FRF Smoothing to Improve Initial Estimates for Transfer Function Identification

Egon Geerardyn, *Student Member, IEEE*, Mikaya L. D. Lumori, *Senior Member, IEEE*, John Lataire, *Member, IEEE*

***Abstract*—Good initial values are crucial to obtain solutions of non-convex optimization problems. When estimating the transfer function of physical systems from measured noisy data, obtaining good initial parameter estimates is therefore a primordial step. In this paper, it is shown that smoothing the measured frequency response function (FRF) of a linear time-invariant system enhances the construction of initial estimates significantly, resulting in the optimization schemes to converge to a better optimum. This is achieved with minimal user interaction.**

Two smoothing techniques, the time-truncated local polynomial method (LPM) and the regularized finite impulse response (RFIR), are compared with the existing generalized total least squares (GTLS) and the bootstrapped total least squares (BTLS) initial estimates. The improvement attributable to smoothing is demonstrated by a simulation and by measurements of an electrical filter. The results ultimately show that the parametric models obtained using the proposed starting values are much more likely to give a good description of the measured system and hence lead to more useful models.

***Index Terms*—Global optimization, Smoothing methods, Estimation, Least squares methods, Frequency response, Measurements, Frequency domain analysis.**

I. Introduction

The task of interpreting measurement data from dynamic systems often involves the estimation of a transfer function (TF). Typical measurement techniques and identification strategies of transfer functions (TFs) or frequency response functions (FRFs) are discussed in [1]–[6]. Application of these methods to real devices and systems is well-known in published literature, including [7]–[9].

Parametric identification of linear time-invariant (LTI) systems from input/output data has been well developed as evidenced by published literature [10]–[13]. Maximizing the likelihood function of the parameters probably yields the most popular, consistent and efficient estimator. However, in a noisy environment, the maximum likelihood estimator (MLE) engenders a nonlinear optimization which, at the very least, requires good initial estimates to avoid local optima. Such local optima have a detrimental effect on the quality of the estimated model.

In the perspective of developing identification tools that require as little user interaction as possible, it is important to increase the probability of the optimization algorithm to produce good estimates, or preferably even the global optimum. However, when noisy input-output measurements are to be processed, the presence of noise obstructs a clear view of the actual system behavior, especially for poor signal-to-noise ratios (SNRs). In non-parametric identification, the presence of noise is typically mitigated by averaging multiple measurements or using smoothing techniques.

The main aim and contribution of this article is to demonstrate that smoothing a measured FRF helps to avoid local optima during the parametric estimation of the MLE of a transfer function, using deterministic optimization algorithms. Hence, such techniques make it possible to increase the "success rate" (i.e. the probability that a good model, or even the global optimum, is obtained) of such a system identification step considerably.

In particular, two different smoothing techniques — the time-truncated Local Polynomial Method (LPM) [14] and the regularized finite-impulse response (RFIR) [15], [16] — are tested for different SNRs and different measurement record lengths of the input/output data of a few single-input-single-output (SISO) LTI systems. These are compared with the existing initialization schemes, namely: (i) the generalized total least squares (GTLS), and (ii) the bootstrapped total least squares (BTLS).

The rest of the paper is structured as follows. Section II covers the problem formulation. Section III discusses the methodology for obtaining the initial estimates, and their influence on the success rate of the MLE. This is followed by Sections IV and V, which are demonstrations of the improved initial estimates by simulation and experimental data, respectively. In Section VI the generality of the results is discussed briefly. Ultimately, concluding remarks are presented in Section VII.

II. Problem formulation

In this section, the assumptions on the system and the noise are presented, together with the associated MLE (formulated in the frequency domain), which turns out to be a non-convex function in the parameters, requiring initial estimates. The procedure for obtaining the initial estimates is discussed in the subsequent sections.

A. System Setup Framework and Model

Fig. 1 depicts a schematic of the output error framework for a generalized (single- or multiple-order) resonating, dynamic

E. Geerardyn and J. Lataire are with the ELEC Department, Vrije Universiteit Brussel (VUB), Pleinlaan 2, 1050 Brussels, Belgium. E-mail: egon.geerardyn@vub.ac.be

M. Lumori is with the Department of Electrical Engineering, , Shiley-Marcos School of Engineering, University of San Diego, CA, USA.

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

LTI discrete-time SISO system, subjected to a known white random noise input. The following mathematical derivations pertain to discrete time systems, but they can be extended to continuous time systems as well. The full mathematical model of the system is

$$y(t) = G_0(q^{-1})u_0(t) + H_0(q^{-1})e(t) \tag{1}$$

where $G_0(q^{-1})$ represents the dynamics of the system to be estimated, $u_0(t)$ is the input signal, $v(t) = H_0(q^{-1})e(t)$ is the noise source at the output, $H_0(q^{-1})$ is the noise dynamics, $e(t)$ is white Gaussian noise, and q^{-m} is the backwards shift operator ($q^{-m}x(t) = x(t - mT_\mathrm{s})$ with m a positive integer and T_s the sampling time). For white noise, $H_0 = 1$ and $e(t) = v(t)$. Note that for simplicity, we only treat the discrete-time case (i.e. $t = n \cdot T_\mathrm{s}$ for integer values of n) theoretically in this paper. However, the generalization to continuous time is straightforward [10, Chapter 6] and has been demonstrated in Section V.

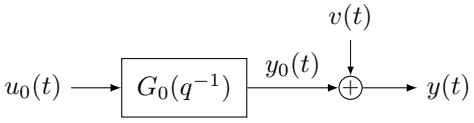

Fig. 1. SISO LTI discrete-time system in an output error setup.

Numerous parametric identification techniques are devoted to the development of parametric plants $G(q^{-1}, \theta)$ and parametric noise models $H(q^{-1}, \theta)$, where θ is the model parameters vector [11], [17].

The output signal is disturbed by white random noise, resulting in noisy error-prone data. It is also possible to apply the estimation procedure in this paper to a system that is disturbed by colored noise, as will be demonstrated in Section V.

B. Parametric System Model

With reference to equation (1), the relation between the noiseless input and the output signals ($v(t) = 0$) is assumed to be of the form

$$A(q^{-1})y_0(t) = B(q^{-1})u_0(t) \tag{2}$$

where A and B are polynomials in q^{-1}. Thus, it follows from equation (1) that

$$G_0(q^{-1}) = \frac{B(q^{-1})}{A(q^{-1})}. \tag{3}$$

From [10, Section 6.3.2.], in the frequency domain, and for the DFTs (Discrete Fourier Transforms) of the windowed signals, equation (2) is of the form

$$A(e^{-j\omega_k})Y_0(k) = B(e^{-j\omega_k})U_0(k) + I(e^{-j\omega_k}) \tag{4}$$

with q^{-1} sampled in $q_k^{-1} = e^{-j\omega_k}$, where $\omega_k = \frac{2\pi k}{N}$ are the DFT frequencies, and I is a polynomial of order $\max(N_a, N_b) - 1$, which depends on the initial and end conditions. In this paper, it is assumed that the orders of the polynomials A, B and I are known.

C. Parametric Identification Algorithm

The maximum likelihood estimate of the parameter vector θ containing the coefficients of the A, B and I polynomials is obtained by solving the optimization problem

$$\hat{\theta} = \arg\min_\theta V(\theta). \tag{5}$$

Since the noise $v(t)$ is assumed to be Gaussian, $V(\theta)$ accords to the weighted least squares cost function [10, Section 9.11]:

$$V(\theta) = \sum_{k=1}^{N/2-1} \frac{|\varepsilon(k,\theta)|^2}{\sigma_\varepsilon^2(k,\theta)} \tag{6}$$

for N measurements and when ε denotes the error in equation (4), *viz.*:

$$\varepsilon(k,\theta) = A(k,\theta)Y(k) - B(k,\theta)U(k) - I(k,\theta). \tag{7}$$

The variance of the error is of the form:

$$\begin{aligned}\sigma_\varepsilon^2(k,\theta) = {} & |A(k,\theta)|^2\sigma_Y^2(k) + |B(k,\theta)|^2\sigma_U^2(k) \\ & - 2\,\mathrm{real}\left(A(k,\theta)\overline{B(k,\theta)}\sigma_{YU}(k)\right)\end{aligned} \tag{8}$$

where σ_Y^2 is independent of k for white noise, and the input error variance is assumed to be zero, i.e. $\sigma_U^2 = 0$, and similarly, the covariance $\sigma_{YU} = 0$. Here, $\overline{B}$ denotes the complex conjugate of B.

Consequently, $V(\theta)$ is a non-quadratic function of θ which, in general, results in a non-convex optimization problem. The Levenberg-Marquardt [18] algorithm (see Algorithm 1) is used to solve this optimization problem deterministically. Such an approach requires good initial estimates of θ to avoid inherent local optima, which is the focus of this paper.

Algorithm 1 Levenberg-Marquardt [18], [10, Sec. 9.L.4]

Cost function (6) is rewritten as $V(\theta) = \epsilon^\mathrm{T}\epsilon$.
The iterations start at $\theta^{[0]} = \theta^{\mathrm{init}}$ and $\lambda = \max(\sigma)/100$.
for i **in** $1 \to \infty$ **do**
  $U\,\Sigma\,W^\mathrm{T} \leftarrow \mathtt{svd}\left(\frac{\partial \epsilon}{\partial \theta}\right)$ with $\mathrm{diag}(\Sigma) = [\sigma_1, \ldots, \sigma_n]$
  $\sigma_k \leftarrow \sigma_k/(\sigma_k^2 + \lambda^2) \quad \forall k \in \{1, \ldots, n\}$
  $\delta\theta \leftarrow -W\Sigma^\mathrm{T}U^\mathrm{T}\epsilon$
  $\theta^{[i]} \leftarrow (\theta^{[i-1]} + \delta\theta)/\|\theta^{[i-1]} + \delta\theta\|_2$
  if $V(\theta^{[i]}) > V(\theta^{[i-1]})$ **then**
    $\lambda \leftarrow 10\lambda$ and then retry at $i \leftarrow i - 1$
  else
    $\lambda \leftarrow 0.4\lambda$
  if $\|\delta\theta\|_2/\|\theta^{[i]}\|_2 < n_\theta \cdot 10^{-6}$ **then return** $\hat{\theta} \leftarrow \theta^{[i]}$

Remark 1. Alternatively, stochastic optimization algorithms [19], [20] are less likely to get stuck in local optima since many randomized initial estimates are tried. This obviously comes with a huge computational cost compared to 'classical' deterministic schemes where only a single or a few initial estimates are used to start the expensive iterative optimization procedure. Stochastic optimization is not discussed in the remainder of this paper.

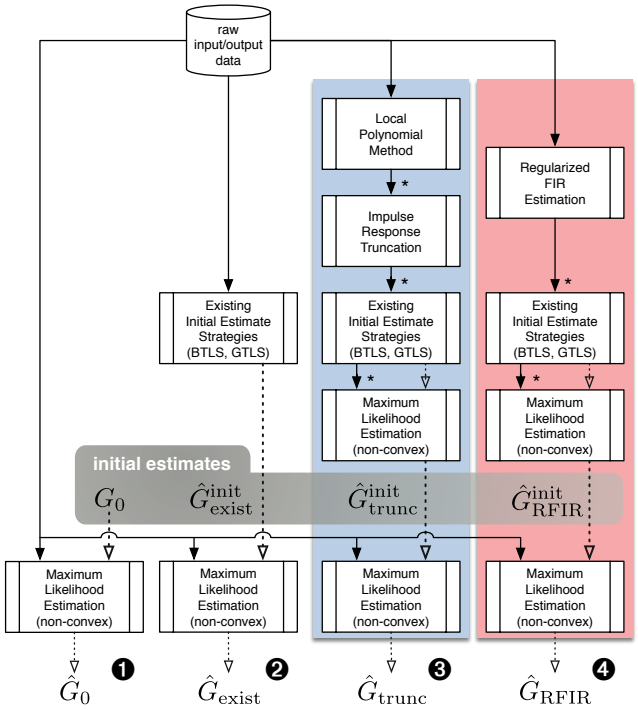

Fig. 2. Flow chart depicting the different estimation procedures, from left to right: (1) an approach using the true model as initial estimate; (2) an existing approach using BTLS and GTLS starting values, and the novel methods outlined in this paper (3, 4). The flow of non-parametric data is depicted by full arrows, with the FRF data marked by asterisks. Parametric models are indicated by dashed, unfilled arrows.

III. Methodology

In this section, the considered methods for obtaining the initial estimates of θ are briefly explained. Then their influence on the success rate of the maximum likelihood estimator is described.

A. Initial Estimates Procedure

The ultimate aim is to find initial estimates that are good enough to steer clear of the local optima during the MLE optimization process in the parametric identification of each system. The requisite procedures for the estimators are as follows, from left to right of the flowchart in Fig. 2:

1) *Using the true model (G_0) as initial estimate*
The estimates from this procedure are for comparison purposes with those from the other procedures. They will be crucial to the computation of the global optimum of the MLE.
2) *Quadratic approximations of the ML*
The Generalized Total Least Squares (GTLS) and the Bootstrapped Total Least Squares (BTLS) have been discussed in [13]. These are modifications of equation (6), which still take into account the noise information while retaining the quadratic nature w.r.t. θ. These estimators preserve consistency, such that the estimates converge asymptotically to the true parameters for $N \to \infty$. Their finite sample behavior is suitable for improvement by the FRF smoothing tools, described below.
3) *Local polynomial method (LPM) with truncation*
A reliably good estimate of the FRF of the chosen system can be obtained via the LPM, which is summarized as follows [14]. Formulated as a nonparametric linear-least-squares-estimate, the LPM is first applied to estimate the FRF from a full data record of a SISO system, systematically expressed in an output-error framework, shown in Fig. 1. The smooth characteristics of both the exact FRF G_0 and the transient term T allow for an optimal application of the LPM, leading to a smooth FRF estimate with the transient suppressed.
With reference to a detailed discussion in [14], the LPM utilizes the following quadratic local polynomials to approximate (in least squares sense) and, thus, smooth G_0 and T around a central frequency Ω_k in a local frequency band $r \in \mathbb{W}_n$

$$G_0(\Omega_{k+r}) \approx \hat{G}_k + g_{1,k}r + g_{2,k}r^2 \tag{9a}$$
$$T(\Omega_{k+r}) \approx T(\Omega_k) + t_{1,k}r + t_{2,k}r^2 \tag{9b}$$

where, in general, $r \in \mathbb{W}_n$, $\mathbb{W}_n = \{-n, -n+1, \ldots, n\}$ and n is a tunable parameter. In this paper, $n = 3$.
The LPM estimate of the FRF at frequency index k is the first estimated local parameter, *viz*:

$$\hat{G}_{\text{poly}}(\Omega_k) = \hat{G}_k. \tag{10}$$

This procedure is repeated for all k in the frequency band of interest.
Impulse response truncation. The estimate is smoothed further by truncating its impulse response function $g_{\text{poly}}(t) = \text{IDFT}\left(\hat{G}_{\text{poly}}(\Omega_k)\right)$ as originally presented in [14].
The impulse response is truncated after the time τ_{trunc} where the signal becomes indistinguishable from the noise, i.e.

$$g_{\text{trunc}}(t) = \begin{cases} g_{\text{poly}}(t) & t \le \tau_{\text{trunc}} \\ 0 & t \ge \tau_{\text{trunc}} \end{cases}. \tag{11}$$

To this end, an estimate of the envelope of impulse response is determined by fitting an exponential function $g_{\text{exp}}(t) = Be^{\beta t}$ to the peaks of the impulse response $|g_{\text{poly}}(t)|$ using a linear least-squares approach. Then, τ_{trunc} is determined as the time instant where this envelope function sinks below the noise level σ_g of the impulse response, i.e.

$$\tau_{\text{trunc}} = \min\{t : g_{\text{exp}}(t) < \gamma\sigma_g\} = \beta^{-1} \ln \frac{\gamma\sigma_g}{B}, \tag{12}$$

where $\gamma = 1$ was used in this paper for simplicity. By changing γ, the user can fine-tune the bias/variance trade-off of the estimated FRF $G_{\text{trunc}}(\Omega_k) = \text{DFT}\left(g_{\text{trunc}}(t)\right)$ further. This can lead to a significant improvement over the classical LPM [14].
4) *Regularized finite impulse response (RFIR)*
The regularized finite impulse response (RFIR) method is a special case of the regularized autoregressive exogenous (RARX) method.
The RFIR estimator is formulated in the time domain. It estimates the impulse response of a discrete time

system as the minimizer of the following regularized least squares objective function:

$$\hat{g}_{\mathrm{RFIR}} = \underset{g}{\operatorname{argmin}} \|y - g * u\|_2^2 + \sigma^2 g^{\mathrm{T}} P^{-1} g \qquad (13)$$

where g is the vectorized impulse response $g(t)$, with $t = 0, 1, \cdots, n_h - 1$, assuming that the impulse response is n_g samples long. Furthermore, $g * u$ is the convolution of g with the input signal u, and $P \in \mathbb{R}^{n_g \times n_g}$ is the kernel matrix. Note that (13) is quadratic in g, such that $\hat{g}_{\mathrm{RFIR}}$ can be computed analytically in a single step. The kernel matrix P embodies prior knowledge on the system to be estimated. In this paper, the Diagonal Correlated (DC) kernel is used. Specifically, the element at t_1, t_2 of P is given by

$$P(\alpha, \beta)_{t_1,t_2} = e^{-\alpha|t_1 - t_2|} e^{-\beta(t_1+t_2)/2} \qquad (14)$$

for $t_1, t_2 \in \{0, 1, \cdots, n_h - 1\}$ which makes the estimated impulse response to be an exponentially decaying function of t with decay rate β^{-1}, and correlation length α^{-1}. The latter *hyperparameters* α and β are determined using Empirical Bayes, i.e. as the maximizers of the log marginal likelihood (LML) of the measured output signal:

$$\mathrm{LML}(y) = -y^{\mathrm{T}} \Sigma(\alpha, \beta)^{-1} y - \log |\Sigma(\alpha, \beta)|, \qquad (15)$$

where $\Sigma(\alpha, \beta) = \sigma^2 \mathrm{I} + \phi^{\mathrm{T}} P(\alpha, \beta) \phi$ and ϕ is a Toeplitz matrix constructed with u. This problem is solved using `fmincon` in MATLAB™.

An implementation of the RFIR estimator is available as the `arx` function in the System Identification Toolbox since MATLAB™ 2013b with `na` $= 0$, `nb` $= n_g$, `nk` $= 0$ and regularization enabled in `arxoptions`. The hyperparameters can be determined via `arxregul`. More information about the RFIR estimator can be found in [15], [16]. The DFT of the obtained regularized estimate of the impulse response $\hat{g}_{\mathrm{RFIR}}$ of the system yields a smoothed estimate of the FRF.

Note that methods 3) and 4) result in a non-parametric estimate of the transfer function, represented as the FRF. A corresponding parametric estimate is obtained as follows. A parametric estimator (GTLS, BTLS, and MLE) is invoked, where the input spectrum is considered to be 1 at all frequencies (Dirac in the time domain), and the output spectrum is set equal to the estimated FRF, with the transient term I set to zero. It is important to note that invoking the GTLS, BTLS and MLE on the FRFs from methods 3) and 4) will yield a different result from applying the GTLS, BTLS and MLE on the raw data. This is because the FRFs from 3) and 4) have been smoothed and, thus, have a significantly lower noise variance.

B. Success Rates of the Initial Estimates

The initial estimates described above are then fed to the MLE together with the raw data to obtain the final estimates.

Formally, let $\hat{G}_\bullet$ be the final parametric estimate of the system, where the subscript $\bullet$ denotes the methods (from left to right in Fig. 2) when an initial estimate $\hat{G}_\bullet^{\mathrm{init}}$ is used, *viz*:

- $\hat{G}_0$ via the true model G_0 as an initial estimate,
- $\hat{G}_{\mathrm{exist}}$ is obtained via GTLS and BTLS,
- $\hat{G}_{\mathrm{trunc}}$ via use of the LPM with truncation as an initial estimate,
- $\hat{G}_{\mathrm{RFIR}}$, initialized by means of the RFIR method.

A particular initial estimate is deemed *successful* if the optimization of the MLE cost with the raw input-output data inserted in equation (6), reaches the *best local optimum* when iteration is initiated with the selected initial value/estimate.

Conjecture 1. *In this paper it is assumed that use of the true model as an initial estimate for the nonlinear identification algorithm (the leftmost path in Fig. 2) is the best possible initial estimate and will allow the optimization to converge to the* best local optimum*, i.e.* $\hat{G}_0$.

From this conjecture, use of the true system as an initial estimate would engender the best final estimate or hopefully even the global optimum. Since the true system is not really known in practice, the work in this paper compares the capability of different possible initial estimates to emulate the result that would be obtained by using the true model as an initial estimate.

The capability is quantified by defining the success rate of the initial estimate as the probability that the identification algorithm reaches the *best local optimum* when the selected initial estimate is used.

The success rate η is expressed mathematically as a probability function, consistent with Conjecture 1, *viz*:

$$\eta = \mathrm{P}\left(\hat{G}_\bullet = \hat{G}_0\right) \qquad (16)$$

In practice, and since the iterative algorithm usually does not reach the local optimum precisely, the above definition is relaxed to the form

$$\eta = \mathrm{P}\left(\left\|\hat{G}_\bullet - \hat{G}_0\right\|_2 < tol\right) \qquad (17)$$

with tol a numerical tolerance that will be specified later.

The distance measure used in equation (17) is defined, for discrete time, as

$$\|G\|_2 = \sqrt{\frac{1}{2\pi} \int_0^{2\pi} |G(e^{-j\omega})|^2 \, \mathrm{d}\omega} \qquad (18)$$

which, in practice, is computed by using the adaptive global quadrature algorithm provided by the MATLAB function `integral`. The success rate in equation (17) is estimated via Monte Carlo simulations in Section IV-C.

IV. Demonstration

A. The System Under Consideration

Two systems are considered. The first, with a quality factor $Q = 6\,\mathrm{dB}$, has the transfer function

$$G_0(z)\Big|_{Q=6\,\mathrm{dB}} = 1.74 \cdot 10^{-3} \frac{z^2 + 2z + 1}{z^2 - 1.93z + 0.94}. \qquad (19)$$

The second system, with a quality factor $Q = 20\,\mathrm{dB}$, has the transfer function

$$G_0(z)\Big|_{Q=20\,\mathrm{dB}} = 309 \cdot 10^{-6} \frac{z^2 + 2z + 1}{z^2 - 1.98z + 0.989}. \qquad (20)$$

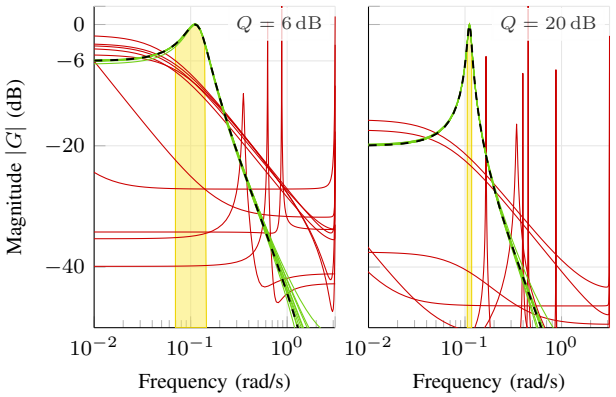

Fig. 3. Illustrative Bode plots of some successful (——) and failed (——) estimates $\hat{G}_\bullet$ for both systems G_0 (- - -) under test. Success or failure is determined using the criterion in equation (17). The 3 dB bandwidth of each system is highlighted ( ).

Both systems are low-pass Chebyshev filters shown in Fig. 3. Their inputs are excited by zero mean white Gaussian noise with unit variance, and the outputs are disturbed by different realizations of white gaussian noise, as in Fig. 1, with a variance σ_v^2, such that a prescribed signal-to-noise ratio (SNR) defined as

$$\mathrm{SNR} = \frac{\sigma_v}{\mathrm{rms}\,(y_0)} \approx \frac{\sigma_v}{\|G_0\|_2\,\mathrm{rms}\,(u_0)} = \frac{\sigma_v}{\|G_0\|_2}. \qquad (21)$$

is attained. $\mathrm{rms}\,(x) = \sqrt{N^{-1}\sum_t x^2(t)}$ denotes the root-mean-squared value of x.

B. Comparison of Successful and Failed Estimates

Fig. 3 depicts the Bode plot of the actual system G_0, in black and different estimates $\hat{G}_{\mathrm{exist}}$ that may or may not approximate $\hat{G}_0$ well, according to the distance defined by equation (17). The systems shown here are drawn randomly from the successes and failures observed in the Monte Carlo simulations, discussed later.

It is clear from Fig. 3 that the successful estimates virtually coincide with the true system. This means that their distances from the true system, given by $\|\hat{G}_\bullet - G_0\|_2^2$, have small values. On the other hand, the 'failed' estimates are inaccurate descriptions of the true system. It is apparent that their associated distances from the true system are larger by, at least, an order of magnitude than for the successful estimates.

Estimation 'failures' can be attributed to the optimization procedure getting stuck in local optima, resulting in very poor estimates of $G_0(\omega)$ over broad frequency ranges, as seen in Fig. 3. Consequently, such estimates are practically worthless approximations of G_0.

C. Computation of the Success Rates

1) Simulation: A Monte Carlo simulation was set up to determine the success rate of the estimator with different initial estimates for the following ranges

- of SNR: −20 dB, −16.7 dB, −13.3 dB, −10 dB, −7.8 dB, −5.6 dB, −3.3 dB, −1.1 dB, 1.1 dB, 3.3 dB, 5.6 dB, 7.8 dB, 10 dB, 13.3 dB, 16.7 dB and 20 dB
- of number of samples N is 3 368 or 4 462 for the first system whose quality factor $Q = 6\,\mathrm{dB}$,
- and of number of samples N is 20 250 or 26 792 for the second system whose quality factor $Q = 20\,\mathrm{dB}$.

One can, equivalently, express the experiment length as the number $N_{3\,\mathrm{dB}}$ of frequency bins that lie within the 3 dB bandwidth of the system. The 3 dB bandwidth is defined as the frequency band where the magnitude of the transfer function is, at most, 3 dB below its maximum as indicated in Fig. 3. For both systems, N was chosen such that $N_{3\,\mathrm{dB}}$ is respectively 36 or 48. For each value-pair (SNR, $N_{3\mathrm{dB}}$), 200 simulations and estimations were performed, for each of which the success (binary result) was determined. The success rate η in equation (17) was then estimated as the proportion of the successes from the 200 runs.

2) Illustrative Example: In particular, the simulation for each pair $(\mathrm{SNR}, N_{3\mathrm{dB}})$ yields 200 estimates per initial value strategy. This is illustrated graphically in Fig. 4 for the value-pair $(-1.1\,\mathrm{dB}, 48)$ of the system with $Q = 6\,\mathrm{dB}$, where the obtained distances $\|\hat{G}_\bullet - \hat{G}_0\|_2$ are shown for the different strategies, and sorted by an ascending norm. This effectively shows an empirical cumulative distribution function of the distance measure $\|\hat{G}_\bullet - \hat{G}_0\|_2$ as sampled by the Monte Carlo algorithm.

3) Pertinent Observations: With reference to Fig. 4, the following observations are worth noting:

- The actual parametric estimates, based on the different initial estimates, have distinct behaviors with respect to their distance from $\hat{G}_0$. Two major classifications can be made:
  1) 'good' estimates which have $\|\hat{G}_\bullet - \hat{G}_0\|_2 < tol \approx 6 \cdot 10^{-5}$,
  2) 'poor' estimates which are much farther from $\hat{G}_0$. This is caused by the local optima.
- The obvious jump of the observed distances of the final parametric estimates shows that both classes can be separated reliably. The index at which this jump occurs is a direct (visual) indication of the success rate η of the corresponding strategy to obtain initial values. This indicates 74% success for the existing methods, 95% for the truncation method and 100% for RFIR in Fig. 4.

D. Improvement Over Existing Techniques

In Fig. 5, the success rate (and its 95% confidence interval) obtained from the Monte Carlo simulations is shown to compare the proposed method with previously existing approaches. The confidence intervals are calculated via the Clopper-Pearson method [21] (as implemented in MATLAB™'s `binofit`). This is known to yield somewhat conservative bounds, especially if only failures or successes are observed [22]. Consequently, the improvements observed in Fig. 5 underestimate the actual improvement slightly.

By aggregating the results for the different simulations and retaining the success rate η only, a clearer view of the performance of the different methods is obtained.

The success rates of $\hat{G}_{\mathrm{exist}}$, $\hat{G}_{\mathrm{trunc}}$ and $\hat{G}_{\mathrm{RFIR}}$ are depicted in Fig. 5 for two different numbers of sample points within

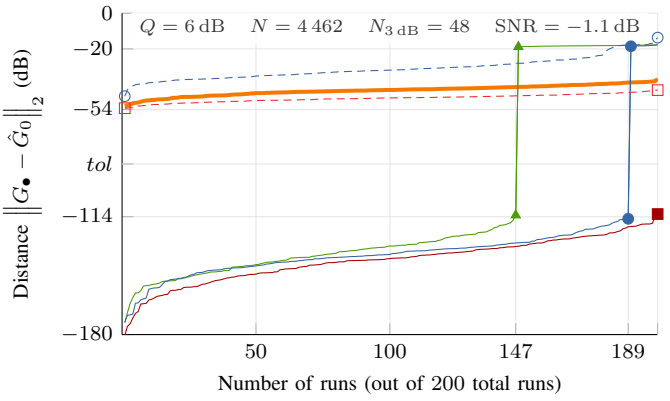

Fig. 4. The distance $\|\hat{G}_\bullet - \hat{G}_0\|_2$ between the different estimates ($\hat{G}_{\mathrm{exist}}$ (—▲—), $\hat{G}_{\mathrm{trunc}}$ (—●—) and $\hat{G}_{\mathrm{RFIR}}$ (—■—)) and the 'best theoretical' estimates ($\hat{G}_0$) shows that the smoothers help the estimate to converge to $\hat{G}_0$. The same measure is shown for the initial values $\hat{G}^{\mathrm{init}}_{\mathrm{RFIR}}$ (- -□- -) and $\hat{G}^{\mathrm{init}}_{\mathrm{trunc}}$ (- -○- -). $\|G_0 - \hat{G}_0\|_2$(——) shows how far the true model and 'best theoretical' estimates are apart as a reference.

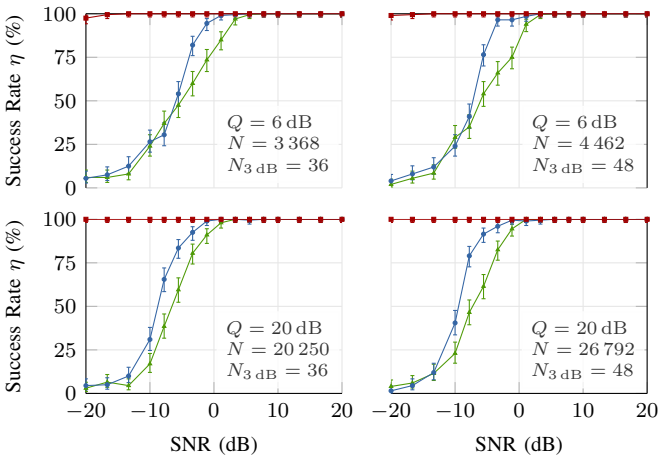

Fig. 5. The success rate η obtained using the smoothers, i.e. $\hat{G}_{\mathrm{RFIR}}$ (—■—) and $\hat{G}_{\mathrm{trunc}}$ (—●—), are significantly higher than for the existing methods $\hat{G}_{\mathrm{exist}}$ (—▲—) in most of the studied range of SNRs, number of samples and for both test systems. The 95% confidence intervals are indicated by bars.

the 3 dB bandwidth, and for a range of SNR values. A close scrutiny of Fig. 5 reveals the following observations:

- The success rates increase with increasing SNR values, as expected.
- For very low SNR values (below 0 dB), $\hat{G}_{\mathrm{exist}}$ and $\hat{G}_{\mathrm{trunc}}$ are unreliable, whereas $\hat{G}_{\mathrm{RFIR}}$ is reliably successful.
- For relatively high SNR (20 dB), all three estimators perform equally well.
- For moderate SNR values, the success rate of $\hat{G}_{\mathrm{trunc}}$ lies above that of $\hat{G}_{\mathrm{exist}}$.
- The RFIR-based estimate exhibits a success rate of, at least, 97% for all studied conditions. This clearly indicates that regularization makes it possible to obtain far more reliable initial estimates than both the LPM-Truncation-based ones and the existing ones. This is an important conclusion.

As far as $\hat{G}_{\mathrm{trunc}}$ and $\hat{G}_{\mathrm{exist}}$ are concerned, the observations on Fig. 5 clearly reveal the extent to which the estimator $\hat{G}_{\mathrm{trunc}}$ allows for the SNR to be decreased before the estimator attains too high a fail rate. In other words, it shows that, at low SNR values $\hat{G}_{\mathrm{trunc}}$ is more reliable than $\hat{G}_{\mathrm{exist}}$.

Fig. 5 shows that, typically, for the system with $Q = 6\,\mathrm{dB}$ when $N_{3\,\mathrm{dB}} = 48$ and $-1.1\,\mathrm{dB}$ SNR, the existing initial values will fail to deliver a good transfer function estimate for one in four trials ($\approx 74\%$ success rate). However, by using the truncated LPM's initial estimates, the number of failures is reduced fivefold ($\approx 95\%$ success rate) over the existing initial estimates. Moreover, the RFIR initial values lead to poor estimates in less than 2% ($> 98\%$ success rate) of the cases, since none were observed in the Monte Carlo simulation.

Extension of the SNR range for $\hat{G}_{\mathrm{trunc}}$ over $\hat{G}_{\mathrm{exist}}$: A close scrutiny of Fig. 5 reveals that the SNR range associated with an acceptable success rate (e.g. success rate $\geqslant 90\%$) is improved for $\hat{G}_{\mathrm{trunc}}$ over $\hat{G}_{\mathrm{exist}}$. The range of improvement is approximately 5 dB where the success rate is above 90%. This result is very useful for a user working with noisy data.

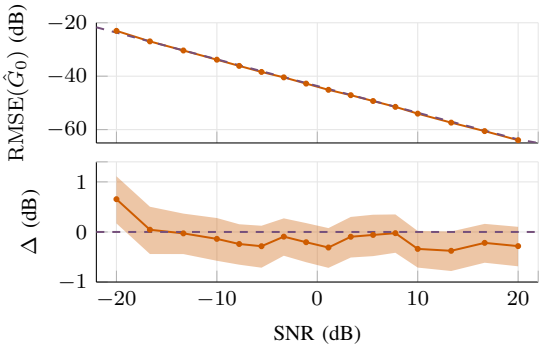

Fig. 6. The accordance between the observed RMSE of $\hat{G}_0$ (——)(with 2σ interval) and the rule-of-thumb (22)(- - -) shows how well $\hat{G}_0$ approximates the true system G_0. In the bottom plot, the difference Δ between both is shown. Consequently, it makes sense to examine the behavior of the models in terms of $\hat{G}_0$ as long as this RMSE meets the user's requirements.

Remark about the model quality: In principle, convergence of the estimate $\hat{G}_\bullet$ to the 'best' estimate $\hat{G}_0$ does not imply that $\hat{G}_\bullet$ will be suitable for a specific purpose as the estimate $\hat{G}_0$ itself might be ill-suited. A straightforward tool to analyze the error associated with $\hat{G}_0$ is to examine its relative RMS error over different realizations of the experiment. In [11], a rule-of-thumb is given that describes the relationship between the SNR of the signals and the quality of the estimated parametric model. In this context this heuristic boils down to:

$$\mathrm{RMSE}(\hat{G}_0) \approx \sqrt{\frac{n_\theta}{N}}\frac{\|G_0\|_2}{\mathrm{SNR}} \tag{22}$$

where $n_\theta = 5$ is the number of estimated parameters in this paper. In Fig. 6, the empirical $\mathrm{RMSE}(\hat{G}_0)$ over the different Monte Carlo runs is displayed against this rule-of-thumb. It can be seen that these runs align almost perfectly, which shows that this rule-of-thumb can be used to approximate the RMSE of the parametric estimate. When the RMSE is then divided by $\|G_0\|_2$, this graph essentially shows the anticipated relative

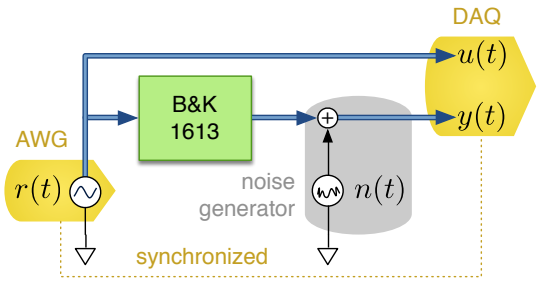

Fig. 7. Schematic representation of the measurement set-up for the Brüel & Kjær 1613 filter. The AWG (Arbitrary Waveform Generator) and DAQ (Data Acquisition) hardware share the same clock and are thus synchronized. Blue arrows indicate coax cables. At the output of the bandpass filter, a noise generator is added that buffers the incoming signal using a TL071 opamp and adds white noise before passing the signal to the DAQ.

error on the transfer function and, hence, gives an indication on its usability.

V. Experimental Results

Measurements were performed on an electrical filter to assess the performance of the initial value generating methods in a physical setting.

A. Experimental Setup

The Brüel & Kjær 1613 Octave Filter Set consists of sixth-order Chebyshev band-pass filters [23], from which we only examined the filter with a center frequency of 4 kHz.

To measure the transfer function, the Brüel & Kjær 1613 Filter was excited by a signal $r(t)$ and both its input $u(t)$ and output $y(t)$ were measured as shown in Fig. 7. The output $y(t)$, however, was disturbed by a noise generator that added white noise over the frequency range $[\mathrm{DC}, 20\,\mathrm{kHz}]$. The excitation signal $r(t)$ was white random noise with a standard deviation $\sigma_r = 0.25\,\mathrm{V}$, consisting of $N_S = 8\,192$ samples, and sampled at 92 kHz. The same noise sequence was repeated $N_R = 100$ times in succession, such that a non-parametric estimate of the noise could be obtained and the performance of the methods could be gauged over these different repetitions. To distinguish the different repetitions, we denote the r^{th} repetition of the u signal as $u^{[r]}(t)$ (and similarly for $y(t)$). The signal generation and acquisition was by means of a National Instruments Elvis II, using the respective `AO` and `AI` pins on the breadboard, which were wired to BNC connectors. Although acquisition and generation were synchronized, there was a small delay $\tau = 9\,\mu\mathrm{s}$ between the acquisition of the input $u(t)$ and the output $y(t)$. The output spectrum $Y(\omega)$ was multiplied by $\exp(-j\omega\tau)$ such that the delay-compensated $y(t)$ could be obtained using the inverse DFT.

B. Identification Procedure

The repeated nature of the experiment makes it possible to estimate the noise level from the signals non-parametrically. The mean signal from the input u is of the form:

$$\tilde{u}(t) = \frac{1}{N_R}\sum_{r=1}^{N_R} u^{[r]}(t) \tag{23}$$

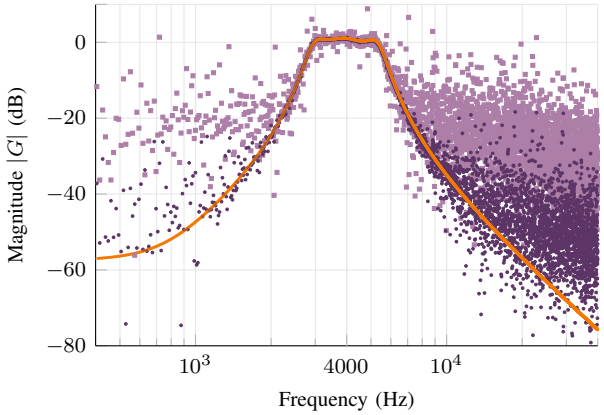

Fig. 8. Transfer function of the considered Brüel & Kjær 1613 filter. The empirical transfer function estimates $Y^{[r]}(\omega)/U^{[r]}(\omega)$ (·) and its average $\tilde{Y}(\omega)/\tilde{U}(\omega)$ (·) obtained from the Elvis measurements are shown together with the estimated reference model G_{VXI} (——).

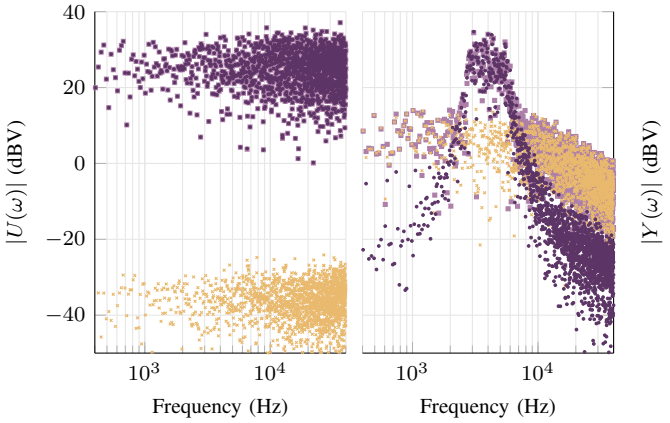

Fig. 9. Input $U^{[r]}(\omega)$ and output $Y^{[r]}(\omega)$ spectra (·) measured using the Elvis. (·): the mean spectra $\tilde{U}$ and $\tilde{Y}$, (·): noise levels $\hat{\sigma}_U$ and $\hat{\sigma}_Y$.

Thus, the reduced noise influence and the approximate noise co-variances are, respectively:

$$\hat{\sigma}_u^2(t) = \frac{1}{N_R - 1}\sum_{r=1}^{N_R}\left(u^{[r]}(t) - \tilde{u}(t)\right)^2 \tag{24}$$

$$\hat{\sigma}_{yu}(t) = \frac{1}{N_R - 1}\sum_{r=1}^{N_R}\left(y^{[r]}(t) - \tilde{y}(t)\right)\overline{\left(u^{[r]}(t) - \tilde{u}(t)\right)} \tag{25}$$

Similar calculations apply to $y(t)$ and as such, $\hat{\sigma}_u(t)$, $\hat{\sigma}_y(t)$ and $\hat{\sigma}_{yu}(t)$ can be estimated. Their frequency domain counterparts $\tilde{U}(\omega)$ and $\tilde{Y}(\omega)$ are obtained by using the DFT. The resulting empirical transfer function estimate $Y^{[r]}(\omega)/U^{[r]}(\omega)$ and its periodic average $\tilde{Y}/\tilde{U}$ are shown in Fig. 8. The noise covariances $\hat{\sigma}_U^2(\omega)$, $\hat{\sigma}_Y^2(\omega)$ and $\hat{\sigma}_{YU}(\omega)$ are calculated as the sample covariance, akin to equation (25), and their values are shown in Fig. 9. Note that the SNR at the output $\mathrm{SNR}_y \approx 14\,\mathrm{dB}$ is much smaller than at the input $\mathrm{SNR}_u \approx 50\,\mathrm{dB}$.

For each measurement repetition, the signals $u^{[r]}(t)$, $y^{[r]}(t)$ and the variances $\sigma_U^2(\omega)$, $\sigma_Y^2(\omega)$ are used in the maximum likelihood cost function given by equation (6). The model chosen for the cost function comprises a numerator of third

order, a denominator of sixth order and transient contribution of a fifth degree. This results in a transfer function of the form:

$$\hat{G}_\bullet(s,\theta) = \frac{\sum_{i=0}^{3} b_i s^i}{\sum_{i=0}^{6} a_i s^i} \tag{26}$$

which is typical of a sixth-order band-pass filter. By processing the different repetitions, a parametric fit per initialization method corresponding to the appropriate branch in Fig. 2, and per repetition can be obtained.

The following equation is constructed from the final estimates for each experiment; it gives an insight into the combined contribution to the 'best' estimate from all the different initialization strategies, *viz*:

$$\hat{G}_{\text{best}} = \arg\min_{\hat{G}} \left\{ V(\hat{G}_{\text{exist}}), V(\hat{G}_{\text{trunc}}), V(\hat{G}_{\text{RFIR}}) \right\} \tag{27}$$

where $V(\hat{G})$ corresponds to the cost function in equation (6) in the parametric estimates of $\hat{G}$. This is also the approach followed in [24] to combine different initial estimates.

C. Reference Model Measurements

Since the true model G_0 is not known for real-life systems, a practically viable reference model is needed. Additional measurements were performed using a VXI measurement setup, which allowed for a signal-to-noise ratio of more than 60 dB. Virtually noiseless, these measurements provided a very high quality model of the system, denoted as G_{VXI}, and used as a reference model. The VXI measurement setup is summarized as follows.

- Signal Generator card: VXI HP E1445A.
- Acquisition cards: VXI HP E1430A.
- Sampling frequency: $f_s = 156\,250$ Hz.
- A total of 558 frequency bins were used for the estimation, in the excited frequency band $[0.1, 20]$ kHz, with a frequency resolution of 35.7 Hz, giving a measurement time of 28 ms. The excitation signal was band-limited periodic noise with an RMS of 100 mV.
- The input and output signals were buffered and anti-alias filtered.

This yielded the model G_{VXI} (shown in Fig. 8) with a relative error of less than 0.3% in the pass-band. Denote $\hat{G}_{\text{VXI}}$ to be the parametric estimate obtained using G_{VXI} as an initial estimate. Furthermore, G_{VXI} was used to play the role of G_0 from the previous section.

D. Model Estimation

The measurements on the NI Elvis II indicate that the high-quality starting value G_{VXI} led to low cost function values as shown in Fig. 10 and Table I. The existing methods exhibit a high spread and a high median cost function, showing that a good estimate is obtained only in about 25% of the cases. The truncation method provides better estimates in many cases, but still suffers from a high variability. On the other hand, the RFIR-based initial values have both a low median cost function and spread, and thus provide better fits in almost all cases. Obviously, $\hat{G}_{\text{best}}$ has the lowest cost function values of all methods. The similarity of the results for $\hat{G}_{\text{RFIR}}$ and $\hat{G}_{\text{best}}$ suggests that the RFIR provides the best estimate in most of the cases. This can indeed be confirmed by inspecting the different estimates per repetition of the experiment.

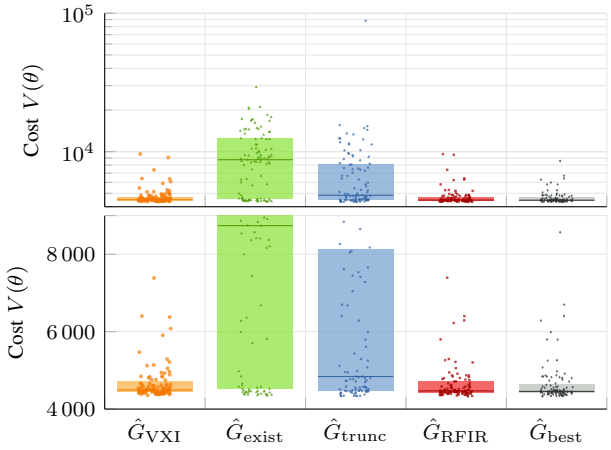

Fig. 10. Cost function values $V(\theta)$ obtained during the different realizations of the measurement. The bottom plot shows a zoom of the top plot. For each method, the inter-quartile range ( ), median (——) and individual values (·) are shown. This implies that a lower (local) minimum of the cost function can be attained using these smoothing techniques.

TABLE I
OBSERVED PERCENTILES OF THE COST FUNCTION $V(\hat{G}_\bullet)$.

METHOD	MIN. 0%	25%	MEDIAN 50%	75%	MAX. 100%
$\hat{G}_{\text{VXI}}$	4 355	4 442	4 494	4 707	9 634
$\hat{G}_{\text{exist}}$	4 345	4 526	8 741	12 413	29 237
$\hat{G}_{\text{trunc}}$	4 336	4 476	4 842	8 130	88 308
$\hat{G}_{\text{RFIR}}$	4 329	4 422	4 469	4 716	9 607
$\hat{G}_{\text{best}}$	4 329	4 415	4 458	4 644	8 570

Remark 2. In the simulations (e.g. Fig. 4), a 60 dB difference between 'good' and 'bad' estimates was observed. Such a large gap is not observed in the measurements (Fig. 10 and Table I). Hence, defining a success rate based on a threshold tolerance would be very sensitive to the specific value of the threshold and hence unreliable. Instead, the statistical location and dispersion are inspected to assess the relative performance of each method. Practically, the median is used as a measure of location and the inter-quartile range (IQR) is used to inspect the spread as these are far more robust to outliers than e.g. the sample mean and variance. These measures also have an easy interpretation: the median (50% percentile) indicates the cost function value that 50% of the repetitions attain. On the other hand, the IQR (25% through 75% percentile) contains exactly half of the observations.

E. Model Validation

1) Cost function limitations: In the previous section, the cost function was studied to determine the effectiveness of the starting values. However, inspecting the cost function

TABLE II
OBSERVED PERCENTILES OF THE VALIDATION DISTANCE $\left\|\hat{G}_\bullet - G_{\mathrm{VXI}}\right\|_2$.

METHOD	MIN. 0%	25%	MEDIAN 50%	75%	MAX. 100%
$\hat{G}_{\mathrm{VXI}}$	0.47	0.92	1.06	1.29	2.41
$\hat{G}_{\mathrm{exist}}$	1.48	1.88	13.72	19.31	26.80
$\hat{G}^{\mathrm{init}}_{\mathrm{trunc}}$	1.35	2.72	14.94	24.72	85.34
$\hat{G}_{\mathrm{trunc}}$	1.27	1.72	2.35	13.03	63.84
$\hat{G}^{\mathrm{init}}_{\mathrm{RFIR}}$	1.11	1.61	1.78	2.00	18.21
$\hat{G}_{\mathrm{RFIR}}$	1.21	1.57	1.72	1.85	4.55
$\hat{G}_{\mathrm{best}}$	1.21	1.56	1.72	1.85	15.01

$V(\theta)$ only accounts for how well an estimated model fits the measured data, which may be misleading. For an example, overfitting a model may result in the absorption of both the systematic behavior and the noise into the model. Consequently, an arbitrarily small cost function may ensue although the estimated model may be virtually useless to predict the system behavior.

2) Validation criterion: To objectively assess the quality of an estimated model, a different criterion than the cost function is inspected. To this end, the model is validated using the 2-norm (or distance) on the model error:

$$\left\|\hat{G}_\bullet - G_{\mathrm{VXI}}\right\|_2. \tag{28}$$

This criterion indicates how well the obtained estimates $\hat{G}_\bullet$ are able to describe the transfer function of the bandpass filter as observed in the validation measurement on the VXI.

3) Validation performance: The observed median and IQR of the distances in Fig. 11 and Table II have a similar qualitative interpretation as the cost function on the estimation data. The existing methods provide good estimates in only 25% of the cases. The truncation method provides a considerable improvement but still suffers from 25% poor estimates. Compared to the existing BTLS and GTLS techniques, the RFIR and hence also $\hat{G}_{\mathrm{best}}$ entail an overall reduction in the observed distance by a factor 8, in most cases. On the other hand, the reference $\hat{G}_{\mathrm{VXI}}$ shows the best global performance. This indicates that the proposed methods do no converge to the same local optimum. Nevertheless, the proposed methods improve the model quality by almost an order of magnitude.

As the criterion in equation (28) no longer depends on the cost function, the initial estimates can also be investigated. The difference between the distance of $\hat{G}^{\mathrm{init}}_\bullet$ and $\hat{G}_\bullet$ indicates how much the final ML estimate from the raw data, in each respective branch of Fig. 2, improves over the initial estimate. Remarkably, on average, the final ML estimation provides only a marginal improvement for the RFIR. However, this final step reduces the spread and hence yields a more reliable estimate. E.g. note in Table II that this step reduces the worst-case distance from 18.21 to 4.55.

VI. REMARK ON GENERALITY

In the previous sections, the usefulness of smoothing techniques to obtain initial values has been illustrated on example

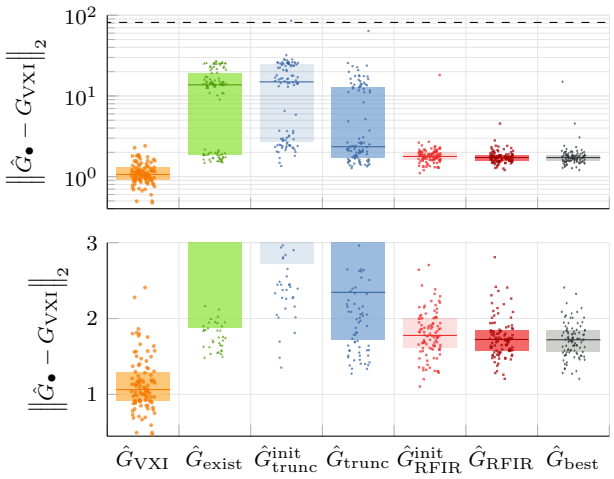

Fig. 11. Validation of the measured models. For each method, the distance (28) between the estimates and $\hat{G}_{\mathrm{VXI}}$ is shown (·) together with the median (——) and inter-quartile range (▭). $\|G_{\mathrm{VXI}}\|_2$(- - -) is shown as a reference. The bottom plot shows a zoom of the top plot. The proposed methods yield models that are closer to $\hat{G}_{\mathrm{VXI}}$ than the existing $\hat{G}_{\mathrm{exist}}$. The results are in line with the values of the cost function in Fig. 10.

systems. Other systems, e.g. discrete-time low-pass, high-pass, band-pass and band-stop filters of tenth order generated by the `cheby1`, `cheby2`, `ellip` and `butter` commands in MATLAB$^{\mathrm{TM}}$ have also been tried.

No cases were observed where using the smoothing to obtain initial values worsens the obtained model quality when the candidate model with the lowest cost function value is used (as in (27)). On the contrary, significant improvements were often achieved by including the initial values from the smoothed FRF and by combining different strategies.

The usefulness of particular initial estimates, e.g. $\hat{G}^{\mathrm{init}}_{\mathrm{RFIR}}$ is obviously linked to the appropriateness of the non-parametric smoother in the experimental conditions at hand. However, it is beyond the scope of this paper to determine which particular smoother is optimal in some specific circumstance.

VII. CONCLUSION

The simulations have demonstrated that use of a smoothed non-parametric estimate of the FRF can improve the success rate of minimizing the ML cost function for the estimation of the transfer functions of linear time-invariant (LTI) systems, subjected to noisy signals. Specifically, two smoothing methods were used and compared: 1) the truncated LPM method, and 2) the RFIR method. The simulation results clearly show that the RFIR method is superior to the truncated LPM method and the existing BTLS/GTLS methods.

The usefulness of the initial values obtained via the proposed smoothing techniques was confirmed on a measurement of a band-pass filter in a noisy environment. In most cases, the proposed initial values in the example made it possible to reduce the cost function by a factor two compared to existing methods. Moreover, the obtained models were validated against a high-quality measurement where the RMS error on the transfer function could be reduced by a factor of eight for most cases, due to the improved initial values.

Thus, the work in this paper has demonstrated the effectiveness of the studied FRF smoothing techniques in enhancing the initial values. Consequently, the ease-of-use of parametric model fitting is increased considerably by making it possible to obtain good parametric models without requiring any user interaction.

REFERENCES

[1] J. Schoukens, Y. Rolain, and R. Pintelon, "Improved frequency response function measurements for random noise excitations," *IEEE Transactions on Instrumentation and Measurement*, vol. 47, no. 1, pp. 322–326, FEB 1998.

[2] ——, "Leakage reduction in frequency-response function measurements," *IEEE Transactions on Instrumentation and Measurement*, vol. 55, no. 6, pp. 2286–2291, DEC 2006.

[3] P. Guillaume, I. Kollar, and R. Pintelon, "Statistical analysis of nonparametric transfer function estimates," *IEEE Transactions on Instrumentation and Measurement*, vol. 45, no. 2, pp. 594–600, APR 1996.

[4] P. Broersen, "A comparison of transfer-function estimators," *IEEE Transactions on Instrumentation and Measurement*, vol. 44, no. 3, pp. 657–661, JUN 1995.

[5] R. Pintelon, J. Schoukens, G. Vandersteen, and K. Barbé, "Estimation of nonparametric noise and FRF models for multivariable systems – Part I: Theory," *Mechanical systems and Signal processing*, vol. 24, no. 3, pp. 573–595, 2010.

[6] 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*, vol. 43, no. 10, pp. 1723 – 1736, 2007.

[7] H. Lim, "Frequency-Domain Analysis of Effects of a Diverged Interconnect Design Involved in High-Speed Front-End Electronics," *IEEE Transactions on Instrumentation and Measurement*, vol. 59, no. 10, pp. 2779–2786, OCT 2010.

[8] S. P. Robinson, D. R. Bacon, and B. C. Moss, "The measurement of the frequency response of a photodiode and amplifier using an optomechanical frequency response calibrator," *Measurement Science and Technology*, vol. 1, no. 11, p. 1184, 1990.

[9] V. Behjat, A. Vahedi, A. Setayeshmehr, H. Borsi, and E. Gockenbach, "Identification of the most sensitive frequency response measurement technique for diagnosis of interturn faults in power transformers," *Measurement Science and Technology*, vol. 21, no. 7, p. 075106, 2010.

[10] R. Pintelon and J. Schoukens, *System Identification: A Frequency Domain Approach*, 2nd ed. Wiley, Mar 2012.

[11] L. Ljung, *System Identification: Theory for the User*, 2nd ed. Upper Saddle River, NJ: Prentice-Hall, 1999.

[12] J. Schoukens, G. Vandersteen, R. Pintelon, and P. Guillaume, "Frequency-domain identification of linear systems using arbitrary excitations and a nonparametric noise model," *IEEE Trans. on Automatic Control*, vol. 44, no. 2, pp. 343 – 347, Feb. 1999.

[13] R. Pintelon, P. Guillaume, G. Vandersteen, and Y. Rolain, "Analyses, development, and applications of TLS algorithms in frequency domain system identification," *SIAM Journal on Matrix Analysis and Applications*, vol. 19, no. 4, pp. 983–1004, Oct. 1998.

[14] M. L. D. Lumori, E. Geerardyn, J. Schoukens, and J. Lataire, "Smoothing the LPM estimate of the frequency response function via an impulse response truncation technique," *IEEE Transactions on Instrumentation and Measurement*, vol. 63, no. 1, pp. 214–220, Jan 2014.

[15] G. Pillonetto and G. D. Nicolao, "A new kernel-based approach for linear system identification," *Automatica*, vol. 46, no. 1, pp. 81 – 93, 2010.

[16] T. Chen, H. Ohlsson, and L. Ljung, "On the estimation of transfer functions, regularizations and gaussian processes—revisited," *Automatica*, vol. 48, no. 8, pp. 1525 – 1535, 2012.

[17] T. Söderström and P. Stoica, *System Identification*. Prentice-Hall, 1989.

[18] D. Marquardt, "An algorithm for least-squares estimation of nonlinear parameters," *Journal of the Society for Industrial and Applied Mathematics*, vol. 11, no. 2, pp. 431–441, 1963.

[19] J. C. Spall, "Stochastic optimization," in *Handbook of Computational Statistics*, 2nd ed., ser. Springer Handbooks of Computational Statistics, J. E. Gentle, W. K. Härdle, and Y. Mori, Eds. Springer Berlin Heidelberg, 2012, pp. 173–201.

[20] W. H. Press, S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery, *Numerical Recipes - The Art of Scientific Computing*, 3rd ed. Cambridge University Press, 2007.

[21] C. J. Clopper and E. S. Pearson, "The use of confidence or fiducial limits illustrated on the case of the binomial," *Biometrika*, vol. 26, no. 4, pp. 404–413, 1934.

[22] T. D. Ross, "Accurate confidence intervals for binomial proportion and poisson rate estimation," *Computers in Biology and Medicine*, vol. 33, no. 6, pp. 509 – 531, 2003.

[23] *Instructions and Applications. Precision Sound Level Meter Type 2203, Octave Filter Set Type 1613*, Brüel & Kjær.

[24] I. Kollár, R. Pintelon, and J. Schoukens. (1994–2015) Frequency domain system identification toolbox. Gamax Ltd.