal processing, and measurement techniques. He has been a Fellow of IEEE since 1997. He was the recipient of the 2002 Andrew R. Chi Best Paper Award of the IEEE Transactions on Instrumentation and Measurement, the 2002 Society Distinguished Service Award from the IEEE Instrumentation and Measurement Society, and the 2007 Belgian Francqui Chair at the Université Libre de Bruxelles (Belgium). Since 2010, he is a member of Royal Flemish Academy of Belgium for Sciences and the Arts. In 2011 he received a Doctor Honoris Causa degree from the Budapest University of Technology and Economics (Hungary). Since 2013, he has been an honorary professor of the University of Warwick.

Keith Godfrey was appointed to the academic staff of the School of Engineering in 1973, and he is now an Emeritus Professor in the School. He received the Doctor of Science degree from the University of Warwick in 1990 for publications with the collective title "Applications of Modeling, Identification and Parameter Estimation in Engineering and Biomedicine". He is author of a book on compartmental modeling published by Academic Press in 1983, and is author, or co-author, of more than 200 papers. He is a member of the International Federation of Automatic Control (IFAC) Technical Committees on Biomedical Engineering and Control, and on Modeling, Identification and Signal Processing. He is currently working on the book "Industrial Process Identification: Perturbation Signal Design and Applications", co-authored by Ai Hui Tan (Multimedia University, Malaysia), to be published by Springer in their book series "Advances in Industrial Control".

Maarten Schoukens received the master's degree in electrical engineering: electronics and information technology and Ph.D. degree from the Vrije Universiteit Brussel (VUB), Brussels, Belgium, in 2010 and 2015, respectively. From 2015 to 2017, he has been a Post-Doctoral Researcher with the ELEC Department, VUB. In October 2017 he joined the Control Systems research group, TU/e, Eindhoven, The Netherlands, as a Post-Doctoral Researcher. His interests include the measurement and identification of linear parameter-varying and nonlinear systems.